Bias Effect on Response Preference

Maryam Mohammadi
University of Konstanz

1 Introduction

Pope (1976) points out that cross-linguistically, polarity particles (PolPrts) may serve two distinct purposes. They can indicate a response as being either positive or negative, referred to as the *polarity* reading. Alternatively, they can express agreement or disagreement with an initiative (including assertions and polar questions), which I call the *conformity* reading.[1] In certain languages, such as English, the same particles can have both readings. For example, *Yes* can indicate a positive response in the polarity reading or agreement with an initiative in the conformity reading, while *No* can express a negative response or disagreement with an initiative in the respective readings. Consequently, PolPrts in either reading express the same proposition in response to positive initiatives, as shown in example (1), but generate different propositions in response to negative forms, as in example (2).

(1) A: Did John come to the party?
B1: Yes, he did.
B2: No, he didn't.

(2) A: Did John not come to the party?
B1: Yes, he DID.
B2: Yes, he didn't.
B3: No, he DID.
B4: No, he didn't.

The interchangeable usage of PolPrts with both positive and negative answers in Example (2) raises three significant concerns. Firstly, it highlights the role of prosody, presented by CAPS LOCK in (B1) and (B3). Secondly, it draws attention to the ambiguity of bare particles in the absence of a short answer, such as *he did*/*he didn't*. Lastly, it emphasizes the challenge of selecting the optimal particle from the possible options. This study specifically focuses on addressing the last issue and presents experimental evidence from Farsi regarding particle preferences (for a comprehensive understanding of the main accounts of PolPrts and discussions about the first two issues, see Kramer & Rawlins 2009 Krifka 2013, Roelofsen & Farkas 2015, Claus *et al.* 2017, Goodhue & Wagner 2018, among others).

Pope (1976) discusses the role of bias in relation to the polarity and conformity readings. She suggests that languages which indicate speaker's expected (biased) answer in polar questions, such as Japanese, tend to utilize the conformity reading system more frequently. Conversely, languages with less explicit expected answers tend to exhibit a preference for the polarity reading system. However, the author

[1] This reading is commonly referred to as the *truth-based* or *(dis)agreement* reading. However, the term truth-based might imply the other reading does not involve true answers. Additionally, the term (dis)agreement can be confusing, as *agreement* and *disagreement* are used to denote the feature values of particles (*Yes* and *No*, respectively). For the sake of clarity, I call it *conformity* reading.

does not specify the optimal reading in languages where the same particles can be used in both readings. Building on this discussion, Mohammadi (2022) presents an observation wherein the stronger the bias for a given alternative, the more the conformity reading is accepted by the speakers. This study aims to substantiate this observation with experimental data. I posit that bias significantly influences the decision-making process when selecting the particle.

While all polar questions, [ϕ?], semantically provide the possible answers, namely $\{\phi, \neg\phi\}$ (Heim 1983), in proper contexts they can exhibit a bias favoring one answer over the other. However, certain types of questions inherently express bias in all contexts (see Ladd 1981, van Rooij & Safarova 2003, Romero & Han 2004, Krifka 2015, Farkas & Roelofsen 2017 amongst all). In other words, some question types can (but not necessarily) signal bias in suitable contexts, but one can also cancel the potential bias. I refer to these questions as weakly biased (WB) questions. Conversely, other question types obligatorily express bias in all contexts, rendering bias cancellation contradictory. I call them as strongly biased (SB) forms.

An experiment was conducted in Farsi to investigate speaker response preferences based on different types of initiatives. I provided plain NPQs as WB questions, NPQs+*dige* as SB forms due to the inclusion of the biased particle *dige*, and (negative) assertions as the extreme case of bias in which the speaker fully commits to an answer. The results revealed a significant impact of bias on the conformity reading. In short, the polarity reading consistently received high acceptability regardless of the type of initiative. However, the acceptability of the conformity reading in response to WB initiatives was only marginally accepted, whereas its acceptability improved in SB forms. The increase in acceptability for the conformity reading can be attributed to the presence of bias in questions, which is obligatory in the case of SB but not in WB. In SB forms, particle *dige* signifies the speaker's expectation for the uttered proposition in the question, allowing the addressee to express agreement or disagreement with the speaker. However, WB forms might not trigger this implicature, resulting in the dispreference of such responses.

This paper is structured as follows. The next section presents the experiment, offering detailed information on the materials in Section 2.1, the procedure in Section 2.2, and the results in Section 2.3. The findings are discussed in Section 3, where a dual-pathway approach to selecting the optimal particle is introduced. Section 4 provides a review of previous accounts, incorporating the findings from our experiment. Finally, the paper concludes in Section 5.

2 The Experiment

This experiment aims to investigate the factors influencing the selection of particles. I employed Farsi as the object language, wherein the particles *âre* 'Yes' and *na* 'No' can be used in both polarity and conformity readings, akin to English. These readings are denoted by superscripts: ^{Pos} and ^{Neg} denote positive and negative features in the polarity reading, while ^{Agr} and ^{DAgr} denote agreement and disagreement features in the conformity reading. Recalling the bias observation presented in Mohammadi (2022), in response to negative polar questions (NPQs) like (3) as WB questions, the polarity responses with *âre*^{Pos} in the Rejecting answer (B1) and *na*^{Neg} in the Accepting answer (B2) were favored ($\succ$) over the alternative particles.

(3) A: Ali mehmuni na-raft?
Ali party NEG-went
'Did Ali not go to the party?'

B1: $\hat{a}re^{Pos} \succ na^{DAgr}$ raft.
yes no went
'Yes ≻ No, he did.'

B2: $\hat{a}re^{Agr} \prec na^{Neg}$, na-raft.
yes no NEG-went
'Yes ≺ No, he didn't.'

On the other hand, in NPQs+*dige* as SB forms, exemplified in (4), particle *dige* obligatorily signifies the speaker's bias towards the negative answer. The observation reveals that the acceptability of na^{DAgr} in the Rejecting answer (B1) and $\hat{a}re^{Agr}$ in the Accepting answer (B2) were comparable (≈) to their corresponding readings.

(4) A: Ali mehmuni na-raft dige?
Ali party NEG-went DIGE
'Did Ali not go to the party?'
⇝ The speaker expects that Ali didn't go.

B1: $\hat{a}re^{Pos} \approx na^{DAgr}$ raft.
yes no went
'Yes ≈ No, he did.'

B2: $\hat{a}re^{Agr} \approx na^{Neg}$, na-raft.
yes no NEG-went
'Yes ≈ No, he didn't.'

It is worth noticing that the studies of biased questions usually consider two types of bias, namely original bias (OB) as the speaker's prior belief and contextual bias (CB) as the contextual evidence in the current setting (see Ladd 1981 Büring & Gunlogson 2000, van Rooij & Safarova 2003, Romero & Han 2004, among others). To avoid any complexity, in this study I excluded CB and offered examples with OB. This experiment addresses the following two questions:

1. Does the bias in the antecedent influence the reading preference of PolPrts?
2. Is the reading preference consistent across accepting and rejecting answers?

2.1 Materials

Since both readings of PolPrts in response to positive initiatives generate the same proposition (e.g., *Did John come? Yes^{Pos}/Yes^{Agr}, he did*), positive initiatives cannot help us to distinguish the reading preference. Therefore, I presented all samples in negative form and included PPQs in the control group, where some response combinations are unacceptable (e.g., *Did John come? #Yes, he didn't.*). I designed the experiment with two main factors. The first factor includes NPQs as WB questions, NPQs+*dige* as SB forms, and negative Assertions (ASS) as the extreme case of bias, indicating the speaker's commitment. The second factor pertains both readings of PolPrts. Thus, we have a 3x2 design as follow:

- **Antecedent type**: WB (NPQs), SB (NPQs+*dige*), (neg) ASS
- **Reading type**: polarity, conformity

All stimuli were constructed using the same structure, including a short scene-setting that provides relevant information about the question and the answer at the

end of the setting. The setting suggests the speaker's prior belief/expectation (OB), as well as the true answer from the addressee's point of view. The response could be either an *Accepting* answer, in which the addressee accepts the uttered proposition, $\neg p$, in the initiative, or a Rejecting answer which rejects the proposition. The concepts are regardless of the polarity of the initiative and the answer. For example, in the sample stimuli shown in Table 1, Sara expects that Leila wouldn't lend her book as usual, but based on the context, the answer that Ali (the addressee) would provide is a rejecting response, in which Leila lent her book.

Context: *Ali wanted to borrow a book from Leila. Leila is quite protective of her books and usually finds excuses not to lend them out. Sara thinks that Leila probably didn't lend her book as usual, even though she did in reality.*	
Sara: *Leila ketâbesh ro behet qarz na-dâd dige*? Did Leila not lend her book DIGE?	**Answer type**
Ali: *âre dâd.* Yes gave ('Yes, she did')	**(i)** polarity: $\hat{a}re^{Pos}$
Ali: *na dâd.* No gave ('No, she did.')	**(ii)** conformity: na^{DAgr}
Ali: OK.	**(iii)** unrelated answer
Ali: She couldn't make any excuse.	**(iv)** indirect answer

Table 1: Sample stimuli: NPQs+*dige* with rejecting answers.

The presented answers include four options: two of them feature the target particles *âre* and *na* in (i) the polarity reading and (ii) the conformity reading. Notice that the particles were followed by a short answer, the predicate in the prejacent.[2] The other two answers serve as the control group by (iii) an unrelated answer, e.g., *xob* or *bâše* 'OK', which is infelicitous as a response to the related initiative (see Krifka 2016), and (iv) an indirect answer that addresses the question, e.g., by adhering to the Maxim of Relevance. Such responses felicitously address the questions.

The last two responses, namely the unrelated and indirect answers, serve the dual purpose of functioning as attention checks to ensure participants' attentiveness during the test, as well as introducing distractors and reinforcing the idea that multiple answers could be considered felicitous. Notably, the indirect answers were intentionally designed to always remain felicitous, irrespective of the speaker's preferred reading of PolPrts. This enables participants to recognize that multiple answers are acceptable to varying degrees.

2.2 Procedure

The experiment consisted of a total of 22 stimuli, including 18 core questions, with 6 trials for each target antecedent type, and 4 control stimuli of PPQs. These stim-

[2]Since the experiment aims to examine the effect of bias on reading preference, the ambiguous bare forms were excluded. The predicate in the short answer clearly indicates the meaning.

uli were distributed across four lists following a Latin-square, pseudo-randomized design. Participants were allocated three tokens for each condition. All examples were chosen from the commonly used predicates in everyday speech. The examples were formulated in the simple past tense and second or third-person singular subjects. The two interlocutors consistently represented a female and a male character throughout all contexts. However, they assumed different arbitrary roles (e.g., sister-brother, wife-husband, mother-son, colleagues, etc.) in each distinct context.

The experiment was administered online via the www.testable.org platform. Participants were initially provided with instructions on the first page, outlining the overall experiment design and specifying the number of trials. They were instructed to attentively read the scene-setting provided and subsequently rate the naturalness of each individual response on a scale of 1-7 (1=the least natural, 7=the most natural). Each trial was presented on a separate page, and following each trial, participants were encouraged to share their feedback on the responses in a "Comment" box. Some participants provided feedback related to their interpretation of the particles, the way they need to pause between the particle and the predicate or the importance of emphasizing the particle for correct interpretation.

2.3 Result

Data were collected from 40 participants, consisting of 23 women and 17 men (AVG age = 36 years).[3] They were randomly divided into six lists. All participants were native speakers of Farsi and had IP addresses limited to Iran. Four participants were excluded from the analysis because they failed more than five attention checks by rating unrelated responses higher than 4. From the data in Figure 1, it is evident that indirect responses to all four types of initiatives were highly accepted (mean=6, median=7), compared to the unacceptability of unrelated ones (mean=1.6, median=1). Additionally, in response to PPQs, participants truly declined [*âre*, ¬*p*] in Accepting answers and [*na*, *p*] in Rejecting answers (see Figure 4, PPQs-plot).

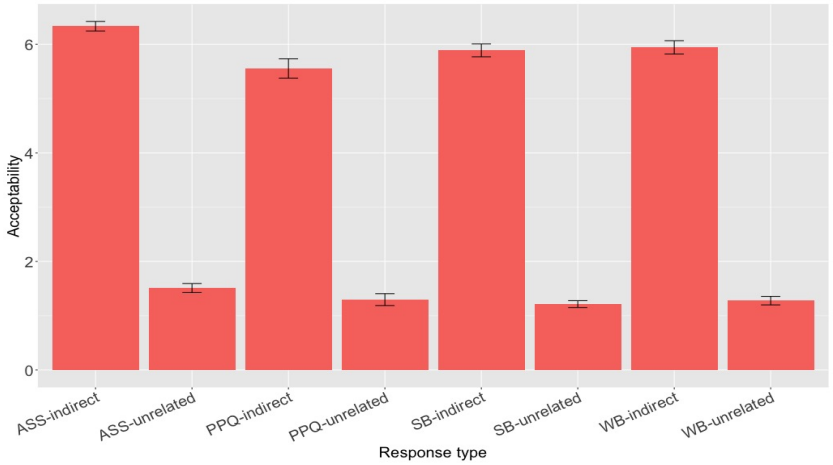

Figure 1: Acceptability judgement of control responses to different types of initiatives.

[3] I extend my sincere gratitude to all the participants who took part in the experiment. The study was conducted during a period when Iran was experiencing significant social unrest.

The data was analyzed using a mixed-effects ordinal regression model fitted with the Cumulative Link Mixed Model (clmm) package in R (ver. 4.1). The model in Figure 2 included the maximal random effects structure that allows for the convergence of the random by-participant intercepts, capturing participants' variability in responses. The results revealed a significant difference between the readings in WB ($\beta = 0.67$, $SE = 0.26$, $p < 0$) and SB ($\beta = 2.73$, $SE = 0.27$, $p < 0.001$), and non-significant difference in Assertions ($\beta = -0.31$, $SE = 0.19$, $p < 0.05$).

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: ResponseValue ~ Bias * Reading + (1 | ParticipantID)
data:    clm_tbl


Random effects:
 Groups        Name        Variance Std.Dev.
 ParticipantID (Intercept) 0.6989   0.836
Number of groups:  ParticipantID 36

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
BiasSB                 -0.6824     0.1846  -3.696 0.000219 ***
BiasWB                 -2.2963     0.1888 -12.165  < 2e-16 ***
ReadingPolarity        -0.3185     0.1853  -1.719 0.085605 .
BiasSB:ReadingPolarity  0.6737     0.2580   2.611 0.009029 **
BiasWB:ReadingPolarity  2.7359     0.2691  10.167  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2  -3.0397     0.2157 -14.089
2|3  -2.6001     0.2105 -12.349
3|4  -2.0025     0.2050  -9.768
4|5  -1.5310     0.2018  -7.588
5|6  -1.0375     0.1993  -5.206
6|7  -0.4636     0.1976  -2.347
```

Figure 2: Cumulative link mixed model for naturalness ratings.

Plot (a) in Figure 3 illustrates the correlation between the reading preference and bias in the initiative. The acceptability of the conformity reading exhibits a gradual increase from weakly biased questions (mean = 3.3, median = 3), to strongly biased ones (mean = 5.1, median = 6), and to assertions (mean = 5.8, median = 7). Conversely, the acceptability of the polarity reading remains consistent across different initiatives (mean = 5.6, median = 7). Statistical analysis reveals the difference in acceptability for the polarity reading between different antecedents is not significant (mean = $5.34 \simeq 5.81 \simeq 5.99$), whereas it is significant for the conformity reading in WB as compared to SB and Assertions (mean = $3.31 < 5.11 \simeq 5.80$, respectively).

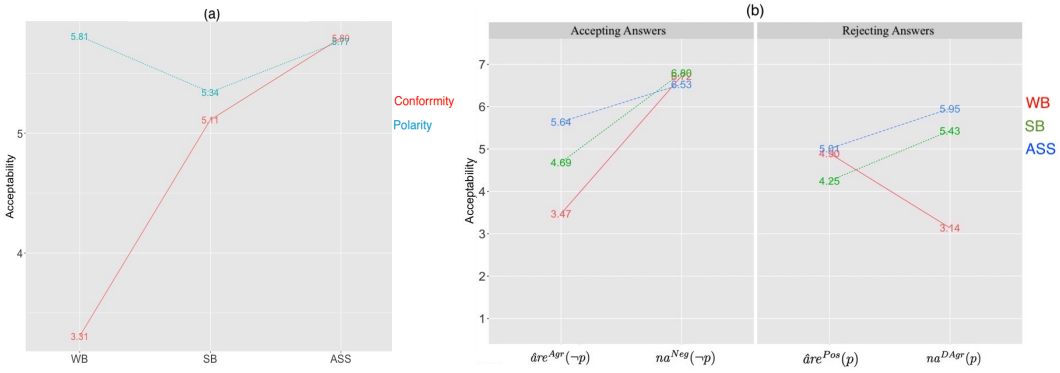

Figure 3: Mean acceptability rating of PolPrts' readings in response to negative initiatives.

Plot (b) in Figure 3 gives us a closer look at the findings in Plot (a), regarding the answer type. First, the polarity reading in both Acceptings and Rejectings consistently received high rating in response to various types of initiatives. Statistically, na^{Neg} in Accepting answers and $\hat{a}re^{Pos}$ in Rejectings have the mean values of 6.6, and 4.7, respectively. Second, the conformity reading in both answer types exhibited similar behavior. The acceptability rating increases with bias in the following order: WB < SB < ASS. In Accepting answers, the mean values of $\hat{a}re^{Agr}$ are 3.5 < 4.7 < 5.6, while in Rejectings, the values of na^{DAgr} are 3.1 < 5.4 < 5.9. Furthermore, the overall rating of *na* is higher than *âre*, which holds true in all cases except in Rejecting to WB questions. That is, while participants prefer *na* in Accepting and Rejecting to both SB and Assertions, they prefer $\hat{a}re^{Pos}$ over na^{DAgr} in Rejecting to WB. Finally, the plot demonstrates that the polarity reading in the rejecting answers ($\hat{a}re^{Pos}$) gains lower rating as compared to its counterpart (na^{Neg}) in the Acceptings.

An interesting observation emerging from the data is the distribution of judgments in Figure 4. In response to PPQs[4] and assertions, participants show a considerable level of convergence. In PPQs, they consistently accept and reject responses based on the related follow-up prejacent. Building on our earlier findings, on one hand, we observe a close competition between both readings in response to assertions, resulting in little variation between the polarity and conformity readings.

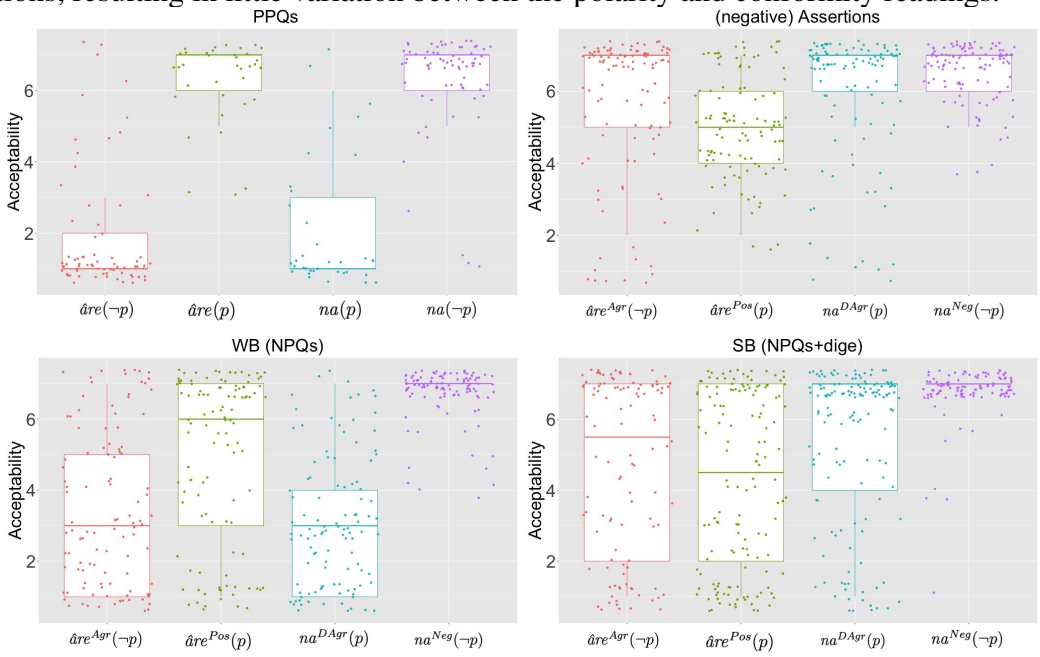

Figure 4: Distribution plots of both readings of *âre* and *na* in response to different initiatives

On the other hand, the distribution in WB and SB plots is not as monotonic as in the earlier types. In WB NPQs, there is a great tendency towards the polarity reading compared to the divergence in the conformity reading (which tends

[4]PPQs were not considered as a factor for the least biased form but rather served as a control group. This choice was made because in PPQs, both readings of PolPrts yield the same proposition, making the readings indistinguishable (as the x-axis of PPQs is not labeled with any superscription).

towards unacceptability). However, in the SB plot, we can observe a gradual shift of judgments towards the pattern seen in assertions. Notably, na^{Neg} exhibits a steady behavior across the figures, showing less variation among participants.

Lastly, Figure 5 presents the acceptability of different readings within and between participants. Overall, there is a notable difference between the conformity reading and the polarity reading across the plots. In the WB plot, the acceptability of the conformity reading significantly deviates from that of the polarity reading. However, in the SB plot, the conformity reading exhibits a slight improvement, and in the ASS(ertion) plot, it reaches a higher level of compatibility. This pattern indicates that speakers generally display a preference for the polarity reading in weakly biased forms, but as the bias strength in the initiatives increases, their dispreference for the conformity reading gradually diminishes. Importantly, this behavior is evident not only between participants but also within individual participants.

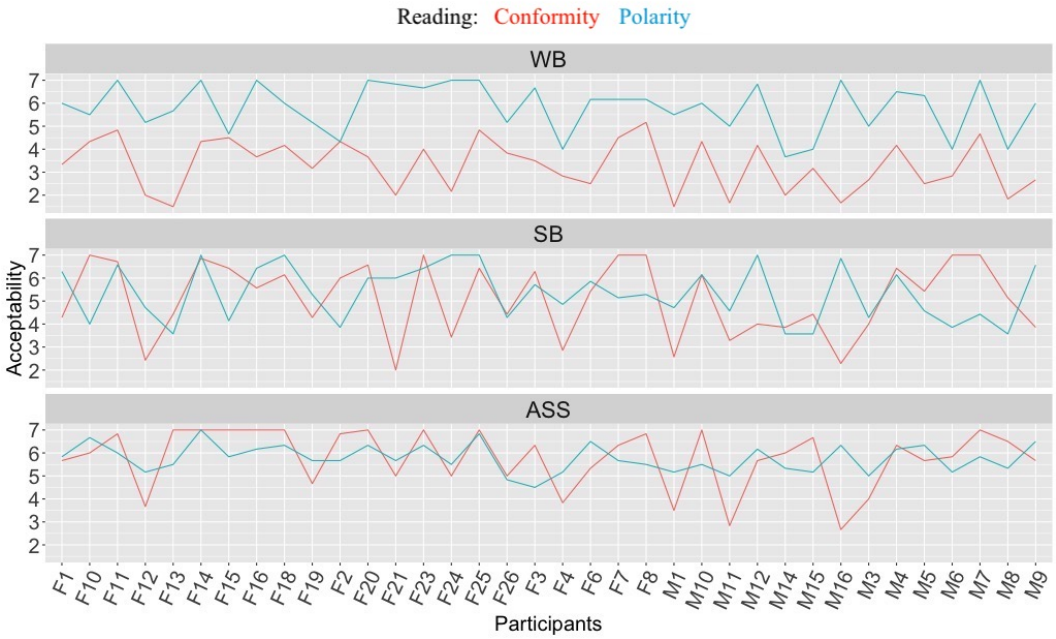

Figure 5: The acceptability judgement of each reading of PolPrts per participant.

To conclude, the results show that bias influences the preference of different readings of PolPrts. The data reveals **(i)** the polarity reading consistently maintains high acceptability across different initiative, **(ii)** the acceptability of the conformity reading improves regarding the presence of bias in the initiative, **(iii)** particle *na* is generally preferred over *âre* in both readings, except for its conformity reading in WB forms, and **(iv)** reading variation occurs between and within speakers.

3 Discussion

Previous research has predominantly concentrated on the selection of an optimal particle for expressing positive and negative answers, often overlooking the type of initiatives. In contrast, the present investigation delves into the influence of the

initiative factor on the preference for particular readings of PolPrts. This approach draws inspiration from Pope (1976)'s work, which underscored the pivotal role of bias in reading systems across languages.

Let's start with the research questions: (1) Does the bias in the antecedent influence the reading preference of PolPrts? (2) Is the reading preference consistent across Accepting and Rejecting answers? The answer to both questions is yes. Regarding the first question (1), the judgments of the conformity reading display a notable trend of improvement from a state of marginal acceptability in WB NPQs to a significantly higher degree of acceptability when bias is obligatorily expressed in NPQs+*dige* and assertions. Conversely, the polarity reading consistently receives high acceptability across responses to various types of initiatives. Concerning the second question (2), the analysis indicates that the observed reading preference patterns, as elucidated in (1), remain consistent for both Accepting and Rejecting answers. Notably, the data reveals no meaningful correlation between the acceptability of the two readings. Specifically, the polarity reading consistently maintains a high level of acceptability across different types of initiatives, irrespective of the gradual improvement observed in the conformity reading—from marginal acceptability in WB forms to high acceptability in SB forms and Assertions.[5]

The result is suggestive for a double-pathway in choosing the response particle. I argue that the decision for the optimal particle is related to the particle's reading and involves a semantic-pragmatic competition. The semantics of polar initiatives presents two possibilities, $\{p, \neg p\}$. Consequently, in the first pathway, the addressee can utilize the particles in the polarity reading, which only indicate the polarity of the answer, resulting in a constant acceptability of the polarity reading as it directly addresses the discussion. In the second pathway, the presence of the speaker's bias in the initiative prompts the addressee to express agreement or disagreement with the speaker's expectation. Hence, the addressee has the option to follow a new pathway. While the addressee may opt for the first path and respond to various initiatives using the polarity reading, the existence of bias opens up the second pathway to respond to the bias in certain initiatives. In WB NPQs, however, the bias implicature is not always in play. Consequently, in such instances, the addressee may find no reason for expressing agreement or disagreement. As a result, the addressee's inclination toward the conformity reading in WB questions exerts the influence on the overall acceptability of the conformity reading in these WB forms.

Furthermore, it is evident that speakers exhibit a consistent particle preference, with *na* being predominantly preferred in both Accepting and Rejecting answers, which results in the lower rating of the polarity reading in the Rejecting. The only significant deviation occurs in Rejecting answers to weakly biased questions, where $\hat{a}re^{Pos}$ is preferred over na^{DAgr}. This finding aligns with the predicted effect of bias, where the conformity reading has low acceptability in WB forms. Consequently, in WB initiatives, the conformity reading is suboptimal for both Accepting and Rejecting responses, leading to a loss of exclusive preference for particle *na* in Rejectings (for more discussions see Section 4). Finally, the overall lower rating

[5]Regarding the feedbacks from the participants as well as my own intuition as a native speaker, I find the conformity reading in response to WB forms hardly acceptable. However, from a statistical perspective, its acceptability mean value is 3.3 out of 7. Thus, when compared to the unacceptability of the control group, which has a mean value of 1.6, I would classify it as marginally acceptable.

of the polarity reading by *âre*^{Pos} in the rejecting answers, as compared to the *na*^{Neg} in the accepting answers, may be somewhat attributed to the principle proposed by Farkas & Bruce (2009), which suggests that rejecting a proposition is more costly than accepting it (cf. the notion of semantic difficulty in studies such as Pope 1976 and Roelofsen & Farkas 2015). However, this raises the question of why this difficulty didn't have a similar impact on the other rejecting answer involving *na*^{DAgr}. I leave this question for further investigation.

The findings also demonstrate reading variations both within and between speakers, which can be attributed to the factor of bias. Within-speaker variations can be traced back to the influence of bias, as previously discussed about different initiatives. Similarly, the reading variation among speakers in response to a specific type of initiative, can be attributed to the varying degrees of bias sensitivity exhibited by different individuals. In essence, speakers have distinct levels of personal sensitivity to bias, leading to instances where one speaker deems the conformity reading felicitous, while another, who exhibits a lower sensitivity to bias, perceives it as infelicitous or unnecessary. This observation substantiates our overarching claim.

4 Previous Accounts

The interchangeable use of PolPrts, which can convey both positive and negative propositions in response to negative initiatives, gives rise a general issue: which particle is the optimal choice for expressing an Accepting and Rejecting answer? We saw the influence of bias in the initiatives on the particle's preference. This section provides an analysis of our empirical findings within the framework of two main approaches to PolPrts: the discourse salient model proposed by Krifka (2013) and the feature marker model by Roelofsen & Farkas (2015) (for a comprehensive analysis of these frameworks, see the relevant studies).

Krifka introduces two constraints, namely **NonSalient* and **DisAgreement*, which penalize the selection of the discourse referent (DR) that is not salient or imply disagreement. He assumes in "typical cases" of negative initiatives, e.g., [$\neg q$?], a positive proposition serves as the salient/highlighted DR, q^h, while in other cases, the negative proposition is highlighted, $\neg q^h$. He also notes that NPQs plausibly express bias, albeit weaker than assertions. Based on that the constraints have different weightings for different initiatives. In the case of SB forms, the **NonSal* constraint incurs a lower penalty compared to the **DisAgr* constraint. However, in the case of WB forms, the costs are reversed, with the **DisAgr* constraint having a lower penalty than the **NonSal* constraint. Applying these constraints to example (5), different predictions arise for negative initiatives with highlighted q^h and $\neg q^h$. When q^h is salient, for Rejecting answer (B1), particle *âre* should be preferred over *na*, as *âre* incurs a penalty for violating the **DisAgr* constraint, while *na* receives penalties for both violating the **NonSal* and **DisAgr* constraints. For Accepting answer (B2), particle *na* should be favored over *âre*, as *âre* violates **NonSal*, while *na* violates no constraints. Conversely, when $\neg q^h$ is salient, the optimal particles are predicted to be vice versa (see Krifka 2013 for details).

(5) A: Ali came to the party. / Did Ali not come to the party?
B1: *âre* / *na* he did. [*âre*^{Pos}/*na*^{DAgr}] (p)
B2: *âre* / *na* he didn't. [*âre*^{Agr}/*na*^{Neg}] ($\neg p$)

Turning to the feature marker account, Roelofsen & Farkas (2015) introduce two constraints: *Markedness* and *Realization*. They argue that the more marked a feature is, the stronger the pressure to be overtly expressed/realized. It is worth mentioning that the authors did not distinguish between different types of initiatives. Following this perspective, in both responses, particle *na* is predicted to be preferred over *âre* due to the greater markedness of the [DAgr] and [Neg] features compared to the [Pos] and [Agr] features in (B1) and (B2), respectively. Neither of the accounts presented predictions that entirely align with our findings. It is important to note that I am not claiming an exact correspondence between the bias implicatures in English NPQs and Farsi NPQs. However, it can be assumed as a fundamental premise that NPQs generally exhibit a lesser degree of bias compared to assertions, regardless of the language. Consequently, I posit that there exists a certain level of overlap in both form and bias cross-linguistically.

The predictions put forth by Krifka regarding initiatives involving q^h are in alignment with our empirical findings for WB NPQs, since they shared the concept of bias effect. While the predictions of $\neg q^h$ differ from our observed data, they still show consistency in terms of the high acceptability of the conformity reading. Hence, we can disregard the **NonSal* constraint in SB forms, assuming bias renders $\neg q$ equally salient as q. Given that particle *na* is a complement function in this framework, I will posit the positive proposition as highlighted one in all negative forms.[6] Therefore, in SB forms, we only have the **DisAgr* constraint. In case of a Rejecting answer p (B1), both *âre* and *na* violate the **DisAgr* constraint, while in an Accepting $\neg p$ (B2), neither of the particles violates any constraint. Hence, both particles are predicted to be equally acceptable, consistent with our data.[7]

Roelofsen & Farkas provide examples including NPQs and assertions. They propose that the same optimal particle, *na*, for both Acceptance and Rejection, regarding the more markedness of [Neg] and [DAgr] features compared to the [Pos] and [Agr] features. While their prediction aligns with our finding that *na* generally exhibits higher acceptability, their account fails to explain why the unmarked [Pos] feature should be preferred over the marked [DAgr] feature in WB forms. Building upon the work of Claus *et al.* (2017), Roelofsen & Farkas (2019) present an update in Optimality Theory. They introduce an additional constraint, *Maximize Marked*, which captures inter-speaker grammatical variation by favoring the realization of relative features over absolute features or vice versa. To accommodate reading variations observed in German speakers, they define *Maximize Marked* and *Maximized Relative* constraints (for detailed information, see R&F 2019). Extending the maximize constraints to our data, we can introduce the *Maximize Absolute* constraint for Farsi speakers, aiming to maximize the realization of absolute polarity features.

[6]One can consider revising the account by treating both particles as identity functions, since this aspect is not crucial. Nonetheless, the revision might affect the rules governing DR selection. It is worth noting that Claus *et al.* (2017) did not observe divergent acceptability of PolPrts in contexts with distinct highlighted propositions among German speakers.

[7]Notice that the **DisAgr* constraint in WB forms seems unnecessary, and we can disregard it without compromising the predictions. In response to WB forms, the speaker simply asks a question, and accepting or rejecting it are both acceptable (similar to PPQs), thus, there is no need to penalize disagreement responses. We still obtain the correct predictions, where *âre* $\succ$ *na* (p) in Rejecting since *âre* does not violate any constraints, while *na* violates **NonSal*. Moreover, *âre* $\prec$ *na* ($\neg p$), as *âre* violates the **NonSal* constraint and *na* receives no penalty.

According to the newly introduced constraint, in WB initiatives, *âre*^{Pos} in (B1) and *na*^{Neg} in (B2) are preferred over the alternative particles in the conformity reading. This preference arises from the tendency of speakers regarding the new rules to prioritize the realization of absolute polarity features over relative polarity features. Furthermore, concerning the data and our analysis, the Maximize Absolute constraint can be disregarded in SB initiatives. In other words, in SB cases, speakers assign similar weights to the realization of [Pos]/[Neg] and [Agr]/[DAgr] features. As a result, both *âre* and *na* in (B1) and (B2) become equally acceptable, providing the flexibility to express either absolute or relative features.

5 Conclusion

The results of our experimental study provide compelling evidence of the impact of bias on reading preference in PolPrts. We have observed that PolPrts consistently exhibit high acceptability in the polarity reading, regardless of the type of initiative. However, in the conformity reading, the acceptability levels vary depending on the presence of bias in the initiative. As a result, it is necessary for previous frameworks to incorporate the type of initiative as an effective parameter to accommodate our empirical findings. This modification would require adjusting the current models, leading to a complex set of weighted parameters. In contrast, the dual-pathway approach not only provides a straightforward (intuitive) model for the particle selection process but also offers an explanation for the observed variation.

References

Büring, D., & C. Gunlogson. 2000. Aren't positive and negative polar questions the same?

Claus, B., A. M. Meijer, S. Repp, & M. Krifka. 2017. Puzzling response particles: An experimental study on the German answering system. *Semantics and Pragmatics* 10.

Farkas, D., & K. Bruce. 2009. On reacting to assertions and polar questions. *Journal of Semantics* 27. 81–118.

Farkas, D. F., & F. Roelofsen. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics* 34. 237–289.

Goodhue, D., & M. Wagner. 2018. Intonation, yes and no. *Glossa:a journal of general linguistic* .

Heim, I. 1983. *File Change Semantics and the Familiarity Theory of Definiteness*, 164–189. Berlin, Boston: De Gruyter.

Kramer, R., & K. Rawlins. 2009. Polarity particles: an ellipsis account. In *Proceedings of the North Eastern Linguistics Society (NELS) 39*, p. 479–492. Amherst, MA:GLSA.

Krifka, M. 2013. Response particles as propositional anaphors. *Semantics and Linguistic Theory* .

Krifka, M. 2015. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. *Semantics and Linguistic Theory* 25. 328.

Krifka, M., 2016. Assertions and questions in commitment space semantics. Handout from TaLK Conference, Keio University, Tokyo.

Ladd, D. 1981. A first look at the semantics and pragmatics of negative questions and tag questions.

Mohammadi, M. 2022. A unified analysis of polar particles in farsi. *Proceedings of the Linguistic Society of America* 7.

Pope, E. 1976. *Question and Answers in English*. The Hague: Mouton.

Roelofsen, F., & D. F. Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91. 359–414.

Roelofsen, F., & D. F. Farkas. 2019. Polarity particles revisited. *Semantics and Pragmatics* 12.

Romero, M., & C. Han. 2004. On negative yes/no questions. *Linguistics and Philosophy* 27. 609–658.

van Rooij, R., & M. Safarova. 2003. On polar questions. *Semantics and Linguistic Theory* 13.