REGISTER
Aspects of
Language
in Situation

The variation of strategic politeness in indirect requests

Roland Mühlenbernd, Mariya Burbelko, Uli Sauerland & Stephanie Solt
Leibniz-Zentrum Allgemeine Sprachwissenschaft

This study empirically examines the register effects of various politeness strategies in English requests. It is based on the premise that linguistic politeness is influenced by specific situational factors, such as the power difference between interlocutors or the degree of imposition of the request. To investigate this, the study compares experimental production data from a request scenario represented in 2 × 3 different experimental conditions, varying in three levels of imposition and two types of power relations between interlocutors. The data were collected through an online experiment with 207 native US English speakers. Next to the main production task, the experiment comprised several follow-up tasks designed to contrast the participants' request formulations with potential alternatives. The goal of the experiment was to identify common politeness strategies, and how they vary across situational factors, as well as to detect motives for (dis)preferring alternative strategies. As key findings, we found out (i) that the variation of two types of politeness strategies – so-called conventional vs. non-conventional indirect requests – strongly correlates with the experimental conditions, and (ii) that potential choice motives strongly differ between these two types.

Keywords: politeness, indirect requests, rank of imposition, power, production experiment

1 Introduction

When a speaker utters a request, they ask someone for something, or they ask someone to do something. In other words, they produce a speech act, that is, they not only present information but perform a communicative act, namely the act of requesting. For example, when a speaker says something such as (1a), they request the listener to close the door. (1a) is a direct request, where the requestive meaning is part of the literal meaning. However, we often find requests, which are not uttered as imperatives, but, for example as questions, such as (1b), (1c),

BerlinUP Journals
Roland Mühlenbernd and Mariya Burbelko and Uli Sauerland and Stephanie Solt. 2026. The variation of strategic politeness in indirect requests. *Register Aspects of Language in Situation (REALIS)*. 4(2), pp. 1–36, doi=10.71717/8h0va035
ISSN 2750-9370

and (1d), or declaratives, such as (1e). Here, the requestive meaning is not part of the literal meaning. For example, the meaning of (1b) is that the speaker asks the listener for their ability to close the door. However, this utterance is generally understood as a request, although the requestive meaning is only indirectly given. Utterances such as (1b) are called indirect requests.[1]

(1) a. Close the door.
b. Can you close the door?
c. Could you close the door?
d. Would it be possible for you to close the door?
e. It is cold in here.

But why do speakers formulate requests indirectly in the first place? Following Brown & Levinson's (1987) Politeness Theory, requests pose a threat to the addressee's negative face since the speaker wants them to perform an action or favor, and therefore restricts their freedom of actions. In other words, requests are face-threatening acts (FTA), and speakers often seek politeness strategies to mitigate the face-threat on the addressee (Brown & Levinson 1987). One very common strategy is to formulate the request indirectly, such as given in the examples (1b–e) (cf. Blum-Kulka 1987).

Indirect requests have been further classified into different types (e. g. Blum-Kulka 1987: 133). For example, (1b–d) are indirect requests that refer to a preparatory condition for the feasibility of the request (Searle 1969), so-called query preparatory questions. It has been shown that this request type is rated as more polite than other even more indirectly phrased request types, such as hints, as given in (1e) (Blum-Kulka 1987). Moreover, previous research has shown that requests such as (1b) and (1c) have a higher level of conventionalization[2] in comparison to requests such as (1d), and are therefore denoted as *conventionally indirect requests* (CIR). For different languages, CIRs are highly frequently produced types of requestive strategies (e. g. Blum-Kulka & House 1989, Ruytenbeek 2019b, Kranich et al. 2021, Ackermann 2023). In addition, CIRs have a high degree of explicitness, which represents "the degree to which the speaker's illocutionary

[1]Note: Politeness markers such as *please* can function as pragmatic cues that favor a request interpretation over a literal ability reading in examples like (1b), e. g. *Can you please close the door?*

[2]For the given phenomenon, conventionalization (also termed as *standardization*) is defined as the diachronically increasing association of a specific form with its indirect request meaning (cf. Clark 1979, Kasper 1995).

intent is apparent from the locution" (Blum-Kulka et al. 1989); higher than the so-called *non-conventionally indirect requests* (NCIR), as given in (1d). Due to their directness, CIRs have been characterized as less polite than NCIRs such as (1d) (cf. Ruytenbeek 2019a, 2021).

When formulating a request, speakers often produce more than solely the actual request utterance (also called the *head act*), but a series of utterances, denoted as a *speech act set* (Murphy & Neu 2006), such as given in (2).

(2)

Could you	possibly close the window?	It's pretty cold in here.
HEAD ACT	INT. MOD.	EXTERNAL MODIFIER

Here, the expression *possibly* is denoted as an *internal modifier*, since it operates on the head act *Could you close the window?*, where it functions as a downtoner. The additional phrase *It's pretty cold in here.* is denoted as *external modifier* and functions as a grounder (explanations/justifications for requests). It has been shown that speakers employ diverse combinations of politeness strategies to formulate requests, including not only the choice of the head act strategy but also the choice of internal and external modifications (Blum-Kulka et al. 1989).

As mentioned in the beginning, politeness strategies are used by the speaker to mitigate the face threat on the addressee. According to Brown & Levinson (1987)'s politeness theory, the weight of a face threat is affected by three situational factors. The first factor is *power*, which denotes the status relationship between the speaker and the addressee. The second factor is the *rank of imposition*, which reflects the fact that some impositions are considered more serious than others. This is particularly relevant for requests, where the rank of imposition is determined by the magnitude of the deed or favor that is requested. The third factor is the *social distance* between interlocutors, determining their degree of unfamiliarity, ranging from a very close person to a stranger. According to Brown & Levinson (1987)'s politeness theory, each of these variables increases the weight of the face-threatening act in the following way:[3]

- weight FTA = power difference + rank of imposition + social distance

Furthermore, under the assumption that an increased weight calls for a stronger need to mitigate the face threat of a request, we would expect a positive correlation of these situational factors with the level of politeness within a

[3]Power difference stands for addressee's status minus speaker's status. In other words, the higher the addressee's status in relation to the speaker's status, the higher the weight of the face threat, all else being equal.

request. In the following, we will summarize several experimental studies that investigate such a correlation.

Holtgraves (1986) studied the role of a speaker's status on the perceived appropriateness of indirectness. He showed that direct requests were perceived as more polite when they were produced by higher-status than by equal-status individuals. In a similar study, Holtgraves (1994) showed that hints such as (1e) are more likely to be understood as directives if the speaker is higher versus equal in status relative to the addressee. Looking at it the other way around, Holtgraves & Yang (1992) showed that a speaker was perceived as lowest in status when the request was a hint such as (1e), and as highest in status when the request was performed with an imperative, as in (1a). In a production study, Ruytenbeek (2019a) showed that participants produced requests more politely (including internal and external modifiers) when the addressee was of high status versus equal status.

Concerning rank of imposition, Freytag (2020) analyzed corpus data of business communications emails and found out that low imposition directives more often include imperatives, whereas, for high imposition directives, query preparatory questions were more frequently used. Production experiments by Kranich et al. (2021) and Ackermann (2023) showed that high imposition contexts give rise to increased politeness within request formulations, which is realized not only by choice of the head act strategy but particularly by (additional) internal and external modification.[4]

With respect to social distance, Blum-Kulka & Olshtain (1984) found that a high social distance is associated with more polite forms including CIRs and negative state remarks, such as (1e). Baxter (1984)'s results support the hypothesis that a higher degree of indirectness corresponds to situations involving higher social distance. By contrast, Holtgraves & Yang (1992) show that the most direct request forms give rise to the greatest perceived distance between interlocutors. Similarly, Freytag (2020) provides evidence that direct requests are more likely to occur in high-distance situations versus low-distance situations. Therefore, the findings of Blum-Kulka & Olshtain (1984) and Baxter (1984) are in line with Brown and Levinson's politeness theory, whereas the findings of Holtgraves & Yang (1992) and Freytag (2020) are at odds with it.

All in all, the presented empirical research confirms the hypothesis that a higher status of the addressee relative to the speaker, as well as a higher rank of imposition, correlates with a more polite formalization of the request. On the other hand, this correlation was not uniformly confirmed for social distance,

[4] Ackermann (2023)'s study involves German native speakers; Kranich et al. (2021)'s study involves German native speakers, as well as native speakers from a range of English varieties.

where quite opposing results were found. Based on these empirical findings, we presume that the factors of power difference and rank of imposition are robust variables for the prediction of the usage of politeness in requests, and we focus on these two within the present study.

Our study aims to empirically investigate how power difference and rank of imposition (RoI) impact the choice of linguistic politeness strategies in formulating requests. As specified above, previous research on this topic attested to a correlation between these situational factors and various politeness strategies, but most of the findings come from rating studies, the results of which do not necessarily track speakers' actual behavior (cf. Baxter 1984, Holtgraves 1986, 1994, Holtgraves & Yang 1992). Furthermore, many of the production studies involve corpus data or relatively uncontrolled experiments, where the tested conditions differ in multiple ways (cf. Blum-Kulka & Olshtain 1984, Freytag 2020). To date, there are only a few studies that manipulate the relevant factors in a controlled way, independently of each other — something that is necessary to fully understand their role, but these studies do not consider the comparison of very particular sets of head act strategies, which were said to differ in their level of politeness (cf. Ruytenbeek 2021), specifically the variation of CIRs vs NCIRs, or the variation of can-CIRs, like (1b), vs could-CIRs, like (1c).[5] We address this research gap by analyzing our data concerning these distinctions. Another novel aspect of this study is that we do not only investigate *how* participants formulate requests, but also *why* participants chose their particular formulation instead of another one.

Our study compares experimental production data from a discourse completion task (cf. Blum-Kulka 1989) represented in 6 different conditions, varying in three levels of imposition and two types of power relations between interlocutors. Furthermore, the experiment consisted of several follow-up tasks that were designed to contrast the participant request formulations with alternative formulations representing typical head act strategies, such as listed in (1). The

[5]There are to our knowledge three noteworthy exceptions: Ruytenbeek (2019a) studies the variation of CIRs vs NCIRs across the variable power in a production study of written requests in French. However, he solely considers one particular NCIR type (Is it possible to VP?), which was produced very infrequently in comparison to CIRs. Kranich et al. (2021) study the variation of CIRs vs NCIRs across the variables power and RoI in a discourse completion task, but they classify all types of query preparatory questions as CIRs; therefore, they do not study the distinction we are particularly interested in, that contrasts types of (1b) with types of (1d). Finally, Ackermann (2023) studies the variation of CIRs vs NCIRs across the variable RoI in a discourse completion task, but in her study, NCIRs play only a subordinate role since they are barely produced (below 10 %), which might be due to a different classification scheme, or due to the particular discourse scenarios in her study.

goal of the follow-up tasks was to detect potential motives for (dis)preferring given alternatives.

The article is structured as follows. Section 2 presents the experimental setup and describes the central task of the experiment: the discourse completion task (Task 1), its results, and the statistical analysis of the data. Section 3 describes the follow-up tasks and their statistical results. Section 4 summarizes the results.

2 Experiment part I: The main task

Our study explores how situational factors impact the formulation of request strategies. For the central task of the study, we employ the discourse completion task methodology (cf. Blum-Kulka 1989), in which participants are exposed to a discourse scenario and have to fill in a missing part of a conversation. Following previous work on English requests, we consider the factors 'rank of imposition' (RoI) and 'power difference' as promising independent variables for investigating politeness in request strategies (e. g. Brown & Gilman 1989, Holtgraves 2005, Biesenbach-Lucas 2007, Kranich et al. 2021).

In our experimental design, we present a scenario where the participant finds themselves in a situation where they need to borrow money from a colleague, whereby they have to formulate the request into a free text box (more details below). For this scenario, we consider (i) two types of power relation, where the colleague is either a coworker (equal power relation) or a supervisor (unequal power relation), and (ii) three levels of imposition, defined via the amount of money that the participant needs to borrow.

2.1 Hypotheses

We first hypothesize about the head act strategies. The examples in (3) represent potential head act strategies for the given scenario.[6]

(3) a. Lend me 20 dollars!
b. Can you lend me 20 dollars?
c. Could you lend me 20 dollars?
d. Would it be possible for you to lend me 20 dollars?

[6]The ordering of the examples in (3) is intended for expository purposes and reflects differences in request type and degree of indirectness. It should not be understood as a strict or fine-grained hierarchy of politeness among individual forms, particularly within the class of non-conventionally indirect requests in (3d–f).

e. Would you be able to lend me 20 dollars?
f. Would you mind lending me 20 dollars?

Here, we anticipate that participants use an indirect form of request as the dominant strategy since we assume that direct requests such as (3a) are considered to be too harsh and impolite. The indirect request forms can be partitioned into CIRs, such as (3b–c), or NCIRs, such as (3d–f). Request types such as (3d–f) are considered less direct and often more polite than CIRs, since the 'request for action' interpretation is not as deeply entrenched (Ruytenbeek 2021). Therefore, we expect NCIRs to be used increasingly more often with increasing power difference or rank of imposition. Our first Hypothesis H1 is as follows:

H1: The frequency of NCIRs relative to CIRs increases with the situational factors power difference and rank of imposition.

The two CIRs, as given in (3b–c), differ in that the latter uses the subjunctive mood. It is assumed that the subjunctive mood makes a request more polite by reducing the expectations as to the fulfillment of the request (Trosborg 1995, Pérez Hernández & Ruiz de Mendoza 2002). Following this conjecture, we expect that could-CIRs (CIRs beginning with *Could*, such as (3c)) are used increasingly more often relative to can-CIRs (CIRs beginning with *Can*, such as (3b)) with increasing power difference or rank of imposition.

H2: The frequency of could-CIRs relative to can-CIRs increases with the situational factors power difference and rank of imposition.

Politeness strategies in requests do not only concern the choice of the head act, but also its modification (cf. Blum-Kulka et al. 1989). In this study, we address as internal modifiers downtoners, such as *by any chance* or *possibly*, and as external modifiers different types of grounders, such as repay promise, reason or purpose for lending money. In line with previous research outlined in the introduction (cf. Ruytenbeek 2019a, Kranich et al. 2021, Ackermann 2023), we believe that these modification strategies correlate significantly with the situational factors of power difference and rank of imposition. We subsume this with the following hypothesis.

H3: The frequency of usage of downtoners (internal modifiers), and grounders (external modifiers) increase with the situational factors power difference and rank of imposition.

2.2 Participants, design, and procedure

A total of 210 self-reported native speakers of English, aged 18-64, with US IP addresses, were recruited via the platform Prolific and paid £1.35 for participation.[7] The study was implemented in a 1-item, fully between-subjects design ($n \approx 35$ per condition), where each participant saw the discourse scenario in one of six (2×3) conditions. The participants completed four experimental tasks, of which we present the first task and its results here, and the remaining tasks in Section 3.

The materials for Task 1 of the experiment were brief scenarios in which the participant finds themselves in a situation where they want to buy an item in a bookstore and have to ask a colleague for money. They had to phrase their request in a free text box. Two factors of the scenarios were manipulated in a 2×3 design:

- POWER relationship: colleague is coworker (CW) or supervisor (SV)
- Rank of imposition (ROI): value of item is \$2 (L), \$20 (M) or \$200 (H)[8]

Figure 1 shows an image of Task 1 where the different condition variables are presented in square brackets.[9]

2.3 Classification and overall frequencies

For the classification of the free text data of Task 1, we distinguish between head act strategies, internal modifiers, and external modifiers as parts of the speaker's politeness strategy (cf. Blum-Kulka et al. 1989).

[7] Due to duplicate participation, 3 data points were excluded, leaving a final total of 207 data points.

[8] L, M, and H stand for low, medium and high rank of imposition.

[9] Note: While the experimental design systematically manipulated the level of imposition and power difference, it is worth noting that participants' understanding of these factors may have been affected by other contextual elements. Specifically, the social distance between coworkers or between a coworker and supervisor may not directly translate to the perceived psychological distance or level of familiarity. Similarly, the level of imposition associated with a request may have been influenced by situational factors like urgency, which were not controlled for separately. As a result, participants may have considered urgency, familiarity, and perceived social roles in addition to the intended experimental variables when choosing their request strategies. The impact of urgency is explored further below, particularly in relation to the low-imposition condition.

You are on a business trip with your **[coworker (cw) / supervisor (sv)]**, and will return home early tomorrow morning. In between your appointments, the two of you stop into a small bookstore. You see a **[birthday card (L) / novel (M) / rare, signed edition of a famous novel (H)]** that would be perfect for your sister, whose birthday is this weekend. It costs **[$2 (L) / $20 (M) / $200 (H)]**. You decide to buy it, but when you get to the cash register, you realize that you've left your wallet back in your hotel room.

Since you won't have time to return to the bookstore before you go home, you decide to ask your **[coworker (cw) / supervisor (sv)]** to lend you the **[$2 (L) / $20 (M) / $200 (H)]** so you can make the purchase.

Task 1: After you have explained the situation to your **[coworker (cw) / supervisor (sv)]**, how would you phrase your request?

Participant answer...

Figure 1: Image of Task 1 with condition variables in bold font and square brackets.

2.3.1 Head act strategies

Initially, we selected a coarse-grained classification for head act strategies by distinguishing between direct requests (DIR), conventionally indirect requests (CIR), and non-conventionally indirect requests (NCIR). We classify requests initially formulated as *Can/Could...* (e. g. *Can you lend me ...*, *Could I borrow ...*) as CIRs, and all other indirect request forms as NCIRs. Following this classification scheme, we found that only five participants produced direct requests (2.4 %). For the following analysis, we excluded these data points since we are interested in the strategic choice of indirect requests. For the remaining data, we found that CIRs were produced around half of the time (51 %), and NCIRs as well (49 %).

In our analysis, however, we also take into account a more detailed classification of indirect head act strategies by carefully considering all initial constructions that account for more than 5 % frequency. Constructions with a frequency below 5 % were grouped into the class OTHER. Table 1 shows the resulting classification scheme with the classes CAN, COULD, MIND, ABLE, POSSIBLE and OTHER, with exemplary beginnings of the request term, mappings to the main classes

CIR and NCIR, and its relative frequency over all indirect request terms. Figure 2 shows the respective bar plot.

Table 1: A fine-grained classification of indirect head act strategies with respect to the initial modal construction of the request term.

Subclass	Exemplary beginning(s)	Main class	Frequency
CAN	Can you lend me... / Can I borrow...	CIR	27.2 %
COULD	Could you lend me... / Could I borrow...	CIR	23.8 %
MIND	Do you / Would you **mind** lending...	NCIR	15.3 %
ABLE	Would you be **able** to lend...	NCIR	9.9 %
POSSIBLE	Would it be / Is it **possible** to lend...	NCIR	7.4 %
OTHER	e. g. May I... / I need... / I would appreciate...	NCIR	16.3 %

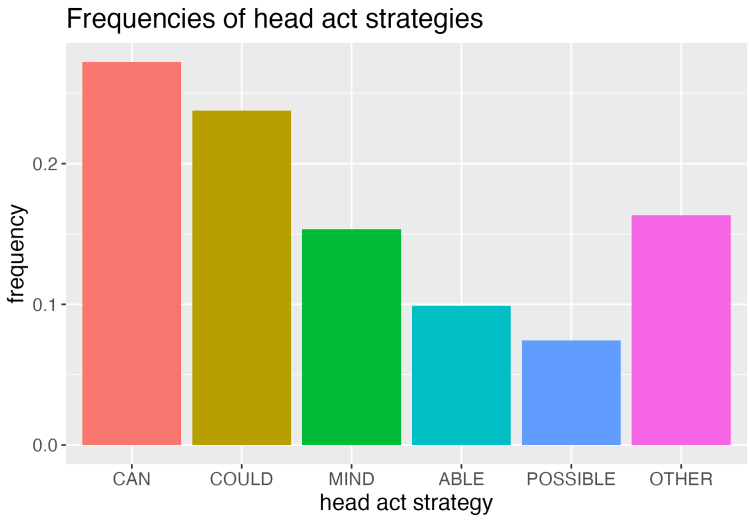

Figure 2: Frequencies of head act strategies according to the fine-grained classification of indirect request terms.

2.3.2 Internal and external modifiers

The modifier categories were determined based on an initial review of responses.[10] The first complete classification scheme contained the classes DOWNTONER (*possibly*, *perhaps*), VOCATIVE (e. g. *bro*, *Madame*) and POLITENESS MARKER (*please*) as internal modifiers, and GROUNDER as an external modifier. However, since *please* occurs exclusively with CIRs in our data (not with NCIRs)[11] and due to the low frequency and great variety of vocatives, we excluded both modification strategies from the further analysis.

For the classification of the remaining modification strategies, we used Boolean values (0 or 1) for the presence or absence of a particular modifier. Requests were annotated with 1 for class DOWNTONER if a downtoner[12] was used anywhere in the head act part of the utterance, and 0, if not. For external modifiers, we found three different types of grounders: repay promise, purpose (what the money is for), and cause (why I don't have the money on me right now). Requests were annotated with 1 for class GROUNDER if at least one of the three types was used as part of the speech act set, and 0, if not. Furthermore, requests were annotated with 1 for class REPAY PROMISE, PURPOSE and/or CAUSE if the respective grounder type was used as part of the speech act set, and 0, if not. Table 2 gives an overview of all annotated modification strategies with exemplary parts of the request term, mappings to the internal or external modifier type, and its frequency over all request terms.

2.4 Statistical analysis and results

The following analysis was performed using R Statistical Software (v4.3.2; R Core Team 2021).[13] For testing Hypothesis 1, we applied a logistic regression model that was fitted to the data, with POWER and ROI (reference level M) as independent variables, and the main type of the head act strategy (CIR or NCIR) as dependent variable. Figure 3 shows the frequencies of main types over ROI (left) and POWER (right). The regression analysis showed significant effects of POWER,

[10] All responses were coded by the 2nd author and reviewed by the 1st author. There were only a few entries where the initial coding decision needed to be updated. All disagreements were discussed and resolved.

[11] Note: While please is not ungrammatical with NCIRs, its distribution is more restricted: in NCIRs it appears only clause-finally, whereas in CIRs it may also occur in medial position.

[12] Ackermann (2023) shows that downtoners often occur in at least two different types, namely as tentativeness marker such as *possibly* or *by any chance* or as casualness marker such as *just* or *shortly* and that their usage correlates quite differently with situational factors. However, in our study, we solely found downtoners that indicate tentativeness in the head act utterances.

[13] See the appendix for statistical details and model output tables.

Table 2: Examples of different modifiers used in the requests.

Modifier	Example	Type	Frequency
DOWNTONER	Could you **possibly** lend me ...	internal	3.9 %
GROUNDER [total]		external	84.1 %
- REPAY PROMISE	I'll **pay you back** as soon as possible.	external	67.1 %
- PURPOSE	I want to **buy a book** for my sister ...	external	50.2 %
- CAUSE	I **left my wallet** in my room ...	external	47.8 %

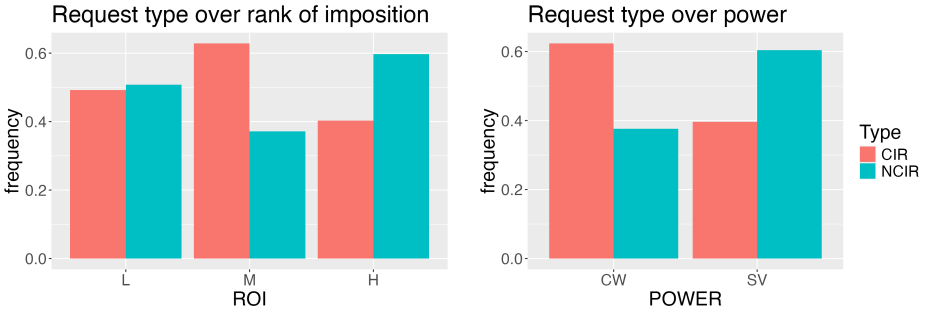

Figure 3: Frequencies of main request types over ROI levels L ($2), M ($20) and H ($200) (left) and POWER levels CW (coworker) and SV (supervisor) (right).

such that all NCIR strategies were used significantly less often in the condition CW ($p < 0.01$). Furthermore, the regression analysis also showed significant effects of ROI, but only with respect to the levels M and H. More concretely, we found that NCIR strategies were used significantly more often in level H in comparison to M ($p < 0.01$). There was no significant effect between the ROI levels L and M. Overall, Hypothesis 1 was fully confirmed with respect to power difference, and partially confirmed with respect to rank of imposition.

For testing Hypothesis 2, we considered a subset of the data where participants used either CAN or COULD as head act strategy. Then we applied a logistic regression model that was fitted to the data, with POWER and ROI (reference level M) as independent variables, and the CIR type of the head act strategy (CAN or COULD) as dependent variable. Figure 4 shows the frequencies of CIR types over ROI (left) and POWER (right). The regression analysis did not show any signifi-

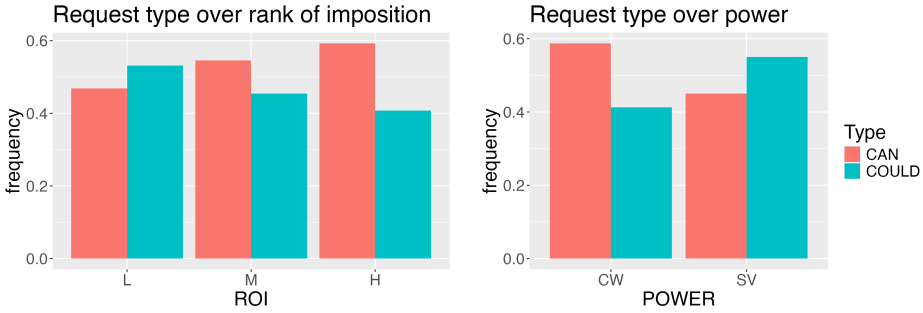

Figure 4: Frequencies of CIR request types CAN/COULD over ROI levels L ($2), M ($20) and H ($200) (left) and POWER levels CW (coworker) and SV (supervisor) (right).

cant effects of POWER, but we found a trend in the expected direction, such that COULD was used less often than CAN in condition CW, but more often in condition SV ($p = 0.17$). The regression analysis did not show any significant effects of ROI. Moreover, we found a trend that goes against the expected direction, where the usage of CAN vs. COULD increases with ROI (cf. Figure 4). Overall, we did not find (reasonable) support for Hypothesis 2.

As an additional analysis, we applied a multinomial logistic regression model that was fitted to the data using the package *nnet* (Venables & Ripley 2002), with the independent variables POWER difference and ROI (reference level M) and the fine-grained set of head act strategy as dependent variable (with strategy types CAN, COULD, POSSIBLE, ABLE and MIND; reference level CAN). In line with the last regression analysis, we did not find any significant effects for CAN vs. COULD, neither of POWER, nor of ROI. However, we found significant effects of power difference for CAN in comparison to POSSIBLE, ABLE, and MIND, such that these three NCIR strategies were used significantly less often in condition CW ($p < 0.05$ each). Furthermore, the multinomial regression analysis also showed significant effects of ROI, but only for some NCIR strategies, and only concerning H but not concerning level L. More concretely, we found that the head act strategies POSSIBLE and ABLE were used significantly less often than CAN in level M vs H ($p < 0.05$ for ABLE, $p < 0.01$ for POSSIBLE). Figure 5 shows the frequencies of the fine-grained request types over ROI (left) and POWER (right).

For testing Hypothesis 3, we applied a logistic regression model for the modifier strategies DOWNTONER and GROUNDER. Figure 6 shows the frequencies of downtoners and grounders (including their subtypes) over ROI (left) and POWER (right). The regression models were fitted to the data, each with the independent variables POWER and ROI (reference level M), and the respective modifier choice

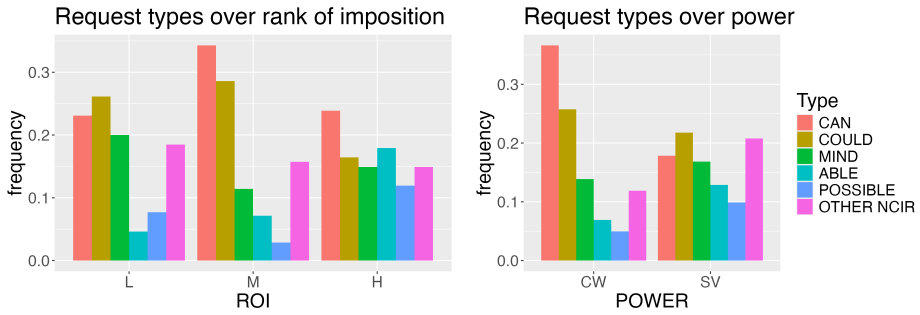

Figure 5: Frequencies of 6 indirect request types over ROI levels L ($2), M ($20) and H ($200) (left) and POWER levels CW (coworker) and SV (supervisor) (right).

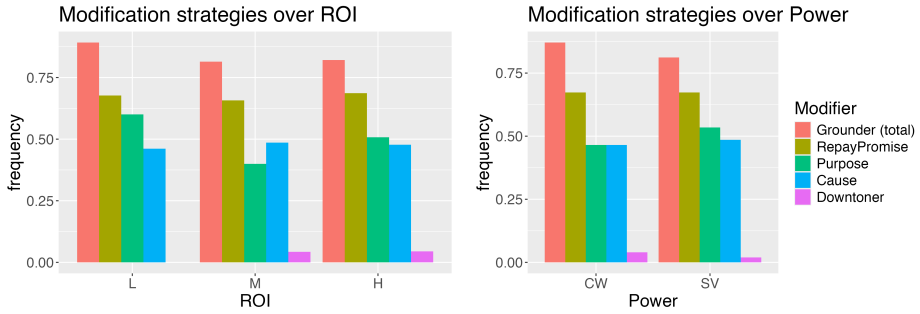

Figure 6: Frequencies of modification strategies over ROI levels L ($2), M ($20) and H ($200) (left) and POWER levels CW (coworker) and SV (supervisor) (right).

(used or not used) as the dependent variable. For both modifier types, downtoners, and grounders, we did not find any significant effect of POWER or ROI.[14] Overall, we did not find any support for Hypothesis 3.

2.5 Discussion

Hypothesis 1 was fully confirmed concerning power: the frequency of NCIR types relative to CIR types was significantly greater in condition SV than in condition CW. This result is in line with Ruytenbeek (2019a), who studied the usage of CIRs (*Can you VP?*) and a particular NCIR type (*Is it possible to VP?*) in a production study of written requests in French. As in the present study, he considered

[14] In a follow-up analysis, we applied logistic regression models for the grounder subclasses REPAY PROMISE, PURPOSE and CAUSE. Here too, we did not find significant effects of both situational factors.

the two power conditions same-status addressee and higher-status addressee. He found that while the usage of CIRs decreases with status (from 62 % to 47 %), the usage of the NCIR type increases with status (from 7 % to 14 %).

Hypothesis 1 was only partially confirmed concerning rank of imposition. The frequency of NCIR types relative to CIR types was significantly greater in condition H than in condition M. However, we did not find a significant effect across the levels M and L. We assume that the issue with condition L is that it is affected by another relevant situational factor, namely *urgency*. While for the conditions M and H, the requester wants to buy a book, which might be difficult to purchase elsewhere before the sister's birthday, for condition L the requester wants to buy a birthday card that might easily be purchased elsewhere shortly afterward. As a result, participants may have perceived the L condition as less *urgent*, allowing for more indirect and less imposing request strategies, contrary to expectations based on rank of imposition alone. This confound limits the interpretability of the results for the L condition and suggests that future experimental designs should more carefully control for urgency, for example by using comparable request goals across imposition levels (e. g. purchasing a book in all conditions, differing only in cost or effort).

Overall, the confirmation of Hypothesis 1 (with an exception for ROI condition L) shows two things. First of all, it seems to be the case that CIR types (CAN, COULD) are considered as less appropriate at mitigating an FTA than many NCIR types, particularly those that are also preparatory questions (e. g. types ABLE, MIND, or POSSIBLE), which has already been assumed by Ruytenbeek (2019a, 2021). Secondly, it seems to be the case that speakers address this distinction in speech production concerning at least two situational factors: power and rank of imposition, such that they tend to use the more polite NCIR type more often with increasing relative status of the addressee or increasing imposition level of the request.

Hypothesis 2 was not confirmed. For the CAN-COULD variation, we did not find a significant effect of the power difference, but a numerical difference that goes in the expected direction ($p = 0.17$). Given that the CAN-COULD distinction arguably reflects a more fine-grained difference in politeness than the broader CIR-NCIR contrast, this null result may be due to limited statistical power. Accordingly, future studies with larger sample sizes are needed to determine whether the observed trend reflects a reliable effect of power difference on the choice between CAN- and COULD-based CIRs.

For the CAN-COULD variation, we also did not find a significant effect of rank of imposition. In this case, the observed numerical trend ran counter to the predicted direction, casting doubt on whether increased statistical power alone

would yield the expected effect. This raises the possibility that the CAN-COULD distinction may interact with power difference but not with rank of imposition, or that it is relatively insensitive to situational factors altogether. Disentangling these possibilities remains an open question for future research.[15]

We did not find support for Hypothesis 3. This result contrasts with findings reported by Ackermann (2023), who noted significant effects of ROI on various modification strategies, including downtoners and grounders.[16] Two factors likely contribute to this discrepancy. First, a simulation-based power analysis reveals that the current study had limited statistical power to detect condition effects for these modification strategies. Downtoners appeared infrequently in our data, while grounders dominated the majority of trials, leading to floor and ceiling effects, respectively. Thus, the lack of significant effects for these strategies does not provide strong evidence against subtle condition-related differences.

Second, the operationalization of ROI varies significantly between the two studies. In Ackermann (2023), low and high imposition correspond to qualitatively distinct request scenarios that likely invite different politeness strategies, particularly regarding modification.[17] In contrast, the present study maintains a constant basic scenario—requesting money in a bookstore—and manipulates ROI more precisely through the item's price. Our results indicate that under such controlled and narrowly defined variations in imposition, modification strategies play a comparatively minor role, while distinctions in ROI are mainly addressed through the choice of head act strategy.

3 Experiment part II: The follow-up tasks

The goal of Task 1 of the experiment was to investigate *how* participants formulate a request given particular situational factors. The experiment is also composed of several follow-up tasks to investigate *why* participants chose their par-

[15] A corpus study of the *Stanford Politeness Corpus* by Danescu-Niculescu-Mizil et al. (2013) has shown that COULD requests were rated higher in politeness scores than CAN requests. However, the corpus consists of two parts (Part I: data of stack exchange users, Part II: data of Wikipedia admins), and a more fine-grained analysis shows that this difference is only significant for Part II of the corpus, whereby in Part I, CAN requests were even rated slightly higher. This result indicates that the hypothesis that COULD requests are more polite than CAN requests does not hold as a general rule.

[16] Moreover, Ruytenbeek (2019a) reports effects of power on the use of formal greetings and the T/V distinction in French, which were not included in the present experimental design.

[17] For instance, the low-imposition scenario involves asking a friend to hold a glass briefly, whereas the high-imposition scenario involves calling a family member after midnight to request a ride.

ticular formulation instead of another one. As the results from Task 1 showed, several typical head act strategies are frequently used by participants (cf. Figure 2) and to some extent correlated with situational factors (cf. Figure 5). In the follow-up task, we presented the participants with several typical types of head act strategies and asked them (i) how appropriate they find each of them in relation to their own request formulation, and (ii) why they found each of them less/equally/more appropriate than their own request.[18] The concrete goal of this follow-up task was to detect the choice motives for the head act strategy used in the participants' own formulation, compared to alternative head act strategies. Some of these choice motives are representative of face threats towards the addressee (e. g. formulation is *too pushy* or *not respectful*), but others pay respect to the speaker's image (speaker appears as *too insecure* or *not friendly*), or to the informativity or length of the request formulation (e. g. formulation is *too short* or contains *no payback promise*).

3.1 Design and procedure

In the first follow-up task (Task 2), we presented four alternatives of request formulations to the participants, which represent the following head act strategies: direct request (ALT DIR), conventionally indirect request that begins with *Can* (ALT CAN), conventionally indirect request that begins with *Could* (ALT CLD), and non-conventionally indirect request (ALT WLD). Concretely, participants saw the following four alternatives of request formulations:[19]

- *Lend me $x, please.* (ALT DIR)
- *Can you lend me $x?* (ALT CAN)
- *Could you lend me $x?* (ALT CLD)
- *Would it by any chance be possible for you to lend me $x?* (ALT WLD)

The participants were then asked for each of these alternatives to give *relative appropriateness* (RA) ratings, which means to indicate how appropriate an alternative is in comparison to their own formulation from Task 1, by addressing a five-point Likert scale with the following levels:

(L1) much less appropriate
(L2) slightly less appropriate
(L3) equally appropriate

[18] In each follow-up task, participants were shown their own formulations from Task 1 to prevent them from forgetting the exact wording they had produced.

[19] Here $x is a placeholder for the actual amount of money in the scenario: $2, $20, or $200.

(L4) slightly more appropriate
(L5) much more appropriate

The next task (Task 3) addresses the evaluations from Task 2. More concretely, participants were asked to give reasons for their relative appropriateness (RA) ratings from Task 2 via a free text entry. Since they had to do it for each alternative, participants had to address the following four subtasks, presented on separate screens and in random order.

- You rated the request *Lend me $x, please.* to be [*relation term*] than/as your own request. Why do you think it is [*relation term*]?
- You rated the request *Can you lend me $x?* to be [*relation term*] than/as your own request. Why do you think it is [*relation term*]?
- You rated the request *Could you lend me $x?* to be [*relation term*] than/as your own request. Why do you think it is [*relation term*]?
- You rated the request *Would it by any chance be possible for you to lend me $x?* to be [*relation term*] than your own request. Why do you think it is [*relation term*]?

Here, [*relation term*] is a placeholder that was set to (i) *less appropriate* in case the participant chose L1 or L2 in Task 2, (ii) *equally appropriate* in case the participant chose L3 in Task 2, or (iii) *more appropriate* in case the participant chose L4 or L5 in Task 2.

In the final task (Task 4) participants were, again, asked to name reasons for their RA ratings from Task 2, but this time via a multiple choice design from a predefined list of reasons. Moreover, participants had to address only those alternatives that were rated as slightly or much less appropriate (L1 or L2) in Task 2. More concretely, participants had to address the following subtask for each alternative that they had rated as less appropriate.

- You rated the request [ALT DIR / ALT CAN / ALT CLD / ALT WLD] to be less appropriate than your own request. Although you might have already mentioned it, which of the following, if any, do you think is a reason for the request being less appropriate?

  ☐ it is too intrusive or pushy (TOO PUSHY)
  ☐ it is too short (TOO SHORT)
  ☐ it makes it less likely that the request will be granted (LESS LIKELY)
  ☐ the speaker does not appear sufficiently friendly or sympathetic (NOT FRIENDLY)

- ☐ the speaker does not show enough respect to their companion (NOT RESPECTFUL)
- ☐ the speaker sounds too unconfident or insecure (TOO INSECURE)
- ☐ none of these

3.2 Classification of Task 3 data

In the following, we consider the data of Task 3 exclusively for the cases where participants found the alternative expression less appropriate. In total, we have 400 data points, since, on average, each participant found around 2 alternatives slightly less or much less appropriate. The reason classes were determined based on an initial review of responses. All responses were coded by an assistant and reviewed by the authors, and any disagreements were discussed and resolved. The first complete classification scheme took into account very fine-grained differences in responses, resulting in more than 20 different classes. In a follow-up review, we found reasonable ways to merge similar response classes, resulting in a second classification scheme that considered the following nine reason classes: TOO PUSHY, MISSING REASON, NOT POLITE, TOO DIRECT, NO PROMISE, CONCRETE WORDING, INAPPROPRIATE AT ALL, NOT NICE and TOO INFORMAL. Our final classification scheme contained only those classes that exceed 10 % of all data points (more than 40 of 400). Table 3 shows the resulting classification with exemplary entries and frequency data.

Table 3: Classification of reasons for finding an alternative request term less appropriate.

Category	**Exemplary entry**	**Freq.**
TOO PUSHY	It seems too forward and a bit demanding.	31 %
MISSING REASON	Because the reason for lending the money isn't stated.	24 %
NOT POLITE	It sounds a bit less polite.	15 %
TOO DIRECT	It's very direct and sounds more like a demand.	14.5 %
NO PROMISE	Because it doesn't say when it will be paid back.	12.8 %
CONCRETE WORDING	I don't see *please*.	12.3 %

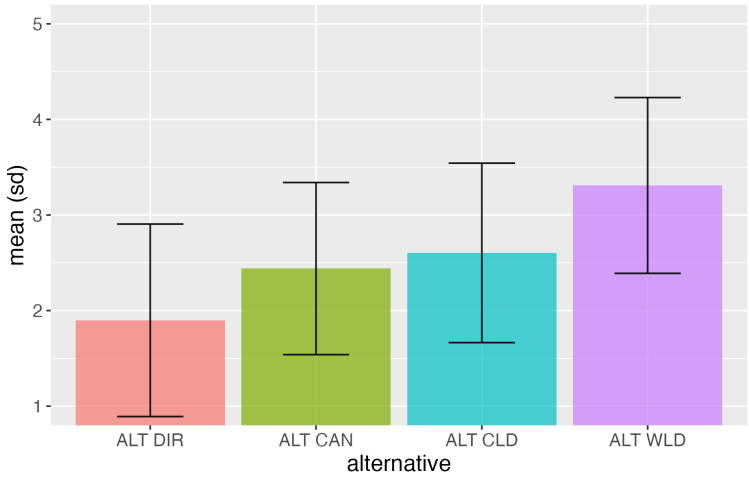

Figure 7: Mean and standard deviation of RA ratings (Task 2) for the four alternative request terms ALT DIR, ALT CAN, ALT CLD and ALT WLD (as listed in Section 3.1).

3.3 Results

The overall rating results of Task 2 are shown in Figure 7. A linear mixed effects model was fitted to the data using the package *lmerTest* (Kuznetsova et al. 2017), with alternative as a fixed factor (reference level ALT CAN) and random intercepts for participant ID.[20] We found highly significantly higher ratings of ALT CAN relative to ALT DIR ($p < 0.001$), highly significantly lower ratings of ALT CAN relative to ALT WLD ($p < 0.001$), and significantly lower ratings of ALT CAN relative to ALT CLD ($p < 0.05$).

One goal of this task was to see if participants connected their own head act choice with the alternative. If this was the case, then we would, e. g. expect participants who chose a head act of type CAN in Task 1 to rate the alternative ALT CAN in Task 2 as more appropriate than the other alternatives. Figure 8 shows the resulting RA ratings (mean and sd) of the four alternative request terms ALT DIR, ALT CAN, ALT CLD and ALT WLD over the choice in Task 1, thereby distinguishing the classes CAN, CLD and NCIR. A linear mixed effects model was fitted to the data using the package *lmerTest* (Kuznetsova et al. 2017), with alternative (reference

[20] Additional complementary analyses addressing alternative modeling approaches and coding schemes are reported in the appendix, together with the corresponding model output tables.

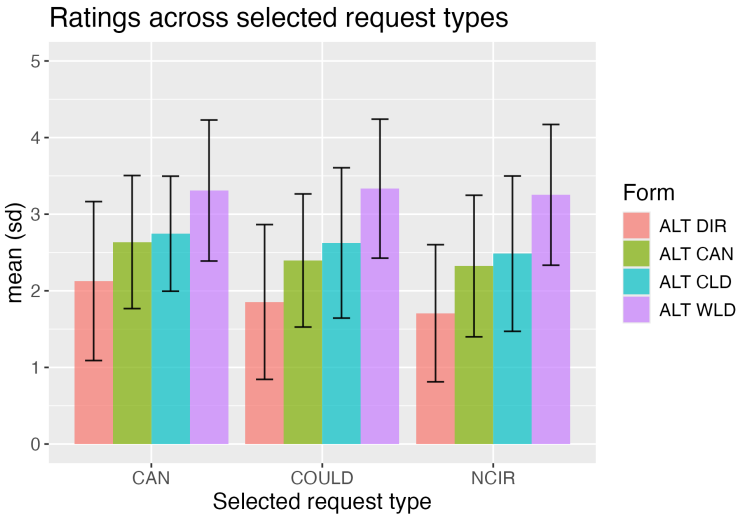

Figure 8: Mean and standard deviation of RA ratings (Task 2) for the four alternative request terms ALT DIR, ALT CAN, ALT CLD and ALT WLD over Task 1 request types, classified as CAN, CLD and NCIR.

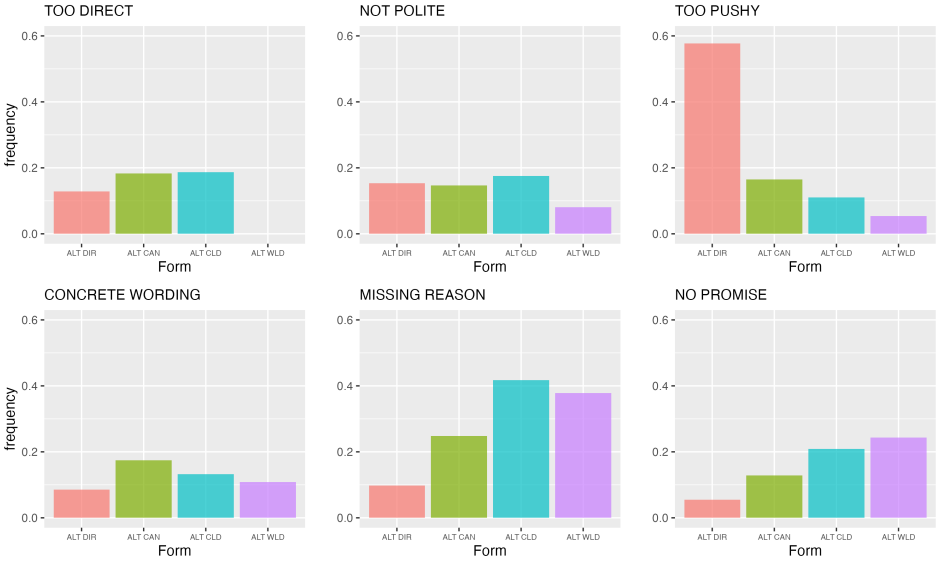

Figure 9: Frequencies of different classified reasons (from Task 3 data) for estimating the alternative forms ALT DIR, ALT CAN, ALT CLD and ALT WLD as less appropriate than the Task 1 request formulation.

level ALT CAN) and selected request types (reference level CAN) and their interaction as fixed factors and random intercepts for participant ID. We did not find any other significant interaction.

The results of Task 3 are given in Figure 9, where we see the frequencies for the top 6 reasons for estimating an alternative form as less appropriate (cf. Table 3) over the four alternative forms. For each reason, a generalized linear mixed effects model was fitted to the data using the package *blme* (Chung et al. 2013), with alternative as a fixed factor (reference level ALT CAN) and random intercepts for participant ID. The statistical analysis showed multiple things. First of all, the reason TOO DIRECT was mentioned significantly less often for finding ALT WLD less appropriate in comparison to ALT CAN ($p < 0.01$). The reason TOO PUSHY was mentioned significantly more often for finding ALT DIR less appropriate in comparison to ALT CAN ($p < 0.001$). The reason CONCRETE WORDING was mentioned significantly less often for finding ALT DIR less appropriate in comparison to ALT CAN ($p < 0.001$). The reason MISSING REASON was mentioned significantly less often for finding ALT DIR less appropriate in comparison to ALT CAN, and significantly more often for finding ALT CLD less appropriate in comparison to ALT CAN (both $p < 0.001$). Finally, the reason NO PROMISE was mentioned significantly less often for finding ALT DIR less appropriate in comparison to ALT CAN, and significantly more often for finding ALT CLD less appropriate in comparison to ALT CAN (both $p < 0.05$).

The results of Task 4 are given in Figure 10, where we see the frequencies for the six multiple choice reasons for estimating an alternative form as less appropriate over the four alternative forms. For each reason, a generalized linear mixed effects model was fitted to the data using the package *lme4* (Bates et al. 2015), with alternative as a fixed factor (reference level ALT CAN) and random intercepts for participant ID. The statistical analysis showed the following. First of all, for all six motives, there was no significant difference in frequency for finding ALT CLD less appropriate than ALT CAN. Moreover, the motives NOT FRIENDLY, TOO PUSHY, NOT RESPECTFUL and TOO SHORT were all mentioned significantly less often for finding ALT WLD less appropriate in comparison to ALT CAN, whereas the motive TOO INSECURE was mentioned significantly more often for finding ALT WLD less appropriate in comparison to ALT CAN. Finally, the motives NOT FRIENDLY, TOO PUSHY, NOT RESPECTFUL were all mentioned significantly more often for finding ALT DIR less appropriate in comparison to ALT CAN, whereas the motive TOO SHORT was mentioned significantly less often for finding ALT DIR less appropriate in comparison to ALT CAN. Exact p-values can be found in the appendix.

We additionally tested whether the effects interacted with the experimental conditions POWER and ROI. For each reason, two generalized linear mixed-effects

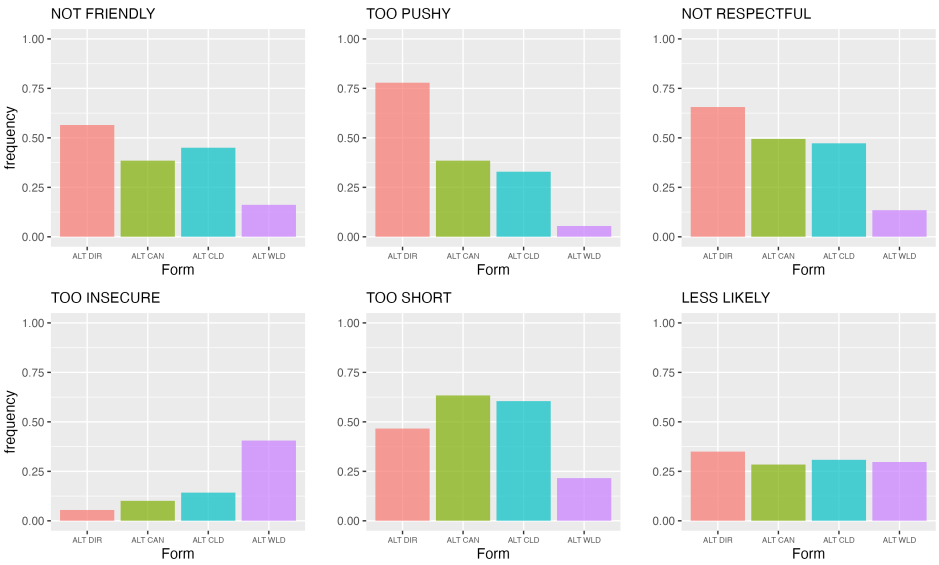

Figure 10: Frequencies of selected predefined reasons (from Task 4) for estimating the alternative form ALT DIR, ALT CAN, ALT CLD and ALT WLD as less appropriate than the Task 1 request formulation.

models were fitted using the *lme4* package (Bates et al. 2015). In both models, alternative (reference level ALT CAN) and either POWER (first model; reference level CW) or ROI (second model) were included as fixed effects, with random intercepts for participant ID.

No significant interactions with ROI were observed. By contrast, significant interactions with POWER were found. First, the effect of NOT FRIENDLY being mentioned significantly less often for the appropriateness of ALT WLD compared to ALT CAN was significantly stronger in the SV condition than in the CW condition ($p < 0.05$). Second, the effect of TOO INSECURE being mentioned significantly more often for the appropriateness of ALT WLD compared to ALT CAN was also significantly stronger in the SV condition than in the CW condition ($p < 0.05$). Figure 11 shows the frequencies for both reasons across POWER.

3.4 Discussion

The results of Task 2 show that it seems to be the case that participants did not connect their own head act strategy (Task 1) to the given alternative, since e. g. CAN users did not rate the alternative form ALT CAN as most appropriate

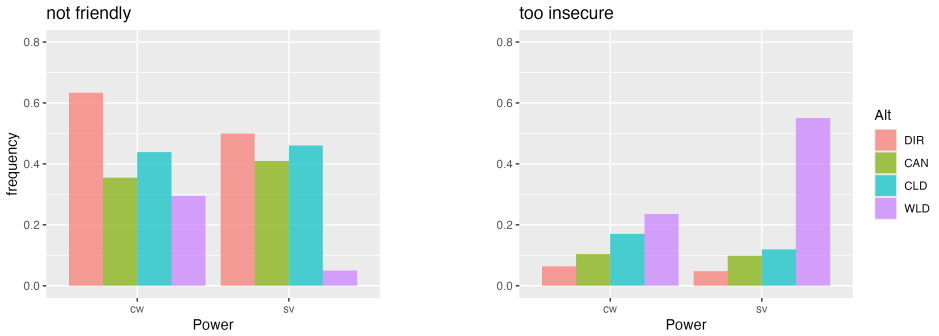

Figure 11: Frequencies of selected reasons NOT FRIENDLY and TOO INSECURE (from Task 4) for estimating the alternative form as less appropriate than the Task 1 request formulation across POWER. In both cases, effects are significantly stronger in the SV condition.

relative to their own request formulation, and COULD users did not rate the alternative form ALT CLD as most appropriate relative to their own request formulation. Rather, the rating order ALT DIR ≺ ALT CAN ≺ ALT CLD ≺ ALT WLD was invariant across the choice classes of Task 1 (cf. Figure 8). We think that this is because the majority of participants (87 %) wrote more than a bare head act phrase as a request, but added some kind of additional explanation, mostly realized by a grounder (84 % of all participants).[21] Therefore, they did not make a direct connection between the head act strategy that they used as part of their often more complex speech act set and the bare head act strategy that was given as a potential alternative. Note that the alternatives are very frequently rated as less appropriate than the Task 1 request formulation (ALT DIR, ALT CAN and ALT CLD have a mean score below 3, and ALT WLD only slightly above 3, which is the score for equally appropriate), which is mainly a consequence of the lack of grounders or other additional supportive moves. This is also supported by the results of Task 3, where two very frequent reasons for rating alternatives as less appropriate were due to the lack of grounders (MISSING REASON, NO PROMISE).

[21]The high frequency of grounders might be due to the domain of the request used in the experiment, namely the requests for money. This context may have heightened participants' sensitivity to justification and commitment, beyond the impact of the mitigation strategies themselves. In other domains, like requests for physical assistance, omitting reasons or promises might have less of an effect on whether people comply, even with more demanding requests. The focus on financial requests may have made certain mitigation strategies seem more important than they would be in other situations, and may have also masked the effects of the variables being studied. Future studies should test a wider variety of request types to see how broadly the findings apply.

The results of Task 3 indicate that participants do not mention all potential reasons for disapproval, but often just the most significant reason(s) that come to their mind. For example, it is the case that all four alternatives are stated as a pure head act strategy, thus without any supportive move, such as a payback promise. Therefore, the lack of a payback promise is a potential reason for disapproval of all four alternatives. However, the reason NO PROMISE was mentioned significantly less often for ALT DIR than for all the other alternatives. We believe that this is because this reason is not the most significant one that comes to mind for finding ALT DIR less appropriate, but probably rather that ALT DIR is too pushy, not polite enough, or too direct. Therefore, we should be careful with drawing conclusions from the data of Task 3, since a low frequency of a reason does not necessarily mean that it is not a frequent reason for disapproval, but that it is very frequently not the most significant reason for disapproval.

This problem is, however, not given in Task 4, where multiple choices are presented to the participants. The results of this task are for the most part in line with what we expected. First of all, we have two reasons that we assume to be connected to negative face threats toward the listener, namely TOO PUSHY and NOT RESPECTFUL. And we see that these reasons were selected significantly more often for ALT DIR in comparison to ALT CAN, and significantly less often for ALT WLD in comparison to ALT CAN. This is in line with the hypothesis that CIRs such as ALT CAN are considered more polite than direct requests, but less polite than NCIRs such as ALT WLD (see our argumentation leading to Hypothesis H1 in Section 2.1). Secondly, the reason TOO INSECURE was selected significantly more often for ALT WLD in comparison to ALT CAN. This result supports the idea that being too oblique or indirect can also have a disadvantage for the speaker's social image (e. g. with respect to determination, confidence, and trustfulness).

4 Summary and Conclusion

Consistent with presumptions in previous work and in line with our predictions, we found that the choice of conventional vs. non-conventional indirect requests strongly correlates with the situational factors of power difference and rank of imposition. We did not find such a correlation for the choice of the more fine-grained distinction CAN vs COULD, which might be due to the small sample size, and we plan to study this distinction with a larger sample size in a subsequent study. We also did not find such a correlation for the modification strategies DOWNTONER and GROUNDER. We assume that, compared to the head act strategies, the strategic usage of most modification strategies plays a subordinate role across situations where the situational factors themselves differ, but

the general scenario remains unchanged. This conjecture still needs to be verified in future research.

In the final follow-up task of our experiment (Task 4), we found significant differences in the choice frequencies of reasons for finding particular types of head act strategies less appropriate than one's own. More precisely, we found that CIR types are more frequently seen as less appropriate due to reasons that are connected to a negative face threat towards the addressee, whereas NCIR types are more frequently seen as less appropriate due to reasons that are connected to the speaker's social image of being confident.[22]

The results of the follow-up tasks showed some limitations of our experimental design, which can inform future work. First, Task 3 found that participants usually only report the most important reasons for their earlier choices, rather than giving a complete list of all the factors they considered. This limitation can be fixed by using a multiple-choice format, like in Task 4, if we have a set of response options that covers all the main points. Second, we found that participants did not consistently connect their own head act strategy from Task 1 to the alternatives in Task 2, for reasons discussed above. This suggests a general limitation of free production tasks when followed by metalinguistic reflection. To address this issue, we propose using a forced-choice design for Task 1, where potential alternatives are built into the production task. We see this forced-choice approach as a key methodological improvement and plan to use it in a future study.

Abbreviations

CIR = conventionally indirect request; NCIR = non-conventionally indirect request; RoI/ROI = rank of imposition; FTA = face-threatening act; RA = relative appropriateness; CW = coworker; SV = supervisor; L/M/H = low/medium/high imposition.

Funding information

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334.

[22]Note that there may also be additional reasons why NCIR is perceived as less appropriate. For example, in a separate follow-up experiment, we found that speakers also judge NCIR forms to be less appropriate because they can be perceived as overly indirect and more easily interpreted as ability questions rather than as genuine requests, which may reduce communicative success.

Acknowledgements

Earlier versions of this work were presented at the Leibniz-MMS Days 2025, the ZAS Internal Semantics & Pragmatics Colloquium, and the CRC 1412 Colloquium. We thank the audiences at these venues for discussion and feedback. We also thank Daniela Palleschi (ZAS, Data Science & Data Management) for consultation on the study. Finally, we are grateful to the editors Jordan Chark and Nico Lehmann, and to two anonymous reviewers for their valuable input.

Competing interests

The authors declare no competing interests.

Supplementary files

The experimental data reported in this paper are openly available on the Open Science Framework (OSF) at https://osf.io/xsq3p/ (DOI: 10.17605/OSF.IO/XSQ3P). The repository contains the anonymised response data from all four experimental tasks as CSV files:

- `task1_anno.csv` (production data, Task 1),
- `task2_result.csv` (appropriateness ratings, Task 2),
- `task3_anno.csv` (open-ended reasons, Task 3), and
- `task4_result.csv` (multiple-choice reasons, Task 4).

Authors' contributions

R. Mühlenbernd: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing, Review & Editing. M. Burbelko: Investigation, Data Curation, Resources, Writing, Review & Editing. U. Sauerland: Conceptualization, Methodology, Review & Editing, Supervision. S. Solt: Conceptualization, Methodology, Review & Editing, Supervision.

References

Ackermann, Tanja. 2023. Mitigating strategies and politeness in German requests. *Journal of Politeness Research* 19(2). 355–389. DOI: 10.1515/pr-2021-0034.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Baxter, Leslie. 1984. An investigation of compliance-gaining as politeness. *Human Communication Research* 10(3). 427–456. DOI: 10.1111/j.1468-2958.1984.tb00026.x.

Biesenbach-Lucas, Sigrun. 2007. Students writing emails to faculty: an examination of e-politeness among native and non-native speakers of English. *Language, Learning and Technology* 11. 59–81. DOI: 10.64152/10125/44104.

Blum-Kulka, Shoshana. 1987. Indirectness and politeness in requests: same or different? *Journal of Pragmatics* 11. 131–146. DOI: 10.1016/0378-2166(87)90192-5.

Blum-Kulka, Shoshana. 1989. Playing it safe: the role of conventionality in indirectness. In Shoshana Blum-Kulka, Juliane House & Gabriele Kasper (eds.), *Cross-cultural pragmatics: requests and apologies*, 37–70. Norwood, NJ: Ablex.

Blum-Kulka, Shoshana & Juliane House. 1989. Cross-cultural and situational variation in requesting behavior. In Shoshana Blum-Kulka, Juliane House & Gabriele Kasper (eds.), *Cross-cultural pragmatics: requests and apologies*, 123–154. Norwood, NJ: Ablex.

Blum-Kulka, Shoshana, Juliane House & Gabriele Kasper. 1989. Investigating cross-cultural pragmatics: an introductory overview. In Shoshana Blum-Kulka, Juliane House & Gabriele Kasper (eds.), *Cross-cultural pragmatics: requests and apologies*, 1–34. Norwood, NJ: Ablex.

Blum-Kulka, Shoshana & Elite Olshtain. 1984. Requests and apologies: a cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics* 5(3). 196–213. DOI: 10.1093/applin/5.3.196.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: some universals in language usage*. Cambridge & New York: Cambridge University Press. DOI: 10.1017/CBO9780511813085.

Brown, Roger & Albert Gilman. 1989. Politeness theory and Shakespeare's four major tragedies. *Language in Society* 18(2). 159–212. DOI: 10.1017/S0047404500013464.

Christensen, Rune H. B. 2025. *Ordinal—regression models for ordinal data*. R package version 2025.12-29. https://CRAN.R-project.org/package=ordinal.

Chung, Yeojin, Sophia Rabe-Hesketh, Vincent Dorie, Andrew Gelman & Jingchen Liu. 2013. A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* 78(4). 685–709. DOI: 10.1007/s11336-013-9328-2.

Clark, Herbert H. 1979. Responding to indirect speech acts. *Cognitive Psychology* 11(4). 430–477. DOI: 10.1016/0010-0285(79)90020-3.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec & Christopher Potts. 2013. A computational approach to politeness with application to social factors. In Hinrich Schuetze, Pascale Fung & Massimo Poesio (eds.), *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 250–259.

Freytag, Vera. 2020. *Exploring politeness in business emails: a mixed-methods analysis*. Bristol: Blue Ridge Summit: Multilingual Matters. DOI: 10.21832/freyta5952.

Holtgraves, Thomas. 1986. Language structure in social interaction. Perceptions of direct and indirect speech acts and interactants who use them. *Journal of personality and social psychology* 51(2). 305–314. DOI: 10.1037/0022-3514.51.2.305.

Holtgraves, Thomas. 1994. Communication in context: effects of the speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20(5). 1205–1218. DOI: 10.1037//0278-7393.20.5.1205.

Holtgraves, Thomas. 2005. Social psychology, cognitive psychology, and linguistic politeness. *Journal of Politeness Research* 1(1). 73–93. DOI: 10.1515/jplr.2005.1.1.73.

Holtgraves, Thomas & Joong-Nam Yang. 1992. Interpersonal underpinnings of request strategies: general principles and differences due to culture and gender. *Journal of Personality and Social Psychology* 62(2). 246–256. DOI: 10.1037/0022-3514.62.2.246.

Kasper, Gabrielle. 1995. Routine and indirection in interlanguage pragmatics. *Pragmatics and Language Learning* 6. 59–78. eric.ed.gov/?id=ED399811.

Kranich, Svenja, Hanna Bruns & Elisabeth Hampel. 2021. Requests across varieties and cultures: norms are changing (but not everywhere in the same way). *Anglistik* 32(1). 91–114. DOI: 10.33675/ANGL/2021/1/9.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. *Lmertest: tests in linear mixed effects models. R package version 2.0-29*. DOI: 10.18637/jss.v082.i13.

Murphy, Beth & Joyce Neu. 2006. My grade's too low: the speech act set of complaining. In Susan M. Gass & Joyce Neu (eds.), *Speech acts across cultures. challenges to communication in a second language*, 191–216. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110219289.2.191.

Pérez Hernández, Lorena & Francisco José Ruiz de Mendoza. 2002. Grounding, semantic motivation, and conceptual interaction in indirect directive speech acts. *Journal of Pragmatics* 34(3). 259–284. DOI: 10.1016/S0378-2166(02)80002-9.

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Ruytenbeek, Nicolas. 2019a. Do indirect requests communicate politeness? An experimental study of conventionalized indirect requests in French email communication. *Journal of Politeness Research* 16(1). 111–142. DOI: 10.1515/pr-2017-0026.

Ruytenbeek, Nicolas. 2019b. Indirect requests, relevance, and politeness. *Journal of Pragmatics* 142. 78–89. DOI: 10.1016/j.pragma.2019.01.007.

Ruytenbeek, Nicolas. 2021. *Indirect speech acts.* 1st edn. (Key Topics in Semantics and Pragmatics). Cambridge: Cambridge University Press. DOI: 10.1017/9781108673112.

Searle, John R. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Trosborg, Anna. 1995. *Interlanguage pragmatics: requests, complaints, and apologies*. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110885286.

Venables, W. N. & B. D. Ripley. 2002. *Modern applied statistics with S*. 4th Edn. New York: Springer. DOI: 10.1007/978-0-387-21706-2.

Appendix: Statistical Details

Task 1 Analyses

For Hypothesis 1 we used the following regression model:

- glm(formula = Type ~ ROI + Power, family = binomial, data = task1_data)

with Type (values CIR, NCIR) as measured variable, and ROI (values L, M, H, reference level M) and Power (values cw, sv, reference level cw) as manipulated variables. The model output is given in Table 4.

Table 4: H1 model output: logistic regression of head act type (CIR vs. NCIR) on ROI and Power.

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	1.0108	0.3006	3.363	0.000771 ***
ROIL	−0.5810	0.3601	−1.613	0.106656
ROIH	−0.9341	0.3608	−2.589	0.009632 **
Powersv	−0.9395	0.2948	−3.187	0.001436 **

For Hypothesis 2 we used the following regression model:

- glm(formula = SubType ~ ROI + Power, family = binomial, data = task1_CIR_data)

with SubType (values CAN, CLD) as measured variable, and ROI (values L, M, H, reference level M) and Power (values cw, sv, reference level cw) as manipulated variables. The data task1_CIR_data contains all data points where Type is CIR. The model output is given in Table 5.

For Hypothesis 3 we used the following two regression models:

- glm(formula = Grounder ~ ROI + Power, family = binomial, data = task1_data)
- glm(formula = Downtoner ~ ROI + Power, family = binomial, data = task1_data)

with Grounder and Downtoner (values USED, NOT_USED), respectively, as measured variables, and ROI (values L, M, H, reference level M) and Power (values

Table 5: H2 model output: logistic regression of CIR subtype (can vs. could) on ROI and Power.

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	0.3891	0.3455	1.126	0.260
ROIL	−0.2853	0.4701	−0.607	0.544
ROIH	0.1669	0.4993	0.334	0.738
Powersv	−0.5291	0.4103	−1.290	0.197

cw, sv, reference level cw) as manipulated variables. The model outputs are given in Tables 6 and 7.

Table 6: H3 model output: logistic regression of grounder use on ROI and Power.

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	1.71278	0.37824	4.528	0.00006 ***
ROIL	0.64332	0.50609	1.271	0.204
ROIH	0.06112	0.44459	0.137	0.891
Powersv	−0.45038	0.39328	−1.145	0.252

Table 7: H3 model output: logistic regression of downtoner use on ROI and Power.

	Estimate	Std. Error	z value	Pr(> |z|)
(Intercept)	-2.80725	0.65710	−4.272	0.0005 ***
ROIL	−17.44398	2180.90054	−0.008	0.994
ROIH	0.07242	0.83778	0.086	0.931
Powersv	−0.74165	0.88408	−0.839	0.402

Note: In these analyses, Rank of Imposition was entered as a categorical predictor using dummy (treatment) coding with the mid level as the reference. Following a reviewer's suggestion, we conducted additional analyses treating the

Rank of Imposition as an ordered factor and applying planned forward contrasts that test for a gradual increase from low to mid to high imposition. Specifically, two orthogonal contrasts compared (a) low versus mid/high and (b) high versus low/mid imposition. Results of these analyses closely mirrored those obtained with dummy coding: the direction, magnitude, and significance of the effects remained unchanged. This convergence indicates that the reported effects are robust with respect to the choice of contrast coding scheme.

We additionally tested whether the effect of Rank of Imposition interacted with Power by including the interaction term between Power and Rank of Imposition in the logistic regression models. These analyses were conducted both for the dummy-coded models reported in the main text and for the alternative models using planned forward contrasts. In neither case did the interaction reach statistical significance, and the inclusion of the interaction term did not alter the pattern or interpretation of the main effects. We therefore report the more parsimonious models without interaction terms.

Task 2 Analyses

For testing effects across ratings, we used the following mixed effects model:

- lmer(formula = Rating ~ ALT + (1 | PID), data = task2_data)

with Rating (values 1 to 5) as measured variable, ALT (values ALT DIR, ALT CAN, ALT CLD and ALT WLD, reference level ALT CAN) as manipulated variable, and PID (participant ID) as random variable. The model output is given in Table 8.

Table 8: Task 2 model output: linear mixed effects model of RA ratings on alternative type.

	Estimate	Std. Error	t value	Pr(> |z|)
(Intercept)	2.43961	0.06549	37.251	$2e^{-16}$ ***
ALT DIR	−0.54106	0.06483	−8.345	$4e^{-16}$ ***
ALT CLD	0.16425	0.06483	2.533	0.0115 *
ALT WLD	0.86957	0.06483	13.412	$2e^{-16}$ ***

Note: In this analysis, ALT was entered as a categorical predictor using dummy (treatment) coding with CAN as reference. Following a reviewer's suggestion, we

conducted an additional analysis treating ALT as an ordered factor and applying planned forward contrasts that test for a gradual increase from ALT DIR to ALT CAN to ALD CLD to ALT WLD. Results of this analysis closely mirrored the one obtained with dummy coding: the direction, magnitude, and significance of the effects remained unchanged. This convergence indicates that the reported effects are robust with respect to the choice of contrast coding scheme.

Furthermore, note that although the main analyses treat evaluation ratings as approximately interval-scaled, the ratings were collected on a 5-point Likert scale and are therefore ordinal in nature. To assess whether this modeling choice affected the results, we additionally fitted the following ordinal mixed-effects regression model using the package *ordinal* (Christensen 2025):

- clmm(formula = Rating ~ ALT + (1 | PID), data = task2_data, link = "logit")

The ordinal analysis produced the same pattern of effects as the linear mixed-effects model, and all substantive conclusions remained unchanged. The model output is given in Table 9.

Table 9: Task 2 ordinal model output: cumulative link mixed model of RA ratings on alternative type.

	Estimate	Std. Error	z value	Pr(> |z|)
ALT DIR_vs_rest	1.8490	0.2186	8.458	$2e^{-16}$ ***
ALT DIRCAN_vs_ALT CLD-WLD	0.5180	0.1933	2.679	0.00738 **
ALT DIRCANCLD_vs_ALT WLD	2.0740	0.2130	9.735	$2e^{-16}$ ***

To test if the ratings interact with the participant's own formulations' head act type from Task 1, we fitted the following model:

- lmer(formula = Rating ~ Type * ALT + (1 | PID), data = task2_data)

with Rating (values 1 to 5) as measured variable, Type (values CAN, CLD, NCIR, reference level CAN) and ALT (values ALT DIR, ALT CAN, ALT CLD and ALT WLD, reference level ALT CAN) as independent variables, and PID (participant ID) as random variable. We did not find any significant interaction.

Task 3 Analyses

For testing effects for reason frequencies across alternatives, we used the following mixed effects model for each reason in {TOO DIRECT, NOT POLITE, TOO PUSHY, CONCRETE WORDING, MISSING REASON, NO PROMISE}:

- bglmer(formula = reason ~ ALT + (1 | PID), data = task3_data, family =binomial(), fixef.prior = normal())

with reason (frequency of the respective reason) as measured variable, ALT (values ALT DIR, ALT CAN, ALT CLD and ALT WLD, reference level ALT CAN) as manipulated variable, and PID (participant ID) as random variable. The resulting significant effects are given in Table 10.

Table 10: Task 3 significant effects: reason frequencies across alternative types (bglmer models).

reason	ALT	Pr(> |z|)
TOO DIRECT	ALT CAN vs ALT DIR	0.006 **
TOO PUSHY	ALT CAN vc ALT DIR	$1.6e^{-09}$ ***
CONCRETE WORDING	ALT CAN vc ALT DIR	0.0005 ***
MISSING REASON	ALT CAN vc ALT DIR	0.00037 ***
	ALT CAN vc ALT CLD	0.00096 ***
NO PROMISE	ALT CAN vc ALT DIR	0.0116 *
	ALT CAN vc ALT CLD	0.0191 *

Task 4 Analyses

For testing effects for reason frequencies across alternatives, we used the following mixed effects model for each reason in {NOT FRIENDLY, TOO PUSHY, NOT RESPECTFUL, TOO INSECURE, TOO SHORT, LESS LIKELY}:

- glmer(formula = reason ~ ALT + (1 | PID), data = task4_data, family = binomial())

with reason (frequency of the respective reason) as measured variable, ALT (values ALT DIR, ALT CAN, ALT CLD and ALT WLD, reference level ALT CAN) as manipulated variable, and PID (participant ID) as random variable. The resulting significant effects are given in Table 11.

Table 11: Task 4 significant effects: reason frequencies across alternative types (glmer models).

reason	ALT	Pr(> |z|)
NOT FRIENDLY	ALT CAN vs ALT DIR	0.00127 **
	ALT CAN vs ALT WLD	0.00851 **
TOO PUSHY	ALT CAN vs ALT DIR	$3.8e^{-09}$ ***
	ALT CAN vs ALT WLD	0.000285 ***
NOT RESPECTFUL	ALT CAN vs ALT DIR	0.002155 **
	ALT CAN vs ALT WLD	0.000145 ***
TOO INSECURE	ALT CAN vc ALT WLD	0.000134***
TOO SHORT	ALT CAN vs ALT DIR	0.0143 *
	ALT CAN vs ALT WLD	$4.5e^{-06}$ ***

We additionally tested whether there are interactions with the experimental conditions POWER and ROI. For testing such interactions, we used the following mixed effects models for each reason in {NOT FRIENDLY, TOO PUSHY, NOT RESPECTFUL, TOO INSECURE, TOO SHORT, LESS LIKELY}:

- glmer(formula = reason ~ ALT * ROI + (1 | PID), data = task4_data, family = binomial())
- glmer(formula = reason ~ ALT * Power + (1 | PID), data = task4_data, family = binomial())

with reason (frequency of the respective reason) as measured variable, ALT (values ALT DIR, ALT CAN, ALT CLD and ALT WLD, reference level ALT CAN), and ROI (values L, M, H, reference level M) or Power (values cw, sv, reference level cw), respectively, as manipulated variables, and PID (participant ID) as random variable.

We found two significant interactions. First of all, we found that the effect of NOT FRIENDLY being mentioned significantly less often for the appropriateness of ALT WLD compared to ALT CAN was significantly stronger in the SV condition than in the CW condition ($p = 0.041$). Second, of TOO INSECURE being mentioned significantly more often for the appropriateness of ALT WLD compared to ALT CAN was significantly stronger in the SV condition than in the CW condition ($p = 0.042$).