OXFORD

Kripkeans of the world, unite!

Farjana Islam

Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, Trondheim

Giosuè Baggio

Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, Trondheim

Abstract

This paper revisits a study by Machery et al. (2004), suggesting that, in experimental versions of Kripke's (1980) fictional cases on the use of proper names, Westerners are more likely than East Asian participants to show intuitions compatible with Kripke's causal-historical (CH) theory of reference. We conducted two experiments, recruiting participants from Norway and Bangladesh, either in English (experiment 1; N = 75) or in the participants' native languages (experiment 2; N = 60), using modified cases and a new approach to data analysis. We replicated the results of Machery et al. (2004), but we show that the residual finding—i.e., that participants who are not aligned with CH produce responses consistent with a definite descriptions (DD) theory of reference—does *not* hold. Most participants in our experiments, and nearly all those who do not provide CH answers, respond as predicted by a theory that accommodates speaker's reference in reasoning about uses of proper names, not according to DD. We suggest that cross-cultural variation in this task is real. However, explanations of variation within or across cultures need not invoke competing theories of reference (CH *vs* DD), and can be unified within a single, broadly Kripkean analysis that honors the basic distinction between semantic reference and speaker's reference.

1. INTRODUCTION

How do proper names pick out their referents? How does 'Gödel' refer to the individual Kurt Gödel? Debates in semantics have focused on two theories of reference. One theory originates in the work of Russell (1905). It holds that a definite description (DD) is associated with each proper name: an individual is the referent of the proper name, if and only if it satisfies the associated DD. For example, 'Gödel' would be associated with the DD 'The man who proved the incompleteness of arithmetic'. The individual who satisfies the DD (Kurt Gödel) is the referent of the name. Kripke (1980) argued that this theory fails to account for the fact or intuition that 'Gödel' would still refer to Kurt Gödel, even if he did *not* satisfy the DD—e.g., if someone else had, in fact, discovered the incompleteness of

arithmetic, and Gödel had stolen the manuscript from them. Kripke offered an alternative causal-historical (CH) account, in which a proper name is introduced to denote a specific individual, baptised using that name, and continues to refer to it, so long as the name's uses are linked to the same individual *causally* in chains of transmission and usage. In the fictional case above, 'Gödel' would still refer to Kurt Gödel, even if he did not actually prove the theorems attributed to him. In the CH theory, in general, a DD plays no role in fixing a proper name's reference.

Kripke's attack on the DD theory is compelling. However, the *method* he used in his argument—i.e., the construction of a fictional case, in which 'Gödel' still refers to Kurt Gödel even though the associated DD is not satisfied—has itself come under attack, most notably by Machery et al. (2004). They presented to participants from Rutgers ('Westerners') and from Hong Kong ('East Asians') vignettes in English modeled after Kripke (1980), describing the hypothetical case of a person, Schmidt, who actually proved the incompleteness theorems, and from whom Gödel stole the proof. They then asked participants whether someone who has learned in school that Gödel proved the incompleteness of arithmetic would, using the name 'Gödel', be speaking of (A) 'The person who really discovered the incompleteness of arithmetic', or rather (B) 'The person who got hold of the manuscript and claimed credit for the work'. Machery et al. regard (A) as the response that participants should give, if their intuitions are constrained by a DD, and (B) as the expected response, if their intuitions are grounded on causal-historical notions. Westerners were more likely than East Asians to select (B) (for replications and extensions in different Western (US, France) and Asian (Hong Kong, Mongolia, India) samples, see Machery et al. 2009, 2010; Beebe & Undercoffer 2015, 2016; Sytsma et al. 2015).

Machery et al. (2004) included a second type of vignette in their study, based on Kripke's (1980) Jonah case. Kripke prompts the reader to suppose that no prophet was swallowed by a big fish or whale, and that nothing in the Biblical account of Jonah's life is true: there is *no individual* who satisfies a DD of that sort. If the Biblical story is based on a real person's life (as opposed to being a fictional account of *no person*), 'Jonah' should still refer to the person 'behind the story', yet no one did the things attributed to him. Machery et al. designed a vignette where the story of Attila is recast as one in which nobody did what is commonly related to him—no warrior expelled the Romans from Germany etc. They asked participants whether someone exposed only to conventional history, in saying 'Attila was the king who drove the Romans from Germany', would be 'talking about' (A) the real person on which Attila's fictional story is based (named 'Raditra'), or (B) 'a fictional person who does not really exist'. Participants whose intuitions are guided by a DD should answer (B)—the DD is *not* satisfied, and 'Attila' fails to refer—, while people whose intuitions are constrained as per the CH theory should choose (A)—one could use 'Attila' to talk about the name's "original bearer" (Machery et al. 2004, p. B4), whether or not the DD is satisfied. Here, there was no difference between Westerners and Asians (however, Beebe & Undercoffer 2016 found the expected effect in one of their replication experiments).

We conducted two experiments with modified versions of the vignettes used by Machery et al. (2004) and a novel approach to data analysis. Our proposed modification concerns the framing of probe questions in Jonah cases (J-cases) to render them fully comparable to Gödel cases (G-cases). One problem with the J-case task of Machery et al. (2004) is that the question which introduces the two alternative responses (A) and (B) contains a quoted sentence ('When a contemporary German high-school student says "Attila was the

king who drove the Romans from Germany", is he actually talking about (. . .)') featuring both the name 'Attila' *and* the associated description 'the king who drove the Romans from Germany'. It is then unclear whether the question bears on the use and reference of the proper name or of the description, or both. Note that this contrasts with the form of the probes in G-cases in Machery et al. (2004), where the question only mentions the proper name (e.g., 'When X uses the name 'Gödel', is he talking about (. . .)'). In our G- and J-cases, we used probes of identical form, i.e., 'When X uses the name 'Y', he/she is talking about [(A) or (B)]', removing the ambiguity inherent in the formulation of J-case probes in Machery et al. (2004). Another difference is that, while Machery et al. used a proper name in (A) answers (e.g., 'Raditra'), we used a definite description, again to render answers structurally similar in G- and J-cases. A further issue in Machery et al. (2004) is that, in both G- and J-cases, by selecting the (A) or (B) responses, participants are simultaneously completing the questions *and* answering them. Here, we opted for a declarative, instead of an interrogative, formulation of probes in both G- and J-cases.

The present study also adopts a different approach to analyzing participants' responses, motivated by a long-running dispute on the meaning of the probes in Machery et al. (2004). Some authors have argued that it is unclear whether those probes bear on *what the proper name refers to* ('semantic reference') or rather on *what a speaker may refer to when using the proper name* ('speaker's reference', Kripke 1977, pp. 256 and 262–264; Ludwig 2007; Deutsch 2009; Genone 2012; Domaneschi et al. 2017; for a related, parallel critique, focused on epistemic perspectives, see Sytsma & Livengood 2011; for more criticism, see Lam 2010; Vignolo & Domaneschi 2017; Izumi et al. 2017). Studies have tried to address these concerns, with some success (e.g., Beebe & Undercoffer 2016), and some have tried to counter the argument that the probe questions in G-cases and J-cases are ambiguous (Machery & Stich 2012). However, Heck (2018) has recently cast doubt on these counter-arguments, showing that the ambiguity remains, even after several attempts at rephrasing the probes.

The gist of Heck's analysis is this. Firstly, it is not sufficient to say, as Machery & Stich (2012) do, that participants *could not* understand the probe question as being about speaker's reference (SR)—their claim being that SR applies to *tokens*, whereas the query 'Who does X refer to when she uses the name 'Y'?' bears upon a *type*, about which only questions of semantic reference could be asked. Heck suggests that that question may also be interpreted as a generic, or a generalisation over uses of a proper name, so that the SR interpretation remains accessible. Secondly, Heck also notes that rephrasing the probe, as in 'When X uses the name 'Y', regardless of who she intends to be talking about, she is actually talking about [(A) or (B)]' (Machery et al. 2015), would not do: the distinction between semantic reference and SR is not between the person about whom one intends to speak and the person about whom one is actually speaking, but between the person about whom one *is intentionally speaking* and the individual to whom a specific expression refers *in the language* (for an early response to this kind of objection, see Machery et al. 2015). Thirdly, according to Heck, it would be a mistake to believe that one who understands the probe question as being about SR *should* give a non-Kripkean answer. In G-cases, even though X intends to refer to the person who became renowned for publishing the proof, it does not follow that X does not *also* intend to refer to the person who proved the theorem, since from X's perspective there just is no difference between the two (Heck 2018). It may be that *all* participants in Machery et al. (2004) and in later experiments have interpreted the probe as being about SR: response differences would show that different intentions (e.g., to

refer to the one who proved the theorem *vs* the one who published it) prevail in each case. If so, the notion of SR may suffice to explain the data.

Our goal here is not to address the points raised by Heck (2018) empirically, but to suggest a new way of disentangling the response patterns predicted by each theory (DD, CH, and SR) without modifying the probe questions, except for the proposed realignment of the task phrasing in G-cases and J-cases. The three theories make different predictions as to the responses subjects should provide, assuming they understand that the questions are about what proper names *refer to* ('... is talking about [(A) or (B)]'):

(1) The DD model predicts a prevalence of (A) responses in G-cases, since the referent of 'Gödel' should satisfy the description 'The person who really discovered the incompleteness of arithmetic'. In J-cases, 'Attila' is associated to the description 'The German warrior who expelled the Romans from Germany'. This description, and the (A) response, imply (following Russell 1905) that *there is* an individual with such and such properties. However, because the story makes it clear that there is no such individual (i.e., that the description is *not* satisfied: "no merciless warrior expelled the Romans from Germany (...)"), participants whose responses are guided by a DD should reject (A) and should accept (B). According to DD, the proper name refers to a fictional individual who does not actually exist (Machery et al. 2004). DD predicts a prevalence of (A) responses in G-cases and of (B) responses in J-cases (DD: A-B).

(2) CH predicts a prevalence of (B) responses in G-cases. 'Gödel' refers to the person bearing that name, regardless of any facts (mis)attributed to him. CH predicts a prevalence of (B) responses in J-cases, too. There is no person whom the name 'Attila' rigidly refers to. Even if a chain of transmission or usage for 'Attila' could be reconstructed, it would *not* terminate in a specific individual. Unlike for 'Gödel' in G-cases, and for 'Jonah' in Kripke's (1980) scenario, there is no "original bearer" *of the name* 'Attila', so 'Attila' fails to refer or refers to a person who does not really exist. CH predicts (B) responses in G- *and* J-cases (CH: B-B).

(3) In the G-case, X believes that Gödel has proved the incompleteness of arithmetic. Therefore, when using the proper name 'Gödel', X *intends to refer* to the person who satisfies the DD expressing this belief. As X lacks knowledge about the story behind the incompleteness theorems, he cannot use 'Gödel' intending to refer to the man who got hold of the manuscript and claimed credit for the proof. SR predicts a majority of (A) responses in G-cases. In the J-case, German students believe there is a warrior, Attila, who expelled the Romans from Germany, and they *intend to refer to him* by using the name 'Attila'. They do believe Attila existed, so they could not, according to SR, use 'Attila' to intentionally refer to a fictional character. SR predicts a prevalence of (A) responses in G- *and* J-cases (SR: A-A).

In two experiments, we tested whether the results obtained by Machery et al. (2004) would replicate in G-cases and modified J-cases. Machery et al. (2004) correctly note that the relevant DD is not satisfied in the J-case. So, according to the DD theory, the name *fails to refer* (B). They also say that "someone can use the name to speak about the name's original bearer, whether or not the description is satisfied" (p. B4). But this cannot mean that, in their/our J-case, (A) is predicted by CH: Raditra is *not* the "original bearer" *of the name* 'Attila', and it is unclear that an "original bearer" exists in J-cases. The vignette rather implies a *broken causal chain*, so 'Attila' fails to refer. Both DD *and* CH predict that (B) is the appropriate response. The different accounts (DD, CH, and SR) can only be

disentangled by analyzing *responses across G-cases and J-cases*: SR (G-cases: A; J-cases: A), DD (G-cases: A; J-cases: B), CH (G-cases: B; J-cases: B).

In experiment 1, we presented 6 G-cases and J-cases in English to a European (Norwegian) and a South Asian (Bangladeshi) groups, following as closely as possible the procedure of Machery et al. (2004). We also included a sample of South Asian participants (from Bangladesh or Nepal) living in Norway, to test whether the expected cross-cultural differences would be robust to exposure to elements of a Western culture (for motivation, see Lam 2010 and Mesoudi et al. 2016). In experiment 2, we administered the test to Bangladeshi and to Norwegian individuals, now in their native languages (Bangla or Norwegian), instead of English, and departing from Machery et al. (2004) in some respects concerning the experimental design and testing procedure (see Methods).

2. METHODS

2.1. *Materials*

The materials were 6 short vignettes describing hypothetical events in English, Norwegian, or Bangla, modeled after the vignettes used by Machery et al. (2004), and inspired by Kripke's original G- and J-cases. Our G-cases G1 and G2 were as in Machery et al. (2004). We included a third G-case (G3) with a similar storyline structure, but content that may be more familiar to participants from Bangladesh. Our J-cases J1 and J2 were based on J-cases by Machery et al. (2004), but we shortened the text and we modified the task structure to render them more directly comparable to G-cases. We included a third J-case (J3) with similar structure, but content that may be more familiar to participants from Bangladesh. Our rationale for using 3, rather than just 2, stories was the following. In experiment 1, we aimed to increase the number of trials per condition, allowing for more reliable reconstructions of response patterns in participants. In experiment 2, we modified one probe from the G-cases and one from the J-cases (i.e., one from G1-G2-G3 and one from J1-J2-J3, counterbalancing across test versions), so that the vignettes were identical as in normal G- or J-cases, but the task was to answer a comprehension question (e.g., in G1, 'According to this story, what is the name of the man who actually proved the incompleteness of arithmetic?', followed by the names 'Gödel' and 'Schmidt'). In experiment 1 we used materials in English, while in experiment 2 we used translations into Bokmål Norwegian (Norwegian participants) and Bangla (Bangladeshi participants). In the Norwegian and Bangla translations, as in the English originals, probe answers must be expressed syntactically as definite NPs (except B in J-cases), to ensure they are interpreted semantically as definites, not just as specifics (all definites are specific, but the converse is not true; see Enç1991, p. 9 for a statement and a discussion). Norwegian has one way of expressing definiteness for unmodified nominals: the definiteness morpheme, in all varieties of Norwegian, as in other Scandinavian languages, is a postnominal suffix (Faarlund 2009). Bangla, instead, as some other Asian languages (e.g., Vietnamese, Hmong, and Cantonese—spoken by participants from Hong Kong; see above), has two ways of encoding definiteness: via bare classifiers or bare nouns (cf. Simpson et al. 2011 and Simpson & Biswas 2016 for examples and discussion). In the Bangla translations, we used bare nouns, which is the appropriate form if the referent is *unique*, as well as definite and specific (see Simpson & Biswas 2016 for an analysis; see Simpson & Biswas 2016 for empirical data). Bare nouns are used if referents are contextually or culturally unique, as in our stories. Bare classifiers impose the "condition that a referent (. . .) be indentifiable/identified in the mind

of the speaker/hearer" (Simpson & Biswas 2016): we avoided bare classifiers also not to bias stimuli in favor of SR responses. The (B) response in J-cases was translated using the appropriate indefinite form in Bokmål Norwegian and Bangla. The materials are provided as Supplementary Information.

2.2. Participants

In experiment 1, we collected data from three groups using materials in English: (i) participants born and living in Norway at the time of testing (NOR1); (ii) first-generation immigrants born in Nepal or Bangladesh, living in Norway at the time of testing (AIM1); and (iii) participants born and living in Bangladesh at the time of the study (BAN1). Each group included 25 participants (33 women, 15 in NOR1, 13 in AIM1, and 5 in BAN1), with mean age 27.7 years (NOR1: 25.4 years, SD = 2.96; AIM1: 29.7, SD = 5.54; BAN1: 27.9, SD = 3.05), 17.4 years of education (NOR1: 16.8, SD = 1.81; AIM1: 18, SD = 0.71; BAN1: 17.3, SD = 0.98), and a mean English proficiency score of 4.03 (NOR1: 4.3, SD = 0.68; AIM1: 3.8, SD = 0.47; BAN1: 4.02, SD = 0.64) as assessed by means of a questionnaire preceding the main task.

In experiment 2, we collected data from two groups with materials translated into Bokmål Norwegian or Bangla: (i) participants born and living in Norway at the time of testing (NOR2); (ii) participants born and living in Bangladesh at the time of testing (BAN2). Each group included 30 volunteers (37 women; 14 in NOR2, 23 in BAN2), with mean age 23.9 years (NOR2: 24.5, SD = 1.83; BAN2: 23.2, SD = 1.14) and 17.9 years of education on average (NOR2: 18.5, SD = 1.83; BAN2: 17.23, SD = 1.14). All NOR2 participants were native speakers of Norwegian, and all BAN2 participants were native speakers of Bangla. All participants included in the sample (NOR2, N = 30; BAN2, N = 30) could answer correctly both comprehension probes. The participant samples (NOR1/NOR2 and BAN1/BAN2) are disjoint (no one participated in both experiments).

2.3. Procedure

The task was administered to participants living in Norway or in Bangladesh via Google forms (experiment 1) or in controlled university classroom environments using printed paper questionnaires (experiment 2). Responses were not timed, and there was no time limit, but participants had to complete the task by responding to all vignettes in one session. Following the procedure in Machery et al. (2004), in both experiments, the order of the vignettes was randomized and counterbalanced across subjects. Moreover, in experiment 2, different test versions were constructed, in which the order of the vignettes and the order of the (A) or (B) responses were randomized and counterbalanced, within and across participants. Modified G- and J-vignettes, used as comprehension controls, were counterbalanced across test versions.

2.4. Data analysis

Participants' responses were analyzed as follows. We first coded each Kripkean (CH-type, or 'B') response as a 1 and each non-Kripkean ('A') response as a 0. Coded binary responses were used in logistic regression analyses, carried out separately for each experiment using a long data format (Baayen 2008) and the R package lme4. For experiment 1, we fitted a mixed effects logistic regression model with fixed factors Group (3 groups) and Case (G- or J-case) and the random factor Subject. Next, we tested the main effects of Group and Case via model comparison, i.e., the full mixed model *vs* models without the Group and Case

Table 1 Summary of descriptive statistics for experiments 1 and 2, reporting the mean (SD) number of Kripkean responses in G-cases, J-cases, and total, and the number of participants producing responses consistent with each theory of reference (abbreviated as SR, DD, CH, or none). The classification of participants by response profiles used pairwise similarity values between the response patterns predicted by each theory and the pattern produced by each participant (see Methods)

Group	G-cases	J-cases	Total	SR	DD	CH	None
NOR1	1.60 (1.38)	1.20 (1.38)	2.80 (2.60)	11	1	10	3
AIM1	0.84 (1.25)	0.68 (0.99)	1.52 (2.02)	18	1	4	2
BAN1	0.64 (1.08)	0.44 (0.96)	1.08 (1.96)	21	0	4	0
NOR2	1.13 (0.94)	0.60 (0.89)	1.73 (1.62)	14	0	9	7
BAN2	1.00 (0.83)	0.13 (0.35)	1.13 (0.90)	18	0	2	10

factors, respectively. Likewise, for experiment 2, we fitted a mixed effects logistic regression model with fixed factors Group (2 groups) and Case (G- or J-case) and the random factors Subject and Vignette (nested; in experiment 2, each random combination of 4 vignettes, out of 6 in total, was seen only by a subset of the participants). We then tested the main effects of Group and Case via model comparison, as for experiment 1.

As a further step in the analysis, we computed pairwise similarities between observed responses (binary vectors of length $L = 6$ in experiment 1; $L = 4$ in 2) and response patterns predicted by each of the three theories under scrutiny (SR = [000000], DD = [111000], and CH = [111111] in experiment 1; SR = [0000], DD = [1100], and CH = [1111] in 2) as 1-D, where D is the normalized Hamming distance. The output is 3 similarity values for each participant, expressing the similarity between the observed response pattern and the patterns predicted by SR, DD, and CH. We used these values for classification of the participants' response profiles: the largest similarity value (between the observed pattern and either SR, DD, or CH) determined whether each participant was classified by the corresponding response profile (SR, DD, or CH); if there was no unique largest value, the response profiles were classified as 'none' (Table 1; but see Discussion). Uncorrected p-values are reported in all cases. The data sets are provided as Supplementary Information.

3. RESULTS

The results of experiment 1 are reported in Table 1, Figure 1, and below. The descriptive statistics show that Kripkean responses are relatively infrequent in G-cases and J-cases, in all three groups. The NOR1 group is split into two subgroups of approximately equal sizes: one producing more frequent B-type (CH) responses, the other producing more A-type answers. This result agrees with the "dichotomous nature of the underlying distributions", also observed by Machery et al. (2004, p. B7). An analysis of response patterns across tasks shows that, in the AIM1 and BAN1 groups, a majority of participants (18 and 21, respectively; Table 1) produce response patterns that are more similar to the pattern predicted by the SR theory, with only 4 participants in each group producing a pattern similar to the one predicted by the CH theory. Instead, in NOR1, about as many participants produced patterns consistent with SR (11) and with CH (10). In all three groups, very few or no participants presented a pattern more similar to DD. The results of logistic
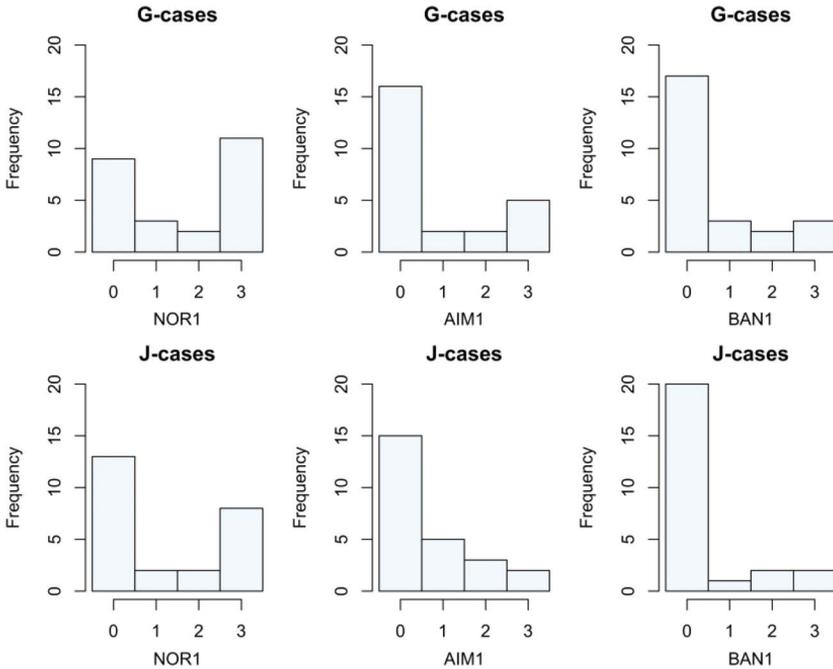
Figure 1 Distribution of the data from experiment 1. The histograms show the frequency of response patterns (i.e., the number of participants) with the number of Kripkean responses (range 0–3) given in the *x*-axis for G-cases and J-cases.

regression show that NOR1 participants are more likely to produce Kripkean (CH-type) responses overall than both BAN1 and AIM1 participants ($\beta = 2.78$, SE $= 1.53$, $z = 1.82$, $p = 0.0694$), and that participants across groups are less likely to give Kripkean responses in J-cases than in G-cases ($\beta = -1.14$, SE $= 0.37$, $z = -3.09$, $p = 0.002$). Moreover, we found main effects of Group ($\chi^2 = 9.4$, $p = 0.009$) and Case ($\chi^2 = 10.5$, $p = 0.0012$), confirming that Kripkean responses are relatively more frequent in the NOR1 group and in G-cases overall (Table 1).

Our results from experiment 1 replicate the original finding by Machery et al. (2004) that Westerners are more likely to produce responses consistent with the causal-historical account, and that Kripkean responses are more frequent in G-cases than in J-cases. Crucially, however, we also found evidence that the remaining responses—in fact, the majority of responses in all three groups—accord best with the SR pattern, and not with the descriptivist (DD) pattern.

Experiment 2 replicated all these findings (Table 1; Figure 2). The descriptive statistics show that CH responses are relatively infrequent in G- and J-cases, in both groups. The NOR2 group is split into two main subgroups. Responses across tasks show that, in NOR2 and BAN2, a majority of participants (14 and 18, Table 1) show patterns that are more similar to the one predicted by SR, with no participants in either group presenting a DD pattern. In NOR2, nearly a third of the participants (9) produce responses most consistent with the CH pattern, in contrast to only 2 in BAN2. NOR2 participants are more likely to produce Kripkean responses than BAN2 ($\beta = 0.78$, SE $= 0.32$, $z = 2.42$, $p = 0.0155$).
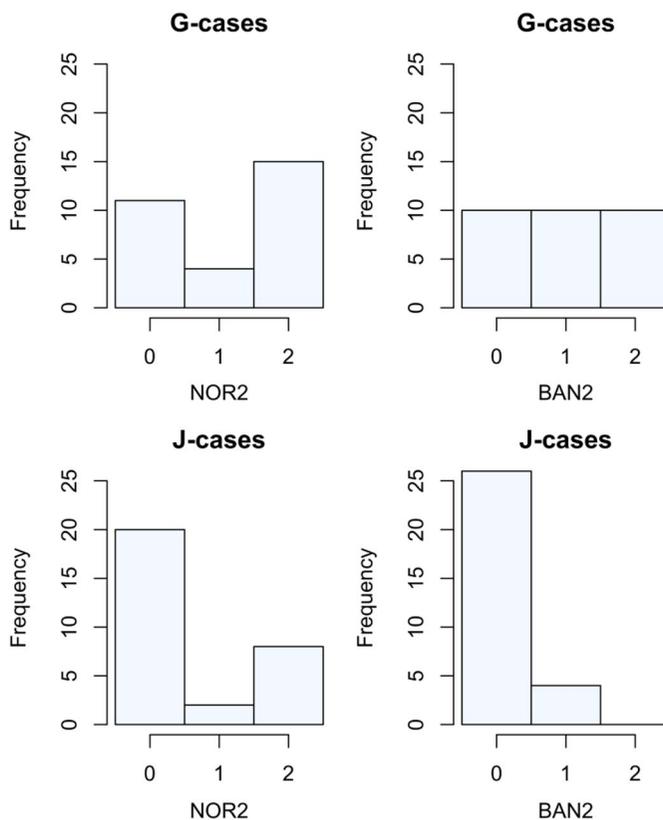
Figure 2 Distribution of the data from experiment 2. The histograms show the frequency of response patterns (i.e., the number of participants) with the number of Kripkean responses (range 0–2) given in the *x*-axis for G-cases and J-cases.

Kripkean responses are less frequent in J-cases than in G-cases across groups ($\beta = -1.71$, SE $= 0.38$, $z = -4.46$, $p < 0.0001$). We found differences between groups ($\chi^2 = 6.84$, $p = 0.009$) and between cases ($\chi^2 = 11.25$, $p = 0.0008$), confirming the picture emerging from Experiment 1.

4. DISCUSSION

In this study, we have provided evidence that: (i) Norwegians are more likely than Bangladeshis to produce Kripkean responses; (ii) participants are more likely to produce Kripkean responses in Gödel cases than in Jonah cases; (iii) the responses of individuals born and living in Bangladesh and the responses of South Asian immigrants living in Norway are comparable; (iv) the majority of responses across groups are consistent with speaker's reference.

We consider (i) an 'approximate replication' of Machery et al. (2004), using a different phrasing of probes in J-cases and an improved experimental design and procedure in experiment 2. This finding is supported by both descriptive and inferential statistics (logistic regression). Similar considerations apply to (ii). In spite of the robustness of the main result

(i), which holds for the entire data set (across case types), J-cases still elicit fewer Kripkean responses than G-cases. This difference cannot be due to the way the questions are framed in the G- and J-cases, since in our experiments those were structurally matched. Overall, both our experiments replicated the principal findings of Machery et al. (2004), while introducing certain desirable modifications to the materials and to the experimental design and procedure.

A second finding from experiment 1 is that the behavior of participants from South Asia appears consistent (iii), whether they are residents of Bangladesh or they have moved to Norway during their lives. The response distributions, descriptive statistics, and the similarity analysis of response patterns support this conclusion. This result, if confirmed by replications (and by experiments on Western immigrants in Asia), may indicate that cross-cultural differences, of the kind observed since Machery et al. (2004), are relatively impervious to late immersion in another cultural sphere (e.g., Western or European). Some behavioral or cognitive patterns may be established early on in childhood (Li et al. 2018), and may not be modifiable by education, enculturation, or other social influences. We refrain from speculating on this topic here. We did not aim to replicate this finding in experiment 2, and we accept that the nature of cross-cultural similarities or differences (e.g., exactly what makes Asians and Westerers respond differently in these tasks) is still poorly understood (see Nisbett 2004 and Lloyd 2007 for discussions). Importantly, we do not aim to generalize our results from Norwegians and Bangladeshis to Westerners and Asians, although our data are consistent with earlier results in other Western and Asian groups (see section 1).

More relevant for our purposes here is (iv). Non-Kripkean response patterns in Norwegians, and the vast majority of responses in Bangladeshis, should be regarded as being compatible with the SR analysis, *not* with the DD theory. In both experiments, this is supported by similarity-based analyses of response patterns and by logistic regressions: given the complementarity of the SR and CH response patterns, the conclusion that Norwegians are more likely than Bangladeshis to produce Kripkean answers trivially entails that Bangladeshis are more likely to produce answers fitting the speaker's reference pattern. In experiment 2, the subgroups of Norwegians and of Bangladeshis that did not behave according to either the CH, DD, or SR patterns were especially visible. This is again due to the fact that G-cases invite more Kripkean responses than J-cases, in both participant groups. But why is speaker reference more salient in J-cases than G-cases? None of the theories examined here seems capable of addressing this question, and further research is needed to clarify the nature of the semantic and pragmatic differences between G- and J-cases.

The theoretical distinction that would explain (iv) is not between supposedly 'competing accounts' of semantic reference, but between notions of semantic reference and speaker's reference—i.e., a within-theory distinction instead of a between-theories divide. If the across-task data analysis approach adopted here is correct, and if the underlying assumption holds (that what matters is the behaviour that participants show *consistently across trials*), then our data do *not* allow us to test or choose among competing theories of reference, and can be accommodated within a single unified account: i.e., a broadly Kripkean analysis of meaning that makes a distinction between semantic and speaker's reference. Such an account could capture the observed differences *within* the Norwegian group (i.e., the CH and SR subgroups) and *between* the Norwegian and the Bangladeshi groups. Other data sets, from experiments with different vignettes and probes, may require different analyses

and explanations of the results. As stated above, we suspend judgment on the causes of the observed effects, but the prevalence of SR-type responses in the Asian groups would be consistent with the idea that members of Asian cultures are more sensitive to *pragmatic factors* (Haberstroh et al. 2002) in tests of this kind than members of Western cultures (Cullen 2010). This presupposes that speaker's reference is understood as a pragmatic concept, not as a semantic concept. This view is endorsed by Kripke (1977) but is not unproblematic (see Lumsen 2010).

Our work supports the conclusion of Machery et al. (2004) that cross-cultural variation in reasoning about proper names exists, and specifically that the CH intuition is not universally shared. Our result that South Asian (Bangladeshi) participants predominantly give SR responses fits with their critique of this universality assumption. Yet, the claim that most Asians are descriptivists (p. B5) is undermined by our study. We remain agnostic as to whether formal or philosophical theories of reference may effectively be tested using empirical data of the kind reported here (Deutsch 2010; Ichikawa et al. 2012). Whether or not that is the case, semantic theories may and should be used to generate models of aspects of cognition or behavior (e.g., models of mental structure), and conversely, properly collected and analyzed data from valid experiments may guide philosophical or linguistic theorizing (Jackman 2009; Devitt 2011; Genone & Lombrozo 2012; Baggio et al. 2012a, 2012b, 2015, 2016; Machery 2017). For example, the theory of definite descriptions may not be viable as a theory *of proper names* (after all, it was not originally conceived by Russell as one), but it might still play a modeling role in the context of cognitive science, where the externalist stance implicit in causal-historical and rigid designator theories is problematic in several respects (Baggio 2018).

Acknowledgements

References

Baayen, H. (2008), *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baggio, Giosuè (2018), *Meaning in the Brain*. MIT Press. Cambridge, MA.

Baggio, Giosuè, Michiel van Lambalgen & Peter Hagoort (2012a), 'The processing consequences of compositionality'. In M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford University Press. 655–72.

Baggio, G., M. Van Lambalgen & P. Hagoort (2012b), 'Language, linguistics and cognition'. *Handbook of the Philosophy of Science* 14: 325–55.

Baggio, Giosuè, Michiel Lambalgen & Peter Hagoort (2015), 'Logic as Marr's computational level: four case studies'. *Topics in Cognitive Science* 7: 287–98.

Baggio, Giosuè, Keith Stenning & Michiel van Lambalgen (2016), 'Semantics and cognition'. In M. Aloni and P. Dekker (eds.), *The Cambridge Handbook of Formal Semantics*. Cambridge: Cambridge University Press. 756–74.

Beebe, James R. & Ryan J. Undercoffer (2015), 'Moral valence and semantic intuitions'. *Erkenntnis* 80: 445–66.

Beebe, James R. & Ryan Undercoffer (2016), 'Individual and cross-cultural differences in semantic intuitions: new experimental findings'. *Journal of Cognition and Culture* 16: 322–57.

Cullen, Simon (2010), 'Survey-driven romanticism'. *Review of Philosophy and Psychology* **1**: 275–96.

Deutsch, Max (2009), 'Experimental philosophy and the theory of reference'. *Mind & Language* **24**: 445–66.

Deutsch, Max (2010), 'Intuitions, counter-examples, and experimental philosophy'. *Review of Philosophy and Psychology* **1**: 447–60.

Devitt, Michael (2011), 'Experimental semantics'. *Philosophy and Phenomenological Research* **82**: 418–35.

Domaneschi, Filippo, Massimiliano Vignolo & Simona Di Paola (2017), 'Testing the causal theory of reference'. *Cognition* **161**: 1–9.

Enç, Mürvet (1991), 'The semantics of specificity'. *Linguistic Inquiry,* **22**, 1. 1–25.

Faarlund, Jan Terje (2009), 'On the history of definiteness marking in Scandinavian'. *Journal of Linguistics* **45**: 617–39.

Genone, James (2012), 'Theories of reference and experimental philosophy'. *Philosophy Compass* **7**: 152–63.

Genone, James & Tania Lombrozo (2012), 'Concept possession, experimental semantics, and hybrid theories of reference'. *Philosophical Psychology* **25**: 717–42.

Haberstroh, Susanne, Daphna Oyserman, Norbert Schwarz, Kühnen Ulrich & Li-Jun Ji (2002), 'Is the interdependent self more sensitive to question context than the independent self? Self-construal and the observation of conversational norms'. *Journal of Experimental Social Psychology* **38**: 323–9.

Heck, Richard G. Jr. (2018), 'Speaker's reference, semantic reference, and intuition'. *Review of Philosophy and Psychology*, **9**, 2. 1–19.

Ichikawa, Jonathan, Ishani Maitra & Brian Weatherson (2012), 'In defense of a Kripkean dogma'. *Philosophy and Phenomenological Research* **85**: 56–68.

Izumi, Yu, Masahi Kasaki, Yan Zhou & Sobei Oda (2017), 'Definite descriptions and the alleged east–west variation in judgments about reference'. *Philosophical Studies* **175**: 1–23.

Jackman, Henry (2009), 'Semantic intuitions, conceptual analysis, and cross-cultural variation'. *Philosophical Studies* **146**: 159.

Kripke, Saul (1980), *Naming and Necessity*. Harvard. Cambridge, MA.

Kripke, Saul (1977), 'Speaker's reference and semantic reference'. *Midwest Studies in Philosophy* **2**: 255–76.

Lam, Barry (2010), 'Are Cantonese-speakers really descriptivists? Revisiting cross-cultural semantics'. *Cognition* **115**: 320–9.

Li, Jincai, Longgen Liu, Elizabeth Chalmers & Jesse Snedeker (2018), 'What is in a name?: The development of cross-cultural differences in referential intuitions'. *Cognition* **171**: 108–11.

Lloyd, Geoffrey (2007), *Cognitive Variations: Reflections on the Unity and Diversity of the Human Mind*. Oxford: Oxford University Press.

Lumsen, David (2010), 'The relationship between speaker's reference and semantic reference'. *Language and Linguistics Compass* **4**: 296–306.

Ludwig, Kirk (2007), 'The epistemology of thought experiments: First person versus third person approaches'. *Midwest Studies in Philosophy* **31**: 128–59.

Machery, Edouard (2017), *Philosophy Within its Proper Bounds*. Oxford: Oxford University Press.

Machery, Edouard, Ron Mallon, Shaun Nichols & Stephen P. Stich (2004), 'Semantics, cross-cultural style'. *Cognition* **92**: B1–B12.

Machery, Edouard, Christopher Y. Olivola & Molly De Blanc (2009), 'Linguistic and metalinguistic intuitions in the philosophy of language'. *Analysis* **69**: 689–94.

Machery, Edouard, Max Deutsch, Ron Mallon, Shaun Nichols, Sytsma Justin & Stephen P. Stich (2010), 'Semantic intuitions: reply to Lam'. *Cognition* **117**: 361–6.

Machery, Edouard & Stephen P. Stich (2012), 'The role of experiment'. In D. G. Fara and G. Russell (eds.), *The Routledge Companion to the Philosophy of Language*. Routledge. New York. 495–512.

Machery, Edouard, Justin Sytsma & Max Deutsch (2015), 'Speaker's reference and cross-cultural semantics'. In A. Bianchi (ed.), *On Reference*. Oxford University Press. Oxford. 62–76.

Mesoudi, Alex, Kesson Magid & Delwar Hussain (2016), 'How do people become WEIRD? Migration reveals the cultural transmission mechanisms underlying variation in psychological processes'. *PloS One* **11**: e0147162.

Nisbett, Richard (2004), *The Geography of Thought: How Asians and Westerners Think Differently... and Why*. New York, Simon and Schuster.

Russell, Bertrand (1905), 'On denoting'. *Mind* **14**: 479–93.

Simpson, Andrew, Hooi Ling Soh & Hiroki Nomoto (2011), 'Bare classifiers and definiteness: a cross-linguistic investigation'. *Studies in Language* **35**: 168–93.

Simpson, Andrew & Priyanka Biswas (2016), 'Bare nominals, classifiers, and the representation of definiteness in Bangla'. *Linguistic Analysis* **40**: 167–98.

Sytsma, Justin & Jonathan Livengood (2011), 'A new perspective concerning experiments on semantic intuitions'. *Australasian Journal of Philosophy* **89**: 315–32.

Sytsma, Justin, Jonathan Livengood, Ryoji Sato & Mineki Oguchi (2015), 'Reference in the land of the rising sun: A cross-cultural study on the reference of proper names'. *Review of Philosophy and Psychology* **6**: 213–30.

Vignolo, Massimiliano & Filippo Domaneschi (2017), 'Referential intuitions are still problematic'. *Analysis* **78**: 472–83.