# 2

# Natural Deduction Systems

## 2.1 Introduction

In this chapter, we will begin our consideration of proof-systems with an eye to how they can be used to articulate an inferentialist account of the meanings of linguistic expressions. As stated in the previous chapter, we will start our journey with inferentialism about specifically *logical* expressions, as this is well-trodden territory and developing formal tools in this context will provide us with a solid foundation for then moving on for the more ambitious *global* inferentialist project. The particular sort of proof system we will consider in this chapter is *natural deduction*. You've likely used this kind of proof system before if you've taken an introductory logic class, but beyond figuring prominently in introductory logic classes, natural deduction has also figured prominently in developments of logical inferentialism since its inception with the work of Gerhard Gentzen, and it is still the most widespread sort of proof system for spelling out a logical inferentialist theory. Thus, it is the natural (pun intended) place to start.

## 2.2 A Negation-Free Natural Deduction System

The basic idea of natural deduction is quite simple. We provide each logical connective with two kinds of rules: *introduction rules*, which enable us to conclude a sentence with that connective from other sentences, and *elimination rules*, which enable us to conclude other sentences from a sentence with that main connective. The introduction rules for a connective can be understood as formally codifying the inferential *conditions* for the correct use of a sentence with that connective: what you need to do prove in order to entitle yourself to conclude a sentence with that main connective. The elimination rules, on the other hand, can be understood as formally codifying the *consequences* of the use of a sentence with that connective: what proving a sentence with that main connective entitles you to conclude.

It will make things cleaner to use logical symbols rather than expressions of natural language for logical operators, and so we'll use the (hopefully) familiar symbols $\wedge$, $\vee$, $\neg$, and $\rightarrow$ for conjunction, disjunction, negation, and the (material) conditional, respectively. Let us start with *conjunction*, typically expressed in English with the word "and," and expressed with in our logical language with the symbol $\wedge$. The rules for conjunction, which we showed in the previous chapter, are the simplest, most straightforward, and least controversial rules of all logical operators. Once again, the *conjunction introduction rule*, enables us to conclude $A \wedge B$ from $A$ along with $B$, and the *conjunction elimination rules* enable us to conclude $A$ from $A \wedge B$ and they also enable us to conclude $B$ from $A \wedge B$. That is, the rules are the following:

$$\frac{A \quad B}{A \wedge B}\wedge_I \qquad\qquad \frac{A \wedge B}{A}\wedge_{E_1} \qquad\qquad \frac{A \wedge B}{B}\wedge_{E_2}$$

As already mentioned in the previous chapter, the inferentialist thesis with respect to the meaning of "and" is that it can be understood in terms of rules of this sort, which may apply in constructing various *proofs* of conclusions from premises in a natural deduction system. In this system, our proofs will take the form of *trees*. We are able to use any premises that we have as *leaves* on the tree, writing them down to conclude things from them with the use of our rules without having to conclude them from anything else. Thus, for instance, the proof of $(p \wedge q) \wedge r$ from $p \wedge r$ and $q$ goes as follows:[1]

$$\dfrac{\dfrac{\dfrac{p \wedge r}{p}\wedge_{E_1} \quad q}{p \wedge q}\wedge_I \quad \dfrac{p \wedge r}{r}\wedge_{E_2}}{(p \wedge q) \wedge r}\wedge_I$$

[1]As a notational convention, I will use uppercase letters such as $A$, $B$, and $C$ to stand for sentences of arbitrary complexity, whereas I will use lowercase letters $p$, $q$, and $r$ to stand for particular atomic sentences. Thus, in the statement of the rules, I use $A$ and $B$, whereas, in a concrete proof using these rules, I use $p$, $q$, and $r$.

Note there we can use the same premise as many times as we like. Here, we use $p \wedge r$ twice, as two different "leaves" of our proof tree, using it once to derive $p$ and a second time to derive $r$. We then combine these two "branches" with conjunction introduction to conclude $(p \wedge q) \wedge r$

Let's now turn to the rules for the *conditional*. Start with the introduction rule. How can we conclude a sentence of the $A \rightarrow B$? The key idea of natural deduction is that in addition to writing down a sentence to express that we *categorically assert* that sentence, we can also write down a sentence (in a special way) to express that we are *hypothetically assuming* that sentence. Some of the rules of a natural deduction system tell us that, if we are then able to conclude something given some assumption, we can *discharge* that assumption, and conclude something else categorically. The introduction rule for the conditional is of the form. It tells us that, in order to prove $A \rightarrow B$, we should hypothetically assume $A$, and then, given that assumption, try to prove $B$. If we can do that, we can discharge our assumption of $A$ (such that we henceforth no longer assume it), and (categorically, rather than merely hypothetically) conclude $A \rightarrow B$. In formal notation, it looks like this:

$$\dfrac{\begin{array}{c}\overline{A}^{\;n} \\ \vdots \\ B\end{array}}{A \rightarrow B} \rightarrow_I^{\,n}$$

Here, the "$n$" signifies the number of the assumption. In this system, when we make a new assumption, we write the formula as a leave of a tree with a line over it, and we signify that assumption with a new number. When we discharge that assumption, we write the number of the assumption we discharge next to the name of the rule we use to show that, downward on the proof tree, we will no longer be using that assumption. We'll consider a a proof wherein we can see this rule in action shortly. First, let us note the corresponding elimination rule, which is quite straightforward:

$$\dfrac{A \rightarrow B \quad A}{B} \rightarrow_E$$

The conditional elimination rule, known as *modus ponens*, tells us that if we have $A \rightarrow B$ and we have $A$ we can conclude $B$.

To see these rules together in action, along with the conjunction rules introduced earlier. Let us prove the conclusion $(p \wedge r) \rightarrow q$ from the premise $p \rightarrow q$:

$$\dfrac{\dfrac{p \rightarrow q \qquad \dfrac{\dfrac{}{p \wedge r}\,1}{A}\wedge_{E_1}}{B}\rightarrow_E}{(p \wedge r) \rightarrow q}\rightarrow_I{}^1$$

Here, we have our premise $p \rightarrow q$ as an underived "leaf" of our proof-tree. In addition, we hypothetically assume $A \wedge C$, the antecedent of the conditional that we want to prove, putting a horizontal line over it to indicate that we are making an assumption and numbering this assumption "1." Then, given this assumption, we use conjunction elimination, followed by conditional elimination to conclude $B$. Given that we've concluded $B$, given our hypothetical assumption of $A \wedge C$, we can now discharge this assumption and use the conditional introduction rule to conclude $(A \wedge C) \rightarrow B$, writing down the "1" next to the rule name to show that this assumption has been discharged. Note that, instead of this proof with a premise, we could have just as well proven the conditional $(A \rightarrow B) \rightarrow ((A \wedge C) \rightarrow B)$ from *no* premises, hypothetically assuming $A \rightarrow B$ (under a different numbered assumption) instead of having it as a premise, and then, in a final step beyond those shown here, discharging this assumption and concluding $(A \rightarrow B) \rightarrow ((A \wedge C) \rightarrow B)$ by way of conditional introduction.

Now, it's worth being clear that, though our conditional introduction rule *allows us* to hypothetical assume the antecedent of the conditional we want to prove and use that sentence in our hypothetical proof of the consequent of the conditional, we don't actually *have to* use the assumption of the antecedent of the conditional. Consider the following proof of $q \rightarrow p$ from $q$:

$$\dfrac{q}{q \rightarrow p}\rightarrow_I{}^0$$

This might look a bit strange. Where did the $q$ come from? It seems to have popped out of nowhere. Perhaps the best way to think about it is to think that it's *there*, but *hiding in the background* as an assumption that we *made* but *never used*. Writing "0" when we apply the conditional introduction rules indicates a *vacuous discharge* of an assumption. We made an assumption of $q$ and subsequently discharged it, without ever using it, and so this discharge was *vacuous*. Now, intuitively, the inference from $p$ to $q \to p$ might not seem like a very good one. After all, even if $p$ is the case, why should it follow that $p$ follows from $q$? The "goodness" of this inference, in the context of classical logic, is one of the "paradoxes of material implication." There are logics in which $q \to p$ does *not* follow from $p$ known as *relevant* logics, and one way to arrive a such a logic is by disallowing vacuous discharges in the context of a natural deduction system. Though there is a lot of to explore here, since our concern is classical logic, the natural deduction system we'll consider here will permit vacuous discharges like this.[2]

Turning now to disjunction, the disjunction introduction rules are very simple. From either disjunct, we can conclude the disjunction:

$$\frac{A}{A \vee B}\ \vee_{I_1} \qquad\qquad \frac{B}{A \vee B}\ \vee_{I_2}$$

The elimination rule for disjunction, however, is a bit tricky. The idea is that if we have a disjunction, we know that *at least one* of the two disjuncts must be true. So, if we can conclude the same thing the from each disjunct separately, we can conclude that thing, even without knowing which disjunct is actually true. So, the way the rule works is that we hypothetically assume each disjunct, and, try to prove some formula from each of these two assumptions separately If we can, then we can discharge those assumptions, and conclude the formula we were able to conclude from each. Thus, the rule is the following:

$$\dfrac{A \vee B \quad \begin{matrix}\overline{A}^{\,n} \\ \vdots \\ C\end{matrix} \quad \begin{matrix}\overline{B}^{\,m} \\ \vdots \\ C\end{matrix}}{C}\ \vee_{E}^{n,m}$$

To see these rules in action, consider the proof of $(p \vee q) \wedge (r \vee s)$ from $(p \wedge r) \vee (q \wedge s)$:

[2]Hopefully, these features of the material conditional of classical logic are familiar enough from an introductory course. If not, see Simonelli 49-53.

$$\dfrac{(p \wedge r) \vee (q \wedge s) \quad \dfrac{\dfrac{\dfrac{\overline{p \wedge r}^{\,1}}{p}\wedge_{E_1}}{p \vee q}\vee_{I_1} \quad \dfrac{\dfrac{\overline{p \wedge r}^{\,1}}{r}\wedge_{E_2}}{r \vee s}\vee_{I_1}}{(p \vee q) \wedge (r \vee s)}\wedge_I \quad \dfrac{\dfrac{\dfrac{\overline{q \wedge s}^{\,2}}{q}\wedge_{E_1}}{p \vee q}\vee_{I_2} \quad \dfrac{\dfrac{\overline{q \wedge s}^{\,2}}{s}\wedge_{E_2}}{r \vee s}\vee_{I_2}}{(p \vee q) \wedge (r \vee s)}\wedge_I}{(p \vee q) \wedge (r \vee s)}\vee_E^{1,2}$$

Here, we use each of our two assumptions twice in the two respective hypothetical proofs. It's important to keep in mind that, while we can use these assumptions as much as we like on the branches corresponding to each of the two hypothetical proofs, it's important that they remain on their respective branches.

We will put off the consideration of the rules for negation for the moment. Having stated the rules for conjunction, the conditional, and disjunction, we can now return to state the logical inferentialist thesis more concretely. Once again, the logical inferentialist thesis is that these rules, which together articulate the *conditions* under which a sentence of the form $A \wedge B$, $A \rightarrow B$, or $A \vee B$ can be concluded and the *consequences* that can be concluded from one of these sentences constitute the meanings of the logical operators they contain. So, to grasp the meaning of $\wedge$, $\rightarrow$, or $\vee$ is to grasp these inferential rules governing the use of them to conclude sentences in which they are the main operator and conclude things from such sentences. The claim of the inferentialist, qua "full-blooded" theorist of meaning, is that articulating these rules in this way really provides an account of the meanings of these connectives, as opposed to the "modest" specification of meaning that one provides in saying, for instance, "$A \wedge B$ is true if $A$ is true and $B$ is true" or "$A \vee B$ is true if $A$ is true or $B$ is true."

***Historical Note:***

I have talked primarily about Gentzen, and he is typically credited as the founder of proof-theoretic semantics in the context of natural deduction set-up. However, natural deduction was simultaneously developed and put forward independently by Stanisław Jaśkowski. Jaskowski's presentation of natural deduction is actually likely closer to the sort of style of proof that you likely

learned in an introductory course, where, rather than using trees and numbering assumptions, one draws a box around a formula to indicate assumptions.

Nowadays, the most common format for natural deduction in introductory courses, known as "Fitch-style" notation (from Friedrich Fitch), is based on Jaskowski's presentation, but one only draws a vertical line rather than the whole box around the assumption. In this notation, which might be familiar, the

| | | |
|---|---|---|
| 1 | $p \to q$ | prem. |
| 2 | $\quad \mid\ p \wedge r$ | asm. |
| 3 | $\quad \mid\ p$ | $\wedge_E$ 1 |
| 4 | $\quad \mid\ p \to q$ | reit. |
| 5 | $\quad \mid\ q$ | $\to_E$ |
| 6 | $(p \wedge r) \to q$ | $\to_I$ 1-2 |

Though this is easier to use, Gentzen's tree-style notation more clearly shows when assumptions are being used where and how often their being used, which permits more flexibility. Accordingly, proof-theorists and proof-theoretic semanticists typically (though not always) prefer Gentzen's notation.

## 2.3 Structural Features of Natural Deduction Consequence

Before we go on, it's worth taking a brief moment to zoom out and talk about the structure of this proof system as a whole. Let us first note explicitly that a proof system generates a *consequence relation* on the language. We'll express proof-theoretic consequence with the single turnstile symbol $\vdash$. If a sentence $A$ can be proven from $X$, we will write $X \vdash A$. We've shown above in our examples that a number of familiar logical consequences obtain in this proof system, for instance, that $p \wedge r, q \vdash (p \wedge q) \wedge r$ and $p \vdash q \to p$. However, beyond these specific logical consequences which obtain in virtue of the specific rules for the logical connectives we've provided, there are certain *structural* features of consequence that obtain simply in virtue of the structure of the proof system that we've laid out. In particular, consider the following three structural features:

**Reflexivity:** $A \vdash A$

**Monotonicity:** If $X \vdash A$, then $X, B \vdash A$

**Transitivity:** If $X \vdash A$ and $Y, A \vdash B$, then $X, Y \vdash B$

It is easy to see the the proof system we've laid out generates a consequence relation with all three of these structural features.

Let us consider first how the consequence generated by this proof system is *reflexive*. Consider just the following proof:

$$p$$

This is a proof of the conclusion $p$, from the premise $p$. Clearly, we can construct a proof of this sort for any sentence at all. Thus, for any sentence $A$, $A \vdash A$. In other words, the consequence relation generated by this proof system is reflexive. Turning to Monotonicity, note that, though we *can* use any premises that we have, we don't *have to* use all of them. So, if we can can prove some conclusion $A$ from some set of of premises $X$, then if we have $X$ along with some further premise $B$,we can still prove $A$, since we can just use the same proof of $A$ from $X$. In general, the consequence relation is monotonic: if $X \vdash A$, then $X, B \vdash A$. Finally, note that if we have a proof of $A$ from $X$ and a proof of $B$ from $Y$ along with $A$, then we can string these two proofs together to produce a proof of $B$ from $X$ along with $Y$ like so:

$$\begin{array}{cc} & X \\ & \vdots \\ Y & A \\ \vdots & \\ B & \end{array}$$

So, the consequence relation is transitive. These structural features of natural deduction consequence are likely obvious, but it is worth noting them explicitly now, since they will be important in the chapters that follow.

## 2.4 Tonk and Tonk

We've now introduced a natural deduction system for classical logic, and this is enough to state the basic inferentialist proposal about the meaning of logical vocabulary. We've provided rules that determine how one may infer to and from sentences containing $\wedge$, $\rightarrow$, and $\vee$ in the context of a proof system. To grasp the meanings of these symbols is to grasp the inferential rules governing their use, which we've specified precisely in terms of introduction and elimination rules in a natural deduction system. Now, it might seem that, for any symbol that we might introduce in accordance with a definite set of introduction and elimination rules, we can think of it as having a particular meaning, one constituted by those inferential rules. A.N. Prior, however, famously showed that we should not be so quick to think that just any set of inferential rules should be counted as constituting a legitimate meaning.

If we think that any set of rules constituted a legitimate meaning, then it seems like there would be no problem with enriching our language with new connectives, with different meanings than the one's we've considered so far, governed by different rules. After all, insofar as there's a legitimate meaning of a German word like "schadenfreude" or a theoretical term like "black hole," there's no problem with enriching English language with it. We might not use some of these terms very often, but, insofar as they have legitimate meanings, there's no reason not to include them in the language. The same, it seems, should be true of logical connectives. Consider, then, the connective that Prior dubbed "tonk," governed by the following introduction and elimination rules:[3]

$$\frac{A}{A \textbf{ tonk } B}\ \textbf{tonk}_{I_1} \qquad\qquad \frac{B}{A \textbf{ tonk } B}\ \textbf{tonk}_{I_2}$$

$$\frac{A \textbf{ tonk } B}{A}\ \textbf{tonk}_{E_1} \qquad\qquad \frac{A \textbf{ tonk } B}{B}\ \textbf{tonk}_{E_2}$$

As you can see **tonk** has the *introduction rules* of *disjunction*, but the *elimination rules* of *conjunction*. So, from $A$, you can infer $A$ **tonk** $B$, and likewise, from $B$, you can infer $A$ **tonk** $B$ , and from $A$ **tonk** $B$ you can infer $A$ and you can also infer

[3]These are, in fact, slightly modified from Prior's original to make them a bit more symmetric. In his original paper, Prior only included $\textbf{tonk}_{I_1}$ and $\textbf{tonk}_{E_2}$, but the same point applies.

$B$. What would happen if we added the connective "tonk" to our language? Quite disastrously, we'd be able to conclude anything and everything! It is quite easy to see that, with the use of these rules, we can introduce a sentence of the form $A$ **tonk** $B$ and subsequently eliminate it to conclude anything from anything else. Concretely, where $p$ and $q$ are arbitrary simple sentences, we can reason as follows:

$$\dfrac{\dfrac{p}{p \textbf{ tonk } q}\ \textbf{tonk}_{I_1}}{q}\ \textbf{tonk}_{E_2}$$

This is obviously a problem. For instance, just suppose $p$ is the sentence "1 + 1 = 2" and $q$ is the sentence "2 + 2 = 5." Clearly, we should not be able to conclude, from the obvious truth that 1+1=2, the obvious falsehood that 2+2=5! Clearly, *something* has gone wrong. But *what* has gone wrong? Prior's challenge to the inferentialist is to say what's defective about these inferential rules such that they don't constitute a legitimate meaning, whereas the inferential rules constitutive of the familiar logical operators *do* constitute legitimate meanings.

Intuitively, it's not hard to state the basic problem with **tonk**: the introduction and elimination rules for **tonk** don't fit together in the nice way that, for instance, the introduction and elimination rules for conjunction fit together. To use a term from Michael Dummett, the introduction and elimination rules don't fit together "harmoniously." In particular, the elimination rules for **tonk** are *too strong*, relative to the introduction rules. Given how little we need to "put in" in order to conclude *A tonk B* via the introduction rules (namely, just $A$ or just $B$), the elimination rules let us "get out" too much. More concretely, insofar as we can conclude both $A$ and $B$ from *A tonk B*, we should need to have both in order to conclude *A tonk B* in the first place. What we are able to deduce, via elimination rules for some connective, shouldn't be more than we are able to deduce from the premises we need to apply the introduction rules for that connective. So, insofar as just $A$ or just $B$ suffices to conclude *A tonk B*, being able to infer both from *A tonk B* violates this constraint.

Another very closely related (but in fact distinct) way to think about the problem with **tonk** is considering that, by introducing **tonk** into a language that doesn't contain already **tonk**, one will be able to prove sentences of that original language that one could not already prove. Contrast this with adding a harmless connective into the language, for instance, the biconditional $\leftrightarrow$, governed by suitable rules (which I'll have you formulate in the exercises). Of course, if one adds $\leftrightarrow$ into the formal language, one will be able to prove formulas one could not before. For instance, one will be able to prove $(p \wedge q) \leftrightarrow (q \wedge r)$. Crucially, however, all of these new sentences one will be able to prove *will contain this new connective*. One won't be able to prove, for instance, $p \wedge q$, a sentence does not contain any biconditionals. We express this by saying that adding a biconditional (governed by suitable rules) to our language constitutes a *conservative extension* of this language. In general terms, the addition of a new connective, governed by a certain set of rules, to a language constitutes *conservative extension* of that language just in case there no sentences of the original language that can be proven (perhaps from other sentences of the original language) in the expanded language that could not already be proven in the original language. Officially, where $\vdash_{\mathcal{L}}$ denotes the provability relation in a language $\mathcal{L}$:

> **Conservativity:** $\mathcal{L}_2 \supseteq \mathcal{L}_1$ is a conservative extension of $\mathcal{L}_1$ just in case, where $A \in \mathcal{L}_1$ and $X \subseteq \mathcal{L}_1$, if $X \vdash_{\mathcal{L}_2} A$, then $X \vdash_{\mathcal{L}_1} A$.

Thus, another way to say what's wrong with **tonk** is that it's not *conservative*. In the next section, we will focus on harmony, but we will return to conservativity shortly.

Now, before we attempt to formulate a criterion of "harmony" that will rule out connectives like **tonk**, it's worth considering the *opposite* of **tonk**: a connective that Nissim Francez [**?**] has dubbed **tunk**.[4] Whereas **tonk** has the introduction rules of disjunction and the elimination rules of conjunction, **tunk** has the introduction rules of conjunction and the elimination rules of disjunction:

$$\frac{A \quad B}{A \mathbin{\textbf{tunk}} B}\ \textbf{tunk}_I \qquad\qquad \frac{A \mathbin{\textbf{tunk}} B \quad \begin{matrix}\overline{A}^{\,n} \\ \vdots \\ C\end{matrix} \quad \begin{matrix}\overline{B}^{\,m} \\ \vdots \\ C\end{matrix}}{C}\ \textbf{tunk}_E^{\,n,m}$$

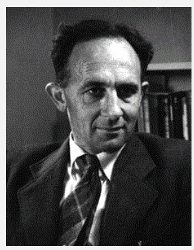
[4]"Tunk" is closely related to Belnap's *plonk*, which has the introduction rule of disjunction and doesn't have any elimination rule at all. Once again, it seems to me that picking a connective that has both introduction and elimination rules is a better way to get to the core of the problem.

Unlike **tonk**, **tunk** doesn't trivialize our logical system, leting us prove anything from anything else. Indeed, we can add it to our language and it won't let us prove any sentence that doesn't contain it that we couldn't already prove. In that way, it's clearly not *as bad* as **tonk**. Still, it seems similarly defective, just defective in an opposite way. Intuitively, whereas the elimination rules for **tonk** let us get out *too much*, given *how little* we need to put in to conclude a sentence of the form $A$ **tonk** $B$ via the introduction rules, the elimination rules for **tunk** let us get out *too little*, given *how much* we need to put in to conclude a sentence of the form $A$ **tunk** $B$ via the introduction rules. In other words, whereas the elimination rules of **tonk** are *too strong* relative to the introduction rules, the elimination rules of **tunk** are *too weak* relative to the introduction rules.

What we need, then, is a criterion of harmony that rules out both **tonk** and **tunk**, ensuring that the introduction rules and elimination rules provided for a given connective fit together just right, with the elimination rules being neither too strong nor too weak relative to the introduction rules. The criterion I'll introduce in the next section, drawn from Pfenning and Davies (expanding on work by Prawitz), is what I'll call *the method of reductions and expansions*.

***Historical Note:***

The connective *tonk* was proposed by Arthur Prior, in his 1960 paper "The Runabout Inference Ticket," just a two-page paper published in the journal *Analysis*. Prior presented the connective as a challenge to theories of meaning that think of the meaning of a logical connective in terms of its proof rules. While Prior's paper did not convince many inferentialists that inferentialism was incorrect, it did show them that they needed to put down constraints on what sorts of inferential rules should count as defining meanings. Shortly after, in an almost equally famous reply, Nuel Belnap proposed the criterion of conservativity as a way of ruling out connectives like *tonk*.



## 2.5 Harmony: Reductions and Expansions

The method of reductions and expansions involves consideration of two kinds of *detours* that might occur in a proof:

1. A detour in which a sentence with a given connective is first introduced via the introduction rules for that connective and subsequently eliminated via the elimination rules.
2. A detour in which a sentence with a given a connective is first eliminated via the elimination rules for that connective and subsequently re-introduced via the introduction rules.

A *reduction* (or a set of a such reductions) shows that detours of the first sort are *never necessary*. That is, in any case in which we can *introduce* a connective and subsequently *eliminate* it to conclude some formula, we could just as well go *directly* from the premises to the conclusion, without the detour of introducing and subsequently eliminating the connective. Such a reduction shows that the elimination rules are not *too strong*, relative to the introduction rules: we can only get out, via the elimination rules, what we already had the resources to derive, given what we put into the introduction rules. An *expansion* (or set of such expansions) shows that detours of the second sort are *always possible*. That is, in any case in which we have some derivation of some connective with some connective and we use the elimination rules, we can then use the introduction rules to recover the sentence we originally had. Such an expansion shows that the elimination rules are not *too weak*, relative to the introduction rules: we can always get enough out, via the elimination rules, to derive what we need to put into the introduction rules to derive the sentence again.

Let us start with reductions. In general, the procedure for a reduction is as follows:

**Reduction Procedure:**

**Step 1:** Assume you have whatever derivations you need in order to apply the introduction rules and subsequently apply the elimination rules to derive some sentence.

**Note:** These could simply be derivations of the premises of an introduction or elimination rule or they could be derivations from sentences to other sentences needed in the context of the hypothetical proof featured in an introduction rule.

**Step 2:** Show that you can use these derivations to prove the ultimately derived sentence *directly*, without the detour through introducing and eliminating the connective.

**Step 3:** Repeat for each possible combination of introduction and elimination rules.

For conjunction, the reduction goes as follows:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ A\end{matrix} \quad \begin{matrix}\mathcal{D}_2\\ B\end{matrix}}{A \wedge B}\wedge_I}{A}\wedge_{E_1} \qquad \rightsquigarrow_r \qquad \begin{matrix}\mathcal{D}_1\\ A\end{matrix}$$

So, to use the introduction rule, we have to assume we have a derivation of $A$, which we name $\mathcal{D}_1$, and that we have a derivation of $B$, which we name $\mathcal{D}_2$. We now show the detour in which we use these derivations to first introduce $A \wedge B$ and then eliminate it using the first elimination rule to conclude $A$. We then sow the reduction in which we conclude $A$ directly by way of $\mathcal{D}_1$. Likewise, if we had instead used the second elimination rule to conclude $B$, we could have just as well concluded $B$ directly by way of $\mathcal{D}_2$.

Let us now consider the conditional. In order, to use both introduction and elimination rules, we need two derivations. First we need a derivation $\mathcal{D}_1$ from $A$ to $B$, which will occur in the context of our hypothetical proof, in order to apply our introduction rule to derive $A \to B$. Second, we need a derivation $\mathcal{D}_2$ of $A$ in order to apply our elimination rule to derive $B$. Our reduction shows, then, that, given these two derivations, we can derive $B$ directly, without the detour through the introduction and elimination of a conditional sentence:

$$\dfrac{\begin{matrix}\mathcal{D}_2\\ A\end{matrix} \quad \dfrac{\begin{matrix}\overline{A}^{\,1}\\ \mathcal{D}_1\\ B\end{matrix}}{A \to B}\to_I^{1}}{B}\to_E \qquad \rightsquigarrow_r \qquad \begin{matrix}\mathcal{D}_2\\ A\\ \mathcal{D}_1\\ B\end{matrix}$$

Similarly, the reduction for the disjunction, using the first introduction rule, goes as follows:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ A\end{matrix}}{A \vee B}\,\vee_{I_1} \qquad \begin{matrix}\overline{A}^{\,1}\\ \mathcal{D}_2\\ C\end{matrix} \qquad \begin{matrix}\overline{B}^{\,2}\\ \mathcal{D}_3\\ C\end{matrix}}{C}\,\vee_{\mathrm{E}}^{1,2} \qquad \rightsquigarrow_r \qquad \begin{matrix}\mathcal{D}_2\\ A\\ \mathcal{D}_1\\ B\end{matrix}$$

The reduction using the second introduction rule is exactly analogous.

Together, these reductions show that none of the elimination rules we've provided for conjunction, the conditional, or disjunction, are too strong, relative to the introduction rules. To see this, consider now **tonk**. Combining the first introduction rule and the second elimination rule, we have the following attempt at a reduction:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ A\end{matrix}}{A\ \textbf{tonk}\ B}\,\textbf{tonk}_{I_1}}{B}\,\textbf{tonk}_{E_2} \qquad \rightsquigarrow_r \qquad ???$$

Clearly, there is no way to use $\mathcal{D}_1$ to derive $B$ directly, without going through the introduction and elimination of **tonk**. Thus, the elimination rules for **tonk** let us *get out* something that we didn't *put in* when using the introduction rules, namely, a derivation of $B$. In the same way, if we use the second introduction rule and the first elimination rule can can derive $A$ given only a derivation of $B$, and, once again, no reduction can be given. Thus, we can rule out **tonk** as a connective whose meaning is given by disharmonious rules on the grounds that we cannot provide a reduction for some combinations of introduction and elimination rules.

Let us now turn to expansions. In general, the procedure for an expansion is the following:

**Expansion Procedure:**

> **Step 1:** Assume you have a derivation of a sentence with that connective.
>
> **Step 2:** Show that you can use this derivation to first *eliminate* the connective, via the elimination rules, and then *reintroduce* it, via the introduction rules.

- **Note:** This will often making assumptions of the premises needed for application of the elimination rules, but, if the rules are harmonious, you should be able to discharge those assumptions.

Alternatively, you can show that you use the elimination rule to reintroduce the connective.

**Step 3:** Repeat for each possible combination of introduction and elimination rules.

For conjunction, the expansion goes as follows:

$$\begin{array}{c}\mathcal{D}_1\\ A \wedge B\end{array} \quad \rightsquigarrow_e \quad \dfrac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ A \wedge B\end{array}}{A}\wedge_{E_1} \qquad \dfrac{\begin{array}{c}\mathcal{D}_1\\ A \wedge B\end{array}}{B}\wedge_{E_2}}{A \wedge B}\wedge_I$$

Note here it's essential here that we have *both* elimination rules. If we only had one of the two elimination rules, we would not be able to regather the materials needed to reintroduce the conjunction. The expansion for the conditional goes as follows:

$$\begin{array}{c}\mathcal{D}_1\\ A \to B\end{array} \quad \rightsquigarrow_e \quad \dfrac{\dfrac{\overline{A}^{\,1} \qquad \dfrac{\mathcal{D}_1}{A \to B}}{B}\to_E}{A \to B}\to_I{}^1$$

Finally, the expansion for disjunction goes as follows:

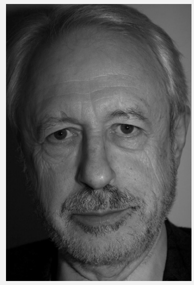
$$\begin{array}{c}\mathcal{D}_1\\ A \vee B\end{array} \quad \rightsquigarrow_e \quad \dfrac{\begin{array}{c}\mathcal{D}_1\\ A \vee B\end{array} \qquad \dfrac{\overline{A}^{\,1}}{A \vee B}\vee_{I_1} \qquad \dfrac{\overline{B}^{\,2}}{A \vee B}\vee_{I_2}}{A \vee B}\vee_E{}^{1,2}$$

Here, this expansion looks a bit different, since we are applying the introduction rules *first* inside the hypothetical proof for the elimination rule, and *then* applying the elimination rule to recover $A \vee B$. Such an expansion, however, still shows that the elimination rules are not too weak, relative to the introduction rule.

With the above expansion in view, consider again **tunk**. The elimination rule for **tunk** is the same as the one for disjucntion. However, unlike disjunction where we only need *either* $A$ or $B$ to conclude $A \vee B$, and thus, can conclude $A \vee B$ under both the assumption of $A$ and the assumption of $B$, for **tunk** we need *both* $A$ and $B$ to conclude $A$ **tunk** $B$, and so there is no way to derive this formula in the context of the subproofs of the elimination rule to re-introduce $A$ **tunk** $B$. This shows that **tunk** is disharmonious, but in the opposite way as **tonk**: the elimination rules are *too weak* relative to the introduction rules.

***Historical Note:***

The core idea of the approach to harmony developed here is owed primarily to the work of Daq Prawitz. In his 1965 book *Natural Deduction*, Prawitz articulated the idea of reductions shown considered here under the heading of the "inversion principle": the idea that applying elimination rules could only restore what one already had in applying the introduction rules. Prawitz used these reductions to prove a normalization theorem, the idea that a proof that contains detours of the first sort could always be turned into a proof in "normal form" which doesn't involve any detours. As we'll see, this theorem is analogous in the natural deduction setting to Gentzen's famous "Cut Elimination" proof for the sequent calculus, which we'll discuss in the next chapter.



In the tradition of proof-theoretic semantics largely owed to Prawitz and Dummett, sometimes, "harmony" is used just to refer to the property established by reductions. Dummett used the term "stability" to refer to opposite property, established by expansions. In contrast to this usage, we'll continue to use "harmony" to refer to the property of the elimination rules being neither too strong nor too weak relative to the introduction rules, established by both reductions and expansions.

## 2.6 The Problem with Classical Negation

We've given rules for conjunction, disjunction, and the conditional, and we've shown that, unlike the rules for **tonk** and **tunk**, the introduction and elimination rules for these connectives fit together harmoniously, with the elimination rules being neither too strong nor too weak relative to the introduction rules. Finally, let us turn to the negation rules. Now, once again, the rules we want are those constitutive of the meaning of *classical* negation. We are thinking of the meaning of classical negation in terms of the inferential rules governing its use, and so let us consider the inferential behavior of classical negation that we want our rules to capture. One important feature of classical negation is that, in classical logic, anything follows from a contradiction. That is, for any sentences $A$ and $B$, we have $A, \neg A \vdash B$. Another feature of classical negation is that it is an *involutive* operator, meaning that applying it to some sentence twice is the same as not applying it at all. That is, for any sentence $A$, we have $A \vdash \neg\neg A$ and $\neg\neg A \vdash A$. Moreover, if we consider the interaction of classical negation with other connectives, two crucial theorems partly constitutive of classical negation are the law of non-contradiction, $\vdash \neg(p \wedge \neg p)$, and the law of excluded middle $\vdash p \vee \neg p$. Let us consider how we can provide harmonious introduction and elimination rules for negation that yield this inferential behavior.

First, how can we conclude a formula of the form $\neg A$? The basic idea is that we hypothetically assume $A$, and if, given this assumption, we can arrive at a contradiction—a pair of formulas of the form $B$ and $\neg B$—we can discharge that assumption, and conclude $\neg A$. Here is the rule:

$$\dfrac{\begin{matrix}\overline{A}^{\,n} \\ \vdots \\ B\end{matrix} \qquad \begin{matrix}\overline{A}^{\,n} \\ \vdots \\ \neg B\end{matrix}}{\neg A}\;\neg_I{}^{n}$$

To see this rule in action, consider the proof of $\neg(p \wedge \neg p)$:

$$\dfrac{\dfrac{\overline{p \wedge \neg p}^{\,1}}{p}\;\wedge_{E_1} \qquad \dfrac{\overline{p \wedge \neg p}^{\,1}}{\neg p}\;\wedge_{E_2}}{\neg(p \wedge \neg p)}\;\neg_I{}^{1}$$

Here, we make use of the same assumption twice, first, to derive both $p$, and second, to derive $\neg p$. Once we derive this pair of contradictory sentences, we discharge this assumption and conclude $\neg(p \wedge \neg p)$. The standard negation elimination rule for a classical natural deduction system is a bit strange. It says that if we have a *double* negation, we can eliminate both of them:

$$\frac{\neg\neg A}{A}\ \neg_E$$

To see both rules together in action, consider the proof of $q$ from $p \wedge \neg p$:

$$\dfrac{\dfrac{\dfrac{p \wedge \neg p}{p}\ \wedge_{E_1} \quad \dfrac{p \wedge \neg p}{\neg p}\ \wedge_{E_2}}{\neg\neg q}\ \neg_I{}^0}{q}\ \neg_E$$

Here, once again, we assumed $\neg q$, but never actually used it in the proof, so we write down 0 to indicate a vacuous discharge.

These rules for negation, along with the rules for conjunction, disjunction, and the conditional we've given above, give us a proof system for classical logic. That is, we can prove a sentence $A$ from some set of sentences $X$ just in case $A$ is a classical consequence of $X$ (in the semantic sense of "consequence" you might establish with truth-tables). It might seem, then, that we've now given a inferentialist, proof-theoretic account of the meanings of the logical connectives belonging to classical logic. There are two major problems. First, the negation rules that we've given *contain negation*. For instance, the negation introduction rules tells us that if, given the assumption of some sentence, we can prove some other sentence and it's negation, then we can conclude the negation of the sentence we assumed. Even though such a rule is clearly sound for classical negation, it's not clear that we should be able to appeal to it in order to provide a non-circular *account* of negation. It seems that we're *appealing* to negation here in the statement of this rule which is supposed to *define* negation. We will consider this sort of problem in the next section. In this section we'll focus on the second problem, which is more serious.

In the previous section we provided reductions for the rules for conjunction, the conditional, and disjunction, showing that the elimination rules were not too strong relative to the introduction rules. Let us now try to provide such a reduction for the negation rules we've just stated. First, to apply the introduction rule and subsequently apply the elimination rule, we need to apply use the introduction rule to $\neg\neg A$. So, we need a derivation going from $\neg A$ to some pair of contradictory sentences. So the proof that needs to be reduced looks like this:

$$\dfrac{\dfrac{\begin{matrix}\overline{\neg A}^{\,1} \\ \mathcal{D}_1 \\ B\end{matrix} \qquad \begin{matrix}\overline{\neg A}^{\,1} \\ \mathcal{D}_2 \\ \neg B\end{matrix}}{\neg\neg A}\,\neg_I{}^1}{A} \qquad\qquad \rightsquigarrow_r \qquad\qquad ???$$

A reduction would have show that we could use $\mathcal{D}_1$ and $\mathcal{D}_2$ to get *directly* to $A$, without the detour through the negation rules. Clearly, however, these derivation, which lets us go from $\neg A$ to $B$ and from $\neg A$ to $\neg B$ are not going to enable us to prove $A$ directly, without using the negation rules to do so. These rules, then, are unharmonious.

Now, in the case of **tonk**, where we could not provide reductions, we noted that this connective was *nonconservative*—adding it to a language that does not contain this connective enables us to prove things in that origin language that we previously couldn't prove (since, of course, it enables us to prove anything). In fact, the same is true about these negation rules. Introducing a negation operator, governed by these rules to the language stated in the first section of this chapter, does *not* constitute a conservative extension of that language. To see this, consider the formula $((p \to q) \to p) \to p$, known as "Pierce's Law" after the philosopher and logician C.S. Pierce. You can check by constructing a truth-table that this is, in fact, a tautology in classical logic. However, given just the rules for the conditional we introduced above, we cannot prove it. Yet, having introduced the rules for negation that we have now introduced, we now can:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overline{(p \to q) \to p}^{\,3} \qquad \dfrac{\dfrac{\dfrac{\overline{p}^{\,1} \quad \overline{\neg p}^{\,2}}{\neg\neg q}\,\neg_I{}^{0}}{q}\,\neg_E}{p \to q}\,\to_I{}^{1}}{p}\,\to_E \qquad \overline{\neg p}^{\,2}}{\neg\neg p}\,\neg_I{}^{1}}{p}\,\neg_E}{((p \to q) \to p) \to p}\,\to_I{}^{3}$$

Since $((p \to q) \to p) \to p$ contains only expressions belonging to the *original* language (it does not contain any negation operators), and we previously couldn't prove it but now we can, this means that the rules by which we've introduced negation are not conservative.

This is a problem. We discounted the connective **tonk** on the grounds that we could not provide a reduction for it and, relatedly, that adding it to a language constitutes a nonconservative extension of the language. We've now seen that precisely the same is the case about negation, as defined by the standard introduction and elimination rules. Of course, unlike **tonk** these rules for negation don't *trivialize* the language, allowing us to prove anything but anything else. Nevertheless, the rules for negation still don't fit together harmoniously, and we've stated that harmony is a non-negotiable criterion for defining a legitimate logical connective. To resolve this problem, it seems to we need to adopt a different approach towards the rules for negation.

## 2.7 The Priority of Minimal and Intuitionistic Negation

Let us start fresh. We'll introduce negation with the help of a special symbol, $\perp$, expressing the occurrence of a contradiction. Now, there are different ways of conceiving this symbol, with both technical and philosophical implications. The standard way of conceiving it is as a 0-place logical connective. That is, whereas $\wedge$ is a 2-place connective which needs to be conjoined with two sentences to constitute a complete formula, and $\neg$ is a 1-place connective, which needs to be concatenated with *one* sentence to constitute a complete formula, $\perp$ is a 0-place connective, which needs to be concatenated with *no* sentences to constitute a complete formula. Conceiving of it this way, it is an embeddable sentence just like any other sentence of the language, and, indeed, negation is often defined in intuitionistic logic as $A \to \bot$. However, if we conceive of it in this way, then we're going to be tasked with giving it harmonious introduction and elimination rules like any other connective, and this is a notoriously hard task. In this book, we will follow Tennant [12], and conceive of this symbol not as a formula, but as a *punctuation mark*, like a special kind of exclamation mark that constitutes part of the structure of our proof system. Understood in this way, like the horizontal line we can place over a formula to indicate an assumption, or the horizontal line we can place under a formula to indicate an inference, we can use the symbol $\bot$ in the context of a proof to indicate absurdity. The crucial distinction is that, unlike a sentence of a language, $\bot$ cannot be embedded in sentences. You can think of $\bot$ as something sort of akin to the English exclamation "Oh no!" which marks the realization of bad event but does not *say* there is some bad event. Notably, the exclamation "Oh no!" cannot really be embedded in English sentences. Whereas "Either $p$ or something bad has happened" is a grammatical sentence, "Either $p$ is true or *Oh no!*" is not (rather than having said a complete sentence, it sounds like I've been cut off from finishing my sentence). Likewise, $\bot$ cannot be embedded in formulas.

The negation introduction rule will now say that if assuming some formula $A$, we are able to arrive at a contradiction (expressed by our new $\bot$ symbol), we can conclude $\neg A$. The negation elimination rule, on the other hand, says that from $A$ along with $\neg A$, we can conclude $\bot$:

$$\dfrac{\begin{matrix}\overline{A}^{\,n} \\ \vdots \\ \bot\end{matrix}}{\neg A}\ \neg_I{}^n \qquad\qquad \dfrac{A \quad \neg A}{\bot}\ \neg_E$$

Now, even if we're clear about the fact that $\bot$ is a structural feature of our proof system and not another logical formula, it's still not too clear to what extent this new approach resolves our first problem. Though negation no longer appears at any place in our introduction rule other than the conclusion, the use of the $\bot$ symbol, understood as expressing the occurrence of a contradiction, still seems to be understood in terms of negation. But then negation is introduced in terms of the $\bot$ symbol. So, it still seems that there's a kind of circularity in our proof-theoretic account of the meaning of negation. Is it a vicious circularity? Perhaps not. Perhaps ¬ and ⊥ are just reciprocally sense-dependent, and thus, are to be defined in relation to one another. You can ruminate on this point, and we'll return to it in the chapter after next, but I'll bracket it for now. The important point is that these new rules *do* enable us to provide a reduction:

$$\dfrac{\begin{matrix}\mathcal{D}_2\\ A\end{matrix}\qquad \dfrac{\begin{matrix}\overline{A}^{\;n}\\ \mathcal{D}_1\\ \bot\end{matrix}}{\neg A}\,\neg_I{}^1}{\bot}\,\neg_E \qquad\qquad \rightsquigarrow_r \qquad\qquad \begin{matrix}\mathcal{D}_1\\ A\\ \mathcal{D}_1\\ \bot\end{matrix}$$

Likewise, we can give an expansion as follows:

$$\begin{matrix}\mathcal{D}_1\\ \hline \neg A\end{matrix} \qquad\qquad \rightsquigarrow_e \qquad\qquad \dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ \neg A\end{matrix}\qquad \overline{A}^{\;1}}{\bot}\,\neg_E}{\neg A}\,\neg_I{}^1$$

So these new negation rules, unlike the original ones we gave, are harmonious. The problem is that we're aspiring to an inferentialist account of the meanings of the *classical* connectives, and these negation rules do not give us classical logic.

There are all sorts of classical tautologies that are not provable with just these negation rules. For instance, we cannot prove $(p \wedge \neg p) \rightarrow q$. The logic we have is called *minimal* logic, which is just intuitionistic logic (which is likely familiar to you) but without the principle of explosion. If we want intuitionistic logic, we can add the following rule pertaining to the ⊥ symbol:

$$\dfrac{\bot}{A}\ \text{Explo.}$$

Once again, insofar as we're thinking about the rules governing ⊥ as structural features of our proof system rather than rules governing a connective like ¬, we can plausibly add this rule without issue. Accordingly, we can arrive at a natural deduction system for intuitionistic logic with harmonious introduction and elimination rules for all the connectives. However, we still don't have classical logic. Consider, for instance, $p \vee \neg p$. This is a classical tautology, and we could prove it with our original negation rules like so:

$$\dfrac{\dfrac{\dfrac{}{\neg(p \lor \neg p)}{}^{1} \qquad \dfrac{\dfrac{\dfrac{}{\neg(p \lor \neg p)}{}^{1} \quad \dfrac{\overline{p}^{\,2}}{p \lor \neg p}\lor_{I_1}}{\neg p}\neg_I{}^{2}}{p \lor \neg p}\lor_{I_2}}{\neg\neg(p \lor \neg p)}\neg_I{}^{1}}{p \lor \neg p}\neg_E$$

However, it cannot be proven with just these rules.

The system for intuitionistic logic that we have is what Gentzen called "NJ." To get classical logic, we can take the system for minimal logic and simply add double-negation elimination as a further rule. Then we'll have the system that Gentzen called "NK." But that's not the only option. Rather than double negation elimination, we could also the first of the following two rules to the system for minimal logic or we could add the second to the system for intuitionistic logic:

$$\begin{array}{c} \overline{\neg A}^{\,n} \\ \vdots \\ \dfrac{\bot}{A}\text{ Cl. Reduc.}^{n} \end{array} \qquad\qquad \dfrac{\begin{array}{c} \overline{A}^{\,n} \\ \vdots \\ C \end{array} \quad \begin{array}{c} \overline{\neg A}^{\,m} \\ \vdots \\ C \end{array}}{C}\text{ Ex. Mid.}^{n}$$

Alternately, rather than adding further negation rules, one could add a new rule for the conditional, known as "Peirce's Rule," to the system for intuitionistic logic:

$$\begin{array}{c} \overline{A \to B}^{\,n} \\ \vdots \\ \dfrac{A}{A}\text{ Peirce}^{n} \end{array}$$

As you can see, this is the *rule* version of Peirce's Law, $((A \to B) \to A) \to A$, which we proved above with our negation rules. So, there are a number of ways to get from minimal or intuitionistic logic to classical logic by adding further rules governing the behavior of negation or the conditional. However, the core issue is that we've already *given* harmonious rules for these connectives, so introducing one of these rules would be an unharmonious add-on. Insofar as we think that harmonious rules are constitutive of the meanings of the connectives, the distinctively classical inferences we make simply won't follow from the meanings of the connectives, but, rather, will be stipulated add-ons.

One response to this issue, adopted by Dummett and Prawitz, is to maintain that insofar as one is an *inferentialist*, one ought to be an *intuitionist* (or, alternatively, a *minimal logician*). By my lights, however, this inference hinges on a far too restrictive conception of the possibilities for proof systems in which an inferentialist account of the classical connectives might be given. Why think natural deduction systems, of the sort laid out in this chapter, are the only proof systems with which one can provide an inferentialist account of the meanings of the classical connectives? In the following three chapters, we'll see that there are other. Before that, however, let us briefly consider some more problematic connectives and some further proof-theoretic constraints on rules that we might want to impose in what follows.

## 2.8 Further Proof-Theoretic Constraints

Reductions and expansions are our criterion of *harmony* between the introduction and elimination rules. Harmony is the most important proof-theoretic criterion for meaningfulness. However, as we've defined it, harmony by itself is not sufficient to ensure *conservativity*. We've considered two nonconservative connectives, **tonk** and classical negation, as defined by the standard rules. We've shown that the rules for these connectives are unharmonious in that the elimination rules are too strong relative to the introduction rules, as demonstrated by the fact that reductions cannot be provided. So, in these cases, failures of conservativity aligned with failures of reductions. Moreover, all the connectives for which the *have* given reductions have been conservativtity. One might think, then, that the ability to provide reductions for a connective and that connective's being conservative are one in the same. However, this is not so. We can define connectives that have harmonious rules, in the sense we've defined, but, like **tonk**, are nonconservative.

Let us suppose that $p$ is some random false atomic sentence, "Neptune is yellow," let's say. Now, consider the following rules, inspired by Stephen Read [7], for introducing and eliminating a 0-place connective $\bullet$:

$$\dfrac{\bullet \to p}{\bullet}\,\bullet_I \qquad\qquad \dfrac{\bullet \qquad \begin{matrix}\overline{\bullet \to p}^{\,n}\\ \vdots\\ A\end{matrix}}{A}\,\bullet_E{}^{n}$$

This is a strange connective, to be sure, but, notably, we *are* capable of providing a reduction for these rules:

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ \bullet \to p\end{matrix}}{\bullet}\,\bullet_I \qquad \begin{matrix}\overline{\bullet \to p}^{\,1}\\ \mathcal{D}_2\\ A\end{matrix}}{A}\,\bullet_E{}^{1} \qquad \rightsquigarrow_r \qquad \begin{matrix}\mathcal{D}_1\\ \bullet \to p\\ \mathcal{D}_2\\ A\end{matrix}$$

We also have an expansion, which is similar to our expansion for disjunction rules:

$$\begin{matrix}\mathcal{D}_1\\ \bullet\end{matrix} \qquad \rightsquigarrow_e \qquad \dfrac{\begin{matrix}\mathcal{D}_1\\ \bullet\end{matrix} \qquad \dfrac{\overline{\bullet \to p}^{\,1}}{\bullet}\,\bullet_I}{\bullet}\,\bullet_E{}^{1}$$

So, by our official criterion, it would seem that the introduction and elimination rules for $\bullet$ are *harmonious*. Intuitively, they *are* harmonious: the elimination rule only lets us get out from $\bullet$ what we need to put into the introduction rule to derive $\bullet$. However, adding this connective with these rules to our language does *not* constitute a conservative extension of that language.

Let us suppose that our original language contains the atomic sentence $p$ and that $p$ is underivable in this language. Now, with the rules for $\bullet$ we can derive $p$ as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overline{\bullet}^{\,2} \qquad \dfrac{\overline{\bullet}^{\,2} \qquad \overline{\bullet \to p}^{\,1}}{p}}{p}\,\bullet_E{}^{1}}{\bullet \to p}\,\to_I{}^{2}}{\bullet} \qquad \dfrac{\dfrac{\overline{\bullet}^{\,4} \qquad \dfrac{\overline{\bullet}^{\,4} \qquad \overline{\bullet \to p}^{\,3}}{p}}{p}\,\bullet_E{}^{1}}{\bullet \to p}\,\to_I{}^{4}}{p}\,\to_E$$

The issue raised by this connective is very closely related to a famous paradox known as *Curry's Paradox*, which is paradox having to do with self-reference quite similar to the more famous *Liar Paradox*, exemplified by the sentence "This sentence is not true." Indeed, if we allow rules to define paradoxical sentences like the liar, we might just consider these rules for a 0-place connective $\star$:

$$\dfrac{\neg\star}{\star}\,\star_I \qquad\qquad\qquad \dfrac{\star}{\neg\star}\,\star_E$$

These rules are *clearly* harmonious, and yet, if you add them the system for intuitionistic logic, you can prove anything like so:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\overline{\star}^{\,1} \quad \dfrac{\overline{\star}^{\,1}}{\neg\star}\,\star_E}{\bot}\,\neg_E}{\neg\star}\,\neg_I}{\star}\,\star_I \qquad \dfrac{\dfrac{\overline{\star}^{\,2} \quad \dfrac{\overline{\star}^{\,2}}{\neg\star}\,\star_E}{\bot}\,\neg_E}{\neg\star}\,\neg_I^{\,2}}{\bot}\,\neg_E}{p}\,\text{Explo.}$$

The issue with these connectives is not harmony but something else. Intuitively, the core issue is that the premise one needs for introducing one of these connectives is a logically complex formula *containing* one of these connectives as a part. It's that feature, familiar from paradoxes of self-reference like the Liar and Curry, that leads to the trouble with these connectives.

There are a number of ways to say, concretely, what is wrong with these connective rule. Consider the following possible proof-theoretic constraints, owed to Dummett:

> **Purity:** The rules for a connective $\circ$ are *pure* if $\circ$ is the only connective that figures in its introduction and elimination rules.
>
> **Simplicity:** The rules for a connective $\circ$ are *simple* if there is no more than one logical connective in any sentence that figures in them.
>
> **Sheerness:** The rules for a connective $\circ$ are *sheer* if $\circ$ figures only in the conclusion of any of the introduction rules and figures only in a premise of any elimination rules.

The rules for $\bullet$ and $\star$ are not *pure*, since they contain other connectives than $\bullet$ and $\star$. In particular, the $\bullet$ rules contain the additional connective $\rightarrow$ and the $\star$ rules contain the additional connective $\neg$. They are also not *simple*, since the $\bullet$ rules contain the sentence $\bullet \to p$, and the $\star$ rules contain the sentence $\neg\star$, which both contain two logical operators. Finally, they are not *sheer*. In the the introduction rule for $\bullet$, $\bullet$ occurs not just in the conclusion, but also in *premise* $\bullet \to p$, from which $\bullet$ can be inferred. Similarly, $\star$ occurs not only in the premise of its introduction rule, but, moreover, occurs in the conclusion of its elimination rule as well.

Many familiar natural deduction rules violate at least one of these properties. For instance, consider the rule of *disjunctive syllogism*, commonly used as a disjunctive elimination rule in introductory textbooks:

$$\frac{A \vee B \quad \neg A}{B}$$

This is *simple* and *sheer*, but not *pure*, since, in addition to disjunction, it features negation. The double negation elimination rule is *pure* and *sheer*, by not *simple*, since it involves as its premise a sentence that has *two* negation operators. Finally, the standard negation introduction rule that does not feature the $\perp$ symbol, which, once again, is the following:

$$\dfrac{\begin{matrix}\overline{A}^{\,n} \\ \vdots \\ B\end{matrix} \qquad \begin{matrix}\overline{A}^{\,n} \\ \vdots \\ \neg B\end{matrix}}{\neg A}\ \neg_I{}^{n}$$

is *pure* (since it only features negation), and *simple* (since each negated formula only has that one negation), but not *sheer* since negation figures not just in the ultimate conclusion of the rule ($\neg A$), but also in the statement of the hypothetical proofs needed to arrive at this conclusion. For lack of a better term, I'll say that rules that are harmonious, pure, simple, and sheer are *immaculate*. Requiring that rules for logical connectives be immaculate is regarded by most proof-theoretic logicians (including Dummett himself) as overkill. However, it's worth emphasizing that the rules for conjunction, the conditional, and disjunction that we've provided all *are* immaculate, and so it's at least not obvious why this shouldn't be our standard.

Though, as we've just seen, adding a connective with harmonious rules to our language still might not constitute a conservative extension of the language, it is very simple to show that adding a connective with *immaculate* rules to our language *will* constitute a conservative extension. I will only sketch the basic proof strategy, rather than giving the full proof.[5] Suppose we introduce some operator $\circ$ into our language, as defined by an immaculate set of introduction and elimination rules, and suppose for reductio that this addition is *not* conservative. That is, suppose there is some proof using the rules for this this operator that enables us to prove some sentence $C$ of the initial language that we could not originally prove. In this case, of course, the rules for $\circ$ must, of course, be used, so we must use the introduction rules for $\circ$ to conclude a sentence of the form $A \circ B$. However, insofar as $C$ sentence of the original language, we must use the elimination rules on $A \circ B$ to arrive at sentence $C$ of the original language that we could not originally prove. But harmony means that we can provide reductions for $\circ$ shows that any such detour is eliminable. Now, in order to show this officially, we need to show that, in any case in which $\circ$ is introduced, and some other inferential steps are taken, we can transform this proof to one in which $\circ$ is introduced and subsequently immediately eliminated. We do this by showing that we can always push the application $\circ$-elimination up the proof tree until it directly follows the introduction rule, at which point the appropriate reduction can be applied. Because the rules are simple and sheer, we apply a reduction to an introduction-elimination detour on $A \circ B$, we strictly decrease the overall "$\circ$-complexity" of the proof (the number occurrences of $\circ$). So, by repeatedly pushing applications $\circ$-elimination upwards until it is immediately below the corresponding $\circ$-introduction and then applying the appropriate reduction, we will eventually obtain a derivation of $C$ in which no formula containing $\circ$ occurs at all, contradicting our assumption. Though full immaculateness is not *required* for this proof, it does make it simpler.

Whether or not immaculate rules are *necessary* do proof-theoretically define the meaning of a logical connectives, it surely seems that such rules are *desirable*. But is it *possible* to provide immaculate rules for the classical connectives in the context of a natural deduction system of this sort? Perhaps, but it's not simple or straightforward.[6] Rather than continuing our attempt to define the classical connectives in this sort of natural deduction system, in following

[5] Versions of such a normalization proof can be given in many textbooks, for instance,

[6] For some attempts,

three chapters we will consider three different kinds of proof systems that *are* capable of yielding immaculate rules for the classical connectives. In the next chapter, we will introduce the *multiple conclusion sequent calculus*, which is technically elegant system, though less straightforward to comprehend as codifying inferential rules than the natural deduction system presented here. In the chapter following next, we will introduce *bilateral natural deduction*, which will solve all of our problems in a natural deduction setting. Finally, in Chapter 5, we will consider a *bilateral sequent calculus* which will combine the bilateralism introduced in Chapter 4 with the sequent calculus approach introduced in Chapter 3.

## 2.9 Exercises

1. We've considered only the connectives of conjunction, the conditional, disjunction, and negation. But what about the biconditional?

   a) Provide pure, simple, and sheer introduction and elimination rules for the biconditional $\leftrightarrow$. The rules should be such that $A \leftrightarrow B$ is equivalent to $(A \to B) \wedge (B \to A)$, but, being pure, they should not use any other connectives.

   b) Provide reductions and expansions to show that these rules are harmonious.

2. We provided the standard elimination rules for conjunction and the conditional. However, conjunction may also be understood as also having an elimination rule of this basic form:

$$\dfrac{A \wedge B \qquad \begin{matrix} \overline{A}^{\,n} \; \overline{B}^{\,m} \\ \vdots \\ C \end{matrix}}{C}\ \wedge_{\mathrm{E}}^{n,m}$$

   This rules tells us that if we have $A \wedge B$ and we can conclude some sentence $C$ from the assumption of $A$ along with the assumption of $B$, then we can discharge these assumptions and conclude $C$. Likewise, an alternative elimination rule for the conditional is the following:

$$\dfrac{A \to B \quad A \quad \begin{array}{c}\overline{B}^{\,n} \\ \vdots \\ C\end{array}}{C} \to_{\mathrm{E}}{}^{n}$$

For this exercise:

a) Derive our conjunction elimination rules as two special cases of this conjunction elimination rule.

b) Derive our conditional elimination rule as a special case of the above conditional elimination rule.

c) Prove that these rule is harmonious with our introduction rules by providing reductions and expansions.

3. We noted that the rule of disjunctive syllogism, $A \vee B, \neg A \vdash B$, which figures in many introductory textbooks is not *pure*. Notably, it is also not *harmonious*, relative to the disjunction introduction rule

4. Use the system for intuitionistic logic along with Peirce's rule to prove $p \vee \neg p$.

5. Consider the following rules and state whether or not they are (1) pure, (2) simple, and (3) sheer.

a) $$\dfrac{A \to B \quad \neg B}{\neg A} \text{ Modus Tollens}$$