1. Implications Are Weird

The "implies" connective $\to$ is one of the stranger connectives. Let P, Q, and R be propositions. Below are a series of statements. For each statement, explain whether it is always true, always false, or depends on the values of the propositions involved. (Hint: draw truth tables!)

a. $P \to \neg P$

> This expression is equivalent to $\neg P$, so its value depends on the value of P.

b. $(P \to Q) \lor (Q \to P)$

> This expression is always true, regardless of what P and Q are.

c. $(P \to Q) \lor (Q \to R)$

> This expression is always true, regardless of what P, Q, and R are.

2. Designing Propositional Formulas

Here are some English descriptions of relationships among propositional variables. For each description, write a propositional formula that means the same thing. Then, briefly explain why your formula works. Try to see if you can come up with the simplest formula possible.

(If you find yourself writing out something extremely long, try rethinking your approach. All of these problems have formulas that can be written out in a single line.)

a. For the variables A, B, and C: Exactly one of A, B, and C is true. Our solution has 8 connectives, not including ¬.

> One option is:
>
> $(A \land \neg B \land \neg C) \lor (\neg A \land B \land \neg C) \lor (\neg A \land \neg B \land C)$
>
> This formula lists all of the possible combinations of how exactly one variable could be true.

b. For the variables A, B, C, and D: If any variable is true, then all the variables that follow it alphabetically are also true. (Hint: Try this out with just two variables first, then three variables.) Our solution has 5 connectives.

> One option is:
>
> $(A \to B) \land (B \to C) \land (C \to D)$
>
> This formula says that if one variable is true, the one alphabetically right after it must also be true. This will guarantee every following letter is true as well.

3. More Translating First-Order Logic to English

Translate these first-order logic statements to English, then say whether or not they are true.

a. **Numbers:**

(1) $\forall n \in \mathbb{N}.\ \exists m \in \mathbb{N}.\ n < m$

> This means "for every natural number n, there is a larger natural number."
>
> This is true! There is no largest natural number.

(2) $\exists n \in \mathbb{N}.\ \forall m \in \mathbb{N}.\ n < m$

> This means "there is a natural number n that is smaller than all natural numbers."
>
> This is not true! Even if we take n to be 0, the smallest natural number, we could also take m to be 0 and the statement would be false.

(3) $\forall n \in \mathbb{N}.\ \exists m \in \mathbb{N}.\ (n < m \land \exists p \in \mathbb{N}.\ (n < p \land p < m))$

> This means "for every natural number n, there is a natural number m where m is larger than n and there is a natural number p where p is larger than n and smaller than m."
>
> This is true! For any n, we could take $p = n + 1$ and $m = n + 2$, for example.

(4) $\forall n \in \mathbb{N}.\ \forall m \in \mathbb{N}.\ (n < m \to \exists p \in \mathbb{N}.\ (n < p \land p < m))$

> This means "for every natural number n, for all natural numbers m where m is larger than n, there is a natural number p where p is larger than n and smaller than m."
>
> This is not true! For any n, if $m = n + 1$, the antecedent of the implication is true but the consequent is false since there is no such p between the numbers n and m.

(5) $\exists n \in \mathbb{Z}.\ (2n > n \leftrightarrow 2n < n)$

> This means "there is an integer n where $2n > n$ if and only if $2n < n$", which is equivalent to saying both $2n > n$ and $2n < n$, or both $2n \leq n$ and $2n \geq n$.
>
> This is surprisingly true! For positive and negative integers, one of the sides of the biconditional is false while the other is true, so the biconditional is false. But for $n = 0$, both sides of the biconditional are false, meaning the biconditional as a whole is true.

4. More Translating English to First-Order Logic

Given these predicates:

- $Old(x)$, which says x is old,
- $Dog(d)$, which says x is a dog,
- $Trick(t)$, which says t is a trick,
- $CanTeach(x, y)$, which says that you can teach y to x

a. Write a first-order logic expression that means "Everything that you can teach to some dog has to be a trick."

> This statement works by saying "for every t, if there's some dog you can teach t to, then t is a trick."
>
> $\forall t. \Big(\exists d. \big(Dog(d) \land CanTeach(d, t)\big) \to Trick(t)\Big)$

b. Write a first-order logic expression that means "Only dogs can learn tricks." (Interpret this as meaning "If there are no dogs, nothing can be taught tricks.")

> Notice how at the highest level, this statement is an implication. The antecedent is "there are no dogs", and the consequent is "for any entity, that entity cannot be taught tricks."
>
> $\forall d. \big(\neg Dog(d)\big) \to \forall x. \forall t. \big(Trick(t) \to \neg CanTeach(x, t)\big)$

c. Write a first-order logic expression that means "There's exactly one old trick."

> This expression works by saying "there is a t that is an old trick, and for anything that is an old trick, it has to be t.
>
> $\exists t. \Big(Old(t) \land Trick(t) \land \forall x. \big((Old(x) \land Trick(x)) \to x = t\big)\Big)$

d. Write a first-order logic expression that means "Every dog can be taught tricks, and some dogs can even be taught all tricks."

> The following works by saying "for any d_1 that is a dog, there must be a trick t_1 that d_1 can learn; and there is a dog d_2 where for any trick t_2, d_2 can learn t_2. (Note: We don't need to use separate variable names here, since d_1 and d_2 are in different scopes. But I think it helps for readability.)

$$\forall d_1. \Big(Dog(d_1) \to \big(\exists t_1. Trick(t_1) \land CanTeach(d_1, t_1) \big) \Big) \land$$
$$\exists d_2. \Big(Dog(d_2) \land \forall t_2. \big(Trick(t_2) \to CanTeach(d_2, t_2) \big) \Big)$$

e. Write a first-order logic expression that means "you can't teach an old dog new tricks". (Assume you can say something is "new" by saying it is "not old".)

This saying is expressing something about all old dogs, namely that they cannot be taught any new tricks. This suggests two universal statements, one for the dog and one for the tricks. The following works by saying "for any d that is an old dog and any t that is a new (non-old) trick, t can't be taught to d".

$$\forall d. \Bigg(\big(Dog(d) \land Old(d) \big) \to \Big(\forall t. \ \big(Trick(t) \land \neg Old(t) \big) \to \neg CanTeach(d, t) \Big) \Bigg)$$

5. More Negations

Look up the first-order logic statement corresponding to each of the statements from the solutions for "Translating English to First-Order Logic", then take the negation.

a. All dogs wear hats.

> Since this takes one of the four Aristotelian forms, we can look up the corresponding negation from the table, which translates to "there is a dog not wearing a hat".
>
> $\exists d.(Dog(d) \land \neg HasHat(d))$

b. There is a dog with a hat.

> Similarly to (1), we have "dogs don't wear hats":
>
> $\forall d.(Dog(d) \to \neg HasHat(d))$

c. Some dogs don't wear hats.

> Similarly to (1), we have "all dogs have a hat":
>
> $\forall d.(Dog(d) \to HasHat(d))$

d. Some dogs wear hats, but not all dogs wear hats.

> The negation of this formula translates to "all dogs wear hats, or no dogs wear hats":
>
> $\forall d.(Dog(d) \to HasHat(d)) \lor \forall d.(Dog(d) \to \neg HasHat(d))$

e. Some robot loves exactly one dog.

> Here's the original expression for reference:
>
> $$\exists r.\Big(Robot(r) \land \exists d.\big(Dog(d) \land Loves(r,d) \land \forall x.\big(Dog(x) \land Loves(r,x) \to d = x\big)\big)\Big)$$
>
> We can define some new "variables" to simplify the original expression without as much nesting.
>
> Let's say that $\bigstar$ represents $\forall x.\big(Dog(x) \land Loves(r,x) \to d = x\big)$, and $\diamondsuit$ represents $\exists d.(Dog(d) \land Loves(r,d) \land \bigstar)$. Then, the original expression can be written as:
>
> $$\exists r.(Robot(r) \land \diamondsuit)$$
>
> The negation of this expression can be written as:

$$\neg\exists r.(Robot(r) \land \diamondsuit)$$
$$\forall r.(Robot(r) \to \neg\diamondsuit)$$

Now, what's $\neg\diamondsuit$? We can simplify this as:

$$\neg(\exists d.(Dog(d) \land Loves(r,d) \land \bigstar))$$
$$\forall d.\neg(Dog(d) \land Loves(r,d) \land \bigstar))$$
$$\forall d.(Dog(d) \land Loves(r,d) \to \neg\bigstar)$$

Finally, we need to find what $\neg\bigstar$ is:

$$\neg\forall x.(Dog(x) \land Loves(r,x) \to d = x)$$
$$\exists x.\neg(Dog(x) \land Loves(r,x) \to d = x)$$
$$\exists x.(Dog(x) \land Loves(r,x) \land d \neq x)$$

Substituting in this expression for $\neg\bigstar$ into $\neg\diamondsuit$ and the expression for $\neg\diamondsuit$ into our original negation, we finally have:

$$\forall r.\Big(Robot(r) \to \forall d.\big(Dog(d) \land Loves(r,d) \to \neg x.(Dog(x) \land Loves(r,x) \land d \neq x)\big)\Big)$$

This can be translated as "for each robot, for all dogs the thing loves, there is some other dog it loves", which is a somewhat convoluted way to say "there is no robot that loves exactly one dog."

f. There are at least two dogs.

We can take the negation of this formula using the negation steps:

$$\neg\big(\exists x.(Dog(x) \land \exists y.(Dog(y) \land y \neq x))\big)$$
$$\forall x.\neg(Dog(x) \land \exists y.(Dog(y) \land y \neq x))$$
$$\forall x.(Dog(x) \to \neg\exists y.(Dog(y) \land y \neq x))$$
$$\forall x.(Dog(x) \to \forall y.(Dog(y) \to \neg(y \neq x)))$$
$$\forall x.(Dog(x) \to \forall y.(Dog(y) \to y = x))$$

This statement translates to "for all pairs of dogs x and y, they are the same object", which is a roundabout way of saying that either there are no dogs or there is only one dog.

It's a fun challenge to come up with different ways of expressing that there is at most one dog!

6. First-Order Logic in Arguments: Epimenides's Meal Plan

Below are some flawed arguments about certain statements. Identify the flaws in each argument. (Hint: Write out each statement in first-order logic.)

a. **Situation:** Epimenides, who is a Cretan, says "all Cretans always lie."

Incorrect Argument: We'll show that Epimenides's statement is a paradox (a statement that cannot be true or false). If Epimenides tells the truth, then all Cretans always lie. Since Epimenides is himself a Cretan, he must be lying, which is impossible because we assumed that Epimenides is telling the truth. This is a contradiction.

If, on the other hand, Epimenides is lying, then his statement is false and all Cretans never lie. Since Epimenides himself is a Cretan, then he must be telling the truth, which is impossible because we know that he was lying. This is also a contradiction. Therefore, this statement cannot be true or false.

The issue is where this argument says that if Epimenides is lying, all Cretans never lie / are always truthful. The actual negation of "all Cretans always lie" is "some Cretan does not always lie", i.e. "some Cretan sometimes tells the truth." So, if Epimenides is lying, then we just know that some Cretan sometimes tells the truth. It doesn't have to be Epimenides at this point in time. Therefore, this statement only leads to a contradiction if Epimenides is telling the truth, so Epimenides is lying.

This statement is known as "Epimenides's paradox"... even though it's not actually a paradox.

b. **Statement:** In every non-empty dining hall, there is someone in the dining hall who could truthfully say "if I am eating, then everyone in this dining hall is eating."

Incorrect Argument: We'll show that this statement is false. Since this statement makes a universal claim, we can disprove it by identifying a counterexample: a dining hall where there is someone who would be lying if they said "if I am eating, then everyone in this dining hall is eating." Consider a dining hall d where some person p in d is eating and some other person in d is not eating. p would be lying if they said "if I am eating, then everyone in d is eating", because the antecedent is true but the consequent is false. Then, there exists a dining hall that does not have the property in the statement, so the statement is false.

If we model a dining hall as a set of people, here's a rough first-order logic translation of this statement. (You can also translate "if I am eating, then everyone is eating" into first-order logic, but it turns out that's not where the issue in the proof lies.)

$$\forall d.\Big(DiningHall(d) \to \big(\exists p \in d.(TellsTruth(p))\big)\Big)$$

The issue is where this argument says that a counterexample would be a dining hall where there is someone who would be lying. However, a counterexample would actually be a dining hall where *everyone* would be lying. In first-order logic terms, the negation of the existential statement $\exists p \in d.(TellsTruth(p))$ is the universal statement $\forall p \in d.(\neg TellsTruth(p))$, rather than the existential statement $\exists p \in d.(\neg TellsTruth(p))$.

It turns out this statement is always true. It's known as the "drinker paradox"... even though it's not actually a paradox.