Conditionals and Curry

1. Introduction

Paradox is one of the driving engines of logic. When we are led from apparently acceptable premises to an apparently unacceptable conclusion, one thing that often strikes us is that we should check the cogency of our reasoning. The paradoxes of greatest interest are, of course, those where the reasoning seems hard to fault. The Liar is a classic example ("This sentence is false") - it seems to require nothing more than the T-scheme and some very appealing principles of quotation and logic. Curry's paradox (e.g. "If this sentence is true then everything is true") is similar, in many respects, but at least in this one: the principles employed to generate the absurd conclusions involve some of the most secure we have in the logic of conditionals, plus some others that are difficult to fault, yet the problem does seem to be in one of these logical principles.

One way to approach this sort of problem is to see what technical solutions are available, preferably those that retain as much as possible of our ordinary principles, and a great deal of very ingenious work has been carried out on both the Liar and Curry paradoxes by philosophical logicians, as well as other paradoxes like the Sorites, the Slingshot, and so on. But of course after the technical options have been examined, our work is not yet finished. Which paradox-blocking option is to be preferred? I think ideally we would resolve such paradoxes by applying an understanding of the phenomena they involve. A philosophical understanding of truth, negation, and representation seem needed to discover the truth about the Liar, and we need a general theory of vagueness to satisfactorily explain what is going on with the Sorites. So to with the Curry paradox, or at least the version stated with "if.. then...". I think the way to understand where the Curry paradox should be blocked can be discovered through an understanding of conditionals. It would also be desirable if the understanding relied upon drew its support from sources other than the need to block a particular paradox under discussion. An account of vagueness otherwise unmotivated apart from the fact that it blocks the Sorites with minimum costs elsewhere, while it would be interesting and perhaps important, seems less prima facie appealing than an account of vagueness that drew a lot of its plausibility from other sources, and on which a solution to the Sorites "dropped out".

That sort of independent motivation is what I claim to offer here for Curry's paradox. The story I shall tell about how the paradox is to be blocked is based on a theory of conditionals developed on entirely different grounds, employing a semantics which treats impossible worlds as seriously as standard counterfactual semantics treats possible ones. Curry conditionals turn out to be a special class of counterpossible conditionals, and can be treated in the same general way as other counterpossibles are treated. That story is one where the defect in the Curry paradox argument, as well as a diagnosis of the status of Curry conditionals themselves, "drops out" in a reasonably straightforward manner.

Since I will be trying to shed light on the version of Curry's paradox that uses "if.. then.." employing a theory of conditionals, my project here will not shed any light directly on the many other versions of Curry's paradox: the material implication version, the strict implication version, the set-membership version, and so on. So it is worthwhile saying something here about why I think the conditionals version of Curry's paradox might deserve a different treatment from some of the other varieties.

The extensional version of Curry's paradox seems to be a close relative of the Liar paradox: the material implication version of the instance given above, in effect, is "either this disjunction is not true or everything is true". Paradoxes involving truth and extensional connectives are difficult to assign a truth-value, even though they seem meaningful: perhaps they have some sort of defective semantic value. In this respect they might be analogous to other apparently meaningful sentences which it is hard to assign a truth-value to: vague expressions when applied to borderline cases, cases of empty proper names in sentences with extensional predicates ("Louis XXIII is bald"), and perhaps others.

Even when sentences are apparently too defective to deserve a truth-value, however, they appear to embed in some more complex structures to deliver claims that are straightforwardly true or false. Even when "The Liar sentence is false" is somehow semantically defective, it may be innocuous to claim that Smith believes that the Liar sentence is false, or that Jones claimed the Liar sentence is false. (Reconstructing debates about the Liar will be particularly difficult if we cannot say what anyone claims or believes about liar sentences.) Other suspicious, and perhaps semantically defective, claims can embed in conditionals with little apparent trouble. Bob is very much a borderline case of baldness, so "Bob is bald" may lack a definite truth value. Nevertheless "If Bob is bald, Bob is balder than Gerald" may be straightforwardly true if Gerald is hairy, or straightforwardly false if Gerald is hairless.

It can be, of course, a puzzle to explain *how* a sentence that cannot easily be assigned a truth value can nevertheless be embedded in more complex constructions to provide sentences that seem to receive straightforward truth-values. Especially if we think that a meaningful sentence requires, on an occasion of utterance, a *truth condition*, where a truth-condition is enough to determine a function from worlds to truth-values: lacking a definite truth-value at the actual world would seem to go along with lacking a definite truth-condition too. Nevertheless, the fact that a theory of this phenomenon is not ready to hand should not make us deny the existence of the phenomenon itself.

One feature shared by "believes that.." "claims that..." and "if.. then.." is that they create non-extensional contexts: the extension of the whole is not a function of the extensions of the clauses embedded in those constructions. Indeed, they create *hyperintensional* contexts: substitution of expressions that necessarily share a truth value cannot always be done *salva veritate* in these constructions. (You can believe one necessary truth without believing them all, make one impossible claim without making them all, and as I will discuss below, substituting necessarily equivalent antecedents into if.. then.. constructions can change their truth value too.) Even when a paradoxical piece of language cannot easily be assigned an extension, or an extension across possible worlds, it seems, on the face of it, that sometimes it can be embedded in these hyperintensional contexts to generate claims that are straightforwardly true (or straightforwardly false). So it seems that resolving what to say about the extensions of paradoxical sentences does not settle the question of how they function in hyperintensional contexts: and in the other direction, we should not expect a story about how they can function in hyperintensional contexts to automatically resolve what to say about extensional puzzles and paradoxes involving them.

Hyperintensional contexts raise paradoxes of their own: if I am right that "if… then..." is hyperintensional the Curry paradox for "if… then..." is one of these. One interesting feature of hyperintensional paradoxes is that they can arise even in formal systems where extensional paradoxes are blocked. Tucker and Thomason 2011 provides a discussion of a number of these paradoxes, and argue convincingly that some approaches to paradoxes such as the Liar and Russell's Paradox gain little traction on them. It is becoming increasingly plausible that we may need to treat hyperintensional paradoxes differently from our approach to more studied paradoxical sentences that do not employ hyperintensional constructions.

This is not supposed to amount to a knock-down case that we can, or must, solve paradoxes that arise using these constructions in distinctive ways, ways that do not generalise to solve paradoxes that arise from our understanding of extensional and (merely) intensional constructions. But the observations above should be enough to motivate the project of at least considering whether we might make progress in understanding the behaviour of these hyperintensional constructions without holding that progress hostage to solutions to other, extensional, semantic paradoxes. Proof that we can make independent progress on puzzles about hyperintensional constructions would be best provided by *in fact making* such progress. To that task I will now turn.

2. Curry's Paradox and Conditional Proof

Curry's Paradox (first presented in Curry 1942, though I will not use the same presentation) uses plausible logical principles and a conditional to produce unacceptable conclusions. Consider, for example, (C):

(C) If C is true, then all humans are ten feet tall.

Assume C is true.

(1) C is true.

It follows from the T-schema, and the fact about what sentence "C" names, that

(2) If C is true, then all humans are ten feet tall.

From (1) and (2) it follows, by modus ponens, that

(3) All humans are ten feet tall.

Discharging the assumption, by conditional proof we have

(4) If C is true, then all humans are ten feet tall.

And now (4) rests on no assumptions. But from (4) and the T-schema we can infer (1) again, and from (1) and (4) we get (3). So we have proved (3) from no assumptions: despite appearances, all humans are ten feet tall. (And that's a theorem, what's more.)

Obviously something goes wrong in the above proof. Not all humans are ten feet tall. Furthermore, if all of those devices were okay, we could use variants of (C) to prove whatever we liked. So it seems we should give up something. We have four main options:

1) Deny there is any such (meaningful) sentence as C.
2) Deny (the truth of all the instances of) the T-schema.
3) Deny the validity of modus ponens.
4) Deny the validity of conditional proof.

(There are more exotic options that have been suggested to deal with one or other logical paradox, such as denying that certain claims can be assumptions in arguments (Geenough 2001), or denying that sub-arguments have the same logic as arguments (Brady 2006), or denying that entailment is transitive (Ripley 2013, Weir 2015) but I will leave aside these exotica here.)

I list these options as involving denial, but in fact that might be stronger than required. Some approaches to "defective" commitments, as I mentioned above, advocate not denial of the problematic claims, but a refusal to either assert or deny them – and not merely from ignorance or being under-opinionated, but because that's the thing to do.

For example, one does not want to deny something with a truth-value gap, if denial is a matter of taking on a commitment to the negation of the thing denied, for that negation is gappy too. So someone who thought that the T-scheme had gappy instances, or that it was neither true nor false that modus ponens was valid, for example, might prefer to refrain from committing themselves to e.g. the T-schema, rather than denying it. With this important caveat, for simplicity I will carry on talking as if the options available all involve denying the relevant principles, though strictly speaking a principled non-acceptance may be enough.

Let me consider each of these four options in turn.

The line that sentences like (C) are somehow meaningless is a popular one: that certain sorts of apparent self-reference mean that a sentence does not express a proposition, for example.[1] Most of the arguments for and against have been rehearsed before, so I will not go through them here. It remains incredible to me that sentences like C are meaningless. I understand the constituents, and I understand how they are put together. We rely on our understanding of the sentence to realise that it is paradoxical in the first place. Some Curry sentences can be translated from one language to another, on the face of it: it is hard to see how to do this if they are meaningless. It is even uncomfortable to say these things if we think the sentence somehow fails to express a proposition even though it is meaningful in some lesser sense: it appears to be saying something (it seems to say that if it is true, then all humans are ten feet tall), and that seems to be enough to express a proposition. Presumably none of these considerations will in fact budge any philosopher who has decided the best way to solve the paradox is to argue that it is not there.

There is one new contribution I want to make to the reasons not to treat Curry paradox sentences as meaningless. Suppose we grant, for the sake of the argument, that sentences like the Liar do not express propositions because they have vicious self-

[1] Tarskian theories of truth based on Tarski 1944 will naturally have this consequence. Broadly "contextualist" approaches such as those of Parsons 1974 and Glanzberg 2001, if extended to the Curry paradox, would ensure that Curry sentences did not express propositions, at least in crucial contexts.

reference, and further that truth and falsehood ascriptions to the Liar do not express propositions for the same reason. Still, we seem able to reason hypothetically about the Liar having various truth-values, and what would happen were it to do so. It looks like we can even make perfect sense of conditionals about them: if Graham Priest's dialetheism were correct about the Liar paradox, a lot of classical logicians would be making a big mistake... and so on. Now, I'm not sure how we are supposed to be able to do that if the Liar does not itself express a proposition, or at least if mentions of it that ascribe it truth-value do not (assuming these can somehow be distinguished). But whatever we say about that challenge, it is just inescapable that a lot of the argument around the Liar employs this sort of hypothetical reasoning and even assertion of conditionals about what would be going on if one or other hypothesis were correct. Even if in our theoretical moments, we say that ascriptions of truth-values to the Liar and connected sentences are meaningless, we can't help thinking hypothetically about what would be the case if it did have one status or another.

The lesson to draw from this observation, I think, is that even if vicious self-reference makes sentences "meaningless" enough to not express propositions when those sentences have only extensional machinery such as truth-predicates, it is not enough to make conditionals embedding those sentences automatically "meaningless" enough to not express propositions. So when we have vicious self-reference in a conditional sentence like a Curry conditional, we should not necessarily think that this viciousness stops the conditional from expressing a proposition, *even if* that is the story we like about the basic Liar sentence.

In the end, I suppose, whether this argument for not applying a "meaningless" approach to the Curry paradox is very persuasive might depend on how good the other alternatives look. The most important argument I want to stress against this and the other rival options is, in the end, the appeal and straightforwardness of my positive proposal. (That's not to say that I think that is the best reason *tout court* to reject the rivals, but just that it is the central one being offered in this paper.) So suppose we do allow that the Curry conditional is meaningful and expresses a proposition. What of our other options?

Option (2) is very tempting to many.[2] If we are not willing to accept all instances of the T-schema, and in particular we are not inclined to accept instances with the sort of paradoxical self-reference we find in the Liar and supposedly in Curry conditionals, then we can block the problematic argument. Of course we have not finished with Curry, because we still would still like to know what its truthvalue is, and so on, but the threat from the paradox would be blocked.

In fact, it is not just the T-scheme that we need to jettison for this strategy to work, but also the admissibility of a T-rule: that from T<p> as a premise we can infer p as a conclusion, and vice versa. Those will enable our argument to go through just as well. Dropping some instances of the T-scheme due to considerations about the conditional would not be enough on its own, then.[3] As well as dropping some T-rules, we would have to go further than giving up the "T-scheme", if that is understood as being stated with "if.. then.." (T<p> iff p). We would also need to drop any variant that would allow us, when taking that variant as a premise, to get from one side of the schema to the other. The most common form of stating the T-scheme, after all, is with a *material* biconditional. We would have to reject, or withhold our assent from, those material biconditionals as well, unless we were persuaded that modus ponens was invalid for the material biconditional (as those who reject disjunctive syllogism will think).

There are too many different critiques of the T-scheme and the T-rules to launch a

---

[2] Kripke 1975 provides constructions where there will be "ungrounded" counterexamples to the T-scheme, and it is natural to treat Curry paradoxes this way in that system. Classical approaches rejecting bivalence include those of Horwich 1998 pp 40-42, which denies an unrestricted T-scheme. But there are many, many other theories that restrict the T-scheme: and armed with this restriction, it is very tempting to apply it to Curry paradoxes.

[3] I am tempted to think that some instances of the T-scheme, considered as a bi*conditional,* are false in some contexts. Consider "If nothing were true, then "nothing is true" would be true." I think in some contexts we should take that conditional to be false. If nothing were true, it may be that dogs would bark and cats would meow, but nothing would be true, including "nothing is true" and the proposition we in fact express with "nothing is true".

general defence of them here. One thing I will point out is that even those who reject the unrestricted T-scheme (for either the material conditional or a less traditional conditional) or the unrestricted T-rules accept that for a great many ordinary instances, T<*p*> is true when, and only when, *p* is. Most would accept that postulating exceptions to that rule is not to be done lightly. And for good reason, too: our ordinary practice of going freely between claims about the truth of a given proposition and asserting that proposition itself would be under suspicion if that was an easily defeasible move.

I think we should only reject instances of the material T-schema when we have excellent reason for doing so. Arguably, we may have that in the case of the Liar – if the Liar itself is somehow defective and unfit for assertion, a truth-functional compound of it may well inherit that trait. On my preferred treatment of Curry conditionals, we do not have to compromise on the material T-scheme for Curry sentences, which seems to be a big advantage, and a significant advantage even by the lights of some of those who reluctantly reject some material T-scheme instances, if they think it is a cost to be paid grudgingly when we have to follow that option. Of course, whether this advantage is outweighed by disadvantages is not settled by the observation that it is better to keep material T-biconditionals when you can.

Option (3) seems to me about the least appealing. Denying the validity of modus ponens for English language "if.. then.." assertions strikes me as desperately unappealing, though it of course has respectable defences (McGee 1985, Lycan 2001 ch 3, Briggs 2012, among others). Even more unappealing is the thought that there's nothing in the vicinity of the natural language conditional that is ponendable and which otherwise shares enough of "if"'s features to allow a Curry sentence to be constructed with it, given that it can be constructed for the English "if". But apart from this unappealingness, the features that give rise to exceptions to modus ponens according to McGee, Lycan or Briggs are missing here. The Curry sentence is not syntactically a nested conditional as in McGee's and Briggs's examples, nor does it seem trade on our not including the actual world in the "real and relevant possibilities" (if any) we envisage when entertaining the antecedent, as in the cases

Lycan argues modus ponens fails. If we want to say that modus ponens is invalid when we consider Curry conditionals, we need a more general philosophical story about why modus ponens would fail here, if the story is to be satisfactory. I cannot see what that story would be, so in addition to my lack of sympathy with abandoning one of our strongest logical intuitions, the fact that there currently seems no prospect of doing this for independently motivated reasons should mean that option (3) is not the way to go.

Option (4) is the option I want to endorse: for the English language "if... then...", the fact that B is a logical consequence of A is not sufficient for "if A then B" to be a logical truth. Indeed, I go further: I think there are counterexamples, where B is a logical consequence of A, but "if A then B" is false. This goes further than merely denying conditional proof, of course, because denying that a given "if A then B" is a logical truth is one thing, but denying that it is true at all is something stronger. If conditional proof is invalid, then the Curry reasoning above is invalid, which would already parry this threat that, for instance, "all humans are ten feet tall" will come out as a logical theorem. Indeed, I will want to argue that what we have here is a case that is itself a counterexample to conditional proof: I am prepared to allow that the proof works up until step three, so the following argument (call it A) is valid:

A:
(C) is true.
Therefore,
All humans are ten feet tall.

(I thus do not need to deny that (C) and "(C) is true" are meaningful, assumable, well-behaved with respect to the T-schema and that applying modus ponens is valid.) Nevertheless, the conditional we get from assuming the first line and then using conditional proof after the third line is (C) itself: and with the other machinery in place, allowing C to obtain produces the paradox we want to avoid. Without the conditional proof, step, though, the argument is stopped in its tracks. From the mere validity of the inference **A** we do not establish any relevant conditionals.

Someone wanting to deny the validity of conditional proof is under a great deal of pressure to say further that the step from maintaining the validity of **A** to the truth of C is itself one of the counterexamples to conditional proof. The corresponding conditional is (C) itself, after all, and the truth of (C) is bad enough. Once one accepts the truth of C, the argument from (C) to the conclusion that all humans are ten feet tall is just the inference from (C) to the truth of (C) and then **A** again. But the conclusion that all humans are ten feet tall is of course entirely unacceptable, even apart from any worries about whether it is a theorem.

So my task, will be to do two things: to argue that it is plausible that conditional proof is not valid, and to argue more particularly that the step from endorsing the validity of (A) to endorsing (C) is a particular case that should seem to us like one of its counterexamples. One of the reasons for accepting this second claim should be obvious: we want to avoid being committed to all humans being 10 feet tall. But in addition to this obvious motivation, I shall also have something to say about what independent reason we might have to find the step from endorsing A to accepting C suspicious, if we accept my semantic story about what is going on here.

I should clarify exactly what I mean when I claim that conditional proof is invalid. I am only claiming that it fails for the English language "if.. then..". In particular, I am not claiming that it fails when we are considering a material conditional. The central deduction theorems for classical logic, for example, is that when B is a logical consequence of A, then it is a theorem that A materially implies B. Steps from logical validities to material conditionals, strict conditionals, relevant conditionals, or whatever are often called "conditional proof", and may have some claim to that label, if only because of historical precedent: nevertheless, they are emphatically not what I am going to be talking about here. Of course, this means that I am not addressing, for example, the "material" version of Curry's paradox: the threat that we can show that an arbitrary *p* is true by considering a claim such as

(CM) CM is true HOOK *p*

CM is not an intensional paradox: it is an extensional paradox, and behaves just like

(ML) Either ML is false or *p*.

which does not, on the surface at least, involve conditionals at all. (ML) and (CM) are of a piece with the other "extensional" semantic paradoxes, it seems to me: at any rate, my discussion of semantic paradoxes employing conditionals is not about them.

I should also point out that the inference I need to deny is so-called "single premise" conditional proof. Conditional proof in full generality, it might be thought, is the step from a valid argument from a set of premises SIGMA, including A, to a conclusion B; to another valid argument from a set of premises SIGMA - A to the conclusion "if A then B". This form of conditional proof is rejected by many formal systems designed to model "if.. then.." or some of its features. It is invalid for (classical) strict implication, for example: the inference from {*p*, ~*p*} to *q* is in general valid, but the inference from {*p*} to ~*p* STRICT *q* is not in general valid. This sort of case also demonstrates the invalidity of multi-premise conditional proof in David Lewis's counterfactual logic: the inference from {*p*, ~*p*} to *q* is again valid, but it does not follow that the nearest ~*p* world to a *p* world has *q* true according to it for arbitrary *q*. (There are no blue elephants, but it is false that were there blue elephants, the Earth would explode.)

Single premise conditional proof, from a valid argument with A as a singleton premise and B as a conclusion, to the validity of the conditional "if A then B", is validated by strict conditional logics and Lewis/Stalnaker logics, and it is this that I must reject if I am to block the Curry argument in the place I want. One common response to the Curry paradox already does this, or something very similar to it. There are a family of logics that attempt to respond to the Curry paradox by denying a principle called Contraction. Let me briefly discuss this approach, since I conceive of it as a way of denying full-blown conditional proof: though I will offer criticisms of it after setting out my own approach.

One limited way to drop conditional proof that has been a popular response to Curry paradoxes is to drop *contraction*. Contraction is a structural rule that allows one to treat any number of uses of an assumption as only invoking that assumption once. It is a standard part of most logics: once I have assumed A and also B, (A&B)&A follows just as much as (A&B), does in classical logic. If we drop contraction, we can keep what is, from the point of view of classical logic, a *limited* form of conditional proof, in which we can infer A→B when GAMMA, A I= B, *provided we only discharge one use of the assumption A each time we perform conditional proof.* (See Restall 1994, or Beall and Murzi 2013 for a recent discussion.) We can keep it, that is, without incurring a Curry paradox, since the proof in effect relies on using the assumption twice and discharging it once. (Strictly, to avoid Curry-style paradoxes one wants a logic that is *robustly-contraction free* – see Restall 1993 – but I will not pursue this subtlety here.) Standard logics without contraction also do not have formulas like (A→(A→B))→(A→B) or (A& (A→B))→B being in general valid, so they block alternative routes to the paradoxical conclusion as well.

Dropping contraction is a significant weakening of one's logic, since it blocks many inferences in first-order logic as well. But the main argument against this approach I want to offer here is that, from my point of view, it does not go far *enough*. The picture I defend rejects many of the conditional theorems endorsed even by standard contraction logics. (See Nolan 1997 p 547 for a discussion of why (A&B)→A plausibly has counterexamples, to take one extreme case.) Another typical concern with contraction-free logics is that they are primarily motivated only by the desire to avoid paradox, rather than possessing an independent philosophical motivation. While I think there is something to this concern, for at least some contraction-free proposals, space precludes a careful examination of the issue. It is only fair to point out that philosophical motivations, rather than merely technical ones, for rejecting contraction are sometimes offered. See Priest 1992 and Restall 1994 for two examples.

3. Understanding Curry's Paradox in my Framework

To have a properly satisfying solution to Curry's Paradox it is not enough just to say which bit of the Curry arguments fail. We would also like the solution, whatever it is, to fit with a more general understanding of conditionals. The ideal, I take it, would be if there was a theory of conditionals we were otherwise attracted to that resolved the paradox in an intuitively satisfying manner, using only resources that were independently motivated: that is, we did not introduce new semantic machinery simply to avoid Curry's paradox.

I think I have the holy grail here. My account of the semantics of conditionals was not constructed with Curry's paradox in mind – indeed, as far as I can remember I only realised that it offered any kind of solution to this problem in 2005.[4] I think the treatment we get of Curry's paradox in this setting is theoretically satisfying: though asking me whether my proposed explanation is satisfying is like asking a parent if their child is cute, so I recommend making up your own mind on that count. The paradox gets resolved in a manner that is appealing in other ways as well: the problematic Curry conditionals turn out to be false (only), so we do not need to relax assumptions of bivalence or excluded middle, or indeed any classical principles to deal with them. My solution only applies to Curry *conditionals*, Curry natural language "if then" sentences in natural language, or in the formal treatment I offer of these sentences. It does not shed light on the material-conditional or strict-conditional versions of the puzzle. I will discuss below how much of an objection this is, but I hope I have already motivated the propriety of treating "extensional" and "intensional" paradoxical statements separately.

Let me provide an outline of my view of the so-called "counterfactual" conditional (detailed in Nolan 1997, though the view presented here is slightly more opinionated than the 1997 paper in some minor respects). I think that the conditional is a "closest world" conditional with a semantics of the general sort due to Stalnaker 1968 and Lewis 1973. I hold that a conditional A→B is true at a possible world w if there is a world w' according to which A and also according to which B, that is closer to w than any world w'' according to which A but *not* according to which B.[5] I think it is most

---

4 I hedge a bit because my Nolan 1997 p 554 points out that statement *modus ponens* fails in my system: and the *presence* of statement *modus ponens* is one of the things I think is a classic symptom of susceptibility to Curry's paradox. But if I suspected in 1997, I forgot again.

5 I should take this opportunity to note a technical error in my statement of the truth conditions of

straightforward to treat a conditional as false otherwise. I prepared to gloss "closeness" as similarity in relevant respects, and like Lewis, I hold that closeness is a contextual matter – the closeness ordering is supplied by a context, by supplying which ordering of relative similarity is needed for each conditional.

My view differs from Lewis's in that there are worlds other than possible worlds. There are impossible worlds as well, and they play a role in the semantics for conditionals. The most obvious place they play a role is with counter-possible antecedents: when A occurs in no possible worlds, Lewis and Stalnaker's accounts say that A→B is automatically true. My account can be more nuanced. Even if it is impossible, A will be true according to some impossible worlds and not according to others, so whether A→B is true is not settled simply by A's being impossible – it depends on whether the impossible worlds according to which A is true and also according to which B is true are more relevantly similar than all the impossible worlds according to which A is true but not according to which B is. "If you square the circle you will amaze mathematicians" is true, but "If you square the circle I'll give you my house" is not.

I am also generous about what impossible worlds there are. I would like to be as generous as I can: impossible worlds are unruly entities, and what is true at them is pretty much arbitrary (*p* and *q* can both be true according to a world without (*p*&*q*) being true according to that world, and conversely, for example). I am inclined to let some impossible worlds be incomplete as well – some might object to calling such things "worlds" if they are incomplete, but this objection is only terminological so far as I can tell. We need to be careful to distinguish what is true about a world (possible or impossible) from what is true according to it, of course: a world can have true according to it ("represent" in some general sense) all sorts of things that are not in fact the case. Many worlds have both *p* and ~*p* true according to them, but this no more means that the law of non-contradiction is in fact violated than if a piece of

---

counterfactual conditionals in Nolan 1997 p 564, where I stated it in terms of A HOOK B holding at spheres. While this definition would have been suitable for possible worlds, whether A HOOK B holds at an impossible world has little to do, in general, with whether A holds or B holds there.

paper has both p and ~p written upon it.

Validity is a matter of truth-preservation in all *possible* worlds, as is common: so the fact that impossible worlds are logically ill-disciplined does not undercut any classical logical principles, or principles of whatever other base logic one chooses to use, if one objects to classical logic, for example because of the extensional semantic paradoxes. Impossible worlds thus only make a difference to the behaviour of the conditional, in the logical system I prefer, though they may have other applications as well outside logic, such as in our treatment of impossible fictions (see e.g. Lewis 2004) and hypothetical reasoning about impossibilities (Nolan 1997 pp 558-561).

The counterfactual, or subjunctive, version of the Curry paradox would probably be enough for one paper, but for what it is worth the account I will give is also the key to resolving the indicative Curry paradox as well. I think indicatives are also closest-worlds conditionals, albeit with different closeness orderings selected by contexts (Nolan 2003). The logic is therefore the same, and according to me at least the closeness relations selected by context will be similar enough that what I say about the counterfactual will apply, mutatis mutandis, to the indicative.[6] I will continue to focus primarily on the counterfactual conditional in what follows, though, for simplicity.

This account allows for failures of conditional proof: just because an inference is valid, so that any *possible* worlds where the premises are true the conclusion is as well, it does not follow that for any *impossible* world where the premises are true the conclusion is also true. Classically, "it is raining and it is not raining" validly entails any consequence you like (e.g. that nuclear war has broken out). Nevertheless, neither "if it is both raining and not raining, then nuclear war has broken out" nor "if it had been raining and not raining, nuclear war would have broken out" are plausibly

[6] It might be otherwise if one rejected my condition (3) on the similarity relation (Nolan 2003 p 218), as for example Brian Weatherson does (see page 339 of Weatherson 2009). More reason not to reject my (3), I say, though no doubt there are other ways to treat the indicative closeness relation that will allow one to solve the Curry paradox in the way I will suggest.

true: nuclear weapons have many effects on climate, but not that one. In this framework, that means that not all the nearest worlds where it is raining and not raining are worlds according to which nuclear war breaks out. Conditional proof fails because of cases like these, independently of what verdicts we might have about Curry conditionals.

Curry conditionals with false consequents are not true. (One tempting direct argument from this is just modus tollens from the negation of the consequent, which shows they are not true even on the assumption that they are!) So, on the impossible worlds framework sketched, this means that among the nearest worlds which represent *(C) is true*, there are worlds that do not represent that all humans are ten feet tall. So not only is the Curry argument blocked, due the independently motivated invalidity of conditional proof, but we have an available diagnosis of what is wrong with Curry conditionals. Paradoxical conditionals like (C) would be true in *impossible* worlds, and the nearest worlds according to which *(C) is true* are impossible. (Furthermore, since presumably (C) and *(C) is true* stand or fall together at possible worlds, the nearest worlds where (C) holds will also be impossible.) I will later argue that conditionals like (C) should be true *only* in impossible worlds. And since these worlds are impossible, and the consequent of (C) does not obtain there, (C) is just plain false at the actual world. The Curry sentence's sting is drawn without compromising the T-scheme or modus ponens, nor even requiring us to abandon bivalence or excluded middle.

Notice that in abandoning conditional proof for the counterfactual conditional, this approach also invalidates the statement forms of contraction: $(A\rightarrow(A\rightarrow B))\rightarrow(A\rightarrow B)$ and so-called "statement modus ponens": $(A\&(A\rightarrow B))\rightarrow B$. While of course many instances of these forms are true, they are not theorems in this framework, and it is plausible, given the framework, that they have counterexamples. Let me spell out how the latter fails in more detail: the explanation for the former failing is similar. There is no general guarantee that the nearest world where $A\&(A\rightarrow B)$ is true will be a possible world: where A is impossible it surely will not be. And impossible worlds are arbitrary, in this system, so from the fact that $A\&(A\rightarrow B)$ is true at one nothing else is guaranteed to be true at it: in particular, not even B. (Not even A, for that matter.) So the nearest A& (A→B) world need not be a B world, so (A& (A→B))→B need not be true.

What might a plausible counterexample look like? Let A be "Modus ponens is invalid" and B be "modus ponens is valid". Of course, one impossible way things to be is still that modus ponens is invalid, and if modus ponens is invalid then modus ponens is valid, and also that modus ponens is valid. Intuitively the most natural way for modus ponens to be invalid would be for it to not also be valid. That does not quite tell us what to think were a world to both have true according to it that modus ponens is invalid *and* that if modus ponens is invalid then modus ponens is valid. But at least plausibly *some* of the impossibilities that might obtain if that bizarre circumstance obtains are ones that do not have true according to them that modus ponens is valid: they are consistent, for a start, in a way that worlds where modus ponens is valid *and* modus ponens is invalid are not.

Treating C in this way requires the verdict that C is straightforwardly *false*. The closest worlds according to which "C is true" are not among that represent "all humans are ten feet tall" as being true. What it takes for a conditional to be true at the actual world is that all the (relevantly) nearest worlds according to which the antecedent is true are worlds according to which the consequent is true, and conditionals are false otherwise: so C is false. While we need this much for the conditional to be false, there are other theoretical choice points to be made about what else to say about the nearest worlds according to which C is true (the "C is true worlds"). For example, is "C is true" true according to those worlds but not C itself (i.e. is "If C is true, then all humans are ten feet tall" true *according to* the nearest worlds according to which "C is true" is true?

The nearest impossible worlds according to which C is true might be ones where T-elimination fails; or they can be worlds where "C is true" and C itself are both true, but which are not closed under modus ponens (so that despite "C is true" and C both being true according to that world, still "all humans are ten feet tall" is not true according to that world). Or we may wish to say there are impossible worlds of both sorts equally close to actuality. We may even wish to say that, along with those impossible worlds, there are equally close impossible worlds that have both the antecedent and the consequent of C true according to them: the semantics requires that *all* the nearest antecedent worlds have the consequent true according to them, so even if we allowed some antecedent-and-consequent worlds into the nearest sphere, the conditional would still be false.

One of the above treatments must be applied to any Curry conditional with a false consequent. If any of those conditionals turned out to be true, after all, an argument to the consequent of that conditional could be constructed from true premises without use of conditional proof. One philosophical challenge my approach faces is making it plausible that the principles of relevant similarity deliver the right results. One response to that challenge is that we come to understand what similarities are relevant through our grasp of the counterfactuals themselves, so it is fair play to "reverse engineer" similarity standards, by doing our best to work out which counterfactuals are true, when we have a grip on that, and then extending the treatment to cover new cases. (This is the approach taken by Lewis 1979: see pp 466-67). Once we notice that the conditional is surely false, (or at least not-true), we can work out that the relevant similarity metric must vindicate this judgement.

Optimistically, I hope there is even more to say here: otherwise the "reverse engineering" response can seem somewhere between unhelpful and ad hoc. I can add a few more tentative things. The first is that many Curry sentences are odd specimens: those with consequents such as "everything is true" are not going to have true antecedents in worlds much like our own, especially in ones where there is no irregularity in the behaviour of the conditional itself. So plausibly, a world where "if this sentence is true then everything is true" obtains will be less impossible if the conditional does not behave in a typical way than those worlds where the conditional and also the consequent is true. What, then, about more innocuous ones, e.g. (C)? It is still a strange, paradoxical conditional. If it gained its truth in the usual way at a world *w*, then something about the human height to be true at every world relatively close to it (in fact, every possible world at which *w* was the closest world where the conditional was true). It causes trouble for our ordinary procedures for evaluating counterfactuals – why would human heights depend on *this* sentence's truth? And if they do not (or would not), why is the sentence true at all? It is easy to get the sense that something funny is going on, in worlds where it obtains, and at least in some of them, not too far from here, the best available explanation of its truth is not one that presupposes that conditionals behave there as they do in possible worlds. It is a world where a paradoxical sentence is true, and so plausibly may be a world where some principles involving truth or conditionals fail. At any rate, that line of thinking leads me to find it plausible that the relevantly similar worlds according to which "(C) is true" may well be ones that do not have (C) being true according to them, or where despite having true according to them "(C) is true" and (C) itself, nonetheless human heights do not vary much from actuality, since the variation is in the behaviour of conditionals.

The "reverse engineering" story is enough to convince me, since I accept it (perhaps reluctantly) elsewhere. So if the "even more" is unconvincing to you, dear reader, I recommend the reverse engineering argument as a fall-back. At the very least, the hope that something like the "even more" story can be offered suggests that there is a place to look for further understanding about how conditionals interact with impossibilities, and the semantic story here is of a piece with a more general semantic story required anyway. It is at worst a small extra step in an already motivated philosophical approach.

4. True Consequents and the SIC

The most damaging Curry paradoxical sentences are those with false consequents. But for a complete account of the Curry paradox, we should give an account of the Curry conditionals with true consequents as well. Consider

(D) "If this conditional is true, then Barak Obama lives on Earth".

Concluding its antecedent from the premise that it is true would not be such a disaster. It would, of course, be absurd to think that its consequent was a theorem, so we still are owed a story of how the original Curry reasoning is to be blocked, but if we allow that conditional proof is not *valid* we block that absurd result. So what truth value should we think (D) has?

The semantic framework might initially suggest that it would be true. Consider the closest world where the conditional is true (whether that world is possible or impossible).[7] Need we adjust the truth-value of "Barak Obama lives on Earth"? Initially we might think not, for two reasons. One is the by now familiar reason that to have "Barak Obama lives on Earth" false but the conditional true, we would need a failure of modus ponens, and ceteris paribus a world is more like ours if there are no such failures. In the previous cases we considered, there were powerful arguments like the *reductio* argument to tell us we should go to a world with a modus ponens failure (or similar), but those arguments seem absent here. The other is that we might think, pending further argument, that fiddling with the truth-conditions of odd self-referential conditionals need not make a difference to where I live: again, everything else being equal, a world where Obama lives on Earth is one more similar to the actual world than one where he does not.

So that seems to me the main planks of the case for taking (D) to be true. Nevertheless, I think the case for taking (D) to be false is a little stronger. The first consideration is a weak one: we get a more uniform story about Curry conditionals if we make them uniformly false. This is weak, because it seems such a silly form of induction elsewhere: we get a more uniform story about the (positive) sentences of Chemistry if we take them all to be false, but this uniformity would be a silly reason to reject the contents of Chemistry textbooks.

A much stronger reason is that it would lead to us having to say very odd things about the closeness relation on possible worlds even as it applies to relationships between

[7] Perhaps there is not a closest, because of ties or infinite sequences of ever closer worlds, but let us not worry about that for now.

possible worlds. We would have to allow that standardly with many Curry conditionals, the relevant closeness relation made an impossible world closer to a possible world than another possible world. The hypothesis that every possible world is closer to every other possible world than any is to any impossible world is one that I have elsewhere called the "Strangeness of Impossibility Condition" SIC (Nolan 1997). If the SIC is in place, the Lewis semantics for counterfactuals is correct (well, at least according to me) in the limiting case where the antecedents of conditionals are possible, and given the plausibility of the Lewis semantics, accuracy in the limiting case might be welcome. On the assumption that closeness is connected to similarity in relevant respects, SIC would also make conditionals with possible antecedents more tractable, if adjudicating relative similarity of impossibilities is often harder than adjudicating relative similarity of possibilities.

I expressed some scepticism about SIC in Nolan 1997, and I am now even more inclined to think it has exceptions (see e.g. the discussion in Vander Laan 2004). But even if SIC does not hold unrestrictedly, the sort of examples required by letting Curry sentences with true consequents be true are still ones I think we should think are unacceptable.

Suppose (D) were in fact the case (and classical logic obtained, and so on.) Consider W, the closest possible world (or one of the closest) where Obama does not live on Earth. It is presumably a lot like ours – very similar or the same laws of nature, quite a bit of similarity of history, and so on. (If you want to think about a concrete case, think about what would have happened if Obama's parents had been selected by a global lottery to be among the first Martian colonists in a possible world with a more advanced space program than ours.) Now consider (D) there. There, it has a false consequent, and so should be false, on pain of contradiction or other logical mishaps. But that would mean that the closest world to W where (D) is true is an impossible world where modus ponens (or the T-scheme, or...) fails but where I still do not live on Earth. That impossible world would have to be nearer to W than the actual world: since if the actual world were closer than any such impossible world, it would not be the case that the closest world to W where (D) was true was an impossible world containing a logical mishap.

This is very strange. The actual world is very similar to W in most respects, including logical ones. Why would (D) produce a context in W where the actual world was less similar to W than a bizarre impossible world with e.g. modus ponens failure? Rather than allow these sorts of failures of SIC, we should rather prefer to say that (D) is false even at the actual world. Generalising, I think it best to say that Curry sentences with true but contingent consequents are all false.

The caveat at the end of the previous paragraph signals that we do not yet have a general story about the truth-conditions of Curry conditionals. The problem with taking one with a contingent consequent to be true does not apply straightforwardly to Curry conditionals with necessarily true consequents. Consider "If this conditional is true, then all bachelors are unmarried." If we allow it is actually true, we will not find a possible world where the consequent is false to run the analogous argument. We will have to say, of the closest impossible world where not all bachelors are married, that some impossible world containing funny business with the conditional is closer to it than the actual world, or alternatively the semantics for the conditional is different in such a counter-analytic world: but why not go ahead and say this? The problem at least looks less acute.

Furthermore, for some Curry-like conditionals with necessary consequents we have some independent temptation to judge them correct. Carrie Jenkins pointed out to me that "If this conditional is true, this conditional is true" is an instance of identity ($p \rightarrow p$), and many have wanted to say instances of identity are not only all true, but all theorems. Indeed, it is hard to get identity to fail when conditionals are treated the way I prefer (see Nolan 1997, esp pp 554-555, 565). This instance of identity is anyway plausible: surely if the conditional is true, it is at least true?

Other Curry conditionals are plausible. "If this conditional is true, then some proposition is true." "If this conditional is true, then some conditional proposition is true." Surely, for example, the closest world where the first is true is not one that fails to represent that some proposition is true? Given that Curry conditionals with necessarily true consequents are not susceptible to the argument against Curry conditionals with contingently true consequents, and given some are plausibly true, what are we to say about them?

I suspect it does not matter very much what we do say about them – the rest of the picture seems not to depend on one choice or another about this case. For what it is worth, I think the best thing to do is to draw a distinction, allowing that some Curry conditionals with necessary consequents are true, some are false, and there may be a vague or otherwise indeterminate border between the two cases. I have already indicated why we might want some of them to come out true: the identity case, and cases where they are intuitively correct, since the nearest impossible worlds where there consequents fail to be true are very far away indeed, plausibly further than the nearest worlds where both they and their consequents are true. Curry conditionals with other necessarily true consequents, though, have consequents that fail at impossible worlds much like some possible ones (e.g. "nobody squares the circle" fails at some mathematically impossible, but perhaps not logically impossible, worlds). So it is tempting to treat them as we treat Curry conditionals with strange but contingently true consequents at the *possible* worlds where those contingent consequents are true.

As I have said, this diagnosis of Curry conditionals with necessarily true antecedents is not crucial to the rest of the analysis, once we have established the invalidity of the central paradoxical argument. If we regarded them all as true, or all as false, would just require some rearrangement of closeness conditions on impossible worlds (albeit ones that would rob us of Identity, if no Curry sentence is true.) Indeed, taking Curry sentences with contingently true consequents to be true (one or all) can be accommodated in the framework at the cost of some curious claims about relative closeness, and more violations of SIC than I think are warranted. My theoretical choices here at least demonstrate that the question of what to do with Curry sentences with true consequents is one that can be approached in a principled way with the conditional and semantic framework I endorse, even if we face choices between principles.

7. Conclusion

The Curry paradox involving a genuine “if.. then...” conditional has a conservative solution. The conditional is meaningful, and says what it appears to say, and it can be treated as having a classical truth value, and without interfering with the T-scheme or the T-rules. Given modus ponens and the T-scheme, disaster would follow from the truth of a traditional Curry sentence, but fortunately there is no real problem with simply allowing that the Curry sentence is false. Conditional proof must be rejected, but fortunately we have an independent understanding of why conditional proof might fail in cases like this. There is a natural philosophical story about conditionals embedding logically false antecedents which delivers a failure of conditional proof while explaining why it seems such an appealing inference: and an application of this story, that was developed independently of any desire to resolve the Curry puzzle, seems to me to straightforwardly explain what is happening here. Technical simplicity, preservation of cherished principles like bivalence and the T-rules, and a satisfying semantic story that follows from a more general, independently supported, philosophical approach to conditionals. I think we have the best solution.[8]

Daniel Nolan
School of Philosophy
Australian National University
Canberra, ACT 0200
Australia
Daniel.Nolan@anu.edu.au

---

[8] Thanks to audiences at the "What If?" conference at the University of Connecticut, the University of Melbourne, the Australian National University (twice), the University of Michigan, the University of St Andrews, and to Jc Beall, Carrie Jenkins, David Ripley and Dustin Tucker for discussion. This research was supported under the Australian Research Council’s *Discovery Projects* funding scheme (project number DP130104665).

References

Beall, Jc and Murzi, J. 2013. "Two Flavors of Curry's Paradox". *Journal of Philosophy* 110.3: 143-165

Briggs, R. 2012. "Interventionist Counterfactuals". *Philosophical Studies* 160.1: 139-166

Curry, H. 1942. "The inconsistency of certain formal logics". *Journal of Symbolic Logic* 7: 115-117

Glanzberg, M. "The Liar in Context". *Philosophical Studies* 103: 217-251

Greenough, P. 2001. "Free Assumptions and the Liar Paradox". *American Philosophical Quarterly* 38.2: 115-135

Horwich, P. 1998. *Truth*, 2nd edition. Oxford: Blackwell.

Lewis, D. 1973. *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. 1979. "Counterfactual Dependence and Time's Arrow". *Noûs* 13: 455-76

Lewis, D. 2004. "Letters to Beall and Preist" in Priest, G., Beall, Jc and Armour-Garb, B. (eds) *The Law of Non-Contradiction: New Philosophical Essays*. Oxford University Press, Oxford pp 176-177

Lycan, W. 2001. *Real Conditionals*. Oxford University Press, Oxford.

McGee, V. 1985. "A Counterexample to Modus Ponens". *Journal of Philosophy* 83: 462-471

Nolan, D. 1997. "Impossible Worlds: A Modest Approach". *Notre Dame Journal of Formal Logic* 38.4: 535-572

Nolan, D. 2003. "Defending a Possible-Worlds Account of Indicative Conditionals". *Philosophical Studies* 116.3: 215-269

Parsons, C. 1974. "The Liar Paradox". *Journal of Philosophical Logic* 381-412

Restall, G. 1993. "How to be Really Contraction Free". *Studia Logica* 52.3: 381-391

Restall, G. 1994. *On Logics Without Contraction*. PhD Thesis, University of Queensland.

Ripley, D. 2013. "Paradoxes and Failures of Cut". *Australasian Journal of Philosophy* 91.1: 139-164

Stalnaker, R. (1968). "A Theory of Conditionals" in N. Rescher (ed), *Studies in Logical Theory*. American Philosophical Monograph Series, no. 2. Oxford: Basil

Blackwell, 98-112.

Tarski, A. 1944. “The semantic conception of truth”. *Philosophy and Phenomenological Research* 4.3: 341–376.

Tucker, D. and Thomason, R. 2011. “Paradoxes of Intensionality”. *The Review of Symbolic Logic* 4.3: 394-411

Vander Laan, D. 2004. “Counterpossibles and Similarity” in Jackson, F. and Priest, G. 2004. *Lewisian Themes: The Philosophy of David Lewis*. Oxford University Press, Oxford pp 258-275

Weatherson, B. 2009. “Conditionals and Indexical Relativism”. *Synthese* 166.2: 333-357

Weir, A. 2015. “A Robust Non-transitive Logic”. *Topoi* 34: 99-107