

## Counterfactual donkeys don't get high

Mike Deigan  
Yale University

**Background:** Much work has been devoted to finding the best way to derive a 'strong' reading of indicative donkey sentences. On this reading, what would be naturally translated as  $\exists xPx \rightarrow Qx$  is equivalent with  $\forall x[Px \rightarrow Qx]$ . Van Rooij (2006) extends this project to *counterfactual* donkey sentences (CDSs) like (1).

(1) If John owned a donkey, he would beat it.

According to van Rooij, there is an analogous reading of these sentences, what Walker and Romero (2015) call the 'high reading', for which the following equivalence holds.

*Counterfactual Donkey Equivalence* (CDE):  $\exists x[Px] > Qx \Leftrightarrow \forall x[Px > Qx]$

To give a semantics which predicts this reading, van Rooij combines a dynamic binding theory of indefinites and donkey pronouns with the standard Lewis-Stalnaker ordering semantics plus one crucial modification: similarity relations between possibilities (world, variable assignment pairs) that hold only when possibilities share assignment functions. Wang (2009) criticizes van Rooij's account: she claims that CDE does not hold in general for any available reading, and provides an assignment-*insensitive* ordering semantics on which it doesn't. Walker and Romero (WR) defend van Rooij on this point, claiming that in some contexts there is a high reading of CDSs, and reject Wang's account on the grounds that fails to predict this.

**Summary:** I defend Wang's approach. After rebutting WR's argument against Wang's account, I show that to avoid being implausibly weak, the  $\forall$  on the righthand side of CDE must be interpreted in a way that ranges over merely possible entities, then argue that this results in it being too strong. I conclude that CDSs do not get high readings.

**WR's case:** WR observe that Wang's account makes predictions equivalent to the high reading (i.e. predicts 'high entailments') of assignment-sensitive theories (only) when the contextually supplied similarity relations have a certain special property: roughly, the world closest to the evaluation world where one individual satisfies the antecedent is just as close as the closest world where any other individual satisfies the antecedent. WR then provide a case which they claim has the high entailments, but where the similarity relation of the context does not have this special feature. The setup is that there are two farmers in the kingdom: the cruel Onophobus, who wants to own a donkey and has nearly the means to do so, and the kind Onophilos, who neither wants to own a donkey nor has nearly enough money to buy one.

(2) a. KING KAKOS: Here's what I think about the farmers in my kingdom. If a farmer in my kingdom owned a donkey, he would beat it.  
b. ADVISOR: No, it's not the case that if a farmer in your kingdom owned a donkey, he would beat it. Onophilos is a vegan and would never do so.

WR judge (2-b) true and say that "Clearly, a world in which Onophobos owns a donkey is more similar to the actual world than one in which Onophilos owns a donkey". They conclude that the ordering cannot have the special feature, since the closest farmer-owning-donkey worlds are ones in which only Onophobos owns a donkey and beats it, and so Wang cannot predict a true reading of (2-a).

This argument, however, depends on assuming that the similarity ordering relevant

for the semantics of counterfactuals aligns with some intuitive notion of similarity. But as has been known since Lewis (1979)'s response to Fine (1975), a similarity-based theory of counterfactuals cannot rely on an intuitive notion of similarity in any direct way. The world where Nixon's button pressing had led to a nuclear holocaust is intuitively more similar to the one where it miraculously malfunctioned, even though it is true (in Fine's scenario) that if he had pressed it there would have been a nuclear holocaust. It is unproblematic, then, for Wang to maintain that the similarity ordering in this scenario has the special property, even if that requires deviation from an ordering of worlds by intuitive similarity. A full defense of a Wang-style account would need to spell out what the similarity ordering must look like in order to generate special orderings where needed, but already we have said enough to see that it has not yet been refuted.

**The vase case:** Now we argue that lacking a high reading is an advantage of Wang's semantics. First we show that the  $\forall$  of CDE must range over merely possible entities, then present evidence that this results in a reading stronger than any that is available.

Suppose Allie and Bert think Mary the potter probably didn't make anything yesterday, but they aren't sure she didn't. And now Allie says the following.

(3) If Mary had made a vase, she would have made it from glass.

Now consider the following scenario.

*Case 1:* Mary didn't make anything, and there is no contextually relevant actual pottery. In this case, (3) does not come out trivially true or give rise to presupposition failure (due to vacuous universal quantification). Nor does it depend on looking at various non-pottery that exists in the world (the high entailment need not imply, for example, that if Mary were a vase that she made, she would have made herself from glass). Instead, what goes into determining the truth or falsity of (3) in Case 1 are some merely possible vases and their composition in worlds where Mary made them—the quantifier in question, then, must be a possibilist one.

If high entailments come from a special similarity ordering, as on Wang's account, we shouldn't expect them to arise when the antecedent is true in the evaluation world, since in such cases there can be no special ordering without violation of strong centering—the requirement that each world must be strictly closest to itself—a requirement assumed by all of van Rooij, Wang, and WR. We now consider a case, Case 2, where some contextually relevant actual entity satisfies the antecedent and consequent but a merely possible one doesn't satisfy the consequent in the nearest world where it satisfies the antecedent.

Suppose the conversation between Allie and Bert continues.

(4) a. *Bert:* No, she could have made it from clay!

b. *Allie:* Oh, I didn't know she had any clay left, nevermind what I just said.

Bert raises a relevant possible way for the antecedent to be made true that wouldn't lead to the truth of the consequent, which gets Allie to retract her claim. This is what we'd expect on a high reading. But now suppose that it turns out that Mary in fact *did* make some vases yesterday.<sup>1</sup>

*Case 2:* Mary made two vases, both of glass.

In this case it seems that Allie's utterance of (3) was true, if only by luck. And the fact that Mary could have made a different vase which she would not have made from glass

---

<sup>1</sup>True antecedents in cases like these are not generally problematic; see Anderson (1951).

has no bearing on its truth. Once the facts about what Mary actually made are known to Allie and Bert, challenging the original assertion again by raising the possibility of the clay vase is bizarre.

- (5) a. *Allie*: Looks like I was right after all.  
b. *Bert*: ??No, even though she *didn't* make any clay vase, she still *could* have made a vase from clay, and she wouldn't have made *that* from glass.

So it seems that in this case, any available reading of (3) is true. There are not high entailments in this test case, as the assignment insensitive semantics theories like Wang's predict, but that theories which allow high readings do not predict. After considering a couple of objections, I conclude that high readings, once they're understood to involve possibilist quantification, are too strong.

**References:** Alan Ross Anderson (1951). "A Note on Subjunctive and Counterfactual Conditionals". In: *Analysis* 12.2, pp. 35–38 ♦ Kit Fine (1975). "Critical Notice of David Lewis's *Counterfactuals*". In: *Mind* 84, pp. 451–458 ♦ David Lewis (1979). "Counterfactual Dependence and Time's Arrow". In: *Noûs* 13, pp. 455–476 ♦ Robert van Rooij (2006). "Free choice counterfactual donkeys". In: *Journal of Semantics* 23.4, pp. 383–402 ♦ Yingying Wang (2009). "Counterfactual donkey sentences: a response to Robert van Rooij". In: *Journal of Semantics* 26.3, pp. 317–328 ♦ Andreas Walker and Maribel Romero (2015). "Counterfactual donkey sentences: A strict conditional analysis". In: *Proceedings of SALT* 25, pp. 288–307