**Supplementary Materials**

**Excerpts from Test Instructions in the True Guess Condition**

The excerpts below illustrate how zero-confidence and guess responses were presented to participants in the instructions for the True Guess condition.

Confidence ratings:

Press 0 if you have no confidence at all (i.e., you are completely guessing and a random computer-generated response is just as likely to be correct).

Source attributions:

Guess = Your decision was based on a guess (i.e., a random response is just as likely to be correct, so you would be fine with the computer generating the answer for you).

Follow-up note during the instructions:

**PLEASE NOTE:** If you indicate that your answer is a guess (i.e., by choosing **zero** **confidence** or **guess** as the basis of your decision), the computer will actually select an answer for you at random. That is, your answer will be replaced with the computer’s answer, with a 50-50 chance of being correct (the same as if you had truly guessed).

Computer response to zero-confidence and guess responses during the test phase:

OK. You have indicated that you were guessing, so the computer has randomly selected an answer for you! Do you accept?

**Interview Questions for the Traditional (TR), True Guess (TG), and No Subjective Measures (NoSM) Groups**

1. [ALL] In the last part of the experiment, what was your reaction when you had to start filling in the blank?
2. [NoSM] Did you ever respond based on a guess?

[TR&TG] Did you ever respond “guess”?

* If YES:
  + How did you feel about it?
  + Around how often did this happen?
  + Can you describe a situation when you were guessing?
* If NO:
  + Why not?

3. [ALL] What do you understand “guessing” to mean in the context of this experiment? (That is, how would you define “guessing”?)

*Note: For NoSM participants, SKIP questions 4-6.*

4. [TR&TG] Did you ever respond “guess” in a way that did **not** match that definition?

* If YES:
  + What was the situation?
  + Why did you respond that way?
  + How often did this happen?

5. [TR&TG] We intended “guessing” to mean that you were completely unsure about how to respond. That is, you really felt as though you might as well have...

* [TR] ...flipped a coin.
* [TG] ...let the computer answer for you.

Did you ever respond “guess” in a way that did **not** match that definition?

* If YES:
  + What was the situation?
  + Why did you respond that way?
  + How often did this happen?

6. [TR&TG] Did you ever **avoid** responding “guess” for any reason (that is, even when you were guessing)?

7. [ALL] OK, now let’s talk about when you were **not** guessing. What criteria did you use to make your choices?

*[If the participant mentions anything related to animacy (e.g., animals vs. objects) or another distinction that fits with the system (e.g., moving vs. stationary), ask]:* At what point did you become aware of that?

8. [NoSM] Would you say that any of your responses were based on a rule of some sort?

[TR&TG] Did you ever respond “rule”?

* If YES:
  + Can you describe the rule?
* If NO (or as a follow-up to elaborate on ‘yes’):
  + Did you use any other criteria to make your choices? Did you focus on any (other) features of the sentences or words within the sentences? Feel free to share anything that came to mind, no matter how hesitant you might have felt about it.
* If STILL NO: Well, in fact, there was a rule. Can you express what it might have been? Did anything come to mind **during the experiment** that you thought might be relevant to responding?

9: [ALL] Explain that *gi* (near) and *ul* (far) were for animate entities, while *ro* (near) and *ne* (far) were for inanimate entities. If the participant did not mention animacy earlier, ask if it ever came to mind **during the experiment**.

10. [ALL] Did you talk with anyone about this experiment or hear anything about it before participating?

**Participant Information**

Table S1.

*Participant Biodata According to Experimental Condition*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Number of Participants** | **Age** | **Sex** | **L1** | **Additional Languages** | **Number of Linguistics Courses** |
| Traditional | 10 | 24.00 | 3 female, 7 male | 8 English, 1 Spanish, 1 English and Hebrew | 9 Spanish, 1 German, 1 English, 1 French,  1 Italian | Mean: 1.00  Range: 0-10 or more  (9 reported 0 courses, while 1 reported “more than 10”) |
| True Guess | 10 | 23.30 | 5 female, 5 male | 8 English, 1 Arabic,  1 Ebonics and Standard American English | 9 Spanish, 2 Arabic, 1 Chinese, 1 English, 1 French, 1 Portuguese | Mean: 1.80  Range: 0-11  (8 reported 0 or 1, 1 reported 4, and 1 listed out 11 by name) |
| No Subjective Measures | 10 | 20.11 | 7 female, 3 male | 8 English, 2 Chinese | 9 Spanish, 2 Chinese, 2 English, 1 French,  1 Japanese, 1 Latin | Mean: 1.50  Range: 0-8  (8 reported 0 or 1, 1 reported 6, and 1 reported 8) |

*Note.* Eleven participants reported having studied more than one additional language. These participants, who had studied between two to four additional languages, were fairly evenly distributed across experimental conditions: three in the Traditional group, three in the True Guess group, and five in the NoSMs group.

**Overall Accuracy of Test Performance Across Groups**

****

*Figure S1.* Boxplot of overall accuracy in the test phase across the True Guess (TG), Traditional (TR), and No Subjective Measures (NoSM) groups. These data show standard performance for the TG and TR groups, but somewhat lower accuracy in the NoSM group, in accordance with other indications that the subjective measures were reactive.

**Accuracy and Proportion Data Across Different Types of Subjective Measures Responses**

Subjective measures are traditionally analyzed by examining the accuracy and proportions of participants’ responses at different reported levels of confidence or with different source attributions (e.g., Hamrick & Rebuschat, 2012, 2014; Rebuschat et al., 2013, 2015; Rebuschat & Williams, 2012). While such analyses are beyond the scope of the main manuscript, we report them here to facilitate comparisons with previous work.

It should be noted that outliers were removed for the inferential statistical analyses presented in the article because normality is important for analyses that involve proportion data as the dependent variable of interest. In this study, a single outlier in the True Guess condition contributed a quarter of the total number of “guess” responses for that group, and 92% of those guesses were reneged, meaning that not only was this participant attributing substantially more decisions to guesses than others in the group, but the outlier was also non-committal as to whether the responses were in fact guesses. The descriptive statistics presented in Tables S2 and S3 below, however, include all participants in order to maintain comparability with what is customarily reported in subjective measures analyses, where participants are typically not omitted for extreme concentrations of a certain type of source attribution or confidence judgment. Thus, Tables S2 and S3 reflect different values than those in the corresponding article due to the outliers being present here and omitted there.

These results demonstrate a fairly standard pattern of performance, with accuracy generally increasing as a function of increased confidence (Table S2) or more explicit source attributions (Table S3).

Table S2.

*Mean Accuracy and Proportion of Responses at Each Confidence Level*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Confidence Rating** | | | |
| **Group** | 0  (Not At All Confident) | 3  (Somewhat Confident) | 6  (Very Confident) | 9  (100% Confident) |
| *True Guess* | 49% (12.69%) | 61% (38.84%) | 81% (33.84%) | 91% (14.61%) |
| *Traditional* | 47% (11.09%) | 51% (34.87%) | 63% (40.24%) | 74% (13.78%) |

*Note.* The first number in each cell represents the mean accuracy rate, while the number in parentheses represents the proportion of trials at a given confidence level.

Table S3.

*Mean Accuracy and Proportion of Responses for Each Type of Source Attribution*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Source Attribution** | | | |
| **Group** | Guess | Intuition | Memory | Rule |
| *True Guess* | 45% (2.62%) | 58% (27.70%) | 62% (29.62%) | 82% (36.11%) |
| *Traditional* | 39% (14.88%) | 55% (45.98%) | 62% (22.83%) | 80% (14.81%) |

*Note.* The first number in each cell represents the mean accuracy rate, while the number in parentheses represents the proportion of trials with a given source attribution.