Priming and Conceptual Pacts in Overhearers' Adoption of Referring Expressions

Abstract

Current theories of communication yield predictions about the expression choice of overhearers as well as primary discourse participants. We discuss three such theories and evaluate them with reference to new data on object naming elicited through a confederate priming paradigm. Our results show that participants adopt primed referring expressions if they are highly involved in the task, but mere exposure to the object labels yields very limited priming effects. Also, common ground is a relatively marginal factor in expression choice here. We interpret these results as supportive of the importance of grounding and challenging for interactive alignment-based accounts of expression choice.

Introduction

Inter-personal communication is customarily taken to involve processes of cooperation and coordination between interlocutors at a number of levels. At the level of the conversational turn, speakers cooperate with hearers by making their contribution appropriate to the current purpose of the talk exchange, as observed by Grice (1975). They also coordinate with hearers by making their intentions understood (Grice 1957). At a discourse level, speaker and hearer work together to achieve conversational goals, which might involve the sharing of information, the making and satisfying of requests, the formation of joint plans, etc. (Clark 1996).

A diverse range of theories have been proposed to account for how speakers and listeners successfully engage in this process of communication, with particular reference to dyadic interactions such as dialogues. An influential account of dialogue, the interactive-alignment model (Pickering & Garrod 2004), places low-level processes of priming at the heart of communication. In this account, interlocutors align their representations as a result of dialogue. This alignment commences at a surface level, in that the dialogue participants converge at a lexical and syntactic level, due to the priming effects exerted by the use of particular words and syntactic forms. The resulting alignment then percolates up through the system, eventually reaching the level of situation models. The goal of communication, on this account, is to accomplish the alignment of situation models. However, the fundamental drivers of this are low-level, automatic and unconscious processes, specifically priming processes. Consequently, this account posits little involvement of strategic factors in the success of dyadic communication.

A contrary viewpoint is that interlocutors are highly aware of each other's mental states and that this awareness informs their behaviour. Clark and Schaefer (1989) argue that successful contributions to a discourse requires grounding; that is, speaker and hearer must mutually believe that the speaker's meaning has been understood. The notion of common ground (CG) – the shared knowledge, beliefs and assumptions of the interlocutors (Clark & Marshall 1981) – thus becomes relevant here. The goal of the interaction involves building and updating CG, and doing this requires consideration of the existing CG state. A simple example is the use of a referring expression: if a speaker predicates a new property of an entity (e.g. "John is away"), this can only be successful as a discourse contribution if the hearer correctly identifies the entity. This requires the speaker to take account of the hearer's knowledge about how this entity is labelled. Such knowledge may be presumed on the basis of linguistic community membership, but it may also arise from previous referential success, or on the basis of the formation of "conceptual pacts" (Brennan & Clark 1996). Distinctively, conceptual pacts involve the establishment of partner-specific labels for entities, which can then be successfully used in interaction with that specific partner but are not preferred for general use with other interlocutors.

An intermediate position between these two viewpoints is occupied by Keysar (2007). He argues that "when people communicate, they do not routinely take into account the mental states of others" (ibid., p.72). Instead, drawing upon evidence from Theory of Mind experiments, he argues for the primacy of egocentric processing, and contends that "one's own perspective is dominant…the consideration of others' beliefs is not automatic" (ibid., p.75). Unlike the interactive-alignment model, this approach entails conscious reasoning about the choice of referring expression, but unlike Clark and colleagues, Keysar considers CG to be a relatively peripheral issue, and the role of the hearer and his/her mental state to be a marginal factor in the speaker's choice of expression. In support of this, Barr and Keysar (2002) provide experimental evidence that people (unconsciously) expect new conversational partners to adhere to conceptual pacts that have previously been established, even though the new partner is not privy to this pact. This in turn suggests that conceptual pacts are not triadic relations between two interlocutors and an entity, in which both agree to refer to this entity in a particular way within their interaction, but rather pairs of relations in which both parties separately agree to refer to this entity by a particular label.

Experimental work on dialogue has been conducted from numerous theoretical perspectives, including those outlined above. However, relatively little attention has been paid in this literature to non-dyadic interactions, for instance those in which a third individual is present but not directly engaged in the conversation[1]. In this paper, we aim to extend findings about the choice of referring expression into the domain of non-dyadic interactions. There are several motivations for this move. First, the potential relevance of third parties in conversation has long been acknowledged (see Clark & Carlson 1982) but the implications of this for expression choice have attracted relatively little attention. Secondly, such

---

[1] A partial exception to this is Keysar and Henly (2002), but their primary focus is on using overhearers to evaluate the likely communicative success of utterances, rather than on examining the effect of dialogue on overhearers themselves.

situations are common in everyday interaction, and understanding the dynamics of conversation in such settings is an end in itself. Thirdly, and perhaps most importantly, non-dyadic interactions represent a testing ground in which the factors governing expression choice can be disentangled to a certain extent, thus offering useful insights as to the relative strengths of the competing factors.

In the following section, we consider how the competing theories of Pickering and Garrod, Clark, and Keysar naturally yield distinctive predictions about the behaviour of overhearers in a non-dyadic setting. We then introduce an experiment to test the effects attributable to priming, egocentricity and conceptual pacts, specifically examining whether and under what conditions overhearers select referring expressions according to their status in a preceding dialogue.

Critical Predictions about Overhearer Behaviour

Although the specific accounts discussed in the previous section are primarily oriented towards explaining dyadic interactions, the mechanisms that they posit should apply also in non-dyadic interactions. If so, predictions can be drawn about overhearer behaviour, as we articulate in the following paragraphs. Of course, the falsification of such a prediction would not imply the incorrectness of the theory in the dyadic case. However, it would suggest that additional machinery would need to be posited to cover non-dyadic interactions. We would interpret it as favourable for a theory if it makes correct predictions about both types of interaction without further stipulation.

In particular, we focus on a specific scenario of non-dyadic interaction. In this scenario, two interactants are playing a game in which they match picture cards that display tangram figures. Both have matching packs of cards. One of them (the 'director') selects a card and describes the figure, and the other (the 'matcher') has to identify which card is being talked about. There is also an overhearer, who does not participate in the game. When the game is completed, the overhearer plays the game, taking the role of director. The question is whether, and to what extent, the overhearer will re-use the descriptions that were used by the original director in the previous phase of the game.

Turning first to Pickering and Garrod's (2004) account, priming is predicted to occur automatically upon exposure to the relevant labels. They predict stronger alignment effects for addressees than overhearers (ibid., 174), on the basis that the former engage their production systems during the interaction (anticipating that they will speak at some point) whereas the latter do not need to. However, overhearers are still expected to exhibit some priming effects. Crucially, this does not depend upon the establishment of full common ground, which is argued only to occur "when radical misalignment becomes apparent" (ibid., 179). Rather, it relies merely on implicit common ground, defined as the information shared between the interlocutors, to which the overhearer might reasonably be supposed to have access. Hence, in this experimental paradigm (where there are no observable failures in communication), their account predicts priming of overhearers, a possible effect of involvement, and no effect of common ground. It further predicts that priming will be boosted if the overhearer's production system is activated.

Contrastingly, for Clark and colleagues, high-level conscious processes are critical to determining whether the overhearer adopts the referring expressions that have been used. These expressions should be used only if they have been observed to be successful, which entails that the overhearer is sufficiently engaged in the dialogic process to determine whether this is the case: merely hearing the expressions will not do. In particular, where conceptual pacts have been formed, the status of the addressee with respect to these pacts should also be relevant. When addressing someone who was involved in the conceptual pact (in our scenario, someone involved in the first phase of the game), the former overhearer is predicted to re-use the established referring expression to a greater extent than they would if addressing a new individual. So in brief, this account predicts no priming unless the overhearer is sufficiently involved in the discourse, and more priming when common ground is also present.

The predictions arising from Keysar's (2007) account differ from Clark's with respect to common ground. According to Keysar, the choice of expression should be egocentrically motivated in the first instance, and therefore it should be irrelevant whether or not the hearer has a prior conceptual pact about that referent. Whether the overhearer should adopt any of the expressions used by the previous director is not clear on this account: as Keysar does not posit a role for low-level priming, this should not occur automatically, although it might be feasible for the overhearer to learn new labels under certain conditions (see General Discussion). In sum, we take this account to predict no priming unless the overhearer is sufficiently involved in the discourse, but no effect of common ground.

Establishing Baseline Naming Probabilities

In order to establish our baseline naming probabilities, we ran an online elicitation study with EFS Survey (http://www.unipark.info). Participants were asked to provide names for 50 configurations of tangram pieces, which were presented in silhouette. One tangram was presented per page, with the task being to give a name to the presented graphical display (no information was given as to whether the display was a picture or representation of a specific object).

331 participants were recruited via the University's mailing list and leaflets around the campus. All were students and native speakers of German. Participants were entered into a prize draw to win €10 cash or one of 10 €10 Amazon vouchers.

For the following experiments, we selected 15 tangrams for each of which a specific response had occurred at rates of 5-15%. These responses could be considered plausible but dispreferred, in that they were neither unique to an individual respondent nor were they the 'obvious' description for the tangram in question. The

use of such items in the following experiments reduces the probability that the participant selects the target expression just by chance. The mean rate of usage for the relevant descriptions across these 15 tangrams was 8.29% (278/3353).

Experiment 1: Effects of Involvement and Common Ground

Experiment 1 was designed to show whether an overhearer's involvement in the interaction, and the extent to which they shared common ground with their subsequent addressee, influenced their uptake of dispreferred referring expressions.

Participants

86 participants (47 female), all native speakers of German, were paid for participation in the experiment. They were divided randomly between the four test conditions.

Materials

Three sets of 15 white cards (74 x 105mm) were used, each with a black tangram on the upper half of the card. The confederate director and the matcher each had one set of cards. Each set of cards showed identical figures: the confederate director's cards also showed the names that were to be used for the tangrams. Video and audio recordings were made of each trial.

Procedure

For each condition, the experimental setting comprised an interacting dyad of director and matcher, plus an overhearer. All three individuals were separated by opaque screens. They were instructed that they were to play a game in which the object was to match the order of 15 cards. The director's cards were arranged in a stack in the correct order, while the matcher's cards were arranged on the table top and all were visible. Both were instructed not to change the orientation of the cards.

The director was instructed to proceed by naming the card on top of the pile so that the matcher could find the corresponding card, using names that were as short and spontaneous as possible but as long as necessary. Matchers were allowed to ask for additional descriptions but were told that they would lose points for doing so. These instructions were devised to avoid the use of detailed descriptions rather than impressionistic names for the tangrams.

The experiment proceeded in two phases, using the confederate priming paradigm. In the first phase, the director was a confederate and used pre-specified descriptions (chosen from the pre-test results as discussed above). The matcher was also a confederate, and the experimental participant was the overhearer. In the second phase, the game was played again, with the participant now playing the role of director, but having no direct access to the list of descriptions that had previously been used.

A 2 x 2 design was used, within which levels of involvement and common ground were manipulated. In the high-involvement conditions, overhearers were

presented with a sheet displaying all 15 tangram shapes prior to the first matching phase of the experiment. In the low-involvement conditions, overhearers were not shown the shapes that were being discussed. Instead, they were asked to count the number of times that /t/ was uttered during the interaction (cf. Bavelas, Coates & Johnson 2000), in order to ensure that they were attending to the linguistic material being uttered. In the high common ground condition, the confederate who was the director in phase 1 of the experiment became the matcher in phase 2 of the experiment, whereas in the low common ground condition, a new confederate who had not participated in phase 1 of the experiment was the matcher in phase 2.

The transcript of phase 2 was analysed in order to establish whether the participant preferentially re-used descriptions that had been used by the confederate in phase 1. The participant's descriptions were considered according to two criteria: a strict lexical priming criterion, in which only identical or similar words (modulo morphosyntactic alternations) were considered to 'match', and a more liberal semantic criterion, in which expressions of similar concepts and synonyms were also considered as matches.

Results

The results are summarised in Table 1, for semantic priming, and Table 2, for lexical priming.

Table 1: % semantic priming effects in Experiment 1

Condition	- CG	+ CG
- involvement	14.8 (49/330)	15.3 (46/300)
+ involvement	41.0 (129/315)	50.4 (174/345)

Table 2: % lexical priming effects in Experiment 1

Condition	- CG	+ CG
- involvement	13.0 (43/330)	14.3 (43/300)
+ involvement	36.2 (114/315)	47.0 (162/345)

In all conditions, the use of primed expressions was significantly higher than their rates of spontaneous use in the pre-study (binomial, all $p < 0.001$). We applied a logistic mixed model with full random slopes to the semantic priming results. This showed a highly significant main effect of involvement ($\beta = 1.88$, SE = 0.259, Z = 7.28, $p < 0.001$), but the main effect of common ground did not reach significance ($\beta = 0.255$, SE = 0.203, Z = 1.26, $p = 0.21$), despite the numerical trend in the high involvement condition. In a second model we also posited an interaction term, but this did not reach significance ($\beta = 0.58$, SE = 0.411, Z = 1.4, $p = 0.16$), while involvement remained significant and common ground non-significant. This pattern of effects was replicated for the lexical priming results.

Discussion

The results of Experiment 1 indicate that the degree of the participants' involvement is highly relevant to their uptake of dispreferred referring expressions. In the conditions in which overhearers were allowed to see the set of tangram figures, they were effective at acquiring the labels used in phase 1 of the experiment. When they were not allowed to see the figures, they exhibited much smaller priming effects, using the primed labels only slightly more frequently than would have been expected in spontaneous, unprimed description.

The presence of common ground led to numerically more frequent reuse of primed descriptions, but this effect did not reach significance in our sample after subject and item effects were taken into consideration.

We interpret these results as potentially supportive of the positions of Clark or Keysar. In particular, it is not the case that overhearers frequently use dispreferred descriptions just as a consequence of having heard these object labels; they must also be aware of the referent picked out by the label. In the terminology of Clark and Brennan (1991), the use of the label must be "grounded". It could of course be argued that the overhearers in our experiment do not have the opportunity to ground the labels with certainty, even in the high involvement conditions, as they cannot be sure which referent is picked out by which expression. Nevertheless, the results suggest that our participants were generally adept in solving this mapping problem, and having done so, used this information to inform their choice of referring expression.

These results can be reconciled with the account of Pickering and Garrod (2004) if we assume that the participant in the high involvement condition is sufficiently engaged in the discourse to have an activated production system, making them effectively a discourse participant rather than merely an overhearer. From that perspective, we could see these results as indicative of the degree of involvement that is required in order for the third individual to be subject to substantial priming effects. On this account, although the priming effect still persists in the absence of the referents (in that primed expressions are used at above-baseline rates), it is very much weakened.

The lack of a strong effect of common ground speaks in favour of the egocentric view proposed by Keysar and colleagues. However, the trend towards greater reuse of priming expression to familiar interlocutors in the high involvement condition suggests that some participants may be influenced by the existence of a prior conceptual pact. If this were the case, it would challenge both the egocentric account and the assumption of Pickering and Garrod (2004) that common ground is only consulted when there is some kind of difficulty in the dialogue, such as deceit or extensive repair. Further work is required to confirm or exclude the existence of this trend.

Experiment 2: Task-Specific Effects

A question arising from the first experiment is whether the manipulation of involvement also influenced the participants' expectations about their task. Could it be the case that the participants who were presented with a copy of the tangram pictures inferred that their task was to learn how to describe these images? This could in turn result in greater activation of their production mechanisms, predicted by Pickering & Garrod (2004) to lead to greater priming effects. To address these possibilities, we conducted a further experiment in which the overhearers were not given access to pictures of the tangrams, similarly to the original low involvement condition, but were told that after the first part of the experiment, they would then be playing the game, in the role of director.

Participants

41 participants (27 female), all native speakers of German, and none of whom participated in Experiment 1, were paid for participation in the experiment. They were assigned to the two new test conditions (common ground and no common ground, as in Experiment 1).

Materials

The same materials were used as in Experiment 1.

Procedure

The same procedure was used as in the low involvement condition of Experiment 1, with the exception that the participants were not asked to perform t-counting, on the basis that this might interfere with their ability to follow the task (and potentially the engagement of their production systems). Instead, they were instructed to listen to what was going on and told that they would be asked about how successful the interaction had been. Before the experiment began, participants were told that they would be taking the role of director in the second part of the experiment.

Results

The results are summarised in Table 3.

Table 3: Results of Experiment 2

Condition	Semantic priming %	Lexical priming %
- CG	16.0 (48/300)	13.0 (39/300)
+ CG	15.9 (50/315)	15.2 (48/315)

In both conditions, the use of primed expressions was significantly higher than their rates of spontaneous use in the pre-study (binomial, both p < 0.001). Comparing these results with the low involvement conditions of Experiment 1, logistic regression analyses showed no significant main effect of task awareness.

Discussion

The results of Experiment 2 suggest that awareness of the potential usefulness of the descriptions that are employed does not suffice, on its own, to enable the overhearer to pick up dispreferred expressions in this paradigm. Without access to depictions of the referents, the participants in this experiment exhibited very limited evidence of priming effects. This suggests that the higher rates of priming

attested in the high involvement condition of Experiment 1 are largely attributable to the perceptibility of the figures, rather than the participants drawing any specific inferences about the way in which they were expected to perform the task.

General Discussion and Conclusions

Our experiments strongly suggest that overhearers are able to acquire dispreferred labels for objects, but that they do so to a very limited degree if they do not have perceptual access to the object that is being referred to.

We take these results to point to limitations in the power of 'pure priming' effects; that is, the view that access to the phonetic content of labels will lead to their adoption by hearers, as a consequence of percolation (Pickering & Garrod 2004). In the confederate priming paradigm, it appears that such access is not enough: the label must also be associated with an object in order for it later to be adopted. This suggests that the process by which speakers align on object labels is not merely bottom-up, but requires the presence of a referent or meaning as well as the verbal label.

Nevertheless, it could be argued that our results do point to non-zero priming effects, with increased uptake of primed expressions even among uninvolved overhearers who do not see the potential referents and are not attending to the dialogic process that is occurring. Such effects could indeed be attributable to the type of processes that Pickering and Garrod (2004) posit. However, at least in this paradigm, these effects are much smaller than the priming effects in the high involvement condition.

A possible explanation of the effect of involvement, within the Pickering and Garrod account, is that the overhearers' production mechanisms are more highly activated in the high involvement condition. Given the results of Experiment 2, we consider this unlikely to be the sole cause of the involvement effect. The results of experiment 2 suggest that, even when participants are explicitly informed that they will later be called upon to describe the same figures, and hence might be assumed to engage their production systems in preparation to participate in a dialogue, they do not exhibit greater uptake of the primed expressions.

Of course, it may be the case that the effect of involvement is a matter of attention, and that overhearers in the low involvement condition are less engaged in the task in general. However, in experiment 1, these overhearers are obliged to attend to the phonetic content of the utterances, which should in principle be sufficient to initiate priming effects via percolation. This explanation might be tenable if we modify the 'pure priming' account to require that lexical items must be heard and understood in their entirety in order to be primed.

The strong effect of involvement is straightforwardly explicable in Clark's approach: according to this view, expressions are re-used as a result of their observable effectiveness in the prior interaction, and it is the highly involved participants who are in a position to discern this. However, it can also be captured by the egocentric approach, articulated by Keysar and colleagues. Here we must also posit that the condition of high involvement –

which presented participants with a visual representation of the tangram figures being talked about – enabled the overhearers to learn the referring expressions corresponding to (some of) these figures. Given that there were no 'right' or 'wrong' answers in our experiment, it is perhaps slightly counter-intuitive that 'egocentric' overhearers should bother to learn the names of tangrams, when they could simply describe them as they saw fit. It is possible that learning the names in this way represents an economical strategy that obviates the need for any decisions about how to describe the tangrams later on (although participants in Experiment 1 were not told that they would need to do so). We cannot, therefore, exclude the possibility that hearing the primed names in the high involvement condition merely shaped the egocentric preferences of the overhearers, and that this was later manifested in their choice of expression.

Our experiments documented a numerical tendency towards common ground effects, but this might be attributable to random variation. If this effect is replicated in further research, it would more seriously challenge Keysar's (2007) claim that the speaker's choice of referent should initially be egocentrically motivated, irrespective of conceptual pacts. Note that, in this experiment, there were no failures of communication (as the matcher was a confederate), hence there was no need for the director to reformulate his or her utterance: purely egocentric behaviour would, to all intents and purposes, have done just as well. The preferential reuse of primed expressions when the matcher was familiar would suggest that awareness of conceptual pacts may, at least for some speakers and on some occasions, be influencing the initial choice of utterance.

In short, our results so far do not permit us to exclude the possibility of egocentrism on the part of our participants, and can be reconciled with a slightly modified version of the form-based priming account of Pickering and Garrod (2004). Nevertheless, the results appear to fit most naturally with the viewpoint articulated by Clark and colleagues. Specifically, in order for expressions to be adopted, it appears to be broadly necessary for overhearers to understand the purpose of the expressions or to experience them being used effectively (our experiments do not distinguish these possibilities). Crucially, it is not sufficient merely to hear expressions that are not grounded, even if you know that you will be called upon to produce similar expressions in the future. However, two aspects of these results are unexpected from the perspective of Clark's approach: firstly, as discussed above, the effect of conceptual pacts is, at most, marginal in these experiments. A definitive absence of such effects would speak in favour of Keysar's view. Secondly, expressions can (occasionally) be picked up spontaneously by uninvolved overhearers without awareness of the current discourse goals, as predicted by the model of Pickering and Garrod. Further investigations might show whether the apparent examples of this in our data are actually attributable to the overhearer attending to the discourse, or whether they should be treated as genuine instances of automatic priming effects that are not predicted by Clark's theory.

References

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391-418.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*, 941-952.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1482-1493.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (eds.), *Perspectives on Socially Shared Cognition*. Washington DC: APA Books.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*, 332-373.

Clark, H. H., & Marshall. C. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (eds.), *Elements of Discourse Understanding*. New York: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259-294.

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*, 377-388.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and Semantics*, vol. 3. New York: Academic Press.

Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, *4*, 71-84.

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13, 207-212.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.