Exploring the existential/universal ambiguity in singular donkey sentences

Chao SUN — *Humboldt-Universität zu Berlin*
Daniel ROTHSCHILD — *University College London*
Richard BREHENY — *University College London*

Abstract. Although widely discussed in the semantics literature, donkey sentences have been the subject of very little experimental study. In this paper, we present experimental work that is aimed at addressing somewhat open questions around donkey sentences with determiners other than 'every' and their susceptibility to both Universal and Existential readings. Our experiments test donkey sentences with 'every', 'no', 'some' and 'more than two'. By using both verification and act-out tasks, we are able to show that Universal readings are available for donkey sentences with existential determiners (specifically, 'more than two'), as well as Existential readings. However, our studies fail to show that sentences with 'no' have a Universal reading, and they also provide some evidence against the idea that 'no' sentences have dual readings.

Keywords: donkey sentences, truth value judgments, act-out task

1. Introduction

Singular donkey sentences are widely recognised as being susceptible to several kinds of construal (Chierchia, 1995; Kanazawa, 1994; 2001; Geurts, 2002; Brasoveanu, 2008). Informants often report an intuition that there is a uniqueness implication. For (1), this would mean that each girl who baked a cake, baked one cake. However, informants are also prepared to accept that a sentence such as (1) could be used to describe a scenario where girls bake more than one cake. In that case, it has been observed that sometimes intuition prefers a so-called Universal reading. For (1), this would be that every girl who baked a cake iced all of the cakes she baked. However at other times, the preference is for an Existential reading. For (1), that would be that each girl who baked cakes ices some of the cakes they baked.

(1) Every girl who baked a cake iced it.

Given the widely accepted duality of readings for donkey sentences with 'every', a variety of proposals have been made to explain how these readings can be derived.[1] Many such proposals entail that donkey sentences headed by determiners other than 'every' should also give rise to both E- and U-readings. However, it is widely recognised that intuitions do not clearly support this prediction nor indeed is it clear the extent to which donkey sentences might have clear, determinate readings in the first place (Rooth, 1987; Kanazawa, 1994; Chierchia, 1995; a.o.). To date, very little literature has systematically explored intuitions about readings of donkey sentences with different determiners. The most widely discussed report on participant intuitions is Geurts (2002). The results of that paper provide clear-cut evidence for two readings for sentences with 'every' and 'not every', while results for sentences with 'some' and 'no' do not provide clear evidence for both. So, the question going forward is whether both readings are possible for these kinds of donkey sentence.

[1] There is not scope in this paper to review these. An excellent and comprehensive overview can be obtained from sources such as Chierchia (1995), Brasoveanu (2008) and Champollion et al. (2018).

 In: M. Franke et al. (eds.)
Proceedings of Sinn und Bedeutung 24, vol. 2, pp.289–305. Osnabrück University.

The primary aim of this paper is to explore the availability of readings experimentally. In particular, our aim is to compare the availability of E- and U-construals for donkey sentences with universal ('every'), negative ('no) and existential ('some', 'more than two') determiners. Unlike previous experimental work on donkey sentences, we combine both truth-value judgement tasks (Experiment 1) and act out tasks (Experiment 2). Because of possible ambiguity in donkey sentences, we have to take account of strategies for responding to different tasks. Critical items in both of our tasks make one reading true and the other false. In a truth-value judgement task, to the extent that both readings are accessible to a participant, a critical item should get a 'true' rating (since the image makes the sentence true on one of its available readings). In an act out task, participants may trade off Caution against Laziness. To the extent that both readings are available, a cautious participant will make the stronger reading true (the one that entails the other). However, if a weaker reading is already true, a lazy respondent may leave things be. To summarise the results below, we find no evidence for two construals of donkey sentences with 'no', but good evidence for two construals of versions with 'every' and 'more than two'. We conclude with a discussion of how to factor in the effects of context and determiner-specific strategies in accounting for why readings may be hard to find.

2. Experiment 1

2.1. Participants

43 participants were recruited from Prolific Academic and were paid 0.4 pound for their participation. All participants speak English as a native language. The experiment was initiated by a consent statement and was approved by the University College London Research Ethics Committee.

2.2. Materials and Procedure

We tested the availability of Universal and Existential readings for donkey sentences with four determiners: universal ('every'), negative ('no') and existential ('some', 'more than two'). For each determiner, we constructed three donkey sentences using three different scenarios: (i) girls baking and icing cakes; (ii) boys making and painting trains; and (iii) monkeys picking and peeling bananas. Each donkey sentence was paired with three types of displays: 'true' controls that made the sentence true on both readings, 'false' controls that made the sentence false on both readings, and target displays that made the sentence true on one of its available readings. Fig. 1 shows example sentences and displays for each determiner. One version of each item was assigned to one of the three lists, with each list containing 12 experimental items, 4 items per condition. In addition, each list contained 24 filler trials. Filler trials did not use donkey sentences. Half of the filler trials contained sentences with determiners, e.g. 'every girl baked a cake', and half contained simple positive/negative sentences, e.g. 'The yellow girl baked/ didn't bake a cookie'. Participants were randomly assigned to one of three lists. A randomized order of presentation of the items was created for each participant.

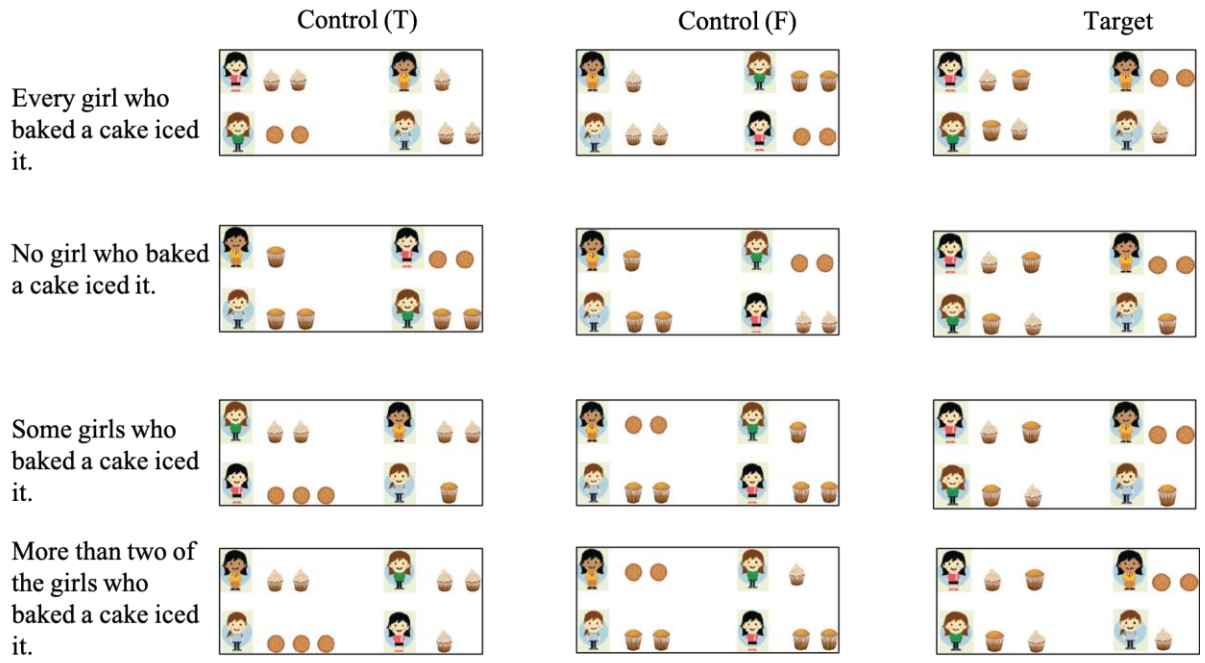

Figure 1. Example sentences and displays for each determiner in Experiment 1.

2.3 Predictions

Before turning to results, it is important to consider what we might expect if a given donkey sentence has both Universal and Existential readings, or if it has only one kind of reading. If one takes the view that a donkey sentence, with a given determiner, is able to give rise to both readings, there are several scenarios that are possible in our experimental setting. It could be that, when the participant reads the stimulus sentence in a trial, both readings are available to them. To give an example of what might happen in this case, let us consider an example of simple ambiguity, from beyond the area of donkey sentence research. Suppose a participant were given a sentence like, 'The woman walked toward a bank' and shown either a picture of a woman approaching a branch of Bank of America on a typical main street, or a picture of a woman walking through a field toward a river bank. In both cases, the expectation is that the combination of the linguistic stimulus and visual stimulus ought to allow the participant to see that there is a parse of the sentence that can make it true, assuming that both possible lexical senses become available to them. Thus, in our study, assuming that both possible construals become available to a participant, we assume that a participant will judge the sentence as true if the accompanying image makes the sentence true on one of those available readings. This could be called a 'charitable' response.

However, what we know about lexical ambiguity is that under certain conditions, even if a word has two senses, a participant may only access one sense – this through a combination of sense frequency and contextual bias (see for example, Duffy et al., 1988). We assume that something similar could carry over to the case of a donkey sentence that has two possible readings: for a combination of factors, on a given trial, only one reading becomes available to the participants and they base their response on that.

The other kind of scenario is that the donkey sentence with a given determiner has only one reading (i.e. only a Universal reading or only an Existential reading) and participants base their response on that.

2.4 Results

The percentages of 'true' responses for each determiner and condition are shown in Fig. 2. The mean accuracy on control items was 96%. Only in the target condition of 'every', the percentage of 'true' responses differed significantly from both 'true' controls (χ^2 (1) = 9.38, p = .002) and 'false' controls (χ^2 (1) = 35.8, p < .001).

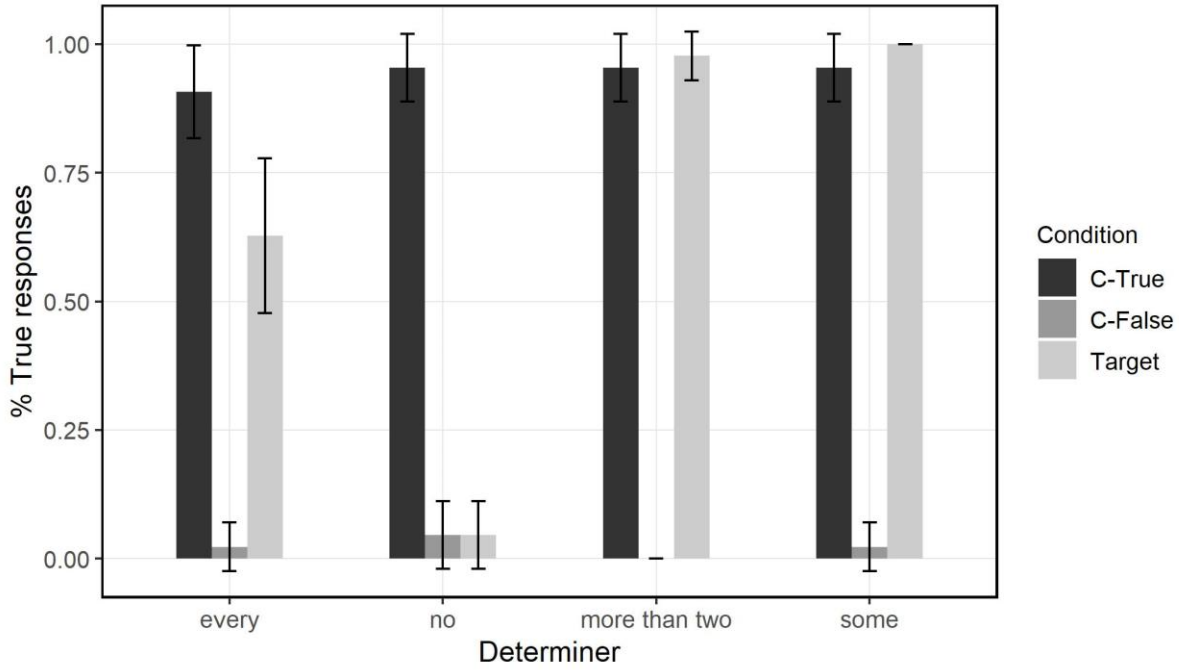

Figure 2. Percentage of 'true' responses for each determiner and condition.

2.5 Discussion

'Every': The fact that rate of 'true' in the target condition differs significantly from both the control conditions tells us that overall, participants did not find these sentences straightforwardly true or false in these cases. If 'every' donkey sentences only had an existential reading, we should expect rates to not differ from the 'true' control condition. If these sentences have only a universal reading, we would expect rates to not differ from the 'false' control condition. Thus, these results disconfirm the single-reading hypothesis for 'every' donkey sentences. In a way, this just confirms what is virtually universally agreed in the literature – that both readings are possible for the 'every' case. Another important observation to make here is that, assuming both readings are possible for these sentences, the existence of 'false' judgements means that, on some trials, only the U-reading is available. This

occurred in 37% of cases. So in these trials we have a situation of the kind mentioned above where, for reasons of frequency or due to contextual cues attended to in the trial, the E-reading does not become available.

'No': In the target condition of donkey sentences with 'no', the display made the U-reading true, yet the rates for target items did not differ from 'false' controls. This suggests either there is only the E-reading or that the U-reading is very inaccessible across virtually all of the trials.

'Some'/'more than two': In the target condition of 'some' and 'more than two', the percentage of 'true' responses did not differ from 'true' controls. Since the display made the E-reading true, these results are of course consistent with the existence of only an E-reading for existential quantifiers. When considering other possible hypotheses, we should note that conditions for these trials differ from the 'no' case. In the 'no' case, participants virtually always responded 'false'. In the case of existential determiners, the high rates of 'true' are consistent not only with a second, universal reading being completely unavailable or inaccessible, but also with the second reading being quite available. In the latter case, we have a situation as in the lexical ambiguity ('bank') case discussed above. It could simply be that with both readings available and an image consistent with one, participants respond 'true'.

In order to tease apart these two scenarios with existential determiners, we turn to an act out task using basically the same sentences and scenarios.

3. Experiment 2

3.1. Participants

42 participants were recruited from Prolific Academic and were paid £0.7 for their participation. All participants speak English as a native language. The experiment was initiated by a consent statement and was approved by the University College London Research Ethics Committee.

3.2 Materials and Procedure

Participants were presented with a display containing four agents and four possible states of affairs. The checked radio button indicates the outcome of each agent's activity. A fairy would give an instruction and the participants' task was to make sure that the outcome is as the fairy wanted. The fairy's instructions were constructed with four determiners: universal ('every'), negative ('no') and existential ('some', 'more than two') determiners, an example sentence for each determiner is given in Table 1. As in Experiment 1, for each determiner, we constructed three donkey sentences using three different scenarios: (i) girls baking and icing cakes; (ii) boys making and painting trains; and (iii) monkeys picking and peeling bananas. Each statement was paired with three situations: 'obligatory act', 'do nothing', and 'optional act', as shown in Fig. 3.

Every	I would like it that every girl who baked a cake iced it.
No	I would like it that no girl who baked a cake iced it.
Some	I would like it that some girls who baked a cake iced it.
More than two	I would like it that more than two of the girls who baked a cake iced it.

Table 1. Example sentence for each determiner in Experiment 2.

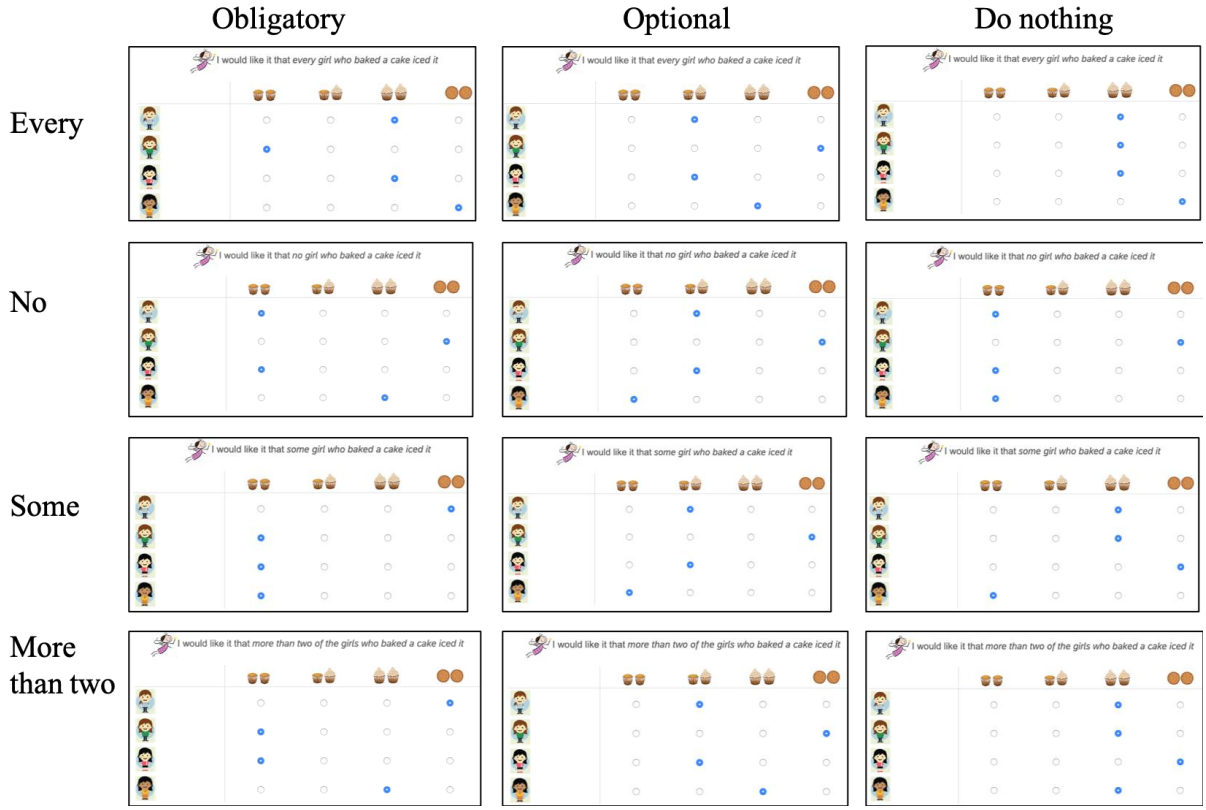

Figure 3. Example situations for each determiner in Experiment 2.

In the 'obligatory act' situation, whether both readings or only one reading were available, participants need to change what an agent has done by clicking a radio button under the desired state of affairs. For instance, when the instruction is 'I would like it that every girl who baked a cake iced it', the image shows one girl as having baked two cakes but having iced none (see Fig.3 top left). If participants access the U-reading, they should change what this girl has done by clicking the radio button under the two iced cakes; whereas if they access the E-reading, they have a choice between clicking the radio button under either both iced cakes or one iced and one un-iced.

In the 'do nothing' situation, the image is such that, whether both readings are available, or only one, the participant need do nothing. Here, the image was compatible with the U-reading in the case of 'every', 'some', and 'more than two'; and it was compatible with the E-reading in the case of 'no'. If these readings are preferred, participants should leave the situation unchanged.

In the 'optional act' situation, the display was compatible with just one reading. As for the test condition in the verification task, the image was compatible with the U-reading in the case of 'no' and the E-reading in the case of 'every', 'some' and 'more than two'.

One version of each item was assigned to one of the three lists, with each containing 12 experimental items, 4 items per condition. In addition, each list contained 24 filler trials. Filler trials did not use donkey sentences. Participants needed to change the display in half of the filler trials. Participants were randomly assigned to one of three lists. A randomized order of presentation of the items was created for each participant.

3.3 Predictions

As in the case of the verification task (Experiment 1), we should consider what could be expected, if a donkey sentence has just one (E- or U-) or two readings. For a given donkey sentence, if we assume that it is in principle possible for both readings to be available, then we still have a question whether both readings will become available on a given trial (due to effects of frequency, the stimulus etc.). Moreover, if both readings are available in a given trial, then what might a participant do? Given that, for each of our determiners, the two readings are ordered by entailment, we might expect that, if a participant exercises *Caution*, then they might base their response on the logically stronger reading. That way, whatever the fairy's actual desire, they can be sure to have satisfied it. In that case, they will make any changes based on that. However, if two readings are available and the visual state of affairs is already consistent with one, then *Economy* of effort might come into play and the participant may leave the situation as it is.

3.4 Results

Obligatory act. 1.2% of the responses were excluded because participants' responses resulted in a display that made the fairy's statement false. 12% of the responses were removed because participants changed the distractor. In most such cases, participants changed the cookie to a cake. Though many such responses could still be evaluated (as being Existential or Universal) we felt that participants who responded this way may have misunderstood the sentence (e.g. as, 'Every girl baked a cake and iced it'). Fig. 4 shows the percentages of U/E-responses for each determiner in the 'obligatory act' situation.

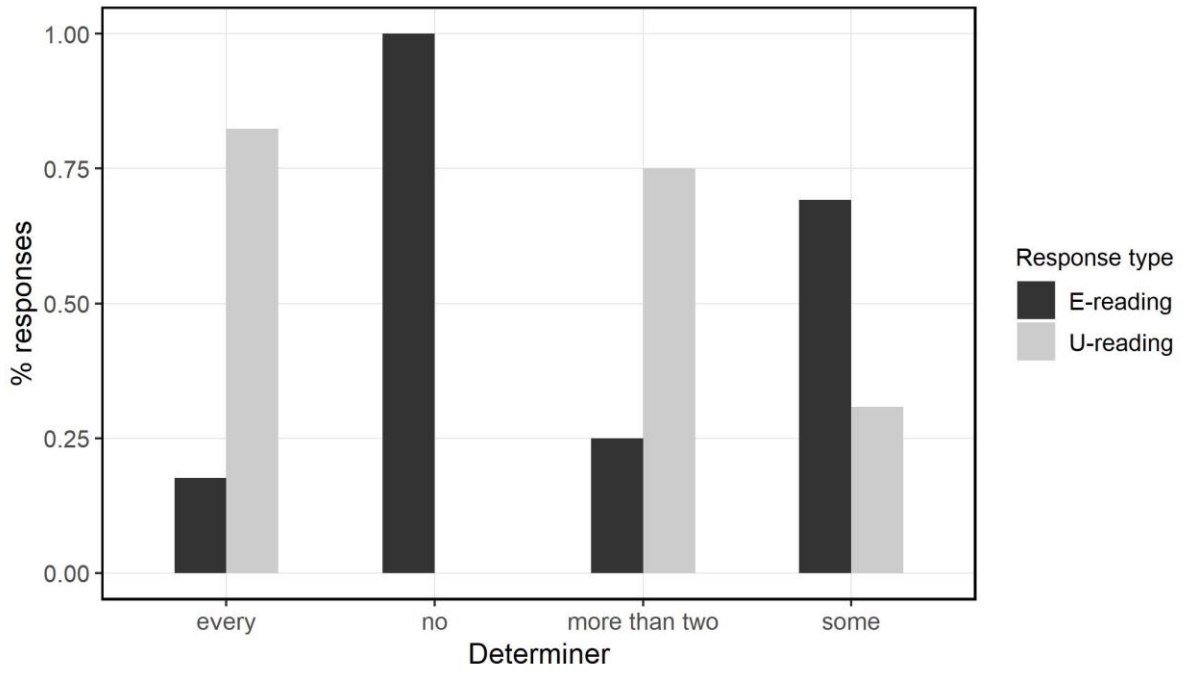

Figure 4. Percentages of U/E-responses for each determiner in the 'obligatory act' situation.

Optional act. 1.2% of the responses were excluded due to false choices. Again 8.9% of the responses were removed due to the change of the distractor. The status quo was compatible with the U-reading in the case of 'no', and it was compatible with the E-reading in the case of 'every', 'some', and 'more than two'. Participants who preferred a different reading would make changes to the display. Fig. 5 shows the percentages of U/E-responses for each determiner in the 'optional act' situation.

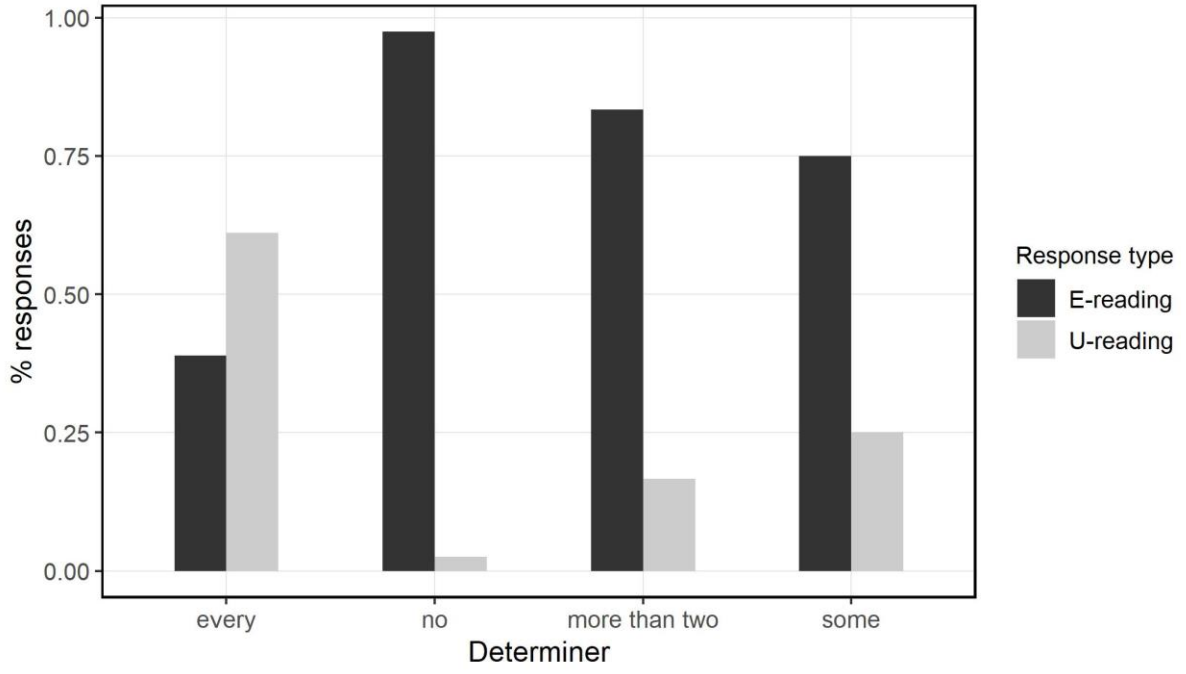

Figure 5. Percentages of U/E-responses for each determiner in the 'optional act' situation.

Do nothing. 1.2% of the responses were excluded due to false choices. 4.8% of the responses were removed due to the change of the distractor. The status quo was compatible with the U-reading in the case of 'every', 'some' and 'more than two' and it was compatible with the E-reading in the case of 'no'. Fig. 6 shows the percentages of act/no-act for each determiner in the 'do nothing' situation.

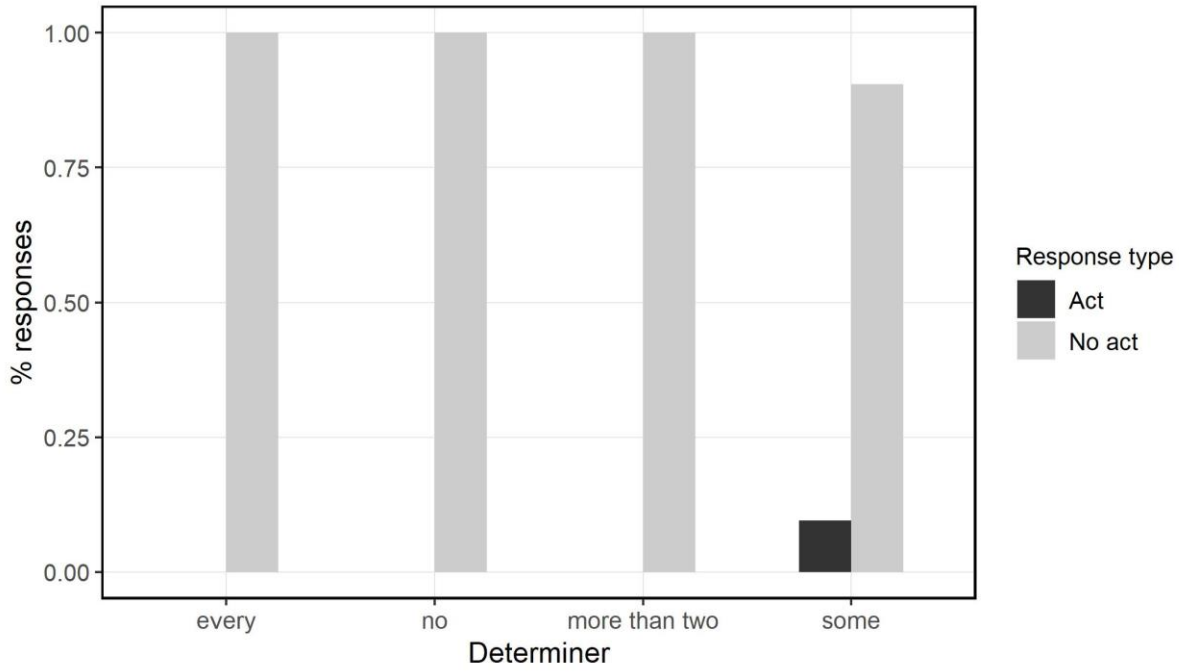

Figure 6. Percentages of act/no-act for each determiner in the 'do nothing' situation.

3.5 Discussion and Further Analysis

'Every': The percentage of U-reading responses in the obligatory act condition was significantly higher than the chance level (χ^2 (1) = 14.24, $p < .001$), whereas this percentage in the optional act condition did not differ from the chance level ($p = .18$). These results suggest that participants were ambivalent about the interpretation of the donkey sentences with 'every'. If there was only a U-reading for these sentences we should expect optional and obligatory responses to not differ and to be overwhelmingly for the universal state. This is clearly not the case. We can see this by looking at the preference for 'existential' response (i.e. choice of radio button with just one iced cake) between the two conditions which marginally differ ($\chi^2(1) = 3.5$, p= .06). To further support a two-reading account, we can consider what an E-reading only hypothesis would predict. On that assumption, we would expect that for both optional and do nothing conditions, no action was required. Thus, any action performed by participants would be for reasons to do with a felt need to do something on a trial. That effect would be the same across conditions. However, rates of action across the two conditions differ markedly (57% vs. 0%).

Thus both act-out and verification results with these items reveal that both U- and E-readings are available when participants encounter 'every' donkey sentences. This bears out the finding in Geurts (2002). It is interesting to now consider the data from 'every' sentences in more detail to see if we can find evidence of *Cautiousness* and *Economy* in participants' responses. If both readings are possible and available, then Cautiousness determines a strong preference for both-cakes-iced response on both optional and obligatory act. However, if both readings are possible and available, then Economy determines doing nothing on optional act (just-one-iced response). Thus if these two conflicting motivations are in operation for 'every' sentences, we would expect to find higher rates of U-reading in obligatory act than optional, and this prediction is borne out.

'No': Participants always preferred the E-reading consistent response (no cakes iced) both in the obligatory act condition (100%) and in the optional act condition (97%). We can easily rule out a hypothesis that says 'no' sentences only have a U-reading by noting that rates of action should be no different between optional act and no act conditions (97% vs. 0%). If 'no' sentences only have E-readings available, then we would expect both optional and obligatory conditions to result in E-reading outcomes above chance – which is the case. So, this study adds further support to the only E-reading hypothesis. In the verification study (Expt. 1), we saw an effective universal rejection of the U-reading state of affairs and this suggested either that these sentences only have an E-reading or that the U-reading was highly inaccessible. We can here further seek some evidence for the multiple reading hypothesis by considering whether we can observe a trade-off between Cautiousness and Economy in the obligatory and optional act trials. If so, we should see a lower rate of E-reading responses in optional act than obligatory, but this is not the case.

'More than two': In the obligatory act condition the percentage of two-iced cake responses was significantly higher than the chance level (χ^2 (1) = 8, p = .005), whereas in the optional act condition the percentage of just-one-iced responses was significantly higher than the chance level (χ^2 (1) = 16, p < .001). There is a clearly different pattern of responses here and this conflicts with a U-reading only view of 'more than two' sentences. As mentioned above, it has been suggested that donkey sentences with existential determiners only have E-readings (e.g. Kanazawa, 1994). However, this data speaks against an E-reading only view since in the obligatory act condition, the rate of two-iced responses should be same as rate of just-one responses. In fact these are markedly different (75% vs. 25%). In line with the 'every' donkey sentences, we see the expected pattern of trade-off between Caution and Economy in the 'more than two' case: more 'universal' responses in the obligatory act than optional. This pattern is predicted on two readings view of this kind of donkey sentence.

'Some': The percentage of just-one-iced responses was significantly higher than the chance level both in the obligatory act condition (χ^2 (1) = 5.77, p = .02), and in the optional act condition (χ^2 (1) = 10, p = .002). Using McNemar's chi-square, we found that the preference for the just-one response did not change significantly between the obligatory act and optional act conditions (p = .58). If 'some' sentences only had U-readings, then we should expect no just-one choices on obligatory trials, and obligatory change on optional. Neither of these outcomes was found.

If 'some' sentences only have an E-reading, then rate of acting on optional trials should be at same 'noise level' as no-action trial, but this is not the case (χ^2 (1) = 4.2, p = 0.04). Another prediction to consider concerns the obligatory act condition where all three girls who baked cakes ice none. If only E-readings are available, then we could assume that choices on how to make the sentence true between the two alternatives (both iced or just one) would be at random. In fact, rates of single icings are clearly greater than 50% (i.e. 72%). We are unsure what can explain this effect. One possibility is that participants somehow are encouraged to treat the indefinite 'a cake' as specific or some kind of singleton. In fact, if we assume all universal responses (28%) were just a result of a random selection between two options for people accessing the E-reading without the additional singleton construal, then around 16% of respondents would have to have acted based on a singleton response. (However we note that, if this were the case, then we could not tell if they access the E-reading or the U-reading for the sentence - see also the discussion of strategies in Section 4).

Another consideration to take into account for 'some' donkey sentences is that they may be more susceptible to a non-quantificational interpretation than other existential quantifiers (Champollion et al., 2018). For example, the subject noun phrase could be construed as introducing a discourse referent, or itself being interpreted specifically – rather than as a generalised quantifier. These choices could impact on readings for the sentence. We set aside resolving what is happening in the 'some' case, noting here only that the results are not straightforward to interpret.

To the extent that we have evidence that participants are able to access both U- and E-readings for both 'every' and 'more than two' sentences, we can consult the act-out results to gain some further information on the extent to which these readings are available. Consider the obligatory act condition. If you access the U-reading you have to click on 'both cakes iced', but if you access the E-reading it does not matter if you click on both iced or just one iced. Now, we see that there are some responses where the person has clicked on the just one iced radio button (not the both iced one). This response must be based on E-reading. So, if we assume that this choice is made at random, then we can infer that if n responses where to click on the 'one-iced/one-not' button, then our best guess would be that around 2n participants access only the E-reading. Our best guess then is that around 50% of participants only access E-readings for 'more than two', while around 36% of participants only access E-readings for 'every'.

4. General Discussion

In summary, the act-out task results confirm that (in some but not all cases) participants respond as if both E- and U-readings of 'every' donkey sentences are available. Given the results of the verification tasks both in our Experiment 1 and Geurts (2002), as well as widely shared intuitions, we have evidence to support a duality of possible construals for this kind of sentence. Importantly, we have indirect evidence for our assumption that competing motivations of Caution and Economy were at play for our participants in the act out task. We can examine the other act-out results in this light. Recall that in Experiment 1 almost all participants responded 'false' to 'no' sentences, and 'true' to 'some' and 'more than two'. This could be because these sentences only give rise to E-readings. However, in the case of 'some' and 'more than two', an

alternative explanation was that participants respond ‘true’ when two readings are available and one is true. This cannot apply to ‘no’ sentences in the verification task. An important role for the act out task was to change the motivations for response when two readings are available from a ‘Charity’ response in verification tasks to the trade off between Caution and Economy – as found in ‘every’ trials. Results for ‘some’ sentences in the act out task were equivocal but for ‘more than two’ we found patterns of response similar to the ‘every’ case and consistent with the two-construal view. In addition results of the ‘more than two’ act out task tended to disconfirm the E-reading-only view.

4.1 Finding missing readings

Thus, contrary to some suggestions in the literature on donkey sentences, we have evidence that (at least some) existential determiners are susceptible to U-readings as well as E-readings. Our investigation, which triangulates on both verification and act out tasks has turned up something akin to the ‘phantom reading’ for modified numerals, which were uncovered in Marty, Chemla and Spector (2015). That is, a reading that has not been immediately apparent to introspection in previous literature but which is shown to be the basis of participants’ responses. In a similar way, these U-readings for donkey sentences with existential determiners seem to lurk not too far below surface awareness. These results tend to undermine suggestions made in Kanazawa (1994) that intersective determiners in general should resist a universal reading (see also Champollion et al., 2018 for further discussion).

But what of negative intersective determiners? Our results for ‘no’ sentences point strongly away from a dual reading hypothesis. The main piece of evidence here is the almost unanimous judgment that ‘no’ sentences were false in scenarios that satisfy the U-reading. The fact that no participant was inclined to take a charitable view, suggests that this reading was unavailable in these items. The same resistance to judge ‘no’ sentences as true was found for visually and situationally distinct items in Geurts (2002), adding to the conclusion that the U-reading is not available. There was further support from the act-out task for an E-reading-only account, as well as an absence of evidence for the dual-reading account. Thus, our studies suggest no U-reading for ‘no’ sentences. So what of the widely shared intuition that there are circumstances under which the U-reading emerges? To date, this intuition is based on examples like in (2):

(2) a. No man who had a credit card failed to use it.
b. No person who had an umbrella would leave it at home on a day like today.

As observed in Geurts (2002) those apparent exceptions to the E-reading generalisation for ‘no’ tend to involve a predicate that is in some ways implicitly negative. If this is the only kind of case where an apparent U-reading is produced for negative determiners, this is consistent with an account for one reading of donkey sentences (the E-reading for ‘every’ and ‘more than some’ sentences) as resulting from a ‘super-narrow’ existential scope with respect to the matrix predicate. How this analysis is spelled out in detail, we leave for another occasion.[2]

[2] The ‘super narrow’ account of E-readings for ‘every’ and existential determiners would have to be combined with a mechanism for deriving their U-reading which in turn only derives an E-reading for ‘no’ and other negative constructions. We believe several such mechanisms are possible but leave consideration of these for another occasion.

4.2 A key role for context?

While some accounts for the readings of donkey sentences have focused on properties of the determiner and their influence on how construals of donkey sentences are derived, other accounts have laid the burden of explanation for the variety of readings on the role of context. To some extent, this begins with early observations that U- and E- readings for singular donkey sentences seem to arise in contexts where the same readings arise for plural definite descriptions or plural donkey pronouns (Krifka, 1996; Yoon, 1996). For example, in (3)a below we can see that the plural definite has a construal which can be glossed as existential ('some of his windows'), while (3)b has one which is universal ('all of his windows') and the idea is that this variation is based in some ways on context:

(3) a. Max left his garage windows open while he was away.
b. Max left his garage windows shut while he was away.

The appeal to context in explaining this difference might go as follows: Thinking about home security issues, it would be relevant to know that Max leaves even some of his windows open while he is out, but hardly more relevant to know that he leaves all of his windows open; by contrast, knowing that he shuts some or all of his windows hardly resolves questions related to security, while the information that he shuts all of them does. Both Krifka and Yoon observe that these factors influence donkey sentence construal in apparently similar fashion:

(4) a. Usually, if a person has a garage with a widow, they leave it open while they are away.
b. Usually, if a person has a garage with a widow, they leave it shut while they are away.

Early proposals that assimilate the interpretation of singular donkey pronouns to plurals were shown to be very problematic (Kanazawa, 2001). Nevertheless, more recent proposals avoid treating the terms themselves as plurals while being able to capture the same intuition. Specifically, Champollion et al. (2018) adapts the treatment of definite plurals found in Kriz (2015) to explain the conditions under which U- and E-readings of singular donkey sentences may arise. The leading idea is that a sentence for which there is not one specified reading poses a dilemma for the language user as to how to understand it. The proposal is that, in that case, the strategy would be to consider the sentence true if it is true on all readings and false if it is false on all readings. In 'mixed' or 'non-homogeneous' cases, no determinate truth value can be assigned. However, following Kriz, Champollion et al. (2018) propose that mixed scenarios can be assimilated to the set of true or false scenarios, depending on the contextual Question Under Discussion (QUD). To be more concrete, let us consider a proposal for an 'every' donkey sentence, based on Champollion et al. (2018):

(5) Every girl who baked a cake iced it.
- True iff every girl who baked a cake iced all of the cakes she baked
- False iff some girl who baked a cake did not ice any of the cakes she baked
- # otherwise

A state of affairs, $w_{\#}$ in which every girl ices some of the cakes she bakes but not all girls ice all is one of these mixed scenarios. Let us assume that (5) addresses a QUD which partitions the context so that true scenarios (w_T) and false scenarios (w_F) lie in different cells. Then we can say that (5) can be treated as true in $w_{\#}$ if $w_{\#}$ belongs to the same cell of the partition determined by the QUD as w_T. For example, if we are interested in how many girls did any icing of their cakes then $w_{\#} \approx w_T$. Likewise, (5) is treated as false in $w_{\#}$ if $w_{\#} \approx w_F$. Based on Champollion et al. (2018), proposals for donkey sentences with 'no' and existential determiner, 'more than two' are given in (6) and (7):

(6) No girls who baked a cake iced it.
- True iff no girl who baked a cake iced any of the cakes she baked
- False iff at least one girl iced all of the cakes they bake
- # otherwise

(7) More than two girls who baked a cake iced it.
- True iff more than two girls who baked a cake iced all of the cakes she iced
- False iff two or fewer girls iced some of the cakes they baked
- # otherwise

Champollion et al. (2018) suggest that a default reading (in the absence of any specific context) could be one based on a maximally inquisitive context, where for all w_i, w_j, w_i is not equivalent to w_j. This means that the default reading for 'every' and existential donkey sentences would look like a U-reading, while for 'no' sentences it look existential (but see footnote 3). We can assume that we move away from the default where information indicates a different kind of QUD. For example, examples sentences in (3)-(4) themselves are suggestive of different kinds of QUDs and, as such, suggest E-readings for the (a) versions and U-reading for the (b) versions.

Before we move on to consider this account in light of our experimental results, we would point out that while we can see how it works very well for plural donkey sentences, our intuitions resist a straightforward application to singular donkey sentences when it comes to the 'no' case. Consider that (8) below suggests a non-homogeneous interpretation. I.e. it could easily be judged true where Smith vaccinated some but not all cows by the deadline:

(8) Farmer Smith was fined because he didn't vaccinate his cows by the government deadline.

As with (3) above, we can explain this intuition as arising from a QUD about whether the explanation for a fine would be not following government rules to vaccinate all of one's livestock. Presumably, this QUD is naturally suggested given the sentence itself and background knowledge. This same combination of sentence and background knowledge seems to readily give rise to the U-reading for the plural donkey sentence in (9) below:

(9) No farmer who owns donkeys vaccinated them before the government deadline.

However, we find a clear contrast in the availability of the U-reading between (9) and (10) below, with only difference being the number on the pronoun:

(10) No farmer who owns a donkey vaccinated it before the government deadline.

Returning to our experimental data, what can we say about our participants' responses in light of these proposals about a role for QUD in determining readings? We can assume that participants may not be able to discern from the items any particular QUD. In that case, following Champollion et al., they would use the default QUD – what is the case? However, as for (3)-(4) and other examples above, participants may imagine more specific contexts for the stimuli. In particular, they may imagine QUDs that would yield U- or E-readings – of the kind suggested below:

- Possible QUDs for our items.
  QE: How many girls who made cakes got to do any icing?
  QU: How many girls who made cakes completed their tasks and iced them all?
  QD: What is the case?

Turning to the verification task with 'every', a 'true' response must have arisen from a participant being able to see a plausible QE. Due to charity, this could have been in addition to seeing a specific QU and/or the QD. A false response for 'every' must have resulted from participants only projecting a QU or potentially also QD[3]. Looking at our results for 'every', we can estimate that in around 60% of critical trials, participants could plausibly project a QE in addition to possibly also projecting QU/QD, while on 40% of trials participants failed to discern a QE. For sentences with 'more than two', verification task results suggest that participants must have seen a QE on all virtually all critical trials. It is interesting that in critical trials for 'every' and 'more than two', items did not have visually different displays. This means that the only difference was the determiner. Following a strong context-based explanation of readings, we would have to assume that different determiners ('every' vs. 'more than two') carry with them different biases (prior probabilities) for different contexts. While this is not completely implausible, it is puzzling why it would be so. As such, it makes sense at this stage to consider that determiner specific strategies could explain this difference.

4.3 Verification strategies as an extra factor

Geurts (2002) observes that, when verifying a sentence with an intersective determiner (this would include both existential determiners and 'no'), it is a generally useful strategy to seek critical evidence in a sub-model of the whole model. In the case of existential determiners, this would be positive evidence. In accounting for the strong tendency for donkey sentences with existential determiners to yield an existential response in his own experiments, Geurts suggests that, whether a participant has derived a U- or E-reading, an application of the strategy of inspecting sub-models could yield an existential response. To illustrate with our own items above, Geurts' idea is that even if a participant has in mind a U-reading it is possible that they

[3] Since test scenarios are 'mixed' scenarios and a QD leads to a homogeneous reading, the 'every' sentence would lack a truth value in such scenarios. However, no third response option is given to our verification task participants, so we could assume they favour a rejection ('false') response to register infelicity. Thus default QUDs would produce outcomes similar to a QU. This is what we assume here, though we are disinclined to put too much weight on this assumption.

selectively ignore un-iced cakes, so long as they can pair some/more than two cake-baking girls with iced cakes. As Champollion et al. note, their own account of donkey sentences can be combined with this account of verification strategies. If we follow this suggestion, it would allow us to resolve the puzzle of the divergence in existential responses between 'every' and existential determiners for the verification task. It could be that on around 40% of 'more than two' trials, participants do not project a QE, leading to only QU/QD contexts and U-readings but in these cases they are strongly biased to adopt the sub-model strategy, yielding what looks like an existential response. While this is not impossible, we should note that our act-out data would pose problems for this position unless we were to assume that the sub-model strategy is only adopted on verifications tasks, not the act-out (see Section 3.5).

4.4 Back to 'no'

If we assume that it is a plausible approach for 'more than two' to adopt a multi-factor account of the readings of donkey sentences, with projected context (QUD) and determiner-specific strategy playing a role, we should also consider what this account implies for our 'no' data. Test items in the verification task for 'no' were minimally different to those for 'every'/'more than two' – a single cake which is un-iced in the latter case is iced in the former (see Fig. 1 above). On the face of it, results for 'no' suggest that QU is not projected at all, with QE (and perhaps also QD) the prevailing contexts. While we cannot rule this out, it would be surprising if no QU is projected for 'no' without also not being projected for other determiners. Since we did not provide a third 'don't know' response option for our trials, we cannot rule out that all rejection responses in the 'every' trials were based on QD rather than QU context (see footnote 3). We note that in Geurts (2002), where participants were given a third response option, very few chose that and this would be consistent with 'false' responders accessing QU contexts rather than QD. So we have reason to doubt that so-called 'default' contexts would dominate in these kinds of sentence-picture verification tasks.

So, let us now suppose a multi-factor account, assuming that QU is sometimes adopted for 'no', but that participants are strongly biased to adopt a sub-model strategy. For the negative quantifiers, this is to seek disconfirming information in sub-models, selectively seeking the iced cakes to falsify the sentence. Again, we do not find this assumption unproblematic since we would have to suppose that the sub-model strategy for positive quantifiers has one aim (to selectively attend to confirming information), while the strategy for negative quantifiers has the opposite aim. The latter aim is at odds with the assumed principle of Charity.

5. Conclusion

Previous experimental research on the U/E ambiguity of donkey sentences has been unable to detect both readings across different classes of determiner. This paper adds to the small body of controlled experimental data on readings of donkey sentences by providing clear evidence in support of analyses of donkey sentences with existential determiners ('more than two') as having both U- and E- readings. In addition, our studies throw up evidence against accounts that predict only E-readings for these sentences. By using both verification and act-out

paradigms, we have been able tease out 'phantom readings' of existential donkey sentences, lying not far from the surface of theorists' introspective access. By contrast, we were unable to unearth a second U-reading for 'no' sentences, replicating results from Geurts (2002); plus we provide some evidence against the idea that 'no' sentences have two readings.

Because there are many potential factors that may contribute to donkey sentence readings, relating to context and determiner specific strategies, further work is required to test these in more detail. However, we remain skeptical that negative determiners like 'no' can give rise to a U-reading.

References

Brasoveanu, A. (2008). Donkey pluralities: plural information states versus non-atomic individuals. *Linguistics and philosophy, 31*(2), 129-209.

Champollion, L., Bumford, D. & Henderson, R. (2018). Donkeys under discussion. *Semantics & Pragmatics, 12*(1).

Chierchia, G. (1995). *Dynamics of meaning*. Chicago: University of Chicago Press.

Duffy, S. Morris R. & Rayner K. (1988). Lexical ambiguity and fixation times in reading, *Journal of Memory and Language, 27*(4), 429–446.

Geurts, B. (2002). Donkey business. *Linguistics and philosophy, 25*(2), 129-156.

Kanazawa, M. (1994). Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. *Linguistics and Philosophy, 17*(2). 109– 158.

Kanazawa, M. (2001). Singular donkey pronouns are semantically singular, *Linguistics and Philosophy, 24*(3), 383-403.

Krifka, M. (1996). Pragmatic strengthening in plural predications and donkey sentences. *6th Semantics and Linguistic Theory Conference (SALT 6)*, 136–153.

Križ, M. (2015). Homogeneity, non-maximality, and all. *Journal of Semantics, 33*(3), 493-539.

Marty, P., Chemla, E. & Spector, B. (2015). Phantom readings: the case of modified numerals. *Language, Cognition and Neuroscience*, *30*(4), 462–477.

Rooth, M. (1987). Noun phrase interpretation in Montague Grammar, File Change Semantics, and Situation Semantics. In Peter Gärdenfors (ed.), *Generalized quantifiers: Linguistic and logical approaches*, 237–268. Dordrecht, Netherlands: Reidel.

Yoon, Y. (1996). Total and partial predicates and the weak and strong interpretations. *Natural Language Semantics, 4*(3), 217–236.