

Counterfactual donkey sentences: A strict conditional analysis *

Andreas Walker
University of Konstanz

Maribel Romero
University of Konstanz

Abstract We explore a distinction between ‘high’ and ‘low’ readings in counterfactual donkey sentences and observe three open issues in the current literature on these sentences: (i) van Rooij (2006) and Wang (2009) make different empirical predictions with respect to the availability of ‘high’ donkey readings. We settle this question in favour of van Rooij’s (2006) analysis. (ii) This analysis overgenerates with respect to weak readings in so-called ‘identificational’ donkey sentences. We argue that pronouns in these sentences should not be analysed as donkey pronouns, but as concealed questions or as part of a cleft. (iii) The analysis also undergenerates with respect to NPI licensing in counterfactual antecedents. We propose a strict conditional semantics for counterfactual donkey sentences that derives the correct licensing facts.

Keywords: donkey sentences, counterfactuals, dynamic semantics, modal horizon, negative polarity items, strict conditional analysis, concealed question, cleft-construction

1 Introduction

Counterfactual donkey sentences are sentences like (3) that have both the ingredients of a classical conditional donkey sentence, i.e., interclausal binding as in (1) (Geach 1962), and the morphology of plain counterfactual sentences, underlined in (2).

- (1) If a farmer owns a donkey, he beats it.
- (2) If Pedro owned Platero, he would beat it/(him).
- (3) If a farmer owned a donkey, he would beat it.

Such sentences raise the question of how to combine a proper semantics for the counterfactual with an adequate semantics for donkey quantification. We first

* We would like to thank the audiences of SALT 25 and the graduate seminar in semantics at Konstanz for their feedback. Special thanks to Irene Heim, Sven Lauer and Brian Leahy for the detailed discussion of our ideas. This research has been supported by the Deutsche Forschungsgemeinschaft (DFG, Project 4247/1-1).

introduce these two ingredients separately. The issue is how the restriction to maximally similar worlds used in the interpretation of counterfactuals should be intertwined with the apparatus needed for donkey quantification.

1.1 Counterfactuals: Variably strict analysis

In the classical variably strict approach developed by [Stalnaker \(1968\)](#) and [Lewis \(1973\)](#), the truth conditions of a plain counterfactual conditional like (2) are relativized to an accessibility function f , mapping each world w in its domain to the set of worlds which make the antecedent clause ϕ true and are otherwise maximally similar to w according to a given ordering relation \leq , as defined in (4).¹ A counterfactual sentence $\phi > \psi$ then asserts that all ϕ -worlds that are maximally similar to the actual world are also ψ -worlds; see (5).

$$(4) \quad f_w(\llbracket \phi \rrbracket^{f, \leq}) = \{v \in \llbracket \phi \rrbracket^{f, \leq} \mid \neg \exists u \in \llbracket \phi \rrbracket^{f, \leq} : u <_w v\}$$

$$(5) \quad \llbracket \phi > \psi \rrbracket^{f, \leq}(w) = 1 \quad \text{iff} \quad \forall w' \in f_w(\llbracket \phi \rrbracket^{f, \leq}) : w' \in \llbracket \psi \rrbracket^{f, \leq}$$

1.2 Donkey sentences: Dynamic predicate logic

We limit our discussion of indicative donkey sentences to a standard dynamic semantics for quantification and pronouns: Dynamic Predicate Logic (DPL, [Groenendijk & Stokhof 1991](#))². The main problem of donkey sentences as perceived from this perspective is the mismatch between the compositionally derived formula in (6) and the reading that donkey sentences are usually taken to have in (7):

$$(6) \quad \exists x Px \rightarrow Qx$$

$$(7) \quad \forall x [Px \rightarrow Qx]$$

The solution advocated by [Groenendijk & Stokhof](#) is a logic that derives the equivalence of those two formulas. That is, in their system there is no difference in the interpretation of (6) and (7):

$$(8) \quad \exists x Px \rightarrow Qx \Leftrightarrow \forall x [Px \rightarrow Qx]$$

The way that DPL derives this equivalence is by moving from a static semantics where the semantic value of an expression is a set of assignments to a dynamic semantics where the semantic value of an expression is a set of pairs of assignments, one ‘input’ pair and one ‘output’ pair. That is, we explicitly record and pass on changes to the assignment functions in moving from static (9) to (10).

¹ For simplicity, we make the limit assumption ([Lewis 1973](#)).

² Dynamic approaches to donkey sentences have been challenged by modern D-type theories that make use of situation semantics ([Heim 1990](#), [Büring 2004](#), [Elbourne 2005](#)). We leave this debate between static and dynamic approaches aside for the purposes of this paper.

$$(9) \quad \llbracket \exists x Px \rrbracket = \{g \mid \exists h: h[x]g \wedge h(x) \in F(P)\}$$

$$(10) \quad \llbracket \exists x Px \rrbracket = \{\langle g, h \rangle \mid h[x]g \wedge h(x) \in F(P)\}$$

Given such a setup, we can now specify which formulas should make permanent or temporary changes in assignments. Groenendijk & Stokhof (1991) define existential quantification and the conditional as follows:

$$(11) \quad \llbracket \exists x \phi \rrbracket = \{\langle g, h \rangle \mid \exists k: k[x]g \wedge \langle k, h \rangle \in \llbracket \phi \rrbracket\}$$

$$(12) \quad \llbracket \phi \rightarrow \psi \rrbracket = \{\langle g, h \rangle \mid h = g \wedge \forall k: \langle h, k \rangle \in \llbracket \phi \rrbracket \rightarrow \exists j: \langle k, j \rangle \in \llbracket \psi \rrbracket\}$$

Putting these together, we can derive the following result for the left side of (8):

$$(13) \quad \begin{aligned} \llbracket \exists x Px \rightarrow Qx \rrbracket &= \{\langle g, h \rangle \mid h = g \wedge \forall k: \langle h, k \rangle \in \llbracket \exists x Px \rrbracket \rightarrow \exists j: \langle k, j \rangle \in \llbracket Qx \rrbracket\} \\ &= \{\langle g, h \rangle \mid h = g \wedge \forall k: (k[x]h \wedge k(x) \in F(P)) \rightarrow \exists j: \langle k, j \rangle \in \llbracket Qx \rrbracket\} \\ &= \{\langle g, h \rangle \mid h = g \wedge \forall k: (k[x]h \wedge k(x) \in F(P)) \rightarrow k(x) \in F(Q)\} \end{aligned}$$

A quick comparison with the result for deriving the right side using the lexical entry for universal quantification in (14) shows that this is indeed equivalent:

$$(14) \quad \llbracket \forall x \psi \rrbracket = \{\langle g, h \rangle \mid h = g \wedge \forall k: k[x]h \rightarrow \exists j: \langle k, j \rangle \in \llbracket \psi \rrbracket\}$$

$$(15) \quad \begin{aligned} \llbracket \forall x [Px \rightarrow Qx] \rrbracket &= \{\langle g, h \rangle \mid h = g \wedge \forall k: k[x]h \rightarrow \exists j: \langle k, j \rangle \in \llbracket Px \rightarrow Qx \rrbracket\} \\ &= \{\langle g, h \rangle \mid h = g \wedge \forall k: k[x]h \rightarrow (\forall u: \langle k, u \rangle \in \llbracket Px \rrbracket \rightarrow \exists m: \langle u, m \rangle \in \llbracket Qx \rrbracket)\} \\ &= \{\langle g, h \rangle \mid h = g \wedge \forall k: k[x]h \rightarrow (k(x) \in F(P) \rightarrow k(x) \in F(Q))\} \\ &= \{\langle g, h \rangle \mid h = g \wedge \forall k: (k[x]h \wedge k(x) \in F(P)) \rightarrow k(x) \in F(Q)\} \end{aligned}$$

Later developments notwithstanding, this proposal provides a solid foundation for the investigation of donkey sentences and other quantificational phenomena.

1.3 Goal and roadmap

The existing analyses of counterfactual donkey sentences in the literature (van Rooij 2006, Wang 2009) have combined a variably strict semantics for counterfactual conditionals with a standard dynamic semantics like DPL. In this paper, we reveal three shortcomings of the current state-of-the-art:

- i) We vindicate van Rooij's (2006) account with respect to Wang's (2009) criticism by showing that Wang (2009) cannot generate all the attested readings of the indefinite NP in counterfactual sentences,
- ii) We show that van Rooij (2006) overgenerates weak readings in identificational sentences, and

- iii)** We observe that [van Rooij \(2006\)](#) undergenerates with respect to the licensing of Negative Polarity Items (NPIs).

To circumvent these problems, we will develop a strict conditional analysis (for problem **(iii)**) that, following [van Rooij \(2006\)](#), builds assignment-sensitivity into the order \leq (for problem **(i)**) and that treats the relevant pronoun in identificational sentences not as a donkey pronoun (for problem **(ii)**).

The paper is organized as follows. Section 2 presents [van Rooij's \(2006\)](#) and [Wang's \(2009\)](#) analyses of counterfactual donkey sentences. Section 3 lays out our criticism: we need the additional readings of the indefinite NP powered by van Rooij's system and excluded by Wang's (3.1), van Rooij's system overgenerates weak readings in identificational sentences (3.2), and van Rooij's apparatus underlicences NPIs (3.3). Our proposal is developed in Section 4. Section 5 concludes.

2 Previous approaches

2.1 Van Rooij (2006)

Starting from the equivalence in (8), [van Rooij \(2006\)](#) argues that the aim of an analysis for counterfactual donkey sentences should be to derive the same equivalence for the counterfactual conditional, i.e., to obtain (16). The intuition is that, assuming the domain of quantification $\{d_1, d_2, d_3\}$, there is a reading of sentence (17) that entails the conjunction of (18a-c). We will label indefinite NPs for which this equivalence is valid as 'high indefinites' and we will call the corresponding reading of the sentence 'high reading'.

$$(16) \quad \exists x Px > Qx \Leftrightarrow \forall x [Px > Qx]$$

(17) If John owned a donkey $_{\{d_1, d_2, d_3\}}$, he would beat it.

(18) a. If John owned d_1 , he would beat d_1 .

b. If John owned d_2 , he would beat d_2 .

c. If John owned d_3 , he would beat d_3 .

As a first step, van Rooij adapts the Stalnaker-Lewis variably strict semantics of counterfactuals from Section 1.1 to a DPL framework. Since we are now dealing not only with worlds, but with pairs $\langle v, h \rangle$ of a world and an assignment function, the ordering relation \leq that the accessibility function f is based on needs to be modified accordingly. In order to derive the equivalence in (16), van Rooij chooses a particular way of doing this: In his modified version of the ordering relation \leq^* , only world-assignment pairs that share an assignment are ranked with respect to each other. This has the effect of making the ordering over world-assignment pairs a partial one.

$$(19) \quad \text{Modified order } \leq^* \text{ (to be revised):}$$

$$\langle v, h \rangle \leq_{\langle w, g \rangle}^* \langle u, k \rangle \quad \text{iff}_{\text{def}} \quad h = k \supseteq g \wedge v \leq_w u$$

It is easy to see that this reduces to the original Lewisian accessibility function in contexts without donkey quantification: If the original assignment is not modified, all pairs can be ranked with respect to each other on the basis of their worlds' similarity. The more interesting effect obtains in contexts with existential quantification: Introducing a new variable into the assignment function means splitting the ordering into smaller sub-orderings, each sub-ordering anchored to a different variable assignment. That is, for each individual that can be assigned as a value to the new variable in the assignment, we end up with one sub-ordering of world-assignment pairs.

Based on \leq^* , a new accessibility function f^* is defined in (20) mapping each world-assignment pair $\langle w, g \rangle$ to the set of world-assignment pairs that satisfy ϕ and that are otherwise maximally similar to $\langle w, g \rangle$. The truth conditions are then set up in a way parallel to the classical Stalnaker-Lewis analysis, except that now we quantify over pairs $\langle v, h \rangle$ rather than over worlds w' . Crucially, since the counterfactual quantifies over the bottom elements $\langle v, h \rangle$ of all the smaller sub-orderings, it will be quantifying over at least one world-assignment pair per individual. But then existential quantification in the antecedent of a counterfactual is the same as universal quantification over the entire counterfactual. This means that the desired equivalence in (16) obtains.

$$(20) \quad f_{\langle w, g \rangle}^*(/\phi/g) =_{\text{def}} \{ \langle v, h \rangle \in /\phi/g \mid \neg \exists \langle u, k \rangle \in /\phi/g : \langle u, k \rangle <_{\langle w, g \rangle}^* \langle v, h \rangle \}$$

$$(21) \quad \llbracket \phi > \psi \rrbracket^{f^*, \leq^*}(\langle w, g \rangle) = 1 \quad \text{iff} \quad \forall \langle v, h \rangle \in f_{\langle w, g \rangle}^*(/\phi/g) : \langle v, h \rangle \in /\psi/g$$

To see this with a concrete example, consider the simple example (22) and the model in Table 1:

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

The selection function in (20) yields the following world-assignment pairs:

$$(23) \quad f_{\langle w, g \rangle}^*(/\text{John owned a donkey}/g) = \{ \langle w_1, g^{x/d_1} \rangle, \langle w_2, g^{x/d_2} \rangle, \langle w_3, g^{x/d_3} \rangle \}$$

That is, although w_1 is the closest world to the actual world, w_2 and w_3 also appear in the selection because their pairs do not share an assignment with $\langle w_1, g^{x/d_1} \rangle$. Note that, while other worlds also make the antecedent true for a given assignment (i.e., w_4 for g^{x/d_1} and g^{x/d_2}), those do not appear because other antecedent worlds are closer and paired with the same assignment.

Since each of the pairs in (23) also makes the consequent true, the counterfactual comes out as true. Worlds that are further away do not matter for the result. But,

	donkey	own	beat
w_0	$\{d_1, d_2, d_3\}$	\emptyset	\emptyset
w_1	$\{d_1, d_2, d_3\}$	$\{\langle j, d_1 \rangle\}$	$\{\langle j, d_1 \rangle\}$
w_2	$\{d_1, d_2, d_3\}$	$\{\langle j, d_2 \rangle\}$	$\{\langle j, d_2 \rangle\}$
w_3	$\{d_1, d_2, d_3\}$	$\{\langle j, d_3 \rangle\}$	$\{\langle j, d_3 \rangle\}$
w_4	$\{d_1, d_2, d_3\}$	$\{\langle j, d_1 \rangle, \langle j, d_2 \rangle\}$	\emptyset

Table 1 A sample model for (22), with worlds ranked as follows: $w_0 < w_1 < w_2 = w_3 < w_4$

for each individual d , we quantify at least over one $\langle v, g^{x/d} \rangle$ pair. This derives the equivalence in (16).

After this first step in adapting the Stalnaker-Lewis semantics to the DPL framework, van Rooij (2006) notes that there is a second reading of indefinite NPs in counterfactual donkey sentences for which the equivalence in (16) does not hold. In identificational counterfactual donkey sentences like (24), it is not asserted that each individual that is an animal and escaped last night in some world is Alex (the Lion) in that world. That is, under the intended reading, (24) does not entail the conjunction of (25a-c). Rather, the sentence expresses that in the closest world in which some animal or other escaped, that animal is Alex.

- (24) If an _{$\{a(\text{lex}), b, c\}$} animal had escaped last night, it would have been Alex.
 (25) a. If a had escaped last night, a would have been Alex.
 b. If b had escaped last night, b would have been Alex.
 c. If c had escaped last night, c would have been Alex.

While van Rooij (2006) treats this as a property of identificational sentences, we note that this reading obtains in other contexts as well. For example, scenario (26a) makes salient a reading of sentence (26b) for which the equivalence (16) does not hold and according to which, in the most likely worlds where John owns some donkey, John happens to own one of Melissa's descendants and thus John beats that donkey. Note that, while the sentence is true in scenario (26a) under this reading, the high reading discussed above would make it false. We will call this second reading 'low reading' and we will label indefinite NPs giving rise to it as 'low indefinites'.

- (26) a. Scenario: Given how poor John's family is, the only realistic chance he ever had to own a donkey was for his grandpa's donkey Melissa to have descendants. Alas, Melissa never has descendants! But, if she had had them, they would have been as stubborn as Melissa herself, so that their

owner would have had to beat them. Excepting stubborn donkeys, John has no inclination to beat donkeys.

- b. If John owned a donkey, he would beat it (because he would own a stubborn descendant of Melissa and he could not help but beat it).

In addition to the low readings of indefinite NPs just discussed for counterfactual donkey sentences, we also encounter a reading familiar from the discussion of indicative donkey sentences: The weak reading of the donkey-pronouns in (27) (Schubert & Pelletier 1987) generalizes to the counterfactual case in (28).

(27) a. If I have a dime, I throw it into the meter.

b. Weak reading:

‘If I have one or more dimes, I throw one of the dimes into the meter.’

(28) If I had a dime, I would throw it into the meter.

In a second step adapting the Stalnaker-Lewis semantics to DPL, van Rooij (2006) tackles both the low and the weak reading at the same time. In order to generate the readings in (24) and (28) (and our (26b)), van Rooij adopts a mechanism originally developed to deal with asymmetric readings in indicative donkey sentences (Root 1986): Instead of relying on unselective quantification, the entire machinery in (19-21) is relativized to a contextually given variable X that contains the variables to be taken into account for the computation.

(29) $h \uparrow^X = k \uparrow^X$ iff $\forall x \in X: h(x) = k(x)$

(30) $\langle v, h \rangle \leq_{\langle w, g \rangle}^{*,X} \langle u, k \rangle$ iff $h, k \supseteq g$, $h \uparrow^X = k \uparrow^X$ and $v \leq_w u$

(31) $f_{\langle w, g \rangle}^{*,X}(/ \phi / g) = \{ \langle v, h \rangle \in / \phi / g : \neg \exists \langle u, k \rangle \in / \phi / g : \langle u, k \rangle <_{\langle w, g \rangle}^{*,X} \langle v, h \rangle \}$

(32) $\langle v, h \rangle \sim^X \langle u, k \rangle$ iff $v = u$ and $h \uparrow^X = k \uparrow^X$

(33) $\llbracket \phi >^X \psi \rrbracket \leq (\langle w, g \rangle) = 1$ iff $\forall \langle v, h \rangle \in f_{\langle w, g \rangle}^{*,X}(/ \phi / g) :$

$\exists \langle u, k \rangle \in f_{\langle w, g \rangle}^{*,X}(/ \phi / g) : \langle u, k \rangle \sim^X \langle v, h \rangle \wedge \langle u, k \rangle \in / \psi / g$

(29-30) tell us that a pair $\langle v, h \rangle$ is closer than $\langle u, k \rangle$ to the actual $\langle w, g \rangle$ if and only if h and k are supersets of g , h and k coincide in the values assigned to the unselectively bound variables in X , and v is closer to w than u . Out of the set of pairs that make the antecedent true, $f_{\langle w, g \rangle}^{*,X}$ selects the ones that are closest to $\langle w, g \rangle$, as defined in (31). The sentence asserts that each such closest pair $\langle v, h \rangle$ is such that there is a variant of it $\langle u, k \rangle$ — with k matching h in the values of all the variables in X and where $v = u$, as defined in (32) — making the consequent true.

The apparatus in (29-33) reduces to the previously defined version in (20-21) if X contains all the variables that have been introduced by ϕ . In that case, we obtain the

	donkey	own	beat
w_0	$\{d_1, d_2, d_3\}$	\emptyset	\emptyset
w_1	$\{d_1, d_2, d_3\}$	$\{\langle j, d_1 \rangle\}$	$\{\langle j, d_1 \rangle\}$
w_2	$\{d_1, d_2, d_3\}$	$\{\langle j, d_2 \rangle\}$	\emptyset
w_3	$\{d_1, d_2, d_3\}$	$\{\langle j, d_3 \rangle\}$	\emptyset

Table 2 A sample model for (34), with worlds ranked as follows: $w_0 < w_1 < w_2 = w_3$

usual high donkey reading. If X lacks one of the variables, we obtain a low reading for the corresponding indefinite and a weak reading for its anaphoric pronoun³.

We demonstrate this, first for the low reading from (26b) and then for the weak reading from (28).

(34) If John owned a donkey ^{x} , he would beat it _{x} .

In the model in Table 2, the high reading with $X = \{x\}$ comes out as false, because John does not beat a donkey in either w_2 or w_3 . For the low reading with $X = \emptyset$, however, we get a different result: Since, according to (30), two world-assignment pairs are now comparable if their assignments are identical with respect to the variables in X (which in this case reduces to a trivial requirement), we can simply rank all candidate world-assignment pairs on the basis of their worlds. Since w_1 is more similar to the actual world than w_2 and w_3 , the output of the selection function defined in (31) returns the following:

$$(35) \quad f_{\langle w, g \rangle}^{*, \emptyset} (/ \text{John owned a donkey} /_g) = \{\langle w_1, g^{x/d_1} \rangle\}$$

Since, in this world-assignment pair, the consequent also holds, the counterfactual comes out as true under the low reading.

For the weak reading of (36), assume the model in Table 3:

(36) If I had a dime, I would throw it into the meter.

The selection function returns the following:

$$(37) \quad f_{\langle w, g \rangle}^{*, \emptyset} (/ \text{I have a dime} /_g) = \{\langle w_1, g^{x/d_1} \rangle, \langle w_1, g^{x/d_2} \rangle, \langle w_1, g^{x/d_3} \rangle\}$$

³ Note that in van Rooij's system, low and weak readings always go together, a feature we will inherit from him. However, it is an open question whether this is empirically the case: It might well be the case that low+strong readings exist. This issue does not arise for the high reading, as here a weak and a strong reading are indistinguishable anyway.

	dime	have	throw-into-the-meter
w_0	$\{d_1, d_2, d_3\}$	\emptyset	\emptyset
w_1	$\{d_1, d_2, d_3\}$	$\{\langle i, d_1 \rangle, \langle i, d_2 \rangle, \langle i, d_3 \rangle\}$	$\{\langle i, d_1 \rangle\}$

Table 3 A sample model for (36), with worlds ranked as follows: $w_0 < w_1$

Not all of these world-assignment pairs make the consequent true. In fact, only $\langle w_1, g^{x/d_1} \rangle$ does. However, (33) only requires that for each world-assignment pair there exists an X -equivalent world-assignment pair, as defined in (32), satisfying the consequent. This is the case here: All three world-assignment pairs in (37) are X -equivalent to $\langle w_1, g^{x/d_1} \rangle$, which in turn satisfies the consequent, making the counterfactual true.

2.2 Wang (2009)

Wang (2009) criticizes the second step by van Rooij (2006), as she sees no principled way of specifying which donkey sentences are high/strong and which ones are low/identifying/weak. While Wang does not explicitly argue for the unavailability of high/strong readings on an empirical basis, she argues that we should give up on the equivalence in (16). In the interest of a more unified analysis, Wang (2009) proposes to treat low donkeys as the standard case instead, i.e., to make the similarity function permanently insensitive to assignments. This essentially amounts to permanently setting $X = \emptyset$ in van Rooij's system.

But, if low readings are the only genuine readings generated by the system, how do we derive van Rooij's (and our) intuition that, at least for some counterfactual donkey sentences, the equivalence in (16) seems to hold? If high indefinite interpretations are to be derived in Wang's system⁴, they would arise as a sub-case of the general low reading just in case the worlds and assignments are set up in the model in a particular way: For any possible values d and d' of the variable, the most likely world that makes the antecedent true for d and the most likely world that makes the antecedent true for d' are equally ranked. This amounts to a rather strong assumption about the model: All individuals are equally likely to fulfill the antecedent, with no difference in similarity for the worlds in which they do so. In the case of our classical donkey sentence in (3), this would require us to assume that all combinations of farmer-donkey ownings are equally likely to obtain. We will

⁴ Wang (2009) does not, in fact, return to the question of high readings. We reconstruct them here in order to show how the system compares to van Rooij's (2006) account. In the interest of comparison, we also neglect the fact that Wang additionally moves the analysis to an update semantics (Veltman 2005), as this seems to have no effect on the pertinent questions here.

come back to this shortly in Section 3.1.

3 Criticism

We advance three points of criticism on the state-of-the-art we just described. The first one concerns the status of the high indefinite interpretation: Is it a genuine reading, as in van Rooij's (2006) approach, or is it a sub-case of the low reading arising only in certain contexts, as expected in Wang's (2009) analysis? We will show that Wang's system makes wrong predictions and thus conclude that the additional resources in van Rooij's system are needed. Then we turn to van Rooij's proposal. As our second criticism, we will show that van Rooij's system, as it stands, overgenerates weak readings for identificational donkey sentences. Our third criticism is that van Rooij's approach underlicenses NPIs in low readings.

3.1 Differing predictions: Wang vs. van Rooij

As just discussed in Section 2.2, Wang (2009) makes the prediction that high donkey interpretations should only be available in very particular circumstances. Obviously, we are not always willing to assume such circumstances. Specifically, consider the following scenario, where not all the individuals quantified over – the farmers in King Kakos's kingdom – are equally likely to own a donkey.⁵:

- (38) Scenario: There are two farmers in the kingdom of King Kakos, called Onophilos and Onophobos. Both are very poor and do not own a donkey. Onophobos is a cruel man who would love to own and beat a donkey. He has been saving money all his life and has nearly enough to buy a donkey. Onophilos is a mild-mannered vegan who has no means or interest in owning a donkey, much less so in beating it. King Kakos only knows Onophobos and is convinced that all inhabitants of his kingdom are just as cruel and evil as Onophobos. He discusses this with his advisor, who is well-informed about all the farmers and their dispositions.
- (39) 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: You are wrong. 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.

In the scenario in (38), we judge (39b) to be true. However, this is not what Wang (2009) predicts. Clearly, a world in which Onophobos owns a donkey is more

⁵ We thank an anonymous reviewer from *Sinn und Bedeutung* 20 for pointing out the correct names for the farmers in our example, and Vasiliki Erotokritou for additional help with Ancient Greek.

similar to the actual world than one in which Onophilos owns a donkey. But then, according to Wang, we should only consider the world-assignment pair in which Onophobos is the donkey-owning farmer in order to evaluate the counterfactual. Since in that world-assignment pair the donkey is beaten, King Kakos's statement is predicted to be true and, thus, his advisor's statement is predicted to be false, contrary to fact.

We conclude that Wang's proposal lacks a way of producing the truth conditions we observe in (38-39).

3.2 Overgeneration: Identificational donkey sentences

van Rooij (2006) aims to account for identificational sentences like (24), repeated below as (40), as a subclass of 'standard' counterfactual donkeys, with *an animal* as a low indefinite and *it* as a donkey pronoun:

(40) If an_{a(lex),b,c} animal had escaped last night, it would have been Alex.

We note that applying van Rooij's analysis as it stands to identificational sentences overgenerates weak readings. To derive the intended reading of (40), van Rooij needs two ingredients: (i) the interpretation template (33), which yields the truth conditions spelled out in (41a) and paraphrased in (41b), and (ii) a uniqueness assumption in the context that guarantees that, in each of the worlds under consideration, exactly one animal escaped (van Rooij 2006:395). The latter assumption has the effect of making the truth-conditional output "one of the escaped animals" contextually equivalent to the perceived interpretation "the (unique) escaped animal".

- (41) a. $\llbracket \text{If an animal had escaped, it would have been Alex} \rrbracket^{\leq}(\langle w, g \rangle) = 1$ iff
 $\forall \langle v, h \rangle \in f_{\langle w, g \rangle}^{*,X} (/ \text{an animal has escaped} /_g) : \exists \langle u, k \rangle \in f_{\langle w, g \rangle}^{*,X} (/ \text{an animal has escaped} /_g) : \langle u, k \rangle \sim^X \langle v, h \rangle \wedge \langle u, k \rangle \in / \text{it was Alex} /_g$
- b. "If an animal out of {a, b, c} had escaped last night, one of the escaped animals would have been Alex."

However, while one might argue that this uniqueness assumption is reasonably invoked in many contexts, it will not always be so. Consider the scenario in (42). Here, we create a context that explicitly removes the uniqueness assumption by making the closest worlds where animal-escaping occurs be worlds where both Alex the Lion and Tara the Lioness escape. Van Rooij's system predicts that in those cases we should obtain the plain weak reading (41) as 'one of the lions would have been Alex', which would render sentence (42b) (felicitous and) true in that scenario. However, this reading is unavailable and the sentence is not judged true.

- (42) a. Scenario: The gate of the lions was left open. Alex the Lion and Tara the Lioness are the most curious among the lions, always tampering with the gates. They are inseparable and always go together.
b. # If an_{alex,b,c,tara} animal had escaped, it would have been Alex the Lion.

While we agree with van Rooij that the NP *an animal* in (40) should be treated as a low indefinite, we will argue in Section 4.1 that a different treatment of the pronoun is needed in identificational sentences.

3.3 Undergeneration: NPI licensing

One of van Rooij's (2006) main goals is to derive the licensing of NPIs in the antecedent of counterfactual donkey sentences, underlined in (43):

- (43) If John owned any donkey, he would beat it.

Van Rooij aims to do so by "dynamicizing" the traditional Stalnaker-Lewis variably strict analysis of counterfactuals, with the result presented in Section 2.1. He derives the NPI licensing on the basis of Kadmon & Landman's (1993) widening analysis. The assumption here is that *any* widens the domain of quantification, and that such a widening is licensed whenever the resulting proposition is logically stronger than that expressed by the same utterance without the domain widening. Van Rooij points out that under his analysis, adding more individuals to the domain has the counterfactual quantifying over more world-assignment pairs (but still including the world-assignment pairs the selection function yields from the original domain). Thus, NPIs are predicted to be licensed by a combination of his analysis with Kadmon & Landman 1993.

However, we note that this only holds as long as we assume the high reading of the counterfactual donkey sentence. Under the high reading, we saw that (17) with a wide domain amounts to the conjunction of (18a), (18b) and (18c). Correspondingly, (44) with a narrower domain amounts to the conjunction of just (18a) and (18b). Since the first conjunction entails the second, NPIs are correctly predicted to be licensed.

- (44) If John owned a_{d₁,d₂} donkey, he would beat it.

But, under a low reading of the indefinite, the situation changes: Depending on the similarity ranking of the world in which members of the widened domain satisfy the antecedent, we might end up with an entirely new set of world-assignment pairs, or possibly with exactly the same set as before. For example, if John is likely to inherit Melissa, whereas acquiring other donkeys in the non-widened domain *D* is less likely, adding more donkeys to the domain of quantification can result in (i)

adding donkeys that are more likely to be acquired than Melissa, (ii) adding donkeys that are less likely to be acquired than Melissa and (iii) adding donkeys that are exactly as likely to be acquired as Melissa. Under a weak reading, (i) will yield quantification over a set of world-assignment pairs that does not include the original world-assignment pair in which John owns Melissa; (ii) will yield quantification over exactly the same world-assignment pair in which John owns Melissa but no other world-assignment pairs; and only in case of (iii) the counterfactual will yield a logically stronger statement. Thus, [van Rooij \(2006\)](#) would predict that we should not be able to license NPIs in low donkeys, except for the specific circumstances of (iii). However, this is not what we observe: NPIs are in fact licensed, as (45) and (46) show.

(45) If I had any dime, I would throw it into the meter.

(46) If any_{a(lex),b,c} animal had escaped last night, it would have been Alex.

4 Proposal

In view of the limitations of Wang's account noted above, we adopt van Rooij's analysis in (29)-(33) and take it as our point of departure. Two refinements will be built on this system. First, it will be argued that the pronoun in identificational sentences is not a donkey pronoun but a concealed question or part of a cleft (subsection 4.1). This minor modification, which we will motivate on morphological data, will derive a mandatory uniqueness effect in identificational sentences. Second, in order to have a general account of NPI licensing in counterfactual antecedents, we will abandon the Stalnaker-Lewis variably strict conditional analysis and assume the strict conditional approach to counterfactuals with a modal horizon defended in [von Stechow \(2001\)](#) (subsection 4.2). Van Rooij's apparatus for donkey quantification will then be recast in the strict conditional analysis (section 4.3). This more substantial change will allow us to license NPIs both for high and for low donkey indefinites.

4.1 The pronoun in identificational sentences

We saw that analyzing (40), repeated below as (47), à la [van Rooij 2006](#) – i.e., treating the indefinite *an animal* as a low indefinite and the pronoun *it* as a donkey pronoun – incorrectly generates a weak reading. While we agree with van Rooij that the indefinite in (47) receives a low reading, we argue that the pronoun should not be analyzed as a donkey pronoun. Donkey pronouns mandatorily agree in gender with their antecedent, as illustrated in (48). In contrast, in the identificational sentences at issue, the pronoun does not agree with the intended antecedent, a fact that becomes apparent as soon as masculine or feminine antecedents are used; witness (49).

- (47) If an_{a(lex),b,c} animal had escaped last night, it would have been Alex.
 (48) If a man owned Platero, he / # it would beat the poor animal.
 (49) a. If Alex were married to a girl from his class, it would be Sue.
 b. # If Alex were married to a girl from his class, she would be Sue.

We propose that the "identificational" pronoun is either the invariable *it*-form of a cleft-construction⁶ or a pronoun referring to a concealed question (of type $\langle s, e \rangle$). This explains its invariably neutral form: In cleft-constructions *it* is a fossilized form, and concealed question NPs are mandatorily pronominalized in neuter form, as shown in (50) (Romero 2004; see also Mikkelsen 2005).

- (50) John guessed the winner of the Oscar for best actress before I guessed it / #her.

More importantly, the proposed treatment of the pronoun rules out weak readings and secures the desired uniqueness effect. Cleft-constructions have been argued to presuppose exhaustivity (Büring 2011, Velleman, Beaver, Destruel, Bumford, Onea & Coppock 2013). This means that the second sentence in (51) cannot be understood as "Sam was one of those who came" and, as a result, the sequence is infelicitous. Similarly, the interrogative in (52) cannot be understood as "Do you know one of the two prices?", which, again, makes the sequence odd.

- (51) Two students came. # It was Sam who came.
 (52) John knows neither the price of milk nor the price of oil. # Do you know it?

Thus, we propose that the static version of the truth conditions for the consequent clause of (47) is as follows: Either the cleft semantics in (53), where ALT is the set of alternative propositions generated by focus (Velleman et al. 2013), or the concealed question semantics in (54) (Romero 2005).⁷

$$(53) \quad \lambda w: \forall q \in \text{ALT}[q \subset [\lambda w'.\text{escape}(a, w')]] \rightarrow \neg q(w)]. \\ \exists q \in \text{ALT}[q(w) \wedge q \subseteq [\lambda w'.\text{escape}(a, w')]]$$

$$(54) \quad \lambda w. [\lambda w'. \text{Ix}[\text{animal}(x, w') \wedge \text{escape}(x, w')]](w) = a$$

⁶ See Büring (1998) for a similar claim concerning anaphora in modal subordination environments.

⁷ We leave open whether *it* corresponds to a cleft or to a concealed question in these examples. In principle, both strategies seem available in the grammar, since, with a plural antecedent, the singular pronoun *it* is possible – suggesting that it is part of a cleft – and the plural form *they* is acceptable as well – which suggests that the pronoun functions as a concealed question.

- (1) Nobody solved the problem. But if two students had solved the problem, it / they would have been Pat and Marisa.

4.2 A strict conditional semantics for counterfactuals and NPI licensing

In plain (non-donkey) counterfactual sentences, the antecedent clause seems, *prima facie*, not to be a downward entailing context: (55a) does not entail (55b) with a strengthened antecedent. Nevertheless, NPIs are licensed here; witness (56).

- (55) a. If I strike this match, it will light.
 b. \nRightarrow If this match gets wet and I strike it, it will light.
- (56) If I strike any match, Mary will be upset.

To solve this puzzle, von Fintel (1999) make two moves. The first one is to refine the notion of downward entailment needed: NPIs are licensed in Strawson-downward entailing contexts, as defined in (57). This modification is independently motivated for NPIs under *only* and other constructions. The second change concerns conditionals: Instead of the traditional variably strict semantic approach, von Fintel advocates for a strict conditional analysis that will make counterfactual antecedents Strawson-downward entailing.

- (57) A function f of type $\langle \sigma, \tau \rangle$ is Strawson-downward-entailing iff for all x and y of type σ such that $x \Rightarrow y$ and $f(x)$ is defined: $f(y) \Rightarrow f(x)$.

Let us see the second change in a bit more detail. In the traditional variably strict semantic approach, the accessibility function f and the antecedent ϕ bundle together to form the restriction of the quantification over possible worlds. This makes conditional antecedents not (Strawson-)downward entailing (see e.g., von Fintel 1999: 136-138). In the strict conditional analysis envisioned in von Fintel 1999 and further developed in von Fintel 2001 and Gillies 2007, the device securing maximal similarity and the antecedent ϕ are separated. More concretely, conditionals are evaluated relative to a domain of quantification D – the so-called ‘Modal Horizon’ – that begins as a set of worlds very similar to the actual world and evolves through discourse. The definedness and truth conditions of a counterfactual sentence $\phi > \psi$ are given in (58). Whenever ϕ and the original domain D do not overlap – as it typically happens with counterfactuals – D is shifted to an expanded domain D' along the lines of (59), so that the presupposition in (58a) is satisfied. Then, restricted to the new domain D' , the sentence asserts that all ϕ -worlds are also ψ -worlds.

- (58) a. Presupposition:
 $\llbracket \phi > \psi \rrbracket^D$ is defined only if $\llbracket \phi \rrbracket \cap D \neq \emptyset$
- b. Assertion:
 $\llbracket \phi > \psi \rrbracket^D(w) = 1$ iff $\forall w' \in D: w' \in \llbracket \phi \rrbracket \rightarrow w' \in \llbracket \psi \rrbracket$
- (59) Modal Horizon expansion:
 $D' = D \cup \{w \mid \forall w' \in \llbracket \phi \rrbracket: w \leq w'\}$

The semantics in (58) makes counterfactual antecedents Strawson-downward entailing. For any sentences ϕ , ψ and χ , it holds that, for any modal domain D for which both $\llbracket \phi > \psi \rrbracket^D$ and $\llbracket (\phi \wedge \chi) > \psi \rrbracket^D$ are defined, the former will entail the latter; see (60a-b). Hence, NPIs are correctly predicted to be licensed in the antecedents of conditionals. As for the inference in (55), the semantics above renders it valid. Its apparent invalidity stems from the fact that, in evaluating the first conditional, the reader picks a minimally extended domain D' (with the match struck and still dry) and, in judging the second, she uses a further extended D'' (with the match struck and wet).

- (60) a. $\forall w' \in D: w' \in \llbracket \phi \rrbracket \rightarrow w' \in \llbracket \psi \rrbracket$
 b. $\forall w' \in D: w' \in \llbracket \phi \wedge \chi \rrbracket \rightarrow w' \in \llbracket \psi \rrbracket$

4.3 Dynamic quantification in a strict conditional analysis of counterfactuals

The strict conditional semantics for counterfactuals described in 4.2 allows us to account for the NPI licensing that was problematic in the variably strict account of van Rooij (2006). Now we are faced with a similar task to that of van Rooij: to combine these semantics with a dynamic semantics for the quantificational phenomena in donkey sentences. In order to adapt the strict conditional analysis of counterfactuals to a dynamic semantics for quantification, we need to reconsider the shape of the modal horizon, similar to the reconstruction of an appropriate similarity function in (30). Since von Fintel (2001) is not dealing with assignment functions, his modal horizon is simply a set of worlds. In order to be compatible with DPL, the modal horizon needs to take the shape of a set of world-assignment pairs. For the definition of auxiliary notions in (61-63), we follow van Rooij (2006). Our innovation is (64-65):

- (61) $h \uparrow^X = k \uparrow^X$ iff $\forall x \in X: h(x) = k(x)$
 (62) $\langle v, h \rangle \leq_{\langle w, g \rangle}^{*,X} \langle u, k \rangle$ iff $h, k \supseteq g, h \uparrow^X = k \uparrow^X$, and $v \leq_w u$
 (63) $\langle v, h \rangle \sim^X \langle u, k \rangle$ iff $v = u$ and $h \uparrow^X = k \uparrow^X$
 (64) Modal Horizon expansion:
 $f^* | \phi >^X \psi |_{g, \leq^*}^{f^*, \leq^*} = \lambda \langle w, g \rangle. f^*(\langle w, g \rangle) \cup$
 $\{ \langle v, h \rangle \in \llbracket \phi \rrbracket_{g, \leq^*}^{f^*, \leq^*} : \neg \exists \langle u, k \rangle \in \llbracket \phi \rrbracket_{g, \leq^*}^{f^*, \leq^*} : \langle u, k \rangle <_{\langle w, g \rangle}^{*,X} \langle v, h \rangle \}$
 (65) Truth conditions:
 $\llbracket \phi >^X \psi \rrbracket_{g, \leq^*}^{f^*, \leq^*} = \{ \langle w, g \rangle : \forall \langle v, h \rangle \in f^* | \phi >^X \psi |_{g, \leq^*}^{f^*, \leq^*} (\langle w, g \rangle) :$
 if $\langle v, h \rangle \in \llbracket \phi \rrbracket_{g, \leq^*}^{f^*, \leq^*}$ then $\exists \langle u, k \rangle \in \llbracket \phi \rrbracket_{g, \leq^*}^{f^*, \leq^*} : \langle u, k \rangle \sim^X \langle v, h \rangle \wedge \langle u, k \rangle \in \llbracket \psi \rrbracket_{g, \leq^*}^{f^*, \leq^*} \}$

	farmer	donkey	own	beat
w_0	$\{a, b\}$	$\{d\}$	\emptyset	\emptyset
w_1	$\{a, b\}$	$\{d\}$	$\{\langle a, d \rangle\}$	$\{\langle a, d \rangle\}$
w_2	$\{a, b\}$	$\{d\}$	$\{\langle b, d \rangle\}$	\emptyset

Table 4 A sample model for (66), with $w_0 < w_1 < w_2$

The update of the modal horizon proceeds as usual. As spelled out in (64), the original modal horizon for the world-assignment pair $\langle w, g \rangle$ is extended to be the union of the original modal horizon plus the set of all the world-assignment pairs $\langle v, h \rangle$ satisfying the antecedent for which there is no other world-assignment pair $\langle u, k \rangle$ satisfying the antecedent which is more similar to $\langle w, g \rangle$.

The truth conditions in (65) tells us that the counterfactual denotes the set of world-assignment pairs $\langle w, g \rangle$ that have the following property: All world-assignment pairs in the updated modal horizon of $\langle w, g \rangle$ are such that if they satisfy the antecedent, there is at least one X -comparable – in the sense of (63) – world-assignment pair in the modal horizon that also satisfies the consequent.

Since this semantics for the counterfactual is Strawson-downward entailing (von Fintel 2001), we obtain the NPI licensing for both high and low readings of counterfactual donkey sentences.

4.4 Generating the high reading

For simplicity, assume an initial modal horizon $f^*(\langle w, g \rangle) = \emptyset$ for any $\langle w, g \rangle$. In the actual world w_0 , no farmer owns a donkey. In w_1 , Onophobos owns a donkey and beats it. In w_2 , Onophilos owns a donkey but doesn't beat it. This is schematized in Table 4. Under these circumstances, we want to judge (66) as true under a high reading where $X = \{x, y\}$.

(66) It is not the case that if a^x farmer owned a^y donkey, he_x would beat it_y.

Then evaluating the counterfactual in (66) requires an update of the modal horizon to include at least some worlds in which a farmer owns a donkey. Using the definition in (64), we obtain the following updated modal horizon:

$$(67) \quad f^* | \text{a farmer owns a donkey} >_{\{x, y\}} \text{he beats it} |_g^{f^*, \leq^*} = \\ \{ \langle w_1, g^{x \setminus a, y \setminus d} \rangle, \langle w_2, g^{x \setminus b, y \setminus d} \rangle \}$$

Quantifying over these worlds, we find that the consequent of (66) is true in w_1 , but not in w_2 , rendering the (negated) counterfactual true as intended.

4.5 Generating the low + weak reading

Now we are aiming for a low plus weak reading of the non-negated version of (66), provided in (68). This means that we make $X = \emptyset$. The model is as above, which makes sentence (68) true.

(68) If a^x farmer owned a y donkey, he $_x$ would beat it $_y$.

We then yield the following modal horizon:

(69) $f^* | a \text{ farmer owns a donkey } >^\emptyset \text{ he beats it} |_{g^*}^{\leq^*} = \{ \langle w_1, g^{x \setminus a, y \setminus d} \rangle \}$

That is, since our similarity function as defined in (30) now ignores the assignments, we end up with a total ordering of the worlds that make the antecedent of (66) true. In this case, w_1 is the bottom element and we only quantify over (69), rendering the counterfactual sentence (68) true, as intended.

5 Conclusions

Contra Wang 2009, our proposal maintains the key insights and chief apparatus from van Rooij 2006 and derives high as well as low readings of indefinite NPs. However, by treating the (relevant) pronoun in identificational sentences not as a donkey pronoun but as a concealed question or part of a cleft, we avoid the problem of overgeneration of weak readings that van Rooij faces. Finally, by moving from a variably strict semantics for the counterfactual to a modal horizon-based strict semantics à la von Stechow 2001, we can account for NPI licensing both in high and low readings of counterfactual donkey sentences, thus improving on the empirical coverage of van Rooij's original system.

Some open questions remain: We would like to know more about which contextual cues determine the content of X , e.g., focus (Heim 1990). In addition, our proposal still keeps the quantificational dynamics quite separate from the counterfactual dynamics, and the consequences of this theoretical choice remain to be explored. As Asher & McCready (2007) point out, there is some interaction between the phenomena described in this paper and the puzzle of modal subordination. We will be exploring some of these interactions in further research.

References

- Asher, Nicholas & Eric McCready. 2007. Were, would, might and a compositional account of counterfactuals. *Journal of Semantics* 24(2). 93–129.
- Büring, Daniel. 1998. Identity, modality, and the candidate behind the wall. In Devon Strolovich & Aaron Lawson (eds.), *Proceedings of Semantics and Linguistic Theory (SALT) 8*, 36–54. Ithaca, NY: CLC Publications.

- Büring, Daniel. 2004. Crossover situations. *Natural Language Semantics* 12(1). 23–62.
- Büring, Daniel. 2011. Conditional exhaustivity presuppositions in clefts (and definites). *Ms. ZAS/Vienna*.
- Elbourne, Paul. 2005. *Situations and Individuals*. MIT Press.
- von Stechow, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. *Journal of semantics* 16(2). 97–148.
- von Stechow, Kai. 2001. Counterfactuals in a dynamic context. *Current Studies in Linguistics Series* 36. 123–152.
- Geach, P. T. 1962. *Reference and Generality*. Cornell University Press.
- Gillies, Anthony S. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360.
- Groenendijk, Jeroen & Martin Stokhof. 1991. Dynamic predicate logic. *Linguistics and Philosophy* 14(1). 39–100.
- Heim, Irene. 1990. E-type pronouns and donkey anaphora. *Linguistics and Philosophy* 13(2). 137–177.
- Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and philosophy* 16(4). 353–422.
- Lewis, David K. 1973. *Counterfactuals*. Blackwell.
- Mikkelsen, Line. 2005. *Copular Clauses. Specification, Predication and Equation*. John Benjamins.
- Romero, Maribel. 2004. Intensional noun phrases with know and be. *Catalan Journal of Linguistics* 3(1). 147–178.
- Romero, Maribel. 2005. Concealed questions and specificational subjects. *Linguistics and Philosophy* 28(6). 687–737.
- van Rooij, Robert. 2006. Free choice counterfactual donkeys. *Journal of Semantics* 23(4). 383–402.
- Root, Rebecca Louise. 1986. *The semantics of anaphora in discourse*. University of Texas Press.
- Schubert, Lenhart K & Francis Jeffrey Pelletier. 1987. Problems in the representation of the logical form of generics, plurals, and mass nouns. In Ernest Lepore (ed.), *New directions in semantics*, 385–451. Academic Press.
- Stalnaker, Robert C. 1968. A theory of conditionals. *American Philosophical Quarterly Monograph Series*, 2. 98–112.
- Velleman, Dan, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coppock. 2013. It-clefts are IT (inquiry terminating) constructions. In Anca Chereches (ed.), *Semantics and Linguistic Theory (SALT)* 22, 441–460.
- Wang, Y. 2009. Counterfactual donkey sentences: A response to Robert van Rooij. *Journal of Semantics* 26(3). 317–328.

Counterfactual donkey sentences

Andreas Walker
University of Konstanz
andreas.walker@uni-konstanz.de

Maribel Romero
University of Konstanz
maribel.romero@uni-konstanz.de