The followingt is a revised version of Chapter 2 in

Bimbó, K., *Proof Theory: Sequent Calculi and Related Formalisms*, CRC Press, Boca Raton, FL, 2015.

Amendment to the *Proof Theory* book

There was a mistake in a definition (Definition 2.25) in Chapter 2, which impacted the listing of cases in the proof of the admissibility of the single cut rule in LK. The new definition follows Bimbó (2015) [*Theoretical Computer Science* 597], and the cases are rearranged (which also saves a few pages). The following pages are the revised version of Chapter 2.

Chapter 2

Classical first-order logic

The first sequent calculus, LK is the example after which many other calculi have been fashioned. We describe this calculus, prove its equivalence to the axiom system K, and provide a sound and complete interpretation too. A major part of Section 2.3 is a presentation of the proof of the *cut theorem* using *triple induction*. We avoid the detour via the mix rule by using a suitable formulation of the *single cut rule*, by adding a *new parameter* to the induction, and by *defining anew* the rank of a cut. Although we do not go through all the details of this proof, we hope that sufficiently many cases are included so that the structure of the proof becomes completely clear. Later on, we will provide more condensed proofs of cut theorems or simply state cut theorems with reference to this proof, possibly, with modifications.

2.1 The sequent calculus LK

The calculus LK was introduced by Gentzen in [104], where it is labeled *logistic* (*logistisch*, in German). We will simply use the general term *sequent calculus* until we introduce *consecution calculi* in Chapter 5. Our presentation of LK is faithful to the original in the shape of the axioms and rules, though we do not always use the same notation or symbols.

The formulas of LK belong to the following language. There are denumerably many *name constants*, and they are denoted by $a_0, a_1, a_2, \ldots$. There are *function symbols*, denoted by $f_0^{n_0}, f_1^{n_1}, f_2^{n_2}, \ldots$, where the subscript identifies the function symbol, whereas the superscript shows the arity of the function symbol. We define the language so that the *arity* of function symbols (and later on, of predicate symbols and of connectives) is always a *positive integer*. The language contains *predicate symbols* that are denoted by $P_0^{m_0}, P_1^{m_1}, P_2^{m_2}, \ldots$. Once again, the subscript identifies the predicate and the superscript indicates the number of arguments the predicate takes. The language includes denumerably many of both function symbols and of predicate symbols. We use $=$ as the symbol for the two-place predicate called *identity*. Identity is considered a *logical* component of the language; therefore, the interpretation of $=$ is fixed, as it will become obvious in Section 2.4. *Propo-*

sitional variables are denoted by $p_0, p_1, p_2, \ldots$, and we stipulate that there are denumerably many of them.

Individual variables are a very important kind of basic expressions. Every first-order language must contain *denumerably many* individual variables, which are the only variables that can be quantified in a first-order language. We denote individual variables by $x_0, x_1, x_2, \ldots$. The set of logical components—beyond identity—includes *connectives*, a pair of *constant propositions* and *quantifiers*. The connectives are *negation* ($\neg$), *conjunction* ($\wedge$), *disjunction* ($\vee$) and *conditional* ($\supset$). The constant propositions are $\boldsymbol{T}$ and $\boldsymbol{F}$, which can be thought of as "the true proposition" and "the false proposition," respectively. The names of these constants clearly show the limited amount of information that is taken into account in classical logic. The *universal* and the *existential* quantifiers are denoted by $\forall$ and $\exists$, respectively.

Occasionally, we may simplify our notation—provided that no confusion is likely. For instance, we might also use a, b and c for name constants, f, g and h for function symbols, P, Q and R for predicate symbols, p, q and r for propositional variables and x, y and z for individual variables. Sometimes we will omit superscripts, when the arity of a symbol can be determined from the (implicit) assumption that an expression is well formed.

So far there is nothing in the set-up of the language that would be specific to a sequent calculus. The next two definitions are standard too.

Definition 2.1. (Terms) The *set of terms* is defined inductively by (1)–(2).

(1) Name constants and individual variables are *terms*;

(2) if $f_i^{n_i}$ is a function symbol and $t_1, \ldots, t_{n_i}$ are terms, then $f_i(t_1, \ldots, t_{n_i})$ is a *term*.

Having said that the definition is inductive, we assume that the set of terms is the *least set* generated from the set of *atomic terms* (i.e., the set specified in clause (1)) by finitely many applications of the inductive clause (i.e., clause (2)). Finitely many includes zero many; that is, all the atomic terms are terms—as their label intended to suggest.[1]

The set of terms could be defined using the *Backus–Naur form* (BNF) (as it is often done nowadays), if we stipulate that $\mathbb{A}$ is a non-terminal symbol that rewrites to an atomic term. Characterizing atomic terms by a *context-free grammar* (CFG) is not difficult, but it is not pretty. We give the following definition as an example, and later on we always assume that a similar definition can be given when there is a base set, the elements of which are indexed by the natural numbers.[2] (The subscript is concatenated to the previous letter and it is slightly lowered for aesthetic reasons only.)

[1] Inductive definitions and inductive proofs are explained in detail, for example, in [21] and in [121]. Here we only rely on a basic understanding of inductive definitions.

[2] Knowledge of formal language theory is not essential for our purposes, though it may be useful. For example, Sipser [189] is a good introductory text on formal language theory.

Definition 2.2. (Atomic terms) Let A, I, S and N be non-terminal symbols. $\mathbb{A}$, the *set of atomic terms* is defined by the next context-free grammar.

1. $\mathbb{A} \ := \ A_I$
2. $A \ := \ a \mid x$
3. $I \ := \ 0 \mid N \mid NS$
4. $S \ := \ 0 \mid N \mid SS$
5. $N \ := \ 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

The *start* symbol is $\mathbb{A}$, which is replaced by an "indexed atomic term" A_I. Since we do not want to have a 0 at the beginning of a string of digits, except when 0 is the whole string, we have to complicate the definition of I (the "index") with N (for "non-zero digit") and S (for "subsequent digit"). Permitting S to be replaced by SS ensures that indices greater than 99 can be generated too. Obviously, the complications in the above CFG stem from the steps that produce a natural number in decimal notation, with positive numbers not starting with 0's.

Definition 2.3. (Terms in BNF) The *set of terms* is defined as

$$T \ := \ \mathbb{A} \mid \mathbb{F}_i^{n_i}(\underbrace{T, \ldots, T}_{n_i}) \ ,$$

where $\mathbb{A}$ rewrites to an atomic term, and $\mathbb{F}_i^{n_i}$ rewrites to a function symbol.

Exercise 2.1.1. Give a definition for $\mathbb{F}_i^{n_i}$ that is similar to 2.2. [Hint: One has to make sure that for each i there is a unique positive integer n_i. Here is an idea that works, though it may not be the simplest solution (and it will not produce f_0, e.g.). Think about the subscript and superscript as the single string in_i, in which there is a substring of the form $\varrho(n_i)n_i$ (or $\varrho(n_i);n_i$ with $;$ to separate the subscript and the superscript), where $\varrho(n_i)$ is the reverse of the string n_i. Palindromes are easy to generate, though a slight complication emerges here, because we do not want n_i (hence, i either) to start with 0. Then i is $m\varrho(n_i)$, where m is any string that does not start with 0. Associating the superscript to the subscript in this way can also ensure that there are denumerably many function symbols of each arity.]

The following definition presupposes that we can refer not only to the set of terms but also to the set of variables as a category of expressions.

Definition 2.4. (Well-formed formulas) The *set of well-formed formulas* is inductively defined by (1)–(5).

(1) $\boldsymbol{T}$ and $\boldsymbol{F}$ are *well-formed formulas*;

(2) propositional variables are *well-formed formulas*;

(3) if $t_1, \ldots, t_{n_i}$ are terms, and $P_i^{n_i}$ is a predicate symbol, then $P_i^{n_i}(t_1, \ldots, t_{n_i})$ is a *well-formed formula*;

(4) if $\mathcal{A}$ and $\mathcal{B}$ are well-formed formulas, then $\neg\mathcal{A}$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$ and $(\mathcal{A} \supset \mathcal{B})$ are *well-formed formulas*;

(5) if $\mathcal{A}$ is a well-formed formula and x_i is an individual variable, then $\forall x_i\, \mathcal{A}$ and $\exists x_i\, \mathcal{A}$ are *well-formed formulas*.

A well-formed formula is an *atomic formula*, if it is by clause (1), (2) or (3).

"Well-formed formula" is a lengthy expression; hence, we will often use instead the abbreviation *wff* or the shorter term "formula."

Note that we take the formation of an identity statement to be a special instance of (3). We introduced $=$ as the identity symbol, but one particular P_i could be fixed as the identity symbol. Then, by a notational convention, we would write $t_1 = t_2$ instead of $P_i(t_1, t_2)$.

Exercise 2.1.2. The expressions in (a) are terms, and those in (b) are wff's. Use the Definitions 2.1 and 2.4 to generate these expressions step by step.

(a) x_3, $f_1^2(a_4, x_6)$, $h(g(x, b), f(a, y, y, z))$

(b) $(p \supset P_1^1(x_0))$, $(\neg\forall x_0\, P_1^1(x_0) \wedge \forall x_1\, P_1^1(x_2))$, $\exists x_1\, \forall x_2\, \exists x_3 (P_1^2(x_1, x_2) \supset (P_2^4(x_1, a, b, x_2) \supset (P_3^1(x_3) \vee P_4^2(x_2, x_4))))$, $(\neg p \supset \neg\neg(Q(b, c, a) \wedge \neg r))$

Exercise 2.1.3. The following expressions are neither terms nor wff's. Explain for each why it does not belong to either of those categories of expressions. (a) $\supset p$, (b) $a_2 p_1 \supset p_2 a_3$, (c) $\forall\neg\,(\neg P(x) \vee \exists Q(p, q))$. [Hint: You may discover that the notation has certain built-in assumptions that are usually left tacit.]

An occurrence of a variable in a formula is just what the informal meaning of the phrase suggests: one can look at the formula and find the variable in it. It is possible to assign unique labels to each occurrence of a variable, and then to define when an occurrence is free or bound. However, for our purposes it is sufficient to note this possibility without giving its details. Similarly, *a subformula* is a wff that occurs in a given wff.

Exercise 2.1.4. Design a schema that assigns a unique numerical label to occurrences of individual variables.

Definition 2.5. (Variable binding) An occurrence of x in $\mathcal{A}$ is *free* iff (i.e., if and only if) it is not within a subformula of $\mathcal{A}$ of the form $\forall x\, \mathcal{B}$ or $\exists x\, \mathcal{B}$. All occurrences of x in $\mathcal{A}$ that are not free are *bound*.

Each displayed quantifier in $\forall x\, \mathcal{A}$ and $\exists x\, \mathcal{A}$ *binds* the immediately following occurrence of x, as well as all the *free occurrences* of x in $\mathcal{A}$.

Exercise 2.1.5. Describe an informal procedure that allows one to decide if an occurrence of a variable is free or bound. Explain how to find the quantifier which binds a particular bound occurrence of a variable.

Every occurrence of every variable is either free or bound, but not both. It is also customary to talk about free and bound variables of a formula. x is a *free variable* of $\mathcal{A}$ iff there is a free occurrence of x in $\mathcal{A}$, and similarly, for a *bound variable*. Then, it is easy to see that, if x occurs in $\mathcal{A}$, then x is a free or a bound variable of $\mathcal{A}$, and possibly, both.

Exercise 2.1.6. List all the variables that occur in the following formulas. Then, for each occurrence determine whether it is free or bound. Lastly, for each formula, list its free and bound variables.

(a) $(\forall x_0\,(P^1_1(x_0) \supset P^2_2(x_0, x_1)) \wedge \exists x_3\,\neg P^1_1(x_3))$

(b) $\forall y\,(Q(z) \vee \neg\exists z\,R(z, y, a_3, z))$

(c) $\forall x_1\,\exists x_1\,(\forall x_2\,P^1_1(x_1) \supset \forall x_3\,\neg\exists x_1\,P^3_4(a_0, x_1, x_3))$

Exercise 2.1.7. Create formulas with the variable binding properties listed.

(a) x has no occurrence, y has one bound and z has one free occurrence.

(b) x has one free, y has two bound and z has two free and one bound occurrences.

A wff of the form $\mathcal{A}(x)$ may have some occurrences of x, and some of those may be free in $\mathcal{A}(x)$. (There are surely no free occurrences of x in $\forall x\,\mathcal{A}(x)$.) Then, in a limited context such as an axiom or a rule, we mean by $\mathcal{A}(y)$, the formula that is obtained from $\mathcal{A}(x)$ by replacing *all* or *all selected* free occurrences of x by y. However, we have to add a proviso to this, because we want all the newly inserted y's to be free in $\mathcal{A}(y)$. (Informally speaking, all the replaced x's were free, hence so should be all the new y's.) The situation in which the new y's are guaranteed to be free is what we call "y is OK for substitution for x in $\mathcal{A}(x)$," and then we say that $\mathcal{A}(y)$ is the result of the *substitution* of y for x in $\mathcal{A}(x)$. Substituting for a subset of the free occurrences of x is simply limited substitution; we scrutinize further the more encompassing concept first.

A slightly more complicated description of when y is OK for substitution for x can be given by characterizing where x occurs in $\mathcal{A}(x)$. x may have both free and bound occurrences in $\mathcal{A}(x)$, and we may forget about all the bound occurrences for now. A free occurrence of x cannot be in the scope of $\forall x$, however, it may be in a subformula of the form $\forall y\,\mathcal{B}$, that is, somewhere inside $\mathcal{B}$. Since we (tacitly) assume that x and y are distinct variables, $\forall y$ leaves the free occurrences of x free. However, if such an x is replaced by a y, then that y becomes bound by $\forall y$. To sum up in a sentence, *y is OK for substitution for x in $\mathcal{A}(x)$* iff no free occurrence of x in $\mathcal{A}(x)$ is within a subformula of the form $\forall y\,\mathcal{B}$. In the case of a limited substitution, the "no free occurrence" is qualified as "no selected free occurrence."

We do not allow substitutions when a variable is not OK for substitution, which makes substitution a partial operation, which is not always defined for a wff and a pair of variables. However, when we use $\mathcal{A}(y)$ in the sense we have just described, we assume that y is a suitably chosen variable. There are other ways to deal with the problem of the "clash of variables," as it is sometimes called. For instance, by renaming the bound variables that would cause the new occurrences of the substituted variable to become bound. However, the problem can always be circumvented by choosing a variable to play the role of y, which has *no occurrences* in $\mathcal{A}(x)$.

Exercise 2.1.8. Decide if the "OK for substitution" conditions are satisfied in the situations (a)–(d) (for unlimited substitution).

(a) y is OK for substitution for x in $(\forall x\, P(x,y) \supset \neg(\forall y\, P(y,x) \supset Q(x)))$

(b) x is OK for substitution for y in $(\forall x\, P(x,y) \supset \neg(\forall y\, P(y,x) \supset Q(x)))$

(c) z is OK for substitution for x in $\forall z\,(P(x,z) \supset \neg\forall x\, P(z,x))$

(d) z is OK for substitution for y in $\forall z\,(P(x,z) \supset \neg\forall x\, P(z,x))$

A quintessential feature of sequent calculi is that they are designed to formalize reasoning *about inferences* rather than simply to be a framework to construct inferences. In the case of classical logic, there can be finitely many premises and finitely many conclusions. This idea is captured in the concept of *sequents*.

Definition 2.6. (Sequents) If $\langle \mathcal{A}_0, \ldots, \mathcal{A}_n \rangle$ and $\langle \mathcal{B}_0, \ldots, \mathcal{B}_m \rangle$ are (possibly empty) sequences of formulas, then $\mathcal{A}_0, \ldots, \mathcal{A}_n \vdash \mathcal{B}_0, \ldots, \mathcal{B}_m$ is a *sequent*. The $\mathcal{A}_0, \ldots, \mathcal{A}_n$ part is the *antecedent* and the $\mathcal{B}_0, \ldots, \mathcal{B}_m$ part is the *succedent* of the sequent. The comma (i.e., $,$), which separates the formulas, is a *structural connective*.

The $\vdash$ symbol *does not* stand for a connective; it forms a sequent from two sequences of formulas. Either of these two sequences of formulas may be *empty*. The *turnstile* (i.e., $\vdash$) is often used to indicate that a formula is a theorem of an axiom system, and we will occasionally use $\vdash$ in that sense too. Thus, $\vdash$ can appear in connection with more than one system, but hopefully, this will not cause confusion, because context will determine which sense is meant. These uses of $\vdash$ are connected; in the sequent calculus LK, $\vdash \mathcal{A}$ will be the analog of $\vdash \mathcal{A}$ in the corresponding axiom system K.

Sequences of formulas will be denoted by capital Greek letters such as Γ and Δ. We omit the angle brackets (as above) in the context of a sequent. By $\Gamma, \mathcal{A}$ and $\mathcal{A}, \Gamma$, we mean the sequence of formulas obtained via appending the formula $\mathcal{A}$ to the end or to the beginning of the sequence Γ, respectively. Sequences are the *structures* in LK.

We explained (on page 15) the concept of limited substitution. In the upper sequent of the quantifier rules, by $\mathcal{A}(y)$, we mean a formula $\mathcal{A}$ with zero or more free occurrences of y selected. Then, $\forall x\, \mathcal{A}(x)$ or $\exists x\, \mathcal{A}(x)$ in the lower sequent is obtained by replacing the selected occurrences of y with x (which must be OK for the selected free occurrences of y in $\mathcal{A}(y)$), and then by attaching a quantifier prefix. Furthermore, substituting y for all free occurrences of x in $\mathcal{A}(x)$ must return $\mathcal{A}(y)$.

LK is a sequent calculus for *classical first-order logic*, FOL. It contains an *axiom* and *rules*. Some of the rules will only change the shape of a structure, and accordingly, they are called *structural rules*. The other rules are connected in an obvious way to connectives and quantifiers, hence, they are called *connective* and *quantifier rules*, or together, *operational rules*. To simplify our notation a bit further, we will omit outside parentheses from wff's. Note that we do not give any special rules involving $\boldsymbol{T}$, $\boldsymbol{F}$ or $=$ now; in effect, we temporarily exclude them from the language.

$$\mathcal{A} \vdash \mathcal{A} \ \text{id}$$

$$\frac{\mathcal{A}, \Gamma \vdash \Delta}{\mathcal{A} \land \mathcal{B}, \Gamma \vdash \Delta} \land\vdash_1 \qquad \frac{\mathcal{B}, \Gamma \vdash \Delta}{\mathcal{A} \land \mathcal{B}, \Gamma \vdash \Delta} \land\vdash_2 \qquad \frac{\Gamma \vdash \Delta, \mathcal{A} \quad \Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \land \mathcal{B}} \vdash\land$$

$$\frac{\mathcal{A}, \Gamma \vdash \Delta \quad \mathcal{B}, \Gamma \vdash \Delta}{\mathcal{A} \lor \mathcal{B}, \Gamma \vdash \Delta} \lor\vdash \qquad \frac{\Gamma \vdash \Delta, \mathcal{A}}{\Gamma \vdash \Delta, \mathcal{A} \lor \mathcal{B}} \vdash\lor_1 \qquad \frac{\Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \lor \mathcal{B}} \vdash\lor_2$$

$$\frac{\Gamma \vdash \Delta, \mathcal{A}}{\neg\mathcal{A}, \Gamma \vdash \Delta} \neg\vdash \qquad \frac{\mathcal{A}, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg\mathcal{A}} \vdash\neg$$

$$\frac{\Gamma \vdash \Delta, \mathcal{A} \quad \mathcal{B}, \Theta \vdash \Lambda}{\mathcal{A} \supset \mathcal{B}, \Gamma, \Theta \vdash \Delta, \Lambda} \supset\vdash \qquad \frac{\mathcal{A}, \Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \supset \mathcal{B}} \vdash\supset$$

$$\frac{\mathcal{A}(y), \Gamma \vdash \Delta}{\forall x\, \mathcal{A}(x), \Gamma \vdash \Delta} \forall\vdash \qquad \frac{\Gamma \vdash \Delta, \mathcal{A}(y)}{\Gamma \vdash \Delta, \forall x\, \mathcal{A}(x)} \vdash\forall^{\oslash}$$

$$\frac{\mathcal{A}(y), \Gamma \vdash \Delta}{\exists x\, \mathcal{A}(x), \Gamma \vdash \Delta} \exists\vdash^{\oslash} \qquad \frac{\Gamma \vdash \Delta, \mathcal{A}(y)}{\Gamma \vdash \Delta, \exists x\, \mathcal{A}(x)} \vdash\exists$$

The rules $(\vdash\forall)$ and $(\exists\vdash)$, which are marked by $^{\oslash}$, are restricted, and they are applicable when the lower sequent *does not contain* the variable y free. An obvious consequence of this stipulation is that there can be no free occurrences of y in the elements of Γ or Δ. Also, all the free occurrences of y in $\mathcal{A}(y)$ must be selected in those two rules.

So far, we have listed the *operational rules* in LK. The next six rules are the *structural rules*.

$$\frac{\Gamma \vdash \Delta}{\mathcal{A}, \Gamma \vdash \Delta} K\vdash \qquad \frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, \mathcal{A}} \vdash K$$

$$\frac{\mathcal{A}, \mathcal{A}, \Gamma \vdash \Delta}{\mathcal{A}, \Gamma \vdash \Delta} W\vdash \qquad \frac{\Gamma \vdash \Delta, \mathcal{A}, \mathcal{A}}{\Gamma \vdash \Delta, \mathcal{A}} \vdash W$$

$$\frac{\Gamma, \mathcal{A}, \mathcal{B}, \Delta \vdash \Theta}{\Gamma, \mathcal{B}, \mathcal{A}, \Delta \vdash \Theta} C\vdash \qquad \frac{\Theta \vdash \Gamma, \mathcal{A}, \mathcal{B}, \Delta}{\Theta \vdash \Gamma, \mathcal{B}, \mathcal{A}, \Delta} \vdash C$$

The labels for the structural rules allude to the Curry–Howard correspondence between implicational formulas and combinators. We will present a more rigorous version of the correspondence between structural rules and combinators in Chapter 5, at which point we will explain the motivation for these labels too.

The $(K\vdash)$ and $(\vdash K)$ rules are called *thinning* or *weakening*. If $\Gamma \vdash \Delta$ expresses that from the premises in Γ, at least one of the conclusions in Δ can be inferred, then adding a premise or a conclusion weakens the original claim about the inference. The $(W\vdash)$ and $(\vdash W)$ rules are called *contraction*, because two occurrences of a formula are reduced to one occurrence. The $(C\vdash)$ and $(\vdash C)$ rules are called *permutation*, *interchange* or *exchange*. These rules allow one to switch a pair of formulas that are adjacent, anywhere within a sequent. Although any of Γ, Δ and Θ may be empty, if they are not, then the shape of the permutation rules guarantees that $\mathcal{A}$ and $\mathcal{B}$ still can be swapped.

Keeping the structural rules in mind, we could (informally) view sequences of wff's as sets, where the order and the number of occurrences of wff's does not matter. Replacing sequences of wff's by other structures will be important in other sequent calculi (e.g., in Chapters 3, 4 and 5). However, in this chapter, we consider *LK* in its original form, where sequents are built from sequences of wff's.

Sequences and sequents imply a linear ordering or a linear succession of formulas; proofs are more complex.

Definition 2.7. (Proofs) A *proof* is a finite tree comprising occurrences of sequents, where each leaf of the tree is an instance of the axiom and all the other nodes are justified by applications of the rules.[3]

We say that a sequent $\Gamma \vdash \Delta$ is *provable* when there is a proof in which $\Gamma \vdash \Delta$ is the end sequent (i.e., the root of the proof tree). To make the definition of proofs more palpable we give two proofs as examples followed by detailed explanations.

Example 2.8. First,we prove an instance of exchanging different quantifiers.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{R(z,v) \vdash R(z,v)}{\forall y\, R(z,y) \vdash R(z,v)}\ \forall\vdash}{\forall y\, R(z,y) \vdash \exists x\, R(x,v)}\ \vdash\exists}{\forall y\, R(z,y) \vdash \forall y\, \exists x\, R(x,y)}\ \vdash\forall}{\exists x\, \forall y\, R(x,y) \vdash \forall y\, \exists x\, R(x,y)}\ \exists\vdash}{\vdash \exists x\, \forall y\, R(x,y) \supset \forall y\, \exists x\, R(x,y)}\ \vdash\supset$$

The first line is an instance of the axiom, from which we can get the next sequent by $(\forall\vdash)$. There is no restriction on this rule, and there is no restriction on the $(\vdash\exists)$ rule either, which yields the next sequent. Each of the variables

[3]The notion of trees as we use it here and elsewhere in this book is explained in Section A.1 of Appendix A. (Our usage is the standard one in logic and in some parts of computer science; however, the term is used with a wider meaning in some other areas.)

z and v has exactly one occurrence in the sequent, which means that the rules $(\vdash\forall)$ and $(\exists\vdash)$ can be applied in the next two steps. The last move is to introduce $\supset$. The proof also shows that Γ and Δ can be empty, for instance, in the last step they are both empty. (Incidentally, note that the labels are not part of the proof, they are added to help to see the proof as a proof.)

Definition 2.9. (Theoremhood) A formula $\mathcal{A}$ is a *theorem* of LK, if there is a proof ending in the sequent $\vdash \mathcal{A}$.

From the previous example, we know that $\exists x\,\forall y\,R(x,y) \supset \forall y\,\exists x\,R(x,y)$ is a theorem of LK.

<u>*Example*</u> 2.10. Now we prove a theorem that involves both connective and structural rules.

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{C}\vdash\mathcal{C}}{\mathcal{A},\mathcal{C}\vdash\mathcal{C}}\,{\scriptstyle K\vdash}}{\mathcal{C},\mathcal{A}\vdash\mathcal{C}}\,{\scriptstyle C\vdash}
\qquad
\dfrac{\dfrac{\dfrac{\mathcal{A}\vdash\mathcal{A}}{\mathcal{A}\vdash\mathcal{A},\mathcal{C}}\,{\scriptstyle \vdash K}}{\mathcal{A}\vdash\mathcal{C},\mathcal{A}}\,{\scriptstyle \vdash C}}{\neg\mathcal{A},\mathcal{A}\vdash\mathcal{C}}\,{\scriptstyle \neg\vdash}}
{\mathcal{C}\vee\neg\mathcal{A},\mathcal{A}\vdash\mathcal{C}}\,{\scriptstyle \vee\vdash}}
{\mathcal{A}\vdash\mathcal{C},\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \vdash\neg}}
{\mathcal{A}\vdash\mathcal{C},\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \vdash\vee}}
{\mathcal{A}\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{C}}\,{\scriptstyle \vdash C}
\qquad \mathcal{B}\vdash\mathcal{B}}
{\mathcal{C}\supset\mathcal{B},\mathcal{A}\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{B}}\,{\scriptstyle \supset\vdash}}
{\mathcal{C}\supset\mathcal{B},\mathcal{A}\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \vdash\vee}}
{\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}),\mathcal{A}\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \wedge\vdash}}
{\mathcal{A},\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B})\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle C\vdash}}
{\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}),\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B})\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}),\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \wedge\vdash}}
{\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}),\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B})\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle \vdash W}}
{\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B})\vdash\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A})}\,{\scriptstyle W\vdash}}
{\vdash(\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}))\supset(\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}))}\,{\scriptstyle \vdash\supset}
$$

The proof is a tree with the *root* being the bottom sequent. There are *three leaves* in this tree, and each leaf is an instance of the axiom with wff's $\mathcal{C}$, $\mathcal{A}$ and $\mathcal{B}$, respectively. This is not the only possible proof of the end sequent $\vdash (\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}))\supset(\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}))$.

Exercise 2.1.9. Take a look at the proof above. Create a couple of other proofs ending in the sequent $\vdash (\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}))\supset(\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}))$. How do the proofs differ from each other?

The nodes in the tree are *occurrences of sequents*, and in this case it so happens that all the occurrences of sequents are occurrences of distinct sequents.

However, this does not need to be so, in general. Here is a simple example.

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A}\vdash\mathcal{A}\qquad \mathcal{A}\vdash\mathcal{A}}{\mathcal{A}\supset\mathcal{A},\mathcal{A}\vdash\mathcal{A}}\;{\scriptstyle\supset\vdash}}{\mathcal{A},\mathcal{A}\supset\mathcal{A}\vdash\mathcal{A}}\;{\scriptstyle C\vdash}}{\mathcal{A}\supset\mathcal{A}\vdash\mathcal{A}\supset\mathcal{A}}\;{\scriptstyle\vdash\supset}}{\vdash(\mathcal{A}\supset\mathcal{A})\supset(\mathcal{A}\supset\mathcal{A})}\;{\scriptstyle\vdash\supset}$$

The two leaves are the same instances of the axiom, and if we would not distinguish between occurrences of sequents, then the tree would have only five nodes. In that situation, the two top lines would look like the following.

$$\dfrac{\mathcal{A}\vdash\mathcal{A}}{\mathcal{A}\supset\mathcal{A},\mathcal{A}\vdash\mathcal{A}}$$

This is *not* an instance of the $(\supset\vdash)$ rule; hence, the lower sequent is unjustified. Having clarified and exemplified when the same sequent appears repeatedly in a proof tree at distinct nodes, now we introduce the *convention* of omitting the "occurrence of" from "occurrence of a sequent," when it is unlikely to cause a confusion.

Exercise 2.1.10. Take a look at the proof of $\vdash(\mathcal{A}\wedge(\mathcal{C}\supset\mathcal{B}))\supset(\mathcal{B}\vee\neg(\mathcal{C}\vee\neg\mathcal{A}))$ above. Suppose that all the occurrences of $\mathcal{A}$ and $\mathcal{B}$ are replaced by $\mathcal{D}$ and that the occurrences of sequents are not distinguished. How would the graph underlying the modified proof tree look? [Hint: Draw the tree corresponding to the proof tree above as an unlabeled graph and modify the graph by collapsing the nodes that stand for the same sequent.]

All the nodes in a proof tree that are not leaves are justified by the rules shown in the *annotation*. As we already noted, the annotation is *not* part of the proof, but it is often useful. Also, once a (purported) proof is given, it is decidable if the tree is indeed a proof. One can imagine how to go through the nodes of a finite tree while verifying that they are justified to be where they are. Annotations are normally intended to be correct, and if they are, then only one rule needs to be checked. In general, it is neither required nor true in all sequent calculi that a sequent in a proof has a *unique* justification.

Exercise* 2.1.11. Determine whether proofs in *LK* always have a unique justification or not. [Hint: Either construct a proof containing a sequent that can be annotated with more than one rule, or sketch an argument that shows that no such proof can be constructed.]

For instance, in the above simple example $\mathcal{A}\supset\mathcal{A}\vdash\mathcal{A}\supset\mathcal{A}$ is an interior node, hence, it must be justified by a rule, and it cannot be justified as an instance of the axiom (though it is an instance of the axiom). Reflecting on this fact, we can see that the same end sequent has a much shorter proof.

Exercise 2.1.12. Prove that the formulas given in (a) and (b) are theorems of *LK*. [Hint: The formula in (a) is an instance of the principal type schema of the combinator S'. That is, not all structural rules will be needed for a proof of the formula. There are various ways to prove the formula in (b). Try to find a proof in which $(C\vdash)$ is the only structural rule used.]

(a) $(\mathcal{A} \supset \mathcal{D}) \supset ((\mathcal{A} \supset (\mathcal{D} \supset (\mathcal{B} \supset \mathcal{D}))) \supset (\mathcal{A} \supset (\mathcal{B} \supset \mathcal{D})))$

(b) $(\neg(\mathcal{A} \wedge \mathcal{B}) \supset (\neg\mathcal{A} \vee \neg\mathcal{B})) \wedge ((\neg\mathcal{A} \vee \neg\mathcal{B}) \supset \neg(\mathcal{A} \wedge \mathcal{B}))$

You might have discovered while trying to prove the two formulas in the previous exercise that it is often helpful to construct a proof (when there is a concrete formula to be proved) from the bottom to the top. Building the proof tree from its root, of course, means that the rules are "applied backward," so to speak. Searching for proofs in a bottom-up fashion can be made precise in the notion of a *proof-search tree*, which however, should not be confused with a *proof tree*. For instance, a proof-search tree may contain no proof at all.

Exercise 2.1.13. Prove the formulas given in (a) and (b). [Hint: The restrictions on some of the quantifier rules mean that occasionally the order in which quantifier rules are applied does matter.]

(a) $\forall x\,(P(x) \supset \forall x\,Q(x,x)) \supset \forall x\,(\exists x\,P(x) \supset Q(x,x))$

(b) $\forall x\,\forall y\,(R(x,y) \supset \forall z\,(R(x,z) \supset R(y,z))) \supset \forall x\,(\exists y\,R(y,x) \supset R(x,x))$

The formula in (b) expresses that if R is *Euclidean*, then it is *end-reflexive* (i.e., a point with an R-predecessor is reflexive). (Such properties are of interest, e.g., in the meta-theory of modal logics.)

Exercise 2.1.14. Determine which of the following formulas are provable. Prove the ones that are provable and explain why the ones that are not provable have no proof. (a) $\forall x\,R(x,x) \supset \forall y\,\forall z\,R(y,z)$, (b) $\forall y\,\forall z\,R(y,z) \supset \forall x\,R(x,x)$, (c) $\exists x\,Q(x,x) \supset \exists y\,\exists z\,Q(y,z)$, (d) $\exists x\,\exists y\,Q(y,x) \supset \exists x\,Q(x,x)$.

Now that we have some sense of how *LK* as a formal system works, let us turn back to exploring Gentzen's original idea that his logistic calculus is a calculus of *logical deduction*.

A system of logic has a notion of proof and deduction associated to it. However, the *properties* of those notions are investigated in the so-called *meta-theory* of the logic. *LK* is intended to be a calculus that allows not simply proofs and derivations, but also reasoning about logical inferences.

Sequents in *LK* have a natural informal interpretation: $\mathcal{A}_1,\ldots,\mathcal{A}_n \vdash \mathcal{B}_1, \ldots, \mathcal{B}_m$ is to be thought of as the inference from *all* the premises $\mathcal{A}_1,\ldots,\mathcal{A}_n$ to *some* of the conclusions $\mathcal{B}_1,\ldots,\mathcal{B}_m$. Using the language of *LK*, we can rephrase this as $\mathcal{A}_1 \wedge \ldots \wedge \mathcal{A}_n$ implies $\mathcal{B}_1 \vee \ldots \vee \mathcal{B}_m$. (To ensure that these expressions are formulas, we may assume that the parentheses are to be restored by association to the left.)

This way of looking at sequents imparts a new interpretation of the rules. Let us start with the three pairs of structural rules.

The left rules, $(K\vdash)$, $(W\vdash)$ and $(C\vdash)$ allow, respectively, the *addition* of a new premise, the *omission* of extra copies of a premise and the *reordering* of the premises. If we think about the premises as joined by $\wedge$, then $(W\vdash)$ and $(C\vdash)$ can be viewed to express half of the *idempotence* of $\wedge$ and its *commutativity*, respectively. The left thinning rule can be seen to express that a conjunction implies its right conjunct. (Sometimes, the wff $(\mathcal{A}\wedge\mathcal{B})\supset\mathcal{B}$ and the wff $(\mathcal{A}\wedge\mathcal{B})\supset\mathcal{A}$ are called *simplification*.)

There is a tendency in everyday reasoning, and even in certain areas of logic, to prefer *single-conclusion* inferences. For instance, syllogisms, no matter if in their ancient or medieval form, contain more than one premise, but only one conclusion. Thus, perhaps, we are less familiar with how sets of conclusions can be manipulated without retracting a conclusion and reexamining the inference. Still, it is not difficult to see that the right structural rules, $(\vdash K)$, $(\vdash W)$ and $(\vdash C)$ are informally justifiable. *Adding* a new conclusion is unproblematic, because if we already have some formulas at least one of which follows, then throwing in one more formula maintains that property, that is, at least one formula still follows. It is even easier to see that *omitting repeated occurrences* of a conclusion or *changing the order* of the conclusions does not affect the correctness of an inference.

Exercise 2.1.15. Prove that $\wedge$ and $\vee$ are idempotent, commutative and associative. Which structural rules are used in the proofs?

To round out our initial look at the rules, we introduce a *cut rule*.

$$\frac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2 \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\ \text{cut}$$

A version of the cut rule (cut_G), which we introduce in Section 2.3.1 on page 30, was included into *LK* in [104]. The cut rule is *admissible* in *LK*, as we prove it in Section 2.3. This rule allows us to *intertwine* two deductions, which is the most sweeping structural change any rule permits.

First, let us consider what happens when Δ_1 and Δ_2 are empty in the sequent $\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2$. Then, in the deduction from $\Theta_1,\mathcal{C},\Theta_2$ to Λ, we can replace $\mathcal{C}$ by Γ, provided that $\mathcal{C}$ follows from Γ, that is, $\Gamma\vdash\mathcal{C}$. This move is very much like how lemmas can be eliminated in favor of their whole proof in mathematical and in real-life reasoning. If the Δ's are not empty, then they must be retained, because we cannot say, on the basis of $\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2$, that $\mathcal{C}$ is *the* consequence of Γ. However, we can still say that if all the wff's in Θ_1,Γ,Θ_2 are available as premises, then at least one of the wff's in $\Delta_1,\Lambda,\Delta_2$ is derivable, when $\Theta_1,\mathcal{C},\Theta_2\vdash\Lambda$.

As it turns out, the most important uses of the cut rule, such as the proof of the *replacement theorem* or the emulation of the *detachment* rule, do not require the full power of the cut rule. However, it is easier to prove a more general form of the cut rule to be admissible.

Lastly, we should reiterate that all sequents are *finite*, that is, they contain finitely many formulas. The informal interpretation turns a sequent into an inference. However, if inferences between infinite sets of formulas are allowed (and they often are), then some of the latter do not have formal equivalents among the sequents in LK.

2.2 An axiom system for FOL

FOL may be axiomatized in more than one way; indeed, it has been axiomatized in many ways.[4] For different purposes, such as to *prove theorems* or to *prove meta-theorems* more easily, various axiom systems are useful.

The language is defined as before, and for the sake of easy translation, we use the same symbols. Then we may assume the identity translation between the formulas that belong to LK and those that belong to K, that is, the axiom system defined below.

Definition 2.11. (Axiom system K) The *axioms* and *rules* of K are (A1)–(A5), (MP) and (UG).

(A1) $(\mathcal{A} \supset (\mathcal{B} \supset \mathcal{A}))$

(A2) $((\mathcal{A} \supset (\mathcal{B} \supset \mathcal{C})) \supset ((\mathcal{A} \supset \mathcal{B}) \supset (\mathcal{A} \supset \mathcal{C})))$

(A3) $((\neg\mathcal{A} \supset \neg\mathcal{B}) \supset ((\neg\mathcal{A} \supset \mathcal{B}) \supset \mathcal{A}))$

(A4) $(\forall x\, \mathcal{A}(x) \supset \mathcal{A}(y))$, where y is OK for substitution for x in $\mathcal{A}(x)$

(A5) $(\forall x\, (\mathcal{A} \supset \mathcal{B}) \supset (\mathcal{A} \supset \forall x\, \mathcal{B}))$, where x is not free in $\mathcal{A}$

(MP) $\mathcal{A}$ and $(\mathcal{A} \supset \mathcal{B})$ imply $\mathcal{B}$

(UG) $\mathcal{A}$ implies $\forall x\, \mathcal{A}$

Axiom (A4) requires that the variable y is OK for x in $\mathcal{A}$—as explained on page 15. The substitution of y is for *all* the free occurrences of x in $\mathcal{A}(x)$, that is, we cannot select a proper subset of the free occurrences of x—unlike we could in the rules of LK.

The language of K contains five other logical constants that have not figured into the axioms and rules. They are viewed as contextually defined by certain formulas.

Definition 2.12. The following *abbreviations* are adopted in K.

(D1) $(\mathcal{A} \vee \mathcal{B})$ for $(\neg\mathcal{A} \supset \mathcal{B})$

---

[4]We present K, which is the axiom system mainly used by Mendelson in [149], where other axiomatizations are presented too. We harmonize the notation with ours.

(D2) $(\mathcal{A} \wedge \mathcal{B})$ for $\neg(\mathcal{A} \supset \neg\mathcal{B})$

(D3) $\exists x\, \mathcal{A}$ for $\neg\forall x\, \neg\mathcal{A}$

(D4) $\boldsymbol{F}$ for $\neg(\mathcal{A} \supset \mathcal{A})$

(D5) $\boldsymbol{T}$ for $(\mathcal{A} \supset \mathcal{A})$

Definition 2.13. (Proofs, theorems) A *proof* is a finite sequence of wff's, in which each formula satisfies at least one of (1)–(3).

(1) The wff is an instance of an axiom;

(2) the wff is obtained from preceding elements of the sequence by an application of the rule (MP);

(3) the wff is obtained from a previous wff in the sequence by an application of the rule (UG).

The last formula in a proof is called a *theorem*. $\vdash \mathcal{A}$ indicates that $\mathcal{A}$ is a theorem of K.

Definition 2.14. (Derivations, consequence) A *derivation* of the wff $\mathcal{A}$ from the set of wff's Γ is a finite sequence of wff's, in which each formula satisfies at least one of (1)–(3).

(1) The wff is a theorem;

(2) the wff is an element of Γ;

(3) the wff is obtained from preceding elements of the sequence by an application of the rule (MP).

The wff $\mathcal{A}$ is a *logical consequence* of the set of wff's Γ, which is denoted by $\Gamma \vdash \mathcal{A}$, when there is a derivation of $\mathcal{A}$ from Γ.

Notice that although both proofs and derivations are *finite* sequences of wff's, the premise set Γ itself does not need to be finite. That is, the consequence relation can hold between an infinite set of wff's and a wff.

Axiom systems have a certain elegance, because often, a few self-evident principles and rules suffice to generate an infinite set of less obvious, or in some sense, more complex theorems. As a way to organize knowledge, axiom systems have had a long and successful history since the 4th century BCE. However, given a wff, even if it is suspected to be a theorem, it may not be easy to find a proof from the axioms; let alone it is straightforward to determine that the wff is not provable using only the axiom system.

For example, in K, an application of the rule (MP) leads to a loss of a subformula, so to speak: $\mathcal{B}$ is obtained from $\mathcal{A} \supset \mathcal{B}$ and $\mathcal{A}$. That is, if we try to construct a proof of $\mathcal{B}$, we have to find an $\mathcal{A}$ such that both the major premise (i.e., the implicational wff) and the minor premise of the detachment rule are theorems. Since the theorems themselves could have been obtained by detachment, and we may not know, to start with, how many steps are needed, the process is challenging. Solving the following exercises may give a flavor of this difficulty.

Exercise 2.2.1. Prove the following wff's in the axiom system K. (a) $(\mathcal{A} \supset \mathcal{A})$, (b) $((\mathcal{C} \supset \mathcal{A}) \supset ((\mathcal{A} \supset \mathcal{B}) \supset (\mathcal{C} \supset \mathcal{B})))$, (c) $((\mathcal{A} \supset \mathcal{B}) \supset ((\mathcal{A} \supset (\mathcal{B} \supset \mathcal{C})) \supset (\mathcal{A} \supset \mathcal{C})))$ and (d) $(((\mathcal{B} \supset \mathcal{C}) \supset \mathcal{B}) \supset \mathcal{B})$. [Hint: The first wff is the principal type schema of the combinator I. The wff's in (b) and (c) are similarly related to B' and S'. The last formula is called Peirce's law.]

Exercise 2.2.2. Prove that the wff's in (a)–(c) are theorems of K, and that the logical consequence stated in (d) obtains. (a) $(\forall x\, \forall y\, \mathcal{A}(x,y) \supset \forall x\, \mathcal{A}(x,x))$, (b) $(\forall x\, \mathcal{A} \supset \forall x\, (\mathcal{A} \vee \mathcal{B}))$, (c) $\exists x\, (\mathcal{A}(x) \supset \forall x\, \mathcal{A}(x))$, (d) $\forall x\, \forall y\, (R(x,y) \supset \forall z\, (R(y,z) \supset R(x,z)))$, $\forall x\, \forall y\, (R(y,x) \supset R(x,y)) \vdash \forall x\, \forall y\, (R(x,y) \supset R(x,x))$.

Exercise* **2.2.3.** Prove that the next wff's are theorems of K. (a) $(\exists x\, (\mathcal{A}(x) \supset \mathcal{B}(x)) \supset (\forall x\, \mathcal{A}(x) \supset \exists x\, \mathcal{B}(x)))$, (b) $(\exists x\, (\mathcal{A}(x) \wedge \mathcal{B}(x)) \supset \exists x\, \mathcal{A}(x))$, (c) $(\exists x\, \forall y\, R(x,y) \supset \forall x\, \exists y\, R(y,x))$.

"Axiom-chopping," as it is sometimes called, helps to develop an understanding of the interactions and relationships between the axioms, rules and theorems. Notwithstanding, our focus is on sequent calculi, another type of proof systems, in this book. Notably, they are much better suited to searching for proofs, because they refine the notion of proofs themselves. There are classic texts that contain excellent examples of how to build up a set of useful theorems from an axiom system.[5] We will freely assume, if we need to, that wff's, which are known to be theorems of FOL, are provable in K. We leave it to the interested reader to find their proofs from the axioms.

2.3 Equivalence of *LK* and *K*

The two proof systems, K and LK, are equivalent, which means that if $\mathcal{A}$ is a theorem of K, then the sequent $\vdash \mathcal{A}$ is provable in LK, and vice versa. It is straightforward to prove that every axiom of K is provable in LK. As an example, we give a proof of axiom (A1).

$$\dfrac{\dfrac{\dfrac{\mathcal{A} \vdash \mathcal{A}}{\mathcal{B}, \mathcal{A} \vdash \mathcal{A}}\; K\vdash}{\mathcal{A} \vdash \mathcal{B} \supset \mathcal{A}}\; \vdash\supset}{\vdash \mathcal{A} \supset (\mathcal{B} \supset \mathcal{A})}\; \vdash\supset$$

[5] See, for example, Church [76] and Kleene [127] as well as Mendelson [149].

Exercise 2.3.1. Construct proofs in LK of axioms (A2) $(\mathcal{A} \supset (\mathcal{B} \supset \mathcal{C})) \supset ((\mathcal{A} \supset \mathcal{B}) \supset (\mathcal{A} \supset \mathcal{C}))$ and (A3) $(\neg\mathcal{A} \supset \neg\mathcal{B}) \supset ((\neg\mathcal{A} \supset \mathcal{B}) \supset \mathcal{A})$.

Exercise 2.3.2. Prove that the quantificational axioms (A4) and (A5) of K are theorems of LK. [Hint: See Definition 2.11 on page 23.]

K did not contain all the logical constants of LK as primitives. Therefore, we have to prove that certain wff's that are counterparts of the Definitions (D1)–(D3) in K are provable in LK. Namely, we have to prove the following six wff's: (1) $(\mathcal{A} \vee \mathcal{B}) \supset (\neg\mathcal{A} \supset \mathcal{B})$, (2) $(\neg\mathcal{A} \supset \mathcal{B}) \supset (\mathcal{A} \vee \mathcal{B})$, (3) $(\mathcal{A} \wedge \mathcal{B}) \supset \neg(\mathcal{A} \supset \neg\mathcal{B})$, (4) $\neg(\mathcal{A} \supset \neg\mathcal{B}) \supset (\mathcal{A} \wedge \mathcal{B})$, (5) $\exists x\, \mathcal{A} \supset \neg\forall x \neg\mathcal{A}$ and (6) $\neg\forall x \neg\mathcal{A} \supset \exists x\, \mathcal{A}$. We prove (4) and leave the other formulas for an exercise.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A} \vdash \mathcal{A}}{\mathcal{A} \vdash \mathcal{A}, \neg\mathcal{B}}\,{\scriptstyle \vdash K}}{\vdash \mathcal{A}, \mathcal{A} \supset \neg\mathcal{B}}\,{\scriptstyle \vdash\supset}}{\neg(\mathcal{A} \supset \neg\mathcal{B}) \vdash \mathcal{A}}\,{\scriptstyle \neg\vdash} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{B} \vdash \mathcal{B}}{\vdash \mathcal{B}, \neg\mathcal{B}}\,{\scriptstyle \vdash\neg}}{\mathcal{A} \vdash \mathcal{B}, \neg\mathcal{B}}\,{\scriptstyle K\vdash}}{\vdash \mathcal{B}, \mathcal{A} \supset \neg\mathcal{B}}\,{\scriptstyle \vdash\supset}}{\neg(\mathcal{A} \supset \neg\mathcal{B}) \vdash \mathcal{B}}\,{\scriptstyle \neg\vdash}}{\neg(\mathcal{A} \supset \neg\mathcal{B}) \vdash \mathcal{A} \wedge \mathcal{B}}\,{\scriptstyle \vdash\wedge}}{\vdash \neg(\mathcal{A} \supset \neg\mathcal{B}) \supset (\mathcal{A} \wedge \mathcal{B})}\,{\scriptstyle \vdash\supset}$$

Exercise 2.3.3. Prove that the wff's listed above as (1), (2), (3), (5) and (6) are also theorems of LK.

The rules of K are detachment, (i.e., (MP)) and universal generalization, (i.e., (UG)). We have to show that if the premises of these rules are provable in LK, then so are their respective conclusions.

For modus ponens, we may assume that there are some proof trees rooted in $\mathcal{A}$ and $\mathcal{A} \supset \mathcal{B}$. We would like to be able to have the following proof tree.

$$\dfrac{\begin{matrix}\vdots \\ \vdash \mathcal{A}\end{matrix} \qquad \begin{matrix}\vdots \\ \vdash \mathcal{A} \supset \mathcal{B}\end{matrix}}{\vdash \mathcal{B}}\ ?$$

The ? indicates that there is no rule (or combination of rules) in LK that we could apply to the premises to obtain the lower sequent. (The horizontal line is thicker to suggest the possibility that several rules had been applied.)

Exercise 2.3.4. The claim we just made may or may not be obvious. Convince yourself that chaining together several rules of LK will not help.

We could add the above pattern as a rule to LK, and then we could try to show that it does not increase the set of provable sequents. Instead, we add the cut rule mentioned on page 22, which is useful for other purposes too.[6]

Sometimes, this rule is called the *single cut* rule, because the premises are only required to contain an occurrence of $\mathcal{C}$, the *cut formula*, and exactly one

[6] This formulation of the cut rule for LK from [39] generalizes the one in [106].

of all the occurrences of $\mathcal{C}$ is affected in each premise. The left premise of the rule has to have an occurrence of $\mathcal{C}$ in the succedent, whereas the right premise has to have an occurrence of $\mathcal{C}$ in the antecedent for the rule to be applicable. ($\mathcal{C}$ may have further occurrences, but those will be ignored from the point of view of an application of this rule; they are merely components in the Δ's and Θ's.) The occurrences of $\mathcal{C}$, which are singled out, are replaced by Γ and Λ, respectively, each of which might comprise several formulas.

If we have the cut rule, then we can show that modus ponens can be emulated in the following way.

$$\dfrac{\begin{array}{c}\vdots\\ \vdash \mathcal{A}\end{array} \qquad \dfrac{\begin{array}{c}\vdots\\ \vdash \mathcal{A} \supset \mathcal{B}\end{array} \qquad \dfrac{\mathcal{A} \vdash \mathcal{A} \qquad \mathcal{B} \vdash \mathcal{B}}{\mathcal{A} \supset \mathcal{B}, \mathcal{A} \vdash \mathcal{B}} \supset\vdash}{\mathcal{A} \vdash \mathcal{B}}\ \text{cut}}{\vdash \mathcal{B}}\ \text{cut}$$

This chunk of the proof shows that if cut would be included in LK, then (MP) would be a *derived* rule. Notice that the first application of the cut rule suggests more uses for the cut rule beyond emulating detachment. If the formula $\mathcal{A} \supset \mathcal{B}$ is a theorem, then the corresponding sequent $\vdash \mathcal{A} \supset \mathcal{B}$ could be obtained only by one of the following two rules: $(\vdash K)$ and $(\vdash \supset)$. We could exclude the first, if we would know that $\vdash$ is not a provable sequent. Notice that there is a huge difference between this sequent and the sequent $\mathcal{A} \vdash \mathcal{A}$, which is the axiom of LK. If the sequent with empty antecedent and empty succedent would be provable, then using the $(\vdash K)$ rule *all wff's* would be theorems. This would be catastrophic for the whole calculus. We will show that LK is consistent, hence, this sequent is not provable. Then the last rule must have been $(\vdash \supset)$. An application of the cut allows us "to take back" this step.

Cut is a very powerful rule, and we look at it in connection to LK in the next section, where we also prove its *admissibility* in LK. Later, in Chapter 7, we consider further versions and other useful consequences of this rule. For now, we only record that the single cut rule suffices to mimic in LK the (MP) rule of K.

The other rule of K is (UG). We have to justify the following step in a proof.

$$\dfrac{\begin{array}{c}\vdots\\ \vdash \mathcal{A}(x)\end{array}}{\vdash \forall x\, \mathcal{A}(x)}\ ?$$

It might seem that we can simply say that we have the rule that we need in place of ?, namely, $(\vdash \forall)$. However, this rule has a side condition, which requires that there are no free occurrences of y anywhere in the lower sequent, which means that the formula $\forall x\, \mathcal{A}(x)$ may not contain such occurrences of y. However, in (UG), the y is x, and $\forall x$ binds all free occurrences of x. We note that x is OK for x in any wff, and substituting x in $\mathcal{A}(x)$ for all free occurrences of x yields $\mathcal{A}(x)$. Lastly, the $(\vdash \forall)$ rule does not require that y and x are distinct. In sum, ? can be replaced by $(\vdash \forall)$, indeed.

In the context of axiomatic calculi, it is usual to prove theorems that license the *renaming of bound variables*.[7] In the context of sequent calculi, the *renaming of free variables*, like y's in and above the $(\vdash\forall)$ and $(\exists\vdash)$ rules, is essential.

Lemma 2.15. *A proof of the sequent* $\Gamma \vdash \Delta$ *may be transformed into a proof of the same end sequent in which* y *in the upper sequent of an application of* $(\exists\vdash)$ *or* $(\vdash\forall)$ *occurs* only in the subtree *rooted in the upper sequent.*

Proof: There are finitely many occurrences of applications of these two rules in any proof, hence, iterating the following step proves the lemma. We select an application of the two rules that either has no application of $(\exists\vdash)$ or $(\vdash\forall)$ above it, or if it has, then the variables of all those rules have already been renamed. We replace all the occurrences of y in the subtree rooted in the upper sequent by z, which occurs nowhere in the proof. The process yields a proof from a proof by the following lemma. qed

Lemma 2.16. *If a* free variable, *which is not the* y *in* $(\exists\vdash)$ *or in* $(\vdash\forall)$ *is* replaced *everywhere in the axiom or in a rule by a variable, which is not the* y *of* $(\exists\vdash)$ *or* $(\vdash\forall)$, *then the result is an instance of the axiom or of the same inference rule, respectively.*

Proof: For most of the rules, the claim is obvious, since nothing depends on the concrete shape of the $\mathcal{A}$'s and $\mathcal{B}$'s. However, the quantifier rules are either not restricted, or by the condition of the lemma, the variable entering the restriction is not renamed. qed

This lemma also shows that we could rename the x in $\mathcal{A}(x)$ in the LK proof imitating (UG), if we would not want to have x in the upper sequent.

Theorem 2.17. (From K to LK) *If* $\mathcal{A}$ *is a theorem of* K, *then* $\mathcal{A}$ *is a* theorem *of* LK.

Proof: We only have to combine the components that we already have. K's axioms are theorems of LK, and K's rules can be simulated in LK. The connectives that were defined in K behave appropriately in LK. qed

To view the axiom system and the sequent calculus as equivalent, we need to show that the converse of the if–then statement in the theorem is true too. The interpretation of sequents, which we introduced somewhat informally earlier, comes handy; we make that precise now.

Definition 2.18. For $\Gamma \vdash \Delta$, we define $\tau(\Gamma \vdash \Delta)$ to be $\bigwedge(\Gamma) \supset \bigvee(\Delta)$, where $\bigwedge$ and $\bigvee$ are as follows.

(1) $\bigwedge(\ ) := (\mathcal{A} \supset \mathcal{A})$

(2) $\bigwedge(\mathcal{A}) := \mathcal{A}$

[7] The ancillary role of bound variables is explained by Schönfinkel [183].

(3) $\bigwedge(\mathcal{A}_{n+1}, \mathcal{A}_n, \ldots, \mathcal{A}_1) := (\mathcal{A}_{n+1} \wedge \bigwedge(\mathcal{A}_n, \ldots, \mathcal{A}_1))$

(4) $\bigvee(\ ) := \neg(\mathcal{A} \supset \mathcal{A})$

(5) $\bigvee(\mathcal{B}) := \mathcal{B}$

(6) $\bigvee(\mathcal{B}_1, \ldots, \mathcal{B}_m, \mathcal{B}_{m+1}) := (\bigvee(\mathcal{B}_1, \ldots, \mathcal{B}_m) \vee \mathcal{B}_{m+1})$

Obviously, $\tau(\Gamma \vdash \Delta)$ is a wff, and $\bigwedge(\Gamma)$ is a conjunction when Γ is not empty, whereas $\bigvee(\Delta)$ is a disjunction when Δ is not empty.

The connectives $\wedge$ and $\vee$ are defined in K, and they may be proved *associative*. This means that we could omit all but the outside parentheses obtained from the wff's $\bigwedge(\Gamma)$ and $\bigvee(\Delta)$. The wff's in (1) and (4) could have an arbitrary $\mathcal{A}$ in it, or $\mathcal{A}$ could be a fixed wff. Alternatively, the empty antecedent of a sequent could be thought of as $\boldsymbol{T}$ and the empty succedent could be taken to be $\boldsymbol{F}$, in accordance with (D4) and (D5).

Theorem 2.19. (From LK **to** K**)** *If* $\mathcal{A}$ *is a theorem of* LK*, then* $\mathcal{A}$ *is a* theorem *of* K*.*

Proof: What we show is that τ takes the axiom of LK into a theorem, and further, if the upper sequent(s) of a rule are theorems, then so is the lower sequent. (The proof is by induction on the height of the proof tree, and we include here only a small selection of the cases.)

1. $\tau(\mathcal{A} \vdash \mathcal{A})$ is $\mathcal{A} \supset \mathcal{A}$, which is a theorem of K.

2.1. If the last rule is $(\wedge\vdash_1)$ (or $(\wedge\vdash_2)$), then $\tau(\mathcal{A} \wedge \mathcal{B}, \Gamma \vdash \Delta)$ should follow from $\tau(\mathcal{A}, \Gamma \vdash \Delta)$ (or from $\tau(\mathcal{B}, \Gamma \vdash \Delta)$). The formula $((\mathcal{A} \wedge \mathcal{C}) \supset \mathcal{D}) \supset ((\mathcal{A} \wedge \mathcal{B} \wedge \mathcal{C}) \supset \mathcal{D})$ (or $((\mathcal{B} \wedge \mathcal{C}) \supset \mathcal{D}) \supset ((\mathcal{A} \wedge \mathcal{B} \wedge \mathcal{C}) \supset \mathcal{D})$) is a theorem of K; hence, if its antecedent is provable, then so is its consequent. (Finding proofs of this and other formulas in K is Exercise 2.3.6.)

2.2. Let us assume that the last rule applied is $(\vdash\neg)$. Then $\tau(\mathcal{A}, \Gamma \vdash \Delta)$ has to imply $\tau(\Gamma \vdash \Delta, \neg\mathcal{A})$. The wff $((\mathcal{A} \wedge \mathcal{C}) \supset \mathcal{D}) \supset (\mathcal{C} \supset (\mathcal{D} \vee \neg\mathcal{A}))$ is a theorem of K, thus, if $\tau(\mathcal{A}, \Gamma \vdash \Delta)$ is a theorem of K, so is $\tau(\Gamma \vdash \Delta, \neg\mathcal{A})$, as needed.

2.3. If the last rule is $(\forall\vdash)$, then the formula that should be shown to be a theorem of K is $((\mathcal{A}(y) \wedge \mathcal{C}) \supset \mathcal{D}) \supset ((\forall x\, \mathcal{A}(x) \wedge \mathcal{C}) \supset \mathcal{D})$. The antecedent of this wff is $\tau(\mathcal{A}(y), \Gamma \vdash \Delta)$, whereas the consequent is $\tau(\forall x\, \mathcal{A}(x), \Gamma \vdash \Delta)$. Again, the hypothesis of the induction together with the theorem of K give the desired conclusion.

2.4. Let us consider $(\vdash\forall)$. $\tau(\Gamma \vdash \Delta, \mathcal{A}(y))$ is $\mathcal{C} \supset (\mathcal{D} \vee \mathcal{A}(y))$ and $\tau(\Gamma \vdash \Delta, \forall x\, \mathcal{A}(x))$ is $\mathcal{C} \supset (\mathcal{D} \vee \forall x\, \mathcal{A}(x))$. There is no reason why the first formula would imply the second, but the $(\vdash\forall)$ rule comes with a restriction that saves the implication. y cannot occur free in the second formula at all, hence, it certainly cannot occur free in $\mathcal{C}$ or $\mathcal{D}$, the antecedent or the disjunct in the consequent of the wff. With this restriction in place, $\forall y\, (\mathcal{C} \supset (\mathcal{D} \vee \mathcal{A}(y))) \supset (\mathcal{C} \supset (\mathcal{D} \vee \forall x\, \mathcal{A}(x)))$ is a theorem of K.

3. The theoremhood of $\mathcal{B}$ in LK is defined as the provability of $\vdash \mathcal{B}$. τ gives $(\mathcal{A} \supset \mathcal{A}) \supset \mathcal{B}$, from which we immediately get $\mathcal{B}$, by detachment, because $\mathcal{A} \supset \mathcal{A}$ is a theorem of K. qed

Exercise 2.3.5. Complete the proof of the previous theorem. [Hint: You may assume that the wff's that are necessary for the proof are provable in K.]

Exercise* 2.3.6. Prove in K the wff's that have to be shown to be theorems in order for the above proof to work. [Hint: Remember that proving theorems in an axiom system sometimes can get tricky and lengthy.]

2.3.1 Cut rules

We used the single cut rule in the previous section with the promise of showing it admissible later. A cut rule was introduced by Gentzen in [104], in fact, he included his cut into LK as a structural rule. To be more precise, the cut rule in [104] is not exactly the same rule that we gave on page 22. Gentzen's cut rule is cut_G:

$$\frac{\Gamma \vdash \Delta, \mathcal{C} \qquad \mathcal{C}, \Theta \vdash \Lambda}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut}_\text{G}$$

$\mathcal{C}$ must have an occurrence in the premises of cut_G, just as we stated for the cut rule. And, only a single occurrence of $\mathcal{C}$ is omitted from $\Delta, \mathcal{C}$ and $\mathcal{C}, \Theta$.

The left and right permutation rules—together with the absence of grouping—guarantee that wff's can be "moved around" both in the antecedent and in the succedent of a sequent. Thus it may appear at first that it is absolutely unimportant where a formula, for example, the cut formula is placed within the antecedent and the succedent. Gentzen seems to have preferred the edges; all the formulas affected by his rules—except those in $(C \vdash)$ and $(\vdash C)$—are either first in the antecedent or last in the succedent. (We defined τ accordingly so that τ yields formulas with a transparent structure.) In the case of the cut rule, however, such a positioning of the cut formulas is clearly problematic, because we want to prove the admissibility of the cut. Gentzen missed this observation, though he obviously realized that the permutation rules must have a more general form than the other rules do. An *insight* that one gains from applications of the single cut rule and from proofs of the admissibility of cut in sequent calculi for non-classical logics is that the *appropriate form* of the single cut rule is the one we gave earlier (and not cut_G). Working with consecution calculi, where the structural connective is not commutative, is especially helpful to reach this view.

Gentzen introduced LK, more or less, starting from scratch. At the same time, when we acknowledge the novelty and the cleverness of his sequent calculus, it may be worthwhile to mention some other problematic features caused by the particular shape of his rules.

Pertaining to the edge positioning of formulas is a peculiarity that became well understood only after the work of Curry [82] (that was later reinforced by Howard [124]). Consider the following proof.

$$\dfrac{\dfrac{\dfrac{\mathcal{A} \vdash \mathcal{A} \qquad \mathcal{B} \vdash \mathcal{B}}{\mathcal{A} \supset \mathcal{B}, \mathcal{A} \vdash \mathcal{B}} \supset\vdash}{\mathcal{A} \vdash (\mathcal{A} \supset \mathcal{B}) \supset \mathcal{B}} \vdash\supset}{\vdash \mathcal{A} \supset ((\mathcal{A} \supset \mathcal{B}) \supset \mathcal{B})} \vdash\supset$$

The formula $\mathcal{A} \supset ((\mathcal{A} \supset \mathcal{B}) \supset \mathcal{B})$ is a theorem (and a valid wff) of classical logic; it is called *assertion*. However, it is not an instance of self-implication $(\mathcal{A} \supset \mathcal{A})$, which is provable from the axiom of LK by one application of the $(\vdash\supset)$ rule. Assertion is the principal type schema of the combinator CI (also denoted by T), with axiom $\mathsf{CI}xy \rhd yx$.[8] Obviously, some permutation is happening in the combinatory axiom, whereas the $(\mathsf{C}\vdash)$ rule has not been applied in the proof displayed above. In other words, the $(\supset\vdash)$ and $(\vdash\supset)$ rules are *mismatched*, when we look beyond mere provability of wff's in classical or intuitionist logic.

Another complaint is that Gentzen considered cut to be a structural rule. It turns out that cut is a special and important rule, but has very different properties than the six other structural rules, which allow the manipulation of the formulas in the antecedent or in the succedent of *one* sequent. Cut combines *two* sequents—while it drops an occurrence of a formula from each.

The appropriately formulated cut rule can be shown to be admissible by a straightforward induction. Gentzen, on the other hand, had to devise a roundabout proof to show the admissibility (really, from his point of view, the *eliminability*) of cut_G. The *mix rule* can be shown to be equivalent to cut_G in LK (but not in some other sequent calculi).

$$\dfrac{\Gamma \vdash \Delta \qquad \Theta \vdash \Lambda}{\Gamma, \Theta^* \vdash \Delta^*, \Lambda} \text{ mix}$$

The rule is applicable if there is a wff $\mathcal{C}$ such that it occurs both in Δ and Θ. Δ^* and Θ^* stand for Δ and Θ, respectively, but with *all the occurrences of* $\mathcal{C}$ omitted.

It is also interesting that this rule is formulated somewhat informally—there is no notation introduced to *locate* the occurrences of $\mathcal{C}$. The rule even may appear not to require the existence of a suitable $\mathcal{C}$ (though it does). Mix is decidedly a strange rule. Instead of the idea of locating one occurrence of $\mathcal{C}$ within Δ and Θ, Gentzen opted for replacing *all* occurrences of $\mathcal{C}$ with *one copy* of each of Γ and Λ. Mix is equivalent to cut_G within LK (and also in LJ), but this equivalence is peculiar to classical (and to intuitionist) logic.

Lemma 2.20. *The* cut_G *rule and the mix rule are* equivalent, *that is, a sequent is provable in* LK *using cut iff it is provable in* LK *using mix.*

[8] We will say more about combinators and types in Chapters 5 and 9.

Proof: Let us assume that $\Gamma,\Theta\vdash\Delta,\Lambda$ is by cut_G from the premises $\Gamma\vdash\Delta,\mathcal{C}$ and $\mathcal{C},\Theta\vdash\Lambda$. If $\Delta,\mathcal{C}$ and $\mathcal{C},\Theta$ have no other occurrences of $\mathcal{C}$, then an application of mix using the same premises yields the same lower sequent. In general, let there be n occurrences of $\mathcal{C}$ in the succedent of the left premise, and m occurrences of $\mathcal{C}$ in the antecedent of the right premise (where $n,m\geq 1$). Then the following proof segment yields the sequent $\Gamma,\Theta\vdash\Delta,\Lambda$. (The thicker horizontal lines indicate that the sequents above and below may be separated by more than one step.) If $m=1$ or $n=1$, then the thinning and permutation steps are altogether omitted.

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Gamma\vdash\Delta,\mathcal{C}\end{array}\qquad\begin{array}{c}\vdots\\ \mathcal{C},\Theta\vdash\Lambda\end{array}}{\Gamma,\Theta^*\vdash\Delta^*,\Lambda}\ \text{mix}}{\Gamma,\Theta\vdash\Delta^*,\Lambda}\ K\vdash(m{-}1\times),\ C\vdash\text{'s}}{\Gamma,\Theta\vdash\Delta,\Lambda}\ \vdash K\ (n{-}1\times),\ \vdash C\text{'s}$$

We cannot specify the exact numbers of the applications of the $(C\vdash)$ and $(\vdash C)$ rules, which are necessary to restore the sequence of formulas Γ,Θ^* to Γ,Θ (and similarly, to derive Δ,Λ from Δ^*,Λ). (The numbers of applications of the thinning rules are determined by the numbers of occurrences of $\mathcal{C}$ in Θ and Δ, respectively.) However, the size of Γ,Θ^* and that of Δ^*,Λ clearly induce an upper bound. If the numbers of wff's in Γ,Θ^* and in Δ^*,Λ are i and j, respectively, then the bounds are $i\cdot(m-1)$ and $j\cdot(n-1)$.

Now let cut_G be given, and let us assume that in the premises $\Gamma\vdash\Delta$ and $\Theta\vdash\Lambda$, there are n and m occurrences of $\mathcal{C}$ (with $n,m\geq 1$). If $m=n=1$, then one application of cut_G yields the same sequent—provided that the cut formulas are in the right positions in the premises. Finitely many applications of $(\vdash C)$ to the left premise and finitely many applications of $(C\vdash)$ to the right premise can ensure that the $\mathcal{C}$'s are on the edges.

Let us consider the more complicated situation when $m\neq 1$ or $n\neq 1$.

$$\dfrac{\vdash W\text{'s }(m{-}1\times)\ \dfrac{\vdash C\text{'s}\ \dfrac{\begin{array}{c}\vdots\\ \Gamma\vdash\Delta\end{array}}{\Gamma\vdash\Delta^*,\mathcal{C}^m}}{\Gamma\vdash\Delta^*,\mathcal{C}}\qquad\dfrac{\dfrac{\begin{array}{c}\vdots\\ \Theta\vdash\Lambda\end{array}}{\mathcal{C}^n,\Theta^*\vdash\Lambda}\ C\vdash\text{'s}}{\mathcal{C},\Theta^*\vdash\Lambda}\ W\vdash\text{'s }(n{-}1\times)}{\Gamma,\Theta^*\vdash\Delta^*,\Lambda}\ \mathrm{cut}_G$$

Applications of the permutation rules at the top may or may not be needed, depending on the concrete shape of Δ and Θ. If $m=1$ or $n=1$, then no contraction is applied on the left or on the right branch. qed

The proof shows not only how powerful the mix rule is, but also that the equivalence of this rule and the cut_G rule relies on *all* the structural rules.

Exercise 2.3.7. Consider the second half of the proof above. One can postpone the applications of the contraction rules. What does the new proof segment look like? [Hint: You may want to work out a small concrete example first, for instance, with the premises $\mathcal{A}\vdash\mathcal{C},\mathcal{C},\mathcal{B}$ and $\mathcal{D},\mathcal{C},\mathcal{C},\mathcal{C}\vdash\mathcal{E}$.]

There are other parts of LK that could be (and later on, were) sharpened. The axiom $\mathcal{A} \vdash \mathcal{A}$ could be stated in the form $p \vdash p$, where p is an atomic formula (that is, a propositional variable or a predicate followed by sufficiently many terms). There is a proof of the sequent $\mathcal{A} \supset \mathcal{A} \vdash \mathcal{A} \supset \mathcal{A}$ on page 20, which is an instance of the original axiom. If we restrict the axiom to atomic formulas, then it is natural to ask if the modified system (let us say, LK_{at}) can prove all the theorems that LK can.

Exercise 2.3.8. Either prove that LK and LK_{at} have the same set of theorems, or give an example, i.e., a wff, that is provable in one but not in the other system. [Hint: The inclusion is obvious in one direction.]

As the last remark concerning certain peculiarities of LK, let us note a discrepancy in the formulation of the two-premise rules. The $(\vdash\wedge)$ and $(\vee\vdash)$ rules assume that the two premises are identical except $\mathcal{A}$ and $\mathcal{B}$. On the other hand, the $(\supset\vdash)$ rule does not prescribe that Γ and Θ, or Δ and Λ are the same. The rule could have been formulated instead as follows.

$$\frac{\Gamma \vdash \Delta, \mathcal{A} \qquad \mathcal{B}, \Gamma \vdash \Delta}{\Gamma, \mathcal{A} \supset \mathcal{B} \vdash \Delta} \supset\vdash_{\text{e}}$$

We switched $\mathcal{A} \supset \mathcal{B}$ to the other edge in the antecedent (in view of our previous complaint related to typing), but the main difference is that now $\mathcal{A}$ and $\mathcal{B}$ must be proved in the same sequents, though on different sides of the turnstile. The subscript $_{\text{e}}$ is to indicate that this is the genuinely *extensional* version of the $(\supset\vdash)$ rule. The distinction between intensional and extensional connectives, which are otherwise alike or similar, will be explained in more detail in Chapter 5. (The $(\supset\vdash)$ rule is due to O. Ketonen.)

Classical logic cannot distinguish between the connectives that are introduced on the left by the two rules $(\supset\vdash)$ and $(\supset\vdash_{\text{e}})$. But Gentzen defined LK with an eye toward LJ (which we will look at in Chapter 3).[9] The rule $(\supset\vdash)$ has an advantage over $(\supset\vdash_{\text{e}})$, if one wants to define LJ by simply restricting the number of wff's on the right-hand side of the $\vdash$ to at most one.

Exercise 2.3.9. Show that $(\supset\vdash_{\text{e}})$ is a derivable rule in LK. Conversely, prove that if the left introduction rule for $\supset$ is $(\supset\vdash_{\text{e}})$, then $(\supset\vdash)$ is derivable (when the rest of LK is kept unchanged).

We listed some features of the original formulation of LK that later turned out to be puzzling, undesirable or suboptimal. To further motivate our formulation of the cut rule, let us consider the sequent calculus again as a system to reason about inferences. The right premise of the cut rule says that Λ can be derived from $\Theta_1, \mathcal{C}, \Theta_2$. If from Γ the formula $\mathcal{C}$ is derivable within Δ_1 and Δ_2 (as given by the left premise), then placing Γ *in the spot* where $\mathcal{C}$ is (that is, starting with $\Theta_1, \Gamma, \Theta_2$), should suffice for the derivation of Λ within Δ_1 and Δ_2 (that is, $\Delta_1, \Lambda, \Delta_2$).

[9]See [168] for more historical considerations.

The two cut rules, cut and cut_G are, obviously, equivalent in LK, and from now on, we always mean our single cut rule (rather than cut_G). There are other versions of the cut rule such as multi-cut (cf. p. 145). The latter version of cut has its own problems with respect to proofs of admissibility.[10] We think that proving the cut rule, which is a single cut rule, admissible by a direct proof—as below—is *superior* to other approaches. In order to prove the admissibility of the cut rule, we introduce a series of new concepts. Some of them characterize the *roles* various formulas play in a rule; others pertain to proofs in which there is an application of the cut rule.

Definition 2.21. (Degree of a wff) The *degree* of a formula $\mathcal{A}$ is denoted by $\delta(\mathcal{A})$, and it is defined inductively by (1)–(7).

(1) $\delta(p) = 0$, where p is an atomic formula;

(2) $\delta(\neg\mathcal{A}) = \delta(\mathcal{A}) + 1$;

(3) $\delta(\mathcal{A} \wedge \mathcal{B}) = \max(\delta(\mathcal{A}), \delta(\mathcal{B})) + 1$;

(4) $\delta(\mathcal{A} \vee \mathcal{B}) = \max(\delta(\mathcal{A}), \delta(\mathcal{B})) + 1$;

(5) $\delta(\mathcal{A} \supset \mathcal{B}) = \max(\delta(\mathcal{A}), \delta(\mathcal{B})) + 1$;

(6) $\delta(\forall x\, \mathcal{A}) = \delta(\mathcal{A}) + 1$;

(7) $\delta(\exists x\, \mathcal{A}) = \delta(\mathcal{A}) + 1$.

The degree of a formula is a natural number assigned to the formula, which indicates the complexity of the formula. (This is not the only possible way though to describe the complexity of a formula by a natural number.) Formulas can be represented as trees with atomic formula occurrences labeling the leaves and the intermediate nodes labeled by occurrences of connectives or quantifiers. Then the degree of a formula corresponds to the height of the formation tree of the formula minus one.

Next we provide an *analysis* (in the sense of Curry [81]) for the rules of the calculus. Clearly, some formulas are simply copied from the upper sequent(s) to the lower sequent, whereas others are combined, modified, or at least, moved around. We introduce three categories of formulas: *principal formulas*, *subalterns* and *parametric formulas*.

Informally, the categories capture certain types of behaviors of formulas that we describe briefly. A principal formula is a formula in the lower sequent of a rule, which is in the center of attention or it is the formula that is introduced by the rule. A subaltern is a formula in an upper sequent in a rule that is intimately connected to the principal formula, most frequently, because it is a proper subformula of the principal formula, but sometimes it is an occurrence of the same formula as the principal formula. A parametric formula is either in the lower or in the upper sequent, but it is neither a

[10] See [8, §61] where there is a work-around, and [141] where difficulties remain.

principal formula nor a subaltern. Parametric formulas are in the sequents, because the sequent calculus is formulated to reason about consequences, which may involve more than two or three wff's; hence, some formulas are not in the spotlight, so to speak, in some proof steps.

Definition 2.22. (Analysis) We consider each rule, and identify the *principal* formulas and the *subalterns*. All the other formulas are *parametric*.

$(\wedge\vdash_1)$, $(\wedge\vdash_2)$ and $(\vdash\wedge)$. The principal formula is $\mathcal{A}\wedge\mathcal{B}$. There is one subaltern in the left introduction rules, depending on whether $\mathcal{A}$ or $\mathcal{B}$ occurs at the edge of the upper sequent. In $(\vdash\wedge)$, both $\mathcal{A}$ and $\mathcal{B}$ are subalterns.

$(\vee\vdash)$, $(\vdash\vee_1)$ and $(\vdash\vee_2)$. The principal formula is $\mathcal{A}\vee\mathcal{B}$. $\mathcal{A}$ and $\mathcal{B}$ are subalterns in $(\vee\vdash)$, and whichever of them occurs at the far right in the succedent is the subaltern in $(\vdash\vee_1)$ and $(\vdash\vee_2)$.

$(\neg\vdash)$ and $(\vdash\neg)$. The principal formula is $\neg\mathcal{A}$, and the subaltern is the displayed $\mathcal{A}$ in both rules.

$(\supset\vdash)$ and $(\vdash\supset)$. The principal formula is $\mathcal{A}\supset\mathcal{B}$ with the subalterns being the two immediate proper subformulas, $\mathcal{A}$ and $\mathcal{B}$, which are displayed in the rules.

$(\forall\vdash)$ and $(\vdash\forall)$. The principal formula is $\forall x\,\mathcal{A}(x)$, whereas the subaltern is the $\mathcal{A}(y)$ at the edge.

$(\exists\vdash)$ and $(\vdash\exists)$. The principal formula is $\exists x\,\mathcal{A}(x)$, and the subaltern is the displayed $\mathcal{A}(y)$.

$(K\vdash)$ and $(\vdash K)$. The principal formula is the newly introduced $\mathcal{A}$. These rules are exceptional in the sense that there is no subaltern in either of them.

$(W\vdash)$ and $(\vdash W)$. The principal formula is $\mathcal{A}$ in the lower sequent. The two displayed occurrences of the same formula in the upper sequent are the subalterns.

$(C\vdash)$ and $(\vdash C)$. The principal formulas are the $\mathcal{B}$ and $\mathcal{A}$ in the lower sequent, and the subalterns are $\mathcal{A}$ and $\mathcal{B}$ in the upper sequent.

The descriptions should make it clear, but it is perhaps useful to emphasize that we label as principal formulas or subalterns only the formula occurrences that are explicit in the formulation of the rules. From the point of view of easy identification of subalterns and principal formulas, it is helpful that the formulas affected by the rules are at the edge of the antecedent or succedent—except in the permutation rules. (Later on, we will see other ways that help with the identification.)

Definition 2.23. (Parametric ancestors) A formula $\mathcal{A}$ (occurring in a sequent higher in the proof tree) is a *parametric ancestor* of a formula $\mathcal{A}$ (occurring in a sequent lower in the proof tree) iff the two formulas are related via the transitive closure of the relation specified in (1)–(3).

(1) A parametric formula $\mathcal{A}$ in an upper sequent of a rule is a parametric ancestor of the same $\mathcal{A}$ in the lower sequent of the rule;

(2) the subalterns in $(C\vdash)$ and $(\vdash C)$ are parametric ancestors of the corresponding principal formulas in the lower sequent of the rules;

(3) the subalterns in the $(W\vdash)$ and $(\vdash W)$ rules are parametric ancestors of the principal formula in the lower sequent of the rules.

Parametric ancestors allow us to trace formulas upward in the proof tree, which provides a more refined view of proofs than if we would rely simply on occurrences of formulas. All the parametric ancestors of $\mathcal{A}$ look like $\mathcal{A}$, and they are called *parametric* because they are often—though not always—parametric formulas with respect to the application of a rule. A useful observation is that all parametric ancestors of a wff $\mathcal{A}$ occur on the same side of the $\vdash$ where $\mathcal{A}$ itself does.

Let us assume that we are given a proof containing an application of the cut rule. We define the three parameters that the inductive proof relies on.

Definition 2.24. (Degree of cut) The *degree* of an application of the cut rule, δ, is the degree of the cut formula, $\mathcal{C}$.

Definition 2.25. (Contraction measure of cut) The *contraction measure* of an application of the cut rule, μ, is the number of applications of $(W\vdash)$ and $(\vdash W)$ rules to parametric ancestors of the cut formula.[11]

Definition 2.26. (Rank of cut) The *left rank* of (an application of) the cut rule, ϱ_l, is $1+$ the maximal number of consecutive sequents (from the left premise upward) in the subtree rooted in the left premise in which parametric ancestors of the cut formula occur in the succedent.

The *right rank* of (an application of) the cut rule, ϱ_r, is $1+$ the maximal number of consecutive sequents (from the right premise upward) in the subtree rooted in the right premise in which parametric ancestors of the cut formula occur in the antecedent.

The *rank* of (an application of) the cut rule, ϱ, is the *sum* of the left and right ranks.

Obviously, δ, μ and ϱ are all natural numbers; $\delta, \mu \geq 0$. Moreover, ϱ is positive, and at least 2, because $\varrho_l \geq 1$ and $\varrho_r \geq 1$. Our definition of rank is different from a similar notion in [104]; but to maintain the least value for $\varrho = 2$, we have "$1+$" in our definition—as if we were counting the cut formula as its own ancestor or the sequent that is the premise of the cut.

Theorem 2.27. (Cut theorem) *The cut rule is* admissible *in* LK.

We will prove that if a sequent is provable with applications of the cut rule, then it is provable without using the cut rule. The cut rule is *not* a derived rule, that is, there is no way, in general, to obtain the lower sequent from the premises by applications of the rules of LK.

[11] The notion of contraction measures was introduced in [39], for $LE^{t}_{\to}$. The idea of having this new parameter in an inductive proof comes from certain observations recorded in [37].

Exercise 2.3.10. Prove the claim in the last sentence. [Hint: Construct a concrete proof with an application of the cut rule.]

The proof we give is *constructive* (or effective), and it is *proof-theoretical*, that is, it does not appeal to the interpretation of the sequents. Effectiveness means that given a proof containing applications of the cut rule, one can produce a proof of the end sequent without cut using appropriate cases from the proof of the cut theorem. There are other methods to prove a cut rule admissible; in Section 7.6, we present a proof of a cut theorem that is semantical and is not effective.

In the steps of the proof, we will specify *local* changes to be made to the given proof tree—unlike in some other proofs, where global modifications of the proof tree are stipulated (cf. Section 7.4). In other words, the steps will not require replacement of formulas throughout a subtree of the proof tree, though we sometimes will assume that it is possible to copy a subtree.

Proof (Cut theorem): The proof is by *triple induction* on δ, μ and ϱ, and by an induction on the number of applications of the cut rule in a proof.[12]

Our strategy is to make clear the *overall structure of the proof*, and to indicate all the cases that need to be considered. However, when we reach a point where a pattern may be discerned, we will insert exercises, and omit writing out all the details. We will give a proof segment with an application of the cut rule, then after a $\rightsquigarrow$, the transformed proof segment.

1.1 $\langle \varrho = 2, \mu = 0, \delta = 0 \rangle$ is the case when each parameter has its least value. There are four pairs of rules or axioms that allow an application of the cut rule.

(a) $\langle \text{id}, \text{id} \rangle$

$$\text{cut}\ \dfrac{p \vdash p \qquad p \vdash p}{p \vdash p} \qquad \rightsquigarrow \qquad p \vdash p$$

(b) $\langle \vdash K, \text{id} \rangle$

$$\text{cut}\ \dfrac{\vdash K\ \dfrac{\begin{array}{c}\vdots\\ \Gamma \vdash \Delta\end{array}}{\Gamma \vdash \Delta, p} \qquad p \vdash p}{\Gamma \vdash \Delta, p} \qquad \rightsquigarrow \qquad \dfrac{\begin{array}{c}\vdots\\ \Gamma \vdash \Delta\end{array}}{\Gamma \vdash \Delta, p}\ \text{cut}$$

(c) $\langle \text{id}, K\vdash \rangle$

$$\text{cut}\ \dfrac{p \vdash p \qquad \dfrac{\begin{array}{c}\vdots\\ \Gamma \vdash \Delta\end{array}}{p, \Gamma \vdash \Delta}\ K\vdash}{p, \Gamma \vdash \Delta} \qquad \rightsquigarrow \qquad \dfrac{\begin{array}{c}\vdots\\ \Gamma \vdash \Delta\end{array}}{p, \Gamma \vdash \Delta}\ K\vdash$$

[12] For a double induction in a similar context, see [127, p. 454]. We provide more details on the structure of the triple inductive proof in Section A.2 in the *Appendix*.

(d) $\langle \vdash K, K\vdash \rangle$

$$\text{cut}\;\dfrac{\vdash K\;\dfrac{\begin{matrix}\vdots\\ \Gamma \vdash \Delta\end{matrix}}{\Gamma \vdash \Delta, p} \qquad \dfrac{\begin{matrix}\vdots\\ \Theta \vdash \Lambda\end{matrix}}{p, \Theta \vdash \Lambda}\;K\vdash}{\Gamma, \Theta \vdash \Delta, \Lambda} \quad\rightsquigarrow\quad \dfrac{\dfrac{\begin{matrix}\vdots\\ \Gamma \vdash \Delta\end{matrix}}{\Gamma, \Theta \vdash \Delta}\;K\vdash\text{'s}, C\vdash\text{'s}}{\Gamma, \Theta \vdash \Delta, \Lambda}\;\vdash K\text{'s}$$

In each pairing, the cut is directly *eliminated* from the proof.

1.2 $\langle \varrho = 2, \mu = 0, \delta > 0 \rangle$ allows more variety in pairing rules and the axiom. There are 40 cases in total. (In order to save space, we will almost always omit—from now on until the end of the section—adding $\vdots$ in place of the subtrees above the premises. However, it should be understood that only instances of the axiom (id) are leaves in a proof tree.)

Exercise 2.3.11. The four pairs considered in **1.1** are possible again. Verify those cases. [Hint: Consider is the degree of the cut formula plays a role.]

(a) $\langle \vdash \wedge, \wedge\vdash_1 \rangle$

$$\text{cut}\;\dfrac{\vdash\wedge\;\dfrac{\Gamma \vdash \Delta, \mathcal{A} \quad \Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \wedge \mathcal{B}} \qquad \dfrac{\mathcal{A}, \Theta \vdash \Lambda}{\mathcal{A} \wedge \mathcal{B}, \Theta \vdash \Lambda}\;\wedge\vdash_1}{\Gamma, \Theta \vdash \Delta, \Lambda} \quad\rightsquigarrow\quad \dfrac{\Gamma \vdash \Delta, \mathcal{A} \qquad \mathcal{A}, \Theta \vdash \Lambda}{\Gamma, \Theta \vdash \Delta, \Lambda}\;\text{cut}$$

(b) $\langle \vdash \wedge, \wedge\vdash_2 \rangle$

$$\text{cut}\;\dfrac{\vdash\wedge\;\dfrac{\Gamma \vdash \Delta, \mathcal{A} \quad \Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \wedge \mathcal{B}} \qquad \dfrac{\mathcal{B}, \Theta \vdash \Lambda}{\mathcal{A} \wedge \mathcal{B}, \Theta \vdash \Lambda}\;\wedge\vdash_2}{\Gamma, \Theta \vdash \Delta, \Lambda} \quad\rightsquigarrow\quad \dfrac{\Gamma \vdash \Delta, \mathcal{B} \qquad \mathcal{B}, \Theta \vdash \Lambda}{\Gamma, \Theta \vdash \Delta, \Lambda}\;\text{cut}$$

Exercise 2.3.12. Work out the details for the pairings of disjunction rules.

(c) $\langle \vdash \neg, \neg\vdash \rangle$

$$\text{cut}\;\dfrac{\vdash\neg\;\dfrac{\mathcal{A}, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg\mathcal{A}} \qquad \dfrac{\Theta \vdash \Lambda, \mathcal{A}}{\neg\mathcal{A}, \Theta \vdash \Lambda}\;\neg\vdash}{\Gamma, \Theta \vdash \Delta, \Lambda} \quad\rightsquigarrow\quad \dfrac{\dfrac{\Theta \vdash \Lambda, \mathcal{A} \quad \mathcal{A}, \Gamma \vdash \Delta}{\Theta, \Gamma \vdash \Lambda, \Delta}\;\text{cut}}{\Gamma, \Theta \vdash \Delta, \Lambda}\;C\vdash\text{'s}, \vdash C\text{'s}$$

(d) $\langle \vdash \supset, \supset\vdash \rangle$

$$\text{cut}\;\dfrac{\vdash\supset\;\dfrac{\mathcal{A}, \Gamma \vdash \Delta, \mathcal{B}}{\Gamma \vdash \Delta, \mathcal{A} \supset \mathcal{B}} \qquad \dfrac{\Theta \vdash \Lambda, \mathcal{A} \quad \mathcal{B}, \Xi \vdash \Psi}{\mathcal{A} \supset \mathcal{B}, \Theta, \Xi \vdash \Lambda, \Psi}\;\supset\vdash}{\Gamma, \Theta, \Xi \vdash \Delta, \Lambda, \Psi} \quad\rightsquigarrow$$

$$\rightsquigarrow\quad \dfrac{\text{cut}\;\dfrac{\text{cut}\;\dfrac{\Theta \vdash \Lambda, \mathcal{A} \quad \mathcal{A}, \Gamma \vdash \Delta, \mathcal{B}}{\Theta, \Gamma \vdash \Lambda, \Delta, \mathcal{B}} \qquad \mathcal{B}, \Xi \vdash \Psi}{\Theta, \Gamma, \Xi \vdash \Lambda, \Delta, \Psi}}{\Gamma, \Theta, \Xi \vdash \Delta, \Lambda, \Psi}\;C\vdash\text{'s}, \vdash C\text{'s}$$

(e) $\langle \vdash\exists, \exists\vdash \rangle$

$$\dfrac{\vdash\exists\ \dfrac{\Gamma \vdash \Delta, \mathcal{A}(y)}{\Gamma \vdash \Delta, \exists x\, \mathcal{A}(x)} \qquad \dfrac{\mathcal{A}(z), \Theta \vdash \Lambda}{\exists x\, \mathcal{A}(x), \Theta \vdash \Lambda}\ \exists\vdash}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\Gamma \vdash \Delta, \mathcal{A}(y) \qquad \mathcal{A}(y), \Theta \vdash \Lambda}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut}$$

(f) $\langle \vdash\forall, \forall\vdash \rangle$

$$\dfrac{\vdash\forall\ \dfrac{\Gamma \vdash \Delta, \mathcal{A}(z)}{\Gamma \vdash \Delta, \forall x\, \mathcal{A}(x)} \qquad \dfrac{\mathcal{A}(y), \Theta \vdash \Lambda}{\forall x\, \mathcal{A}(x), \Theta \vdash \Lambda}\ \forall\vdash}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\Gamma \vdash \Delta, \mathcal{A}(y) \qquad \mathcal{A}(y), \Theta \vdash \Lambda}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut}$$

The cases **(a)–(f)** are justified by a *reduction in the degree* of the cut formula. For **(e)** and **(f)**, we note that Lemma 2.15 ensures that the variable z can be renamed in the subtrees above $\mathcal{A}(z)$.

There are four connectives and two quantifiers, which gives 7 right rules and 7 left rules. They can be paired with the axiom or a thinning rule. We summarize the 28 cases in the next four patterns, assuming $\divideontimes \in \{\wedge, \vee, \neg, \supset, \exists, \forall\}$.

(g) $\langle \vdash\divideontimes, \mathrm{id} \rangle$

$$\dfrac{\vdash\divideontimes\ \dfrac{\vdots}{\Gamma \vdash \Delta, \mathcal{C}} \qquad \mathcal{C} \vdash \mathcal{C}}{\Gamma \vdash \Delta, \mathcal{C}}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\vdots}{\Gamma \vdash \Delta, \mathcal{C}}\ \vdash\divideontimes$$

(h) $\langle \mathrm{id}, \divideontimes\vdash \rangle$

$$\dfrac{\mathcal{C} \vdash \mathcal{C} \qquad \dfrac{\vdots}{\mathcal{C}, \Gamma \vdash \Delta}\ \divideontimes\vdash}{\mathcal{C}, \Gamma \vdash \Delta}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\vdots}{\mathcal{C}, \Gamma \vdash \Delta}\ \divideontimes\vdash$$

(i) $\langle \vdash\divideontimes, K\vdash \rangle$

$$\dfrac{\vdash\divideontimes\ \dfrac{\vdots}{\Gamma \vdash \Delta, \mathcal{C}} \qquad \dfrac{\overset{\vdots}{\Theta \vdash \Lambda}}{\mathcal{C}, \Theta \vdash \Lambda}\ K\vdash}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\dfrac{\overset{\vdots}{\Theta \vdash \Lambda}}{\Gamma, \Theta \vdash \Lambda}\ K\vdash\text{'s}}{\Gamma, \Theta \vdash \Delta, \Lambda}\ \vdash K\text{'s}, \vdash C\text{'s}$$

(j) $\langle \vdash K, \divideontimes\vdash \rangle$

$$\dfrac{\vdash K\ \dfrac{\overset{\vdots}{\Theta \vdash \Lambda}}{\Theta \vdash \Lambda, \mathcal{C}} \qquad \dfrac{\vdots}{\mathcal{C}, \Gamma \vdash \Delta}\ \divideontimes\vdash}{\Theta, \Gamma \vdash \Lambda, \Delta}\ \text{cut} \quad\rightsquigarrow\quad \dfrac{\dfrac{\overset{\vdots}{\Theta \vdash \Lambda}}{\Theta \vdash \Lambda, \Delta}\ \vdash K\text{'s}}{\Theta, \Gamma \vdash \Lambda, \Delta}\ K\vdash\text{'s}, C\vdash\text{'s}$$

The cut rule is *eliminated* in each case in **(g)–(j)**.

Exercise 2.3.13. Choose at least 4 logical constants and write out an instance of the last four cases. [Hint: Convince yourself that the patterns given capture all the 28 cases.]

1.3 $\langle \varrho > 2, \mu = 0, \delta = 0 \rangle$ can happen when $\varrho_l > 1$ or $\varrho_r > 1$. We assume the former; then we have 23 cases, depending on which rule was applied in the left premise of the cut. In the operational rules, even if the rule is a right rule, the principal formula cannot be the cut formula, because $\varrho_l > 1$. The principal formula of $(\vdash W)$ cannot be the cut formula, because $\mu = 0$.

(a) $\langle \vdash K, \text{--} \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A}}\vdash K \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}}\text{cut} \quad \rightsquigarrow \quad \dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\text{cut}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}}\vdash K$$

(b) $\langle \vdash C, \text{--} \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \mathcal{D}, \Delta_2}{\Gamma \vdash \Delta_1, \mathcal{D}, \mathcal{C}, \Delta_2}\vdash C \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \mathcal{D}, \Lambda, \Delta_2}\text{cut} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \mathcal{D}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \mathcal{D}, \Delta_2}\text{cut}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \mathcal{D}, \Lambda, \Delta_2}\vdash C\text{'s}$$

Exercise 2.3.14. Create two other cases with the $(\vdash C)$ rule.

(c) $\langle \vdash \wedge, \text{--} \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A} \quad \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{B}}{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A} \wedge \mathcal{B}}\vdash\wedge \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A} \wedge \mathcal{B}}\text{cut} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A} \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}} \quad \dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{B} \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{B}}\text{cuts}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A} \wedge \mathcal{B}}\vdash\wedge$$

(d) $\langle \vdash \vee_1, \text{--} \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A}}{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A} \vee \mathcal{B}}\vdash\vee_1 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A} \vee \mathcal{B}}\text{cut} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A} \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}}\text{cut}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A} \vee \mathcal{B}}\vdash\vee_1$$

Exercise 2.3.15. Detail the case with the $(\vdash \vee_2)$ rule.

(e) $\langle \vdash \supset, \_\_ \rangle$

$$\text{cut}\ \dfrac{\vdash\supset\ \dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{B}}{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A}\supset\mathcal{B}} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{A}\supset\mathcal{B}} \quad\rightsquigarrow$$

$$\rightsquigarrow\quad \dfrac{\dfrac{\dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{B}\quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{B}}\ \text{cut}}{\mathcal{A},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{B}}\ C\vdash\text{'s}}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{A}\supset\mathcal{B}}\ \vdash\supset$$

(f) $\langle \vdash \neg, \_\_ \rangle$

$$\text{cut}\ \dfrac{\vdash\neg\ \dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2}{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\neg\mathcal{A}} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\neg\mathcal{A}} \quad\rightsquigarrow$$

$$\rightsquigarrow\quad \dfrac{\dfrac{\dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2\quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\ \text{cut}}{\mathcal{A},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\ C\vdash\text{'s}}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\neg\mathcal{A}}\ \vdash\neg$$

(g) $\langle \vdash \exists, \_\_ \rangle$

$$\text{cut}\ \dfrac{\vdash\exists\ \dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,A(y)}{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\exists x\,A(x)} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\exists x\,A(x)} \quad\rightsquigarrow$$

$$\rightsquigarrow\quad \dfrac{\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,A(y)\quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,A(y)}\ \text{cut}}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\exists x\,A(x)}\ \vdash\exists$$

(h) $\langle \vdash \forall, \_\_ \rangle$

$$\text{cut}\ \dfrac{\vdash\forall\ \dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,A(z)}{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\forall x\,A(x)} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\forall x\,A(x)} \quad\rightsquigarrow$$

$$\rightsquigarrow\quad \dfrac{\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,A(y)\quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,A(y)}\ \text{cut}}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\forall x\,A(x)}\ \vdash\forall$$

Exercise 2.3.16. Scrutinize and explain the transformation in **(h)**.

(i) $\langle \vdash W, \_\_ \rangle$

$$\text{cut}\ \dfrac{\vdash W\ \dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A},\mathcal{A}}{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A}} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{A}} \quad\rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2, \mathcal{A}, \mathcal{A} \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}, \mathcal{A}}\ \text{cut}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2, \mathcal{A}}\ {\vdash}W$$

(j) $\langle W{\vdash}, \_\_\rangle$

$$\text{cut}\ \dfrac{W{\vdash}\ \dfrac{\mathcal{A}, \mathcal{A}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\mathcal{A}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2} \quad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A}, \mathcal{A}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A}, \mathcal{A}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ \text{cut}}{\mathcal{A}, \mathcal{A}, \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ C{\vdash}\text{'s}}{\mathcal{A}, \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ W{\vdash}}{\Theta_1, \mathcal{A}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ C{\vdash}\text{'s}$$

(k) $\langle K{\vdash}, \_\_\rangle$

$$\text{cut}\ \dfrac{K{\vdash}\ \dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\mathcal{A}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2} \quad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ \text{cut}}{\mathcal{A}, \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ K{\vdash}}{\Theta_1, \mathcal{A}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ C{\vdash}\text{'s}$$

(l) $\langle C{\vdash}, \_\_\rangle$

$$\text{cut}\ \dfrac{C{\vdash}\ \dfrac{\Gamma_1, \mathcal{A}, \mathcal{B}, \Gamma_2 \vdash \Delta_1, \mathcal{C}, \Delta_2}{\Gamma_1, \mathcal{B}, \mathcal{A}, \Gamma_2 \vdash \Delta_1, \mathcal{C}, \Delta_2} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma_1, \mathcal{B}, \mathcal{A}, \Gamma_2, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2} \quad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\Gamma_1, \mathcal{A}, \mathcal{B}, \Gamma_2 \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma_1, \mathcal{A}, \mathcal{B}, \Gamma_2, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ \text{cut}}{\Theta_1, \Gamma_1, \mathcal{B}, \mathcal{A}, \Gamma_2, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\ C{\vdash}$$

(m) $\langle \wedge{\vdash}_1, \_\_\rangle$

$$\text{cut}\ \dfrac{\wedge{\vdash}_1\ \dfrac{\mathcal{A}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\mathcal{A} \wedge \mathcal{B}, \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A} \wedge \mathcal{B}, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2} \quad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2 \quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\text{cut}}{\mathcal{A},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;C\vdash\text{'s}}{\mathcal{A}\wedge\mathcal{B},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\wedge\vdash_1}{\Theta_1,\mathcal{A}\wedge\mathcal{B},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;C\vdash\text{'s}$$

Exercise 2.3.17. Write out the details for the case with rule $(\wedge\vdash_2)$. [Hint: It is practically the same as **(m)**.]

(n) $\langle\vee\vdash, \text{--}\rangle$

$$\dfrac{\vee\vdash\;\dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2 \quad \mathcal{B},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2}{\mathcal{A}\vee\mathcal{B},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A}\vee\mathcal{B},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\text{cut} \qquad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2 \quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}}{\mathcal{A},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2} \qquad \dfrac{\dfrac{\mathcal{B},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2 \quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{B},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\text{cuts}}{\mathcal{B},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;C\vdash\text{'s}}{\mathcal{A}\vee\mathcal{B},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\vee\vdash}{\Theta_1,\mathcal{A}\vee\mathcal{B},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;C\vdash\text{'s}$$

(o) $\langle\neg\vdash, \text{--}\rangle$

$$\dfrac{\neg\vdash\;\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A}}{\neg\mathcal{A},\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\neg\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\text{cut} \qquad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A} \quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{A}}\;\text{cut}}{\neg\mathcal{A},\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;\neg\vdash}{\Theta_1,\neg\mathcal{A},\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2}\;C\vdash\text{'s}$$

(p) $\langle\supset\vdash, \text{--}\rangle$

$$\dfrac{\supset\vdash\;\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A} \quad \mathcal{B},\Xi\vdash\Psi}{\mathcal{A}\supset\mathcal{B},\Gamma,\Xi\vdash\Delta_1,\mathcal{C},\Delta_2,\Psi} \qquad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\mathcal{A}\supset\mathcal{B},\Gamma,\Xi,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\Psi}\;\text{cut} \qquad \rightsquigarrow$$

$$\rightsquigarrow \qquad \dfrac{\dfrac{\text{cut}\;\dfrac{\Gamma\vdash\Delta_1,\mathcal{C},\Delta_2,\mathcal{A} \quad \Theta_1,\mathcal{C},\Theta_2\vdash\Lambda}{\Theta_1,\Gamma,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\mathcal{A}} \qquad \mathcal{B},\Xi\vdash\Psi}{\mathcal{A}\supset\mathcal{B},\Theta_1,\Gamma,\Theta_2,\Xi\vdash\Delta_1,\Lambda,\Delta_2,\Psi}\;\supset\vdash}{\Theta_1,\mathcal{A}\supset\mathcal{B},\Gamma,\Xi,\Theta_2\vdash\Delta_1,\Lambda,\Delta_2,\Psi}\;C\vdash\text{'s}$$

(q) $\langle \supset\vdash, \_\_ \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta, \mathcal{A} \quad \mathcal{B}, \Xi \vdash \Psi_1, \mathcal{C}, \Psi_2}{\mathcal{A} \supset \mathcal{B}, \Gamma, \Xi \vdash \Delta, \Psi_1, \mathcal{C}, \Psi_2}\,{\scriptstyle \supset\vdash} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A} \supset \mathcal{B}, \Gamma, \Xi, \Theta_2 \vdash \Delta, \Psi_1, \Lambda, \Psi_2}\,{\scriptstyle \mathrm{cut}} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\Gamma \vdash \Delta, \mathcal{A} \qquad \dfrac{\dfrac{\mathcal{B}, \Xi \vdash \Psi_1, \mathcal{C}, \Psi_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{B}, \Xi, \Theta_2 \vdash \Psi_1, \Lambda, \Psi_2}\,{\scriptstyle \mathrm{cut}}}{\mathcal{B}, \Theta_1, \Xi, \Theta_2 \vdash \Psi_1, \Lambda, \Psi_2}\,{\scriptstyle C\vdash\text{'s}}}{\mathcal{A} \supset \mathcal{B}, \Gamma, \Theta_1, \Xi, \Theta_2 \vdash \Delta, \Psi_1, \Lambda, \Psi_2}\,{\scriptstyle \supset\vdash}}{\Theta_1, \mathcal{A} \supset \mathcal{B}, \Gamma, \Xi, \Theta_2 \vdash \Delta, \Psi_1, \Lambda, \Psi_2}\,{\scriptstyle C\vdash\text{'s}}$$

(r) $\langle \exists\vdash, \_\_ \rangle$

$$\dfrac{\dfrac{\mathcal{A}(z), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\exists x\, \mathcal{A}(x), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}\,{\scriptstyle \exists\vdash} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \exists x\, \mathcal{A}(x), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \mathrm{cut}} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A}(y), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A}(y), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \mathrm{cut}}}{\mathcal{A}(y), \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle C\vdash\text{'s}}}{\exists x\, \mathcal{A}(x), \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \exists\vdash}}{\Theta_1, \exists x\, \mathcal{A}(x), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle C\vdash\text{'s}}$$

(s) $\langle \forall\vdash, \_\_ \rangle$

$$\dfrac{\dfrac{\mathcal{A}(y), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}{\forall x\, \mathcal{A}(x), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2}\,{\scriptstyle \forall\vdash} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \forall x\, \mathcal{A}(x), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \mathrm{cut}} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\dfrac{\dfrac{\mathcal{A}(y), \Gamma \vdash \Delta_1, \mathcal{C}, \Delta_2 \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \mathcal{A}(y), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \mathrm{cut}}}{\mathcal{A}(y), \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle C\vdash\text{'s}}}{\forall x\, \mathcal{A}(x), \Theta_1, \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle \forall\vdash}}{\Theta_1, \forall x\, \mathcal{A}(x), \Gamma, \Theta_2 \vdash \Delta_1, \Lambda, \Delta_2}\,{\scriptstyle C\vdash\text{'s}}$$

The transformations in **(a)–(s)** are justified by a *reduction in* ϱ_l. The shape of the cut formula has not changed, hence, δ is unchanged. For the same reason, the decrease in rank cannot change μ.

Exercise 2.3.18. Take a look at the cases in **1.3**. Do the transformations introduce applications of a rule that was not used in the given proof segment?

Exercise 2.3.19. Again, look at the cases. Could they be presented more compactly by grouping some of them together? [Hint: You may consider whether a one-premise or a two-premise rule was applied, or whether the rule is a left rule or a right rule.]

1.4 $\langle \varrho > 2, \mu = 0, \delta > 0 \rangle$ with $\varrho_l > 1$ gives rise to a similar set of cases as **1.3** did. (Incidentally, we used $\mathcal{C}$ for the cut formula in **1.3** without placing any restrictions on $\mathcal{C}$.)

Exercise 2.3.20. Explain why there is no difference between the cases when $\delta = 0$ and when $\delta > 0$. [Hint: You probably will have to appeal to the definitions of rank, contraction measure and the cut being a single cut.]

1.5 $\langle \varrho > 2, \mu > 0, \delta = 0 \rangle$ and $\varrho_l > 1$ leads to similar cases as in **1.3** with a new case added when the cut formula is the principal formula of $(\vdash W)$.

Exercise 2.3.21. Create all the possible cases "from scratch" and compare your list with the cases in **1.3**. [Hint: LK is defined on page 17.]

(a) $\langle \vdash W, \_\_ \rangle$

$$\dfrac{\dfrac{\Gamma \vdash \Delta, \mathcal{C}, \mathcal{C}}{\Gamma \vdash \Delta, \mathcal{C}}\,{\scriptstyle \vdash W} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta, \Lambda}\,{\scriptstyle \text{cut}} \quad \rightsquigarrow$$

$$\rightsquigarrow \quad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, \mathcal{C}, \mathcal{C} \quad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta, \Lambda, \mathcal{C}}\,{\scriptstyle \text{cut}} \qquad \Theta_1, \mathcal{C}, \Theta_2 \vdash \Lambda}{\Theta_1, \Theta_1, \Gamma, \Theta_2, \Theta_2 \vdash \Delta, \Lambda, \Lambda}\,{\scriptstyle \text{cut}}}{\Theta_1, \Gamma, \Theta_2 \vdash \Delta, \Lambda}\,{\scriptstyle C\text{'s},\, W\text{'s}}$$

The transformation is justified by a decrease in the contraction measure of the cut—for both cuts. The first cut, in addition, has a lower rank, but the second cut may have the same rank as the cut in the given proof chunk. However, that is not a problem, because the contraction measure is decreased by at least one.

1.6 $\langle \varrho > 2, \mu > 0, \delta > 0 \rangle$ with $\varrho_l > 1$ generates the same pairings as in **1.5**.

Exercise 2.3.22. Verify that the greater degree of the cut formula is not an impediment to the transformations in **1.5**.

1.7 $\langle \varrho > 2, \mu = m, \delta = d \rangle$, where $\varrho_r > 1$ leads to duals of the cases in **1.3–1.6**. The cases can be dealt with similarly as before.

This concludes the proof. qed

We contend, that ours is the *most elegant proof* of the admissibility of the *single cut rule* for LK.

Exercise* 2.3.23. Write out in detail all the cases. [Hint: If you do not bunch cases together, then there are *many* cases.]

Exercise 2.3.24. Once you are certain that the cut proof works, design a compact way to present the cases.

Exercise 2.3.25. Having the details of the triple inductive proof of cut, explain why a double induction (without μ) works for mix but not for cut_G (or the cut rule).

The transformations in the proof could be viewed as a non-deterministic algorithm, because if $\varrho_l > 1$ and $\varrho_r > 1$, then no order is specified on applicable subcases. This is not how Gentzen specified his double induction in the proof of the admissibility of the mix rule. In the case of LK, we could require either that ϱ_r must be reduced first, and then the assumption $\varrho_l > 1$ may be strengthened with adding $\varrho_r = 1$, or the other way around. Imposing an order on the steps in cut elimination process may be advantageous in some other sequent calculi, which have exactly one formula on the right-hand side of the $\vdash$ and have a richer language than classical logic does. However, LK enjoys an almost complete left–right symmetry.

2.4 Interpretations, soundness and completeness

The previous sections have not given a rigorous interpretation of the formulas and sequents. We used notation that may be familiar from other presentations of classical logic or suggestive of the intended meaning of the formal expressions, but it is time to put a formal interpretation in place. The interpretation that we present follows the approach that is usual nowadays: it defines inductively the notion of a structure, also called a *model*, *satisfying* (or making true) a wff. Then we extend this notion to sequents in a straightforward way, which is in accordance with viewing the comma as $\wedge$ or $\vee$—as in Definition 2.18 or in the informal sense we introduced on page 21.

Definition 2.28. (Classical models) A classical *model* $\mathfrak{M}$ is a triple $\langle D, I, v\rangle$, where D is a non-empty set of objects, the *domain of interpretation*; I is a function, the *interpretation function* satisfying conditions (1)–(4) below; and v is a function, the *valuation function* that interprets individual variables into D.

(1) If t is a name constant, then $I(t) \in D$;

(2) if f^n is an n-ary function symbol, then $I(f) \in D^{D^n}$;

(3) if p is a propositional variable, then $I(p) \in \{T, F\}$ (where $T, F \notin D$; they stand for the two truth values);

(4) if P^n is an n-place predicate, then $I(P) \subseteq D^n$.

This definition does not specify the cardinality of D, which is sometimes called the *universe of discourse*. Nor does it define a unique structure; in other words, there is more than one, indeed, there are infinitely many models—though some wff's are true only in finitely many models. The variability of the interpretations of non-logical components is in contrast with the interpretation of the logical particles, which are fixed by Definition 2.30 below.

In order to make the definition of satisfaction smooth, we introduce a function, denoted by $\mathfrak{i}$, which gives the *denotation* of a term. An atomic term can be a constant or a variable, and a function symbol can take either kind of atomic term as its argument, as well as it can have a complex term as its argument. Clearly, I and v are sufficient to compute the denotation of a term in a bottom-up fashion, thus $\mathfrak{i}$'s role is merely to make the presentation easier, rather than to add something new to an interpretation.

Definition 2.29. (Denotation of terms) If t is a term, then the *denotation of* t, denoted by $\mathfrak{i}(t)$, is determined by a classical model $\mathfrak{M} = \langle D, I, v \rangle$ in accordance with (1)–(3).

(1) If t is a name constant, then $\mathfrak{i}(t) = I(t)$;

(2) if t is a variable, then $\mathfrak{i}(t) = v(t)$;

(3) if t is of the form $f^n(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are terms and f^n is a function symbol, then $\mathfrak{i}(f(t_1, \ldots, t_n)) = I(f)(\mathfrak{i}(t_1), \ldots, \mathfrak{i}(t_n))$.

It is an easy induction on the structure of terms to prove that $\mathfrak{i}$ is well-defined, that is, $\mathfrak{i}$ is a total function on the set of terms, and maps each term into an element of D.

We introduce a further notational device to indicate the *pointwise modification* (including null or no modification) for valuation functions. By $v[x : d]$ (where $d \in D$) we mean a function such that for any y that is (syntactically) not identical to the variable x, $v(y) = v[x : d](y)$, whereas $v[x : d](x) = d$. Obviously, if $v(x) = d$, then literally $v[x : d]$ does not modify v at all, and we label this as null modification. Otherwise (i.e., when $v(x) \neq d$), $v[x : d]$ differs from v in exactly one place, in the value for x.

Definition 2.30. (Truth in classical models) Let $\mathfrak{M} = \langle D, I, v \rangle$ be a classical model. The relation $\mathfrak{M} \vDash_v \mathcal{A}$ is defined recursively by (1)–(10). (We omit $\mathfrak{M}$, as well as $_v$, when they are unaltered in a clause.)

(1) $\vDash \boldsymbol{T}$ and $\nvDash \boldsymbol{F}$ (i.e., $\boldsymbol{F}$ is never satisfied);

(2) $\vDash p$ iff $I(p) = T$;

(3) $\vDash t_1 = t_2$ iff $\mathfrak{i}(t_1) = \mathfrak{i}(t_2)$;

(4) if P^n is an n-place predicate (other than $=$) and $t_1, \ldots, t_n$ are terms, then $\vDash P(t_1, \ldots, t_n)$ iff $\langle \mathfrak{i}(t_1), \ldots, \mathfrak{i}(t_n) \rangle \in I(P)$;

(5) $\vDash \neg\mathcal{A}$ iff $\nvDash \mathcal{A}$, that is, not $\vDash \mathcal{A}$;

(6) $\vDash \mathcal{A} \wedge \mathcal{B}$ iff $\vDash \mathcal{A}$ and $\vDash \mathcal{B}$;

(7) $\vDash \mathcal{A} \vee \mathcal{B}$ iff $\vDash \mathcal{A}$ or $\vDash \mathcal{B}$;

(8) $\vDash \mathcal{A} \supset \mathcal{B}$ iff $\nvDash \mathcal{A}$ or $\vDash \mathcal{B}$;

(9) $\vDash_v \forall x\, \mathcal{A}(x)$ iff for any $v[x:d]$, $\vDash_{v[x:d]} \mathcal{A}(x)$;

(10) $\vDash_v \exists x\, \mathcal{A}(x)$ iff there is a $v[x:d]$ such that $\vDash_{v[x:d]} \mathcal{A}(x)$.

$\mathfrak{M} \vDash_v \mathcal{A}$ is read as "the model $\mathfrak{M}$ with the valuation v *satisfies* (or makes *true*) the wff $\mathcal{A}$." In most of the clauses, $_v$ may be left implicit, because $\mathfrak{M}$ includes v. The definition makes clear that given a model $\mathfrak{M}$, the truth of a formula $\mathcal{A}$ is completely determined by $\mathfrak{M}$ itself. Although v's pointwise modification must be considered in (9) and (10), which yield models of the form $\mathfrak{M}' = \langle D, I, v[x:d]\rangle$, the d's are chosen from D, which is a component of $\mathfrak{M}$. (9) and (10) differ from the rest of the clauses, because they show that quantification is *not compositional*. In the case of quantificational formulas, the semantic values of the component expressions are not sufficient to compute the semantic value of the complex formula.

Exercise 2.4.1. Are the next four formulas satisfiable or not? (a) $(\exists x\, \forall y\, R(x,y) \supset \forall y\, \exists x\, R(x,y))$, (b) $(\forall x\, (P(x) \supset \exists y\, P(y)) \supset (\exists x\, P(x) \supset \exists y\, P(y)))$, (c) $(\forall x\, (\forall x\, \neg P(x) \supset Q(x)) \supset \forall y\, (\neg Q(y) \supset P(y)))$, (d) $(\forall x\, \forall y\, \forall z\, ((R(x,y) \wedge R(y,z)) \supset R(x,z)) \wedge (\forall x\, \exists y\, R(x,y) \wedge \forall x\, \forall y\, (R(x,y) \supset \neg R(y,x))))$. [Hint: Use the semantic notions introduced so far to substantiate your claims.]

Definition 2.31. (Validity in classical models) $\mathcal{A}$ is *logically valid* (or *valid*, for short) in classical models iff $\mathfrak{M} \vDash_v \mathcal{A}$ for all $\mathfrak{M}$, that is, the wff $\mathcal{A}$ is true in all classical models.

The definition can be rephrased—equivalently—to state that $\mathcal{A}$ is valid when given any domain D and any interpretation function I, all valuations v make $\mathcal{A}$ true.

Exercise 2.4.2. Determine if the following wff's are logically valid or not. (a) $(\exists x\, (P(x) \wedge \forall y\, (Q(y) \supset R(x,y))) \supset \forall x\, (Q(x) \supset \exists y\, (P(y) \wedge R(y,x))))$, (b) $(\forall x\, (P(x) \supset Q(x)) \supset \exists y\, (P(y) \wedge Q(y)))$, (c) $((\forall x\, P(x) \vee \forall y\, Q(y)) \supset \forall z\, (\neg P(z) \supset Q(z)))$, (d) $\exists x\, \exists y\, (x = y \wedge (P(x) \supset P(y)))$. [Hint: The negation of a valid wff is not satisfiable.]

Definition 2.18 allows us to view a sequent as a wff. If we think of $\mathcal{A}$ in clauses (1) and (4) of the definition as a fixed wff, then τ is a *function*, that is, every sequent is mapped into a unique formula.

Exercise 2.4.3. Prove that τ is a well-defined function, that is, for each sequent $\Gamma \vdash \Delta$, $\tau(\Gamma \vdash \Delta)$ is a wff. Is it true that if $\tau(\Gamma \vdash \Delta)$ is the same wff as $\tau(\Gamma' \vdash \Delta')$, then $\Gamma \vdash \Delta$ is $\Gamma' \vdash \Delta'$?

The soundness and completeness theorems guarantee that the proof system is adequate with respect to the intended interpretation. Our focus in this book is on proof systems; hence, we only outline the proofs of the next three theorems and do not give many details. (We exclude $\boldsymbol{T}, \boldsymbol{F}$ and $=$, as well as the $f_i^{n_i}$'s from consideration, as before.)

Theorem 2.32. (Soundness, 1) *If* $\Gamma \vdash \Delta$ *is a provable sequent, then* $\tau(\Gamma \vdash \Delta)$ *is a* logically valid *formula.*

Proof: We outline the structure of the proof and detail two steps. The rest of the proof is relegated to an exercise.

The proof is by induction on the height of the proof tree. If the height is 1, then the sequent is provable, because it is an instance of the axiom. Otherwise, the bottom sequent in the proof is by a rule, and one has to consider the possible cases rule by rule.

1. If the sequent is of the form $\mathcal{A} \vdash \mathcal{A}$, then $\tau(\mathcal{A} \vdash \mathcal{A})$ is $\mathcal{A} \supset \mathcal{A}$. For this wff not to be logically valid, there should be a model $\mathfrak{M}$ such that $\mathfrak{M} \nvDash_v \mathcal{A} \supset \mathcal{A}$. The latter, by clause (8) from Definition 2.30, means that $\mathfrak{M} \vDash_v \mathcal{A}$ and $\mathfrak{M} \nvDash_v \mathcal{A}$. This is obviously impossible in a classical model.

2. Let us assume that the last rule applied in the proof is $(K\vdash)$, in particular, let the shape of the lower sequent be $\mathcal{A}, \Gamma \vdash \Delta$. The hypothesis of the induction is that $\tau(\Gamma \vdash \Delta)$ is valid. By Definition 2.18, $\tau(\Gamma \vdash \Delta)$ is $\mathcal{C} \supset \mathcal{D}$, and $\tau(\mathcal{A}, \Gamma \vdash \Delta)$ is $(\mathcal{A} \wedge \mathcal{C}) \supset \mathcal{D}$. If $\mathfrak{M} \nvDash_v (\mathcal{A} \wedge \mathcal{C}) \supset \mathcal{D}$, for some $\mathfrak{M}$, then $\mathfrak{M} \vDash_v \mathcal{A} \wedge \mathcal{C}$ but $\mathfrak{M} \nvDash_v \mathcal{D}$. However, then also $\mathfrak{M} \vDash_v \mathcal{C}$, which is a contradiction. qed

Exercise 2.4.4. Consider the 19 other rules (counting each of $(\wedge\vdash)$ and $(\vdash\vee)$ as two rules). [Hint: The steps are straightforward.]

Informally, the idea behind the soundness proof is that an instance of the axiom is *logically valid*, and the rules—whether operational or structural—*preserve validity*.

The soundness theorem is often summarized by saying that theorems are valid. Definition 2.9 introduced the notion of theoremhood into LK.

Theorem 2.33. (Soundness, 2) *If* $\mathcal{A}$ *is a theorem of* LK*, then* $\mathcal{A}$ *is* valid*.*

Exercise* 2.4.5. Prove the previous theorem. [Hint: Notice that no instance of the axiom—without further applications of rules—yields a theorem.]

The converse of the soundness theorem is the completeness theorem.

Theorem 2.34. (Completeness) *If* $\mathcal{A}$ *is a logically valid formula without name constants, then* $\vdash \mathcal{A}$ *is a* provable sequent*.*

Proof: We sketch the proof of the contrapositive of the claim, that is, if $\mathcal{A}$ is not a theorem of LK, then there is a model $\mathfrak{M}$ such that $\mathfrak{M} \nvDash \mathcal{A}$.

We outline some steps, and leave filling in the details as exercises. If there are free variables in $\mathcal{A}$, then we change each of them to a new name constant, and from now on, we take $\mathcal{A}$ to be this closed wff. We also exclude function symbols and $=$ from the language in order to provide a brief general description of the construction. We take $\forall$ as a defined logical constant, that is, $\forall x\, \mathcal{B}(x)$ is $\neg\exists x\, \neg\mathcal{B}(x)$. We consider sets of wff's starting with $\{\neg\mathcal{A}\}$. This is obviously a finite set, hence there are finitely many non-logical symbols that occur in $\neg\mathcal{A}$. Nonetheless, there are denumerably many closed formulas that can be generated using those non-logical components. We fix an enumeration of those wff's. Originally, we assumed that we have a denumerable sequence of name constants in the language; now, we take a "fresh" sequence of name constants, let us say, $b_0, b_1, b_2, \ldots$. These name constants enter into terms and wff's exactly as the name constants, which are in the original language, do.

The following construction is usually called *Lindenbaumizing*.[13] Starting with a set of wff's Γ_0 (i.e., $\{\neg\mathcal{A}\}$), we consider the wff's that we stipulated to be enumerated, and at each stage we expand the set if we can do so without introducing a contradiction.

Some of the enumerated wff's are of the form $\exists x\, \mathcal{C}(x)$, and we want to make sure that each of these formulas is instantiated. We define $\mathcal{C}^{b/x}$ to be $\mathcal{C}$, if $\mathcal{C}$ is not an existentially quantified formula. Otherwise, $\mathcal{C}^{b/x}$ is the formula obtained from $\mathcal{C}$ by omitting the quantifier prefix and substituting for the previously quantified variable the next unused name constant from the sequence of b's.

1. $\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\mathcal{B}_{n+1}, \mathcal{B}_{n+1}^{b/x}\}, & \text{if } \Gamma_n \vdash \neg\mathcal{B}_{n+1} \text{ is not provable;} \\ \Gamma_n, & \text{otherwise.} \end{cases}$

2. $\Gamma = \bigcup_{n \in \omega} \Gamma_n$

The construction *preserves consistency*, that is, if there was a wff $\mathcal{E}$ such that $\Gamma_0 \vdash \mathcal{E}$ was not provable, then there is no $\Gamma' \subseteq \Gamma$ such that $\Gamma' \vdash \mathcal{E}'$ is provable, for all $\mathcal{E}'$. The starting set $\{\neg\mathcal{A}\}$ also has this property, because if $\neg\mathcal{A} \vdash \mathcal{B}$, for any $\mathcal{B}$, then $\vdash \mathcal{A}$ is provable.

The next step is to use the set Γ to *define a model* $\mathfrak{M}$ with the property that—due to the definition of $\mathfrak{M}$ itself—$\mathfrak{M} \vDash \mathcal{C}$, for all $\mathcal{C} \in \Gamma$. The wff's in Γ (may) contain some name constants, and if they do, then we take D to be the set of those name constants. If no wff in Γ contains a name constant, then we take an object d to be the only element of D.

Lastly, it remains to be shown that $\mathfrak{M}$ indeed has the property that it makes all the elements of Γ *true*. qed

The next exercises ask you to fill in the details in the above proof sketch.

[13]The verb is derived from the name of A. Lindenbaum who was a logician in Warsaw.

Exercise 2.4.6. Prove that if $\mathcal{A}$ is not a theorem of LK, then $\{\neg\mathcal{A}\}$ is consistent. [Hint: Find a suitable wff in place of $\mathcal{B}$, and show that $\vdash \neg\neg\mathcal{A}$ is provable, hence, by cut, $\vdash \mathcal{A}$ is provable.]

Exercise 2.4.7. Show that Γ, obtained by Lindenbaumizing, is consistent, provided that Γ_0 is consistent.

Exercise 2.4.8. Make the definition of $\mathfrak{M}$ precise. [Hint: The atomic formulas determine the model, because Γ contains all the formulas that it can, in the restricted fragment of the language.]

Exercise 2.4.9. Prove that the model $\mathfrak{M}$ defined in the previous exercise has the desired property, that is, for all $\mathcal{C} \in \Gamma$, $\mathfrak{M} \vDash \mathcal{C}$.