

Local similarity and the two readings of counterfactual donkey sentences

[A] The problem. Counterfactual donkey sentences vary systematically between two readings (van Rooij 2006, Walker & Romero 2015): for a sentence like (1), we usually obtain the entailments in (2), that is, for any donkey d in the domain of individuals we obtain a counterfactual that asserts that if John owned d , John would beat d , and the equivalence in (3) – familiar from indicative donkeys (\rightarrow), but with a counterfactual conditional ($>$) – holds.

- (1) If John had owned a donkey, he would have beaten it.
- (2) a. If John had owned donkey a , John would have beaten a .
b. If John had owned donkey b , John would have beaten b .
c. If John had owned donkey c , John would have beaten c ...

(3) $\exists x[Px] > Qx \Leftrightarrow \forall x[Px > Qx]$ (cf. Groenendijk & Stokhof 1991, van Rooij 2006)

This is the *high* reading; however, there is also another (*low*) reading that is most readily (but not exclusively) available in identificational sentences like (4), which do not entail (5).

- (4) If John had married a girl from his class, it would have been Sue.
- (5) a. If John had married Ann, Ann would have been Sue.
b. If John had married Betty, Betty would have been Sue.
c. If John had married Carol, Carol would have been Sue ...

Elbourne (2013) proposes a semantics for counterfactual donkey sentences that amounts to mixing standard D-type theory (Elbourne 2005, 2013) with a Stalnaker-Lewis semantics for counterfactuals (Stalnaker 1968, Lewis 1973). However, unlike dynamic accounts, this only derives the *low* reading.

[B] The aim. We argue that Elbourne's account undergenerates as it provides only one of the two empirically attested readings. To generate the other reading without giving up his general framework, we modify and combine the proposed analysis with an independently motivated analysis of counterfactuals that uses local, rather than global, similarity (Arregui 2009). We show that this provides a plausible connection between the high/low distinction and Gettier-type judgements on counterfactuals embedded under knowledge, which the dynamic literature cannot readily account for.

[C] Elbourne's semantics. Elbourne (2013) proposes the following semantics:

- (6) *Counterfactual (adapted from Elbourne 2013: 147)*
 $[[CF]] = \lambda p_{st}. \lambda q_{st}. \lambda s_s.$ for every minimal situation s' such that $p(s') \ \& \ w_s$ is otherwise as similar as possible to w_s , there is a situation s'' such that s'' is a minimal situation such that $s' \leq s'' \ \& \ q(s'')$

That is, the counterfactual quantifies over all minimal antecedent-situations in the closest antecedent-worlds and asserts that they can all be minimally extended into consequent-situations. It is easy to see that this does not derive the entailments in (2): if the closest antecedent-world is one in which John owns a , then (2a) will be entailed, but (2b) and (2c) will not. This is roughly equivalent to a dynamic account without the additional machinery introduced by van Rooij (2006) and, like such an account, can only yield the low reading. The problem is that Elbourne's semantics yields too few worlds: if the possible referents for the indefinite are in less similar worlds, we can't quantify over them.

[D] Local similarity. Independently of this debate, Arregui (2009) argues for a modified notion of similarity. Instead of the Stalnaker-Lewis “global” similarity (i.e. overall similarity between worlds), she develops a notion of “local” similarity between situations: the past tense of the counterfactual identifies a past (actual) situation (the *res*, identified with s in (7), which essentially collects the fact to be kept constant in the evaluation of the counterfactual), and the counterfactual quantifies over all counterparts of that situation in other possible worlds:

- (7) *Counterfactual (adapted from Arregui 2009: 7)*
 $[[\text{would}]] = \lambda p_{st}. \lambda q_{st}. \lambda s_s.$ for every situation s' such that $s \leq_m s' \ \& \ p(s')$, there is a situation s'' such that $s' \leq s'' \ \& \ q(s'')$
- (8) *Modal parthood (adapted from Arregui 2009: 7)*
 $s_i \leq_m s_j$ iff there is an s_t such that s_t is a counterpart of s_i and $s_t \leq s_j$

In addition to plausibly connecting the semantics of the counterfactual to the additional layer of past tense in counterfactual antecedents, this semantics provides an account of counterfactual pairs like (9) and (10), which can both be judged as true within the same world when presented independently, but which cannot simultaneously be true:

- (9) If New York had been in Georgia, New York would have been in the South.
- (10) If New York had been in Georgia, Georgia would have been in the North.

A counterfactual assertion *de re* of a particular situation (e.g. the location of Georgia) can make only one of these counterfactuals true at the same time (e.g. (9) in the above case).

[E] Knowledge diagnostics. So-called Gettier-cases (Gettier 1963) can be used to probe the content of the *res* (Arregui 2009): embedding counterfactuals under knowledge yields judgements that show that we only attribute knowledge of a counterfactual to an agent acquainted with the *res* that makes it true (see section H for an application of this test to our data.)

[F] The proposal. Our proposal brings together the key elements of both approaches in a novel way: we make use of Arregui's *de re* account of counterfactuals, yielding a larger number of worlds for quantification – since any world that contains a counterpart of the *res* will be quantified over – but constrain the situations quantified over by importing Elbourne's minimality requirements on situations, in order to guarantee uniqueness for the definite descriptions in D-type theory.

(11) *Counterfactual (proposal)*

$[[CF]] = \lambda p_{st}. \lambda q_{st}. \lambda s_s.$ for every minimal situation s' such that $p(s')$ and $s \leq_m s'$,
for every minimal situation s^* such that $p(s^*)$ and $s^* \leq s'$, there is a minimal
situation s'' such that $s^* \leq s''$ and $q(s'')$

Our proposal (11) differs from (6) in the following key feature: Elbourne completely separates counterfactual quantification from the topic situation, using s only to identify the actual world. We follow Arregui in assuming that s is bound by a higher past tense that identifies a specific temporally prior *res* situation. The counterfactual then quantifies over situations that satisfy the antecedent and, crucially, that are minimal antecedent-situations which are part of minimal antecedent-satisfying extensions of some counterpart of the *res*.

[G] High and low readings in the proposal. Consider the scenario and sentence in (12)-(13):

(12) John loves donkeys except for his grandfather's mean old donkey, Platero, which John hates. Last week, John's grandfather died and John was supposed to inherit Platero. But he decided not to accept the inheritance.

(13) If John had inherited a donkey, he would have hated it.

(13) is true in (12) under a low reading (because John would have hated Platero), but false under a high reading (because John would not have hated other donkeys.) That is, there is a *res* that verifies the counterfactual – for the low reading, see (14) – and a *res* that doesn't – for the high reading, see (15).

(14) *res*₁: John loves donkeys, except for Platero. John would have inherited Platero.

(15) *res*₂: John loves donkeys, except for Platero.

(14) and (15) differ from one another in that (14) contains additional information about the most likely referent of the indefinite in the counterfactual. In lawfully extended counterparts of *res*₁, John inherits Platero and hates it. In lawfully extended counterparts of *res*₂, John inherits different donkeys and loves most of them.

[H] Applying knowledge diagnostics. We can apply Arregui's knowledge diagnostics to check whether the proposed *res* in (14)-(15) are actually responsible for the observed readings.

(16) **Scenario** John's grandfather has a lot of donkeys, and John grew up loving donkeys, except for Platero. Everyone knows this, including Bill. Bill also recently learned that John would inherit Platero after his grandfather's death. What Bill doesn't know is that there is a second donkey John hates, Grisella, and that the grandfather changed his will, making John heir to Grisella instead of Platero. When the grandfather dies on Tuesday, John declines the inheritance and the will is never opened.

(17) If John had inherited a donkey on Tuesday, he would have hated it.

(18) Bill knows that if John had inherited a donkey on Tuesday, he would have hated it.

Here, we judge (17) as true – under a low reading –, but still reject (18): Bill doesn't have knowledge of the counterfactual, because he is not acquainted with the correct *res* (containing Grisella).

[I] Conclusions. A global similarity approach to counterfactuals is insufficient, because it only produces low readings when combined with any standard semantics for donkey sentences, either dynamic or D-type. We need to implement a more fine-grained notion of similarity that takes into account assumptions about possible referents for indefinite expressions; either via partialization by assignments as in dynamic semantics, or – as in our proposal – by encoding the relevant information in the *res* of a local similarity approach. Our proposal additionally explains the Gettier intuitions about embedded counterfactuals. As the mechanism that generates the high and low reading (different *res* contents) is in principle independent of the particular mechanisms for donkey anaphora, it also opens up the possibility of giving an account that does not commit us to a particular (D-type or dynamic) theory of donkey sentences in order to account for the behaviour of the counterfactual.

Selected References. Arregui, A. 2009. On similarity in counterfactuals. *L&P* 32 ■ Elbourne, P. 2005. *Situations and individuals*. MIT Press ■ Elbourne, P. 2013. *Definite descriptions*. OUP ■ Lewis, D. *Counterfactuals*. Blackwell ■ van Rooij, R. 2006. Free choice counterfactual donkeys. In *JoS* 23 ■ Walker, A. & M. Romero. 2015. Counterfactual donkey sentences. A strict conditional analysis. *SALT* 25