

PHIL 50 – INTRODUCTION TO LOGIC

MARCELLO DI BELLO – STANFORD UNIVERSITY

MODAL LOGIC – WEEK #9

1 IMAGINE

Imagine the world was different from what it actually is. Imagine there were no wars. Imagine hatred did not exist. Imagine women were not discriminated against. Imagine we were all poor, brutish, and needy. And so on. We can imagine infinite ways the world we live in could be different from what it actually is. It is an interesting question to what extent we can imagine the world to be different from what it actually is. Can we truly imagine that cows were made of metal? Can we truly imagine ourselves to be insects just like in Kafka's *Metamorphosis*? These questions aside, the topic of this note is modal logic. Modal logic allows us to reason about alternative ways the world we live in could be (or could have been).¹

2 POSSIBLE WORLDS

A key notion in modal logic is that of a *possible world*. We can represent a possible world simply as a dot at which some formulas are true and some formulas are false. Below you see the possible world labeled w and represented by \circ , as follows:

$$w \circ$$

Suppose formula p is true at w while formulas q and r are false at w . We now have:

$$w \circ p, \not{q}, \not{r}$$

The convention is to put the formulas to the right of the possible world and the label to the left of the world. This is a convention to avoid confusion. Note that the same formula can be true at one possible world and false at another. For example, in another possible world, say v , the same formulas p , q , and r could all be true:

$$v \circ p, q, r$$

¹To be sure, modal logic is not only that. It also allows us to reason about beliefs, knowledge, preferences, and obligations. But since this note will be short and introductory, it is useful to think of modal logic as a branch of logic that is primarily concerned with reasoning about alternatives to the world we live in.

This variability in truth across different possible worlds is a way to capture the idea that things could be different from one world to another. To illustrate, suppose r stands for the statement ‘*Roger Federer is a tennis player from Switzerland*’ which is true in the actual world, say world v . But Roger Federer, in another possible world, could have been a writer, so that the same formula r , in another world, say w , is false.

We should not place too much emphasis on the distinction between actual world and possible worlds. In fact, the actual world is one among the many possible worlds. The actual world is just one world we have selected as the world we live in. How many possible worlds are there? There are many, and potentially infinite, possible worlds. We shall refer to the set of all possible worlds by W and to individual possible worlds by w , v , or u ,

3 NECESSITY AND POSSIBILITY

Given a certain world, say w , some formulas will be true in it and some formulas will be false. More formally, we shall use the following notation:

$$\begin{aligned} w \models p & \text{ iff } p \text{ is true at } w \\ w \not\models p & \text{ iff } p \text{ is false at } w \end{aligned}$$

If we look at all possible worlds, we will find that:

- a. some formulas are true at all possible worlds; and
- b. some formulas are false in all possible worlds.

The formulas that are true at all possible worlds are *necessarily true*. One of the insights from modal logic is that statements that are necessarily true can be conceptualized as statements that are true at all possible worlds. Are there examples? Logical principles such as $\neg(\varphi \wedge \neg\varphi)$ are an example of statements that are necessarily true. Also, statements about the meaning of words such as ‘*all bachelors are unmarried men*’ are another example of necessarily true statements. The language of modal logic allows us to express that a formula φ is necessarily true by writing $\Box\varphi$ (read ‘box φ ’). More formally:

$$w \models \Box\varphi \text{ iff for all possible worlds } v, \text{ it holds that } v \models \varphi$$

In other words, $\Box\varphi$ is true at a world w if and only if φ is true at all possible worlds. Now, while some formulas are true at all possible worlds, some formulas are false at all possible worlds, i.e. they are *necessarily false*. Contradictory statements are an example of statements that are necessarily false. We can say that a formula φ is necessarily false by writing $\Box\neg\varphi$. (This presupposes that if a formula is necessarily false, its negation is necessarily true.)

Now, besides formulas such $\Box\varphi$, the language of modal logic also contains formulas of the form $\Diamond\varphi$ (read ‘diamond φ ’). While \Box is meant to capture the idea of necessity, the symbol \Diamond is meant to capture the idea of possibility. Formally, we have:

$$w \models \Diamond\varphi \text{ iff for some possible worlds } v, \text{ it holds that } v \models \varphi$$

While $\Box\varphi$ is true provided φ is true in all possible worlds, $\Diamond\varphi$ is true provided φ is true in at least one possible world or in some possible worlds.

Note a couple of things. First, there is a parallelism between \Box and \forall and there is a parallelism between \Diamond and \exists . Second, $\Diamond\varphi$ and $\Box\varphi$ need not be incompatible. The same formula φ can be both necessarily true and possibly true, i.e. both $w \models \Box\varphi$ and $w \models \Diamond\varphi$ can hold. After all, if φ is true in all possible worlds, this does not exclude that φ is true in some possible worlds or in at least one possible world.

4 TRUTH

One might wonder, how can we tell whether a formula is true in a world or not? Well, this is partly the result of a stipulation we make. This depends on how we construct our model M . For now, a model M consists of a non-empty set W of possible worlds. Further, relative to our model M , we define which atomic formulas are true at which worlds and which atomic formulas are false at which worlds. We can make any assignment of truth or falsity that we please. The only constraint is that the same atomic formulas cannot be both true and false at the same possible world.

Assigning truth values to atomic formulas can be done by defining a valuation V_w that for each possible world w and for each atomic formula assigns the value 1 (i.e. the atomic formula is true at w) or 0 (i.e. the atomic formula is false at w). Similarly, in propositional logic, we had a valuation V that assigned 1 or 0 to every atomic formula, although in propositional logic V did not take into account possible worlds.

Once we have assigned a truth value to all atomic formulas relative to each possible world, what about the more complex formulas that are not atomic? We can use a recursive definition of truth to determine the truth value of more complex formulas. So, let p be an atomic formula, and let φ and ψ be placeholders for formulas of arbitrary complexity. The truth conditions for formulas in modal logic are recursively defined as follows:

$M, w \models p$	iff	p is true at w
$M, w \models \neg\varphi$	iff	it is not the case that $M, w \models \varphi$
$M, w \models \varphi \wedge \psi$	iff	$M, w \models \varphi$ and $M, w \models \psi$
$M, w \models \varphi \vee \psi$	iff	$M, w \models \varphi$ or $M, w \models \psi$
$M, w \models \varphi \rightarrow \psi$	iff	$M, w \models \varphi$ implies $M, w \models \psi$
$M, w \models \Box\varphi$	iff	for all possible worlds v , it holds that $M, v \models \varphi$
$M, w \models \Diamond\varphi$	iff	for some possible worlds v , it holds that $M, v \models \varphi$

We shall now illustrate how the above definition works. Suppose we have a model M with $W = \{w, v\}$ and we consider only two atomic formulas p and q . Suppose we have defined that p is true at w but false at v , while q is false at w but true at v . In other words, we have:

$$w \circ p, \not q$$

$$v \circ \not p, q$$

Given these initial stipulations for the atomic formulas, we can check that the following hold for more complex formulas:

$$\begin{aligned} M, w &\models p \vee q \\ M, w &\models \neg q \\ M, v &\models p \rightarrow q \\ M, v &\models \neg p \wedge q \\ M, v &\models \diamond \neg p \\ M, v &\models \Box(p \vee q) \\ M, v &\models \Box \diamond p \end{aligned}$$

The last three modal formulas require some explaining. First, $M, v \models \diamond \neg p$ holds because there is some possible world, namely v itself, such that $M, v \models \neg p$. Second, $M, v \models \Box(p \vee q)$ because $p \vee q$ is true in all possible worlds, and in our case all possible worlds are simply v and w . Note that $M, v \models p \vee q$ and $w \models p \vee q$. Third, $M, v \models \Box \diamond p$ because $\diamond p$ is true in all possible worlds, for $M, v \models \diamond p$ and $M, w \models \diamond p$.

Bear in mind that our reasoning so far was relative to model M . We could have defined a different model M' with a different set W' of possible worlds such that $W' = \{w', v'\}$, where

$$w' \circ \not p, \not q$$

$$v' \circ \not p, q,$$

Relative to the new model M' , we can check that

$$\begin{aligned} M', v' &\not\models \Box(p \vee q) \\ M', v' &\not\models \diamond p \end{aligned}$$

Now, $M', v' \not\models \Box(p \vee q)$ because $p \vee q$ is not true in w' where both p and q are false. Also, $M', v' \not\models \diamond p$ because there is no world in which p is true, so $\diamond p$ is not true at v' .

5 CONTINGENTLY TRUE AND CONTINGENTLY FALSE

Using \Box , \Diamond , and the connectives from propositional logic, we can express the fact that a formula φ is true in a world, say w , but it could be false, as follows:

$$M, w \models \varphi \wedge \Diamond \neg \varphi$$

In other words, φ is true at w and there is a world, say v , such that $\neg \varphi$ is true at v . In this case, we say that φ is *contingently true*. We can also express the fact that a formula φ is false, say in w , but it could be true, as follows:

$$M, w \models \neg \varphi \wedge \Diamond \varphi$$

In other words, $\neg \varphi$ is true at w and there is a world, say v , such that φ is true at v . In this case, we say that φ is *contingently false*.

6 VALIDITY

Just as in proposition logic, there are formulas that are true in all models, i.e. valid formulas, in modal logic there are formulas that are true in all models and in all possible worlds. We shall write:

$$\models \varphi \text{ iff for all } M \text{ and all } w, \text{ it holds that } M, w \models \varphi$$

Suitable examples of formulas that are valid in modal logic are the logical principles that are valid in propositional logic, such as $\neg(\varphi \wedge \neg \varphi)$, $\varphi \vee \neg \varphi$, $(\varphi \wedge (\varphi \rightarrow \psi)) \rightarrow \psi$, etc.

7 CONSTANTS, PREDICATES, VARIABLES, AND QUANTIFIERS IN MODAL LOGIC

So far our language consisted of the propositional connectives with the modal operators \Box and \Diamond . We shall now add constant, variable, and predicate symbols as well as quantifiers. You should be familiar with these linguistic ingredients from studying predicate logic. In order to interpret this richer and more expressive language, we need to refine our earlier notion of a model M . So far a model was simply a set of possible worlds where a truth value was assigned to each atomic formula at each world. We shall now make this simple notion of a model more complex. For each possible world w , we have

- a domain D_w of objects relative to w ;
- an interpretation function I_w that assigns constant symbols to objects in D_w and assigns one-place predicate symbols to sets of objects in D_w ; and
- an assignment function g_w that assigns variable symbols to objects in D_w .

Note that we have a significant degree of complexity here. Each world w has associated its own D_w , I_w , and g_w . A model M now is a set W of worlds where each world w is associated to its own D_w , I_w , and g_w as defined above.

Here is a simple illustration. Consider a predicate modal language consisting of the predicates ‘Player’, ‘Writer’ and the constant symbol ‘federer’ and ‘wallace’. Now, consider the model where:

$$\begin{array}{l}
 W = \{w, u\} \\
 D_w = \{ \text{Head 1}, \text{Head 2}, \text{Head 3} \} \quad D_u = \{ \text{Head 1}, \text{Head 2}, \text{Head 3}, \text{Person 1} \} \\
 I_w(\text{Player}) = \{ \text{Head 1} \} \quad I_u(\text{Player}) = \{ \text{Head 1}, \text{Person 1} \} \\
 I_w(\text{Writer}) = \{ \text{Head 2}, \text{Head 3} \} \quad I_u(\text{Writer}) = \{ \text{Head 2}, \text{Head 3} \} \\
 I_w(\text{wallace}) = \{ \text{Head 1} \} \quad I_u(\text{wallace}) = \{ \text{Head 2} \} \\
 I_w(\text{federer}) = \{ \text{Head 2} \} \quad I_u(\text{federer}) = \{ \text{Person 1} \} \\
 g_w(x) = \text{Head 2} \quad g_u(x) = \text{Head 2} \\
 g_w(y) = \text{Head 3} \quad g_u(y) = \text{Person 1}
 \end{array}$$

Let’s call the above model **TW**. We use the notation **TW** to distinguish the above model from a generic model M . We will refer to **TW** for purpose of illustration later on. Note that each possible world gets assigned its own domain of objects, and each predicate symbol gets assigned a set of objects relative to each possible world, and each constant symbol gets assigned an object relative to each possible world. Variable symbols x and y also get assigned objects relative to each possible world.

How can we determine the truth of formulas relative to a model? Below are the recursive truth conditions for formulas with predicates, constants, quantifiers, and modal operators \Box and \Diamond :

$M, w \models P(a)$	iff	$I_w(a) \in I_w(P)$
$M, w \models P(x)$	iff	$g_w(x) \in I_w(P)$
$M, w \models a = b$	iff	$\langle I_w(a), I_w(b) \rangle \in I_w(=)$
$M, w \models x = y$	iff	$\langle g_w(x), g_w(y) \rangle \in I_w(=)$
$M, w \models \neg\varphi$	iff	<i>it is not the case that</i> $M, w \models \varphi$
$M, w \models \varphi \wedge \psi$	iff	$M, w \models \varphi$ and $M, w \models \psi$
$M, w \models \varphi \vee \psi$	iff	$M, w \models \varphi$ or $M, w \models \psi$
$M, w \models \varphi \rightarrow \psi$	iff	$M, w \models \varphi$ implies $M, w \models \psi$
$M, w \models \Box\varphi$	iff	for all possible worlds v , it holds that $M, v \models \varphi$
$M, w \models \Diamond\varphi$	iff	for some possible worlds v , it holds that $M, v \models \varphi$
$M, w \models \forall x\varphi(x)$	iff	for all d , if $d \in D_w$, then $\langle D_w, I_w, g_{w[x:=d]} \rangle \models \varphi(x)$
$M, w \models \exists x\varphi(x)$	iff	for some d , $d \in D_w$ and $\langle D_w, I_w, g_{w[x:=d]} \rangle \models \varphi(x)$

The parallelism with the truth conditions in predicate logic should be apparent. Truth conditions for formulas in predicate logic are analogous to the ones above, with the most notable differences being, first, that the modal formulas containing \Diamond and \Box are missing from predicate logic, and second, that in predicate logic the truth conditions are not relativized to a possible world.²

We can now check that:

- i. $\mathbf{TW}, u \models \text{Writer}(\text{wallace})$ because $I_u(\text{wallace}) \in I_u(\text{Writer})$
- i. $\mathbf{TW}, w \models \text{Player}(\text{wallace})$ because $I_w(\text{wallace}) \in I_w(\text{Player})$
- iii. $\mathbf{TW}, u \models \text{Player}(\text{federer})$ because $I_u(\text{federer}) \in I_u(\text{Player})$
- iv. $\mathbf{TW}, w \models \text{Writer}(\text{federer})$ because $I_w(\text{federer}) \in I_w(\text{Writer})$
- v. $\mathbf{TW}, u \not\models \text{Player}(\text{wallace})$ because $I_u(\text{wallace}) \notin I_u(\text{Player})$
- vi. $\mathbf{TW}, u \not\models \text{Writer}(\text{federer})$ because $I_u(\text{federer}) \notin I_u(\text{Writer})$
- vii. $\mathbf{TW}, u \models \Diamond\text{Writer}(\text{federer})$ because there is a world w such that $\mathbf{TW}, w \models \text{Writer}(\text{federer})$

²Here are the truth conditions for formulas in predicate logic:

$M \models P(a)$	iff	$I(a) \in I(P)$
$M \models P(x)$	iff	$g(x) \in I(P)$
$M \models a = b$	iff	$\langle I(a), I(b) \rangle \in I(=)$
$M \models x = y$	iff	$\langle g(x), g(y) \rangle \in I(=)$
$M \models \neg\varphi$	iff	<i>it is not the case that</i> $M \models \varphi$
$M \models \varphi \wedge \psi$	iff	$M \models \varphi$ and $M \models \psi$
$M \models \varphi \vee \psi$	iff	$M \models \varphi$ or $M \models \psi$
$M \models \varphi \rightarrow \psi$	iff	$M \models \varphi$ implies $M \models \psi$
$M \models \forall x\varphi(x)$	iff	for all d , if $d \in D$, then $\langle D, I, g_{[x:=d]} \rangle \models \varphi(x)$
$M \models \exists x\varphi(x)$	iff	for some d , $d \in D$ and $\langle D, I, g_{[x:=d]} \rangle \models \varphi(x)$

viii. $\mathbf{TW}, u \models \Diamond \text{Player}(\text{wallance})$ because there is world w and $\mathbf{TW}, w \models \text{Player}(\text{wallace})$

Let's focus on world u , which looks like the actual world. In world u , we have that Federer is a Player and Wallace is a writer; see i. and iii. above. Also, in world u Federer is not a writer and Wallace is not a player; see v. and vi. above. Finally, again in u , it is possible that Federer is a writer and that Wallace is a player; see vii. and viii. above.

We now consider a more complicated formula, namely $\Box \forall x (\text{Player}(x) \vee \text{Writer}(x))$. The formula, intuitively, means it is necessary that every object is a player or a writer. Is that true? We can check that

ix. $\mathbf{TW}, u \models \Box \forall x (\text{Player}(x) \vee \text{Writer}(x))$.

By the truth conditions, we have that:

$\mathbf{TW}, u \models \Box \forall x (\text{Player}(x) \vee \text{Writer}(x))$

IFF for all v , it holds that $\mathbf{TW}, v \models \forall x (\text{Player}(x) \vee \text{Writer}(x))$

IFF for all v , for all d , if $d \in D_v$, then $\langle D_v, I_v, g_{v[x:=d]}(x) \rangle \models \text{Player}(x) \vee \text{Writer}(x)$

IFF for all v , for all d , if $d \in D_v$, $g_{v[x:=d]}(x) \in I_v(\text{Player})$ or $g_{v[x:=d]}(x) \in I_v(\text{Writer})$

IFF (*) for all v , for all d , if $d \in D_v$, $d \in I_v(\text{Player})$ or $d \in I_v(\text{Writer})$.

We can check that (*) holds in model \mathbf{TW} . We need to consider all possible worlds, namely w and u . As far as w is concerned, every object d in D_w is such that it either belongs to $I_w(\text{Player})$ or to $I_w(\text{Writer})$. As far as u is concerned, every object d in D_u is such that it either belongs to $I_u(\text{Player})$ or to $I_u(\text{Writer})$. Another way to see this is that $I_w(\text{Player}) \cup I_w(\text{Writer}) = D_w$ and $I_u(\text{Player}) \cup I_u(\text{Writer}) = D_u$. So, since (*) holds, $\mathbf{TW}, u \models \Box \forall x (\text{Player}(x) \vee \text{Writer}(x))$.

By contrast, we can check that

x. $\mathbf{TW}, u \not\models \forall x \Box (\text{Player}(x) \vee \text{Writer}(x))$.

The formula $\forall x \Box (\text{Player}(x) \vee \text{Writer}(x))$, intuitively, means that every object is necessarily a player or a writer. Contrast this formula with the earlier $\Box \forall x (\text{Player}(x) \vee \text{Writer}(x))$, whose intuitive meaning is that, necessarily, every object is a player or a writer.

By the truth conditions, we have that:

$\mathbf{TW}, u \models \forall x \Box (\text{Player}(x) \vee \text{Writer}(x))$

IFF for all d , if $d \in D_u$, it holds that $\langle D_u, I_u, g_{u[x:=d]}(x) \rangle \models \Box (\text{Player}(x) \vee \text{Writer}(x))$

IFF for all d , if $d \in D_u$, for all v , $\langle D_v, I_v, g_{v[x:=d]}(x) \rangle \models \text{Player}(x) \vee \text{Writer}(x)$




IFF for all d , if $d \in D_u$, for all v , $g_{v[x:=d]}(x) \in I_v(\text{Player})$ or $g_{v[x:=d]}(x) \in I_v(\text{Writer})$


IFF (**) for all d , if $d \in D_u$, for all v , $d \in I_v(\text{Player})$ or $d \in I_v(\text{Writer})$.

We can check that (**) does not hold in model \mathbf{TW} . To see why, we should consider all

objects d in D_u where $D_u = \{ \text{head with question mark}, \text{head with crown}, \text{head with crown and wings}, \text{person with crown} \}$. This is what (**) requires us to do.

More precisely, we should check that every object in D_u is such that it either belongs to the interpretation of the predicate *Writer* or it belongs to the interpretation of the predicate *Player* relative to each possible world.

Now, consider the object  and world w . Clearly,  $\notin I_w(\textit{Writer})$ and  $\notin I_w(\textit{Player})$.

So, in world w , object  belongs to the interpretation of neither predicate. So (**) does not hold, whence $\mathbf{TW}, u \not\models \forall x \Box (\textit{Player}(x) \vee \textit{Writer}(x))$. It is crucial here to understand the difference between (*) which holds in \mathbf{TW} and (**) which does not hold in \mathbf{TW} .

8 ARE DOMAINS OF OBJECTS THE SAME ACROSS POSSIBLE WORLDS?

We have just shown that

$\mathbf{TW}, u \models \Box \forall x (\textit{Player}(x) \vee \textit{Writer}(x))$; and

$\mathbf{TW}, u \not\models \forall x \Box (\textit{Player}(x) \vee \textit{Writer}(x))$.

This means that

$\mathbf{TW}, u \not\models (\Box \forall x (\textit{Player}(x) \vee \textit{Writer}(x))) \rightarrow (\forall x \Box (\textit{Player}(x) \vee \textit{Writer}(x)))$.

and therefore that

$\not\models (\Box \forall x (\textit{Player}(x) \vee \textit{Writer}(x))) \rightarrow (\forall x \Box (\textit{Player}(x) \vee \textit{Writer}(x)))$.

Despite this result, the logician and philosopher Ruth Barcan Marcus believed that

$\models \Box \forall x \varphi(x) \leftrightarrow \forall x \Box \varphi(x)$

A justification for believing that $\Box \forall x \varphi(x) \leftrightarrow \forall x \Box \varphi(x)$ is valid is by postulating that the domains of objects are the same across different possible worlds. Indeed, if the domains are the same across all possible worlds, then $\Box \forall x \varphi(x) \leftrightarrow \forall x \Box \varphi(x)$ comes out valid. But is it plausible to believe that domains of objects are the same across possible worlds? Isn't it the case that an object can cease to exist in some possible world or that a new and different object comes into existence in another possible world? This is currently terrain of dispute among philosophers and metaphysicians; see, for instance, the recent book by the philosopher Timothy Williamson, *Modal Logic as Metaphysics*, Oxford University Press, 2013.

9 IS PAIN JUST A BRAIN STATE?

Another area of controversy is philosophy of mind. In his legendary lectures, *Naming and Necessity*, the philosopher and logician Saul Kripke showed how we can use logic and philosophy of language for formulating arguments in philosophy of mind. Kripke argued that proper names are rigid designators, i.e. proper names refer to the same object across all possible worlds. His argument is simple. We can say things like 'imagine Roger Federer was not a tennis player' or 'imagine Christine Lagarde was not the head of the International

Monetary Fund.’ Now, in order to make sense of these statements, the proper names ‘*Federer*’ and ‘*Lagarde*’ must refer to the same object or individual across all possible worlds. If not, it would make no sense to imagine a possible world in which Federer (that Federer!) was not a tennis player or a possible world in which Lagarde (that Lagarde!) was not the head of the IMF.

Now, constant symbols such as a , b , c are the logical equivalent of proper names. So, the rigidity of proper names becomes:

RIGIDITY. In any model M , for any possible worlds w and v , and for any constant symbol c , it holds that $I_w(c)$ is the same as $I_v(c)$. In other words, constant symbols refer to the same object across all possible worlds in all models.

If RIGIDITY holds, one can show that (with a and b constant symbols):

$$\models (a = b) \rightarrow \Box(a = b)$$

That is, the formula $(a = b) \rightarrow \Box(a = b)$ is valid. This means that if $a = b$ is true in any world and in any model, it is necessarily true. Given RIGIDITY, statements of identity become necessarily true. (To check that this is indeed the case is left for you as an exercise.)

Here is how the claim that $(a = b) \rightarrow \Box(a = b)$ becomes relevant for philosophy of mind. Suppose ‘*pain*’ is the constant symbol referring to the mental state of feeling pain and ‘*C-fiber-firing*’ is the constant symbol referring to a particular brain state. (The constant symbol ‘*C-fiber-firing*’ is just a name for some brain state corresponding to a neural state of the brain; there is no need to enter into the details here.)

Those in philosophy of mind who believe that mental states, such as pain, are nothing other than brain states, such as the firing of C-fibers, will believe (roughly) that pain is identical to C-fibers firing. So, they will believe that, in the world we live in, it is true that $(\textit{pain} = \textit{C-fiber-firing})$. Interestingly enough, because of rigidity, it holds that:

$$\models (\textit{pain} = \textit{C-fiber-firing}) \rightarrow \Box(\textit{pain} = \textit{C-fiber-firing})$$

In other words, if we grant that pain is identical to the firing of C-fibers, then by RIGIDITY, it follows that the identity must be necessary.

Now, Kripke argued that identities such as $(\textit{pain} = \textit{C-fiber-firing})$ cannot be necessary. How so? For Kripke, we can easily imagine a possible world in which pain and the firing of C-fibers are different. Maybe this is a possible world in which there are creatures different from us who experience pain without any underlying brain state. Call this imaginary world v and call the actual world we live in w . If Kripke is right, $M, v \not\models (\textit{pain} = \textit{C-fiber-firing})$ with v an imaginary world. This means that the identity $(\textit{pain} = \textit{C-fiber-firing})$ cannot be necessary. So, in the actual world w we live in, $M, w \not\models \Box(\textit{pain} = \textit{C-fiber-firing})$.

But here is the surprising conclusion. If $M, w \not\models \Box(\textit{pain} = \textit{C-fiber-firing})$ and further if $\models (\textit{pain} = \textit{C-fiber-firing}) \rightarrow \Box(\textit{pain} = \textit{C-fiber-firing})$, it follows by simple *modus tollens* that

$M, w \not\models (\text{pain} = \text{C-fiber-firing})$. This means that in the actual world w we live in, pain is not identical to the firing of C-fibers. Those who wish to reduce the mental state of pain to some brain state are wrong! End of Kripke's argument.

What does the argument show? It shows that the identity between brain states and mental states must be false in the actual world we live in. More carefully, the argument shows that if it is possible that brain states are not identical to mental states and RIGIDITY holds, then the identity between mental states and brain states cannot hold in the actual world we live in either. What's interesting—and somewhat surprising—about this argument is that it uses modal logic to derive conclusions about the nature of reality. You might think this is a dubious way to proceed. And yet, it is one of the merits of Anglo-American philosophy to use logic and the workings of language as ways to uncover the mysteries of reality. But this is another story . . .