# Dimensions of judgment in stigmatized and non-stigmatized variation

Gert-Jan Schoenmakers (Radboud University)

Linguistic judgment experiments typically elicit responses in terms of the acceptability or surface probability of a sentence. Evidence that the conceptual dimension of the judgment scale influences the outcome of the experiment exists, but is only limited. This study investigates whether the scale dimension affects judgments for stigmatized (prescriptive norm violations) and for non-stigmatized (scrambling) variation in Dutch. Sentences are judged in one of three dimensions, *viz.* acceptability, surface probability, or aesthetic quality. The results indicate that participants take into account the scale dimension, but the effects depend on the type of variation. The findings are related to Schütze's (1996) theory of judgments.

## 1 Introduction

Linguistic judgments are increasingly considered the basis for inferences about linguistic representation (Goodall 2021, Schindler, Drożdżowicz, and Brøcker 2020, Schütze 1996). However, as Featherston (2021) points out, it is not exactly clear what these data quantify. Judgments are commonly interpreted as some sort of window into "grammaticalness" (Chomsky 1965), although it is well known that many other underlying factors contribute to a reported judgment as well. Sprouse (2020), for example, argues on the basis of Schütze's (1996) seminal work that linguistic judgments are the conscious reports of automatic responses to a stimulus. Such responses result from a composite of various considerations, yet they also contain an acceptability core that can sometimes serve as a proxy for grammaticality.

That it is difficult to determine what linguistic judgments are judgments of raises the question whether the instructions in judgment experiments can affect their outcome. In most cases, experiments prompt responses in a particular dimension through the experimental instructions, e.g. the *acceptability* or *naturalness* of a stimulus sentence. The present paper addresses the question to what extent linguistically naïve participants take into consideration the dimension of the judgment scale, and to what extent the scale dimension contributes to the acceptability core of linguistic judgments. Furthermore, the present study contrasts item sets of stigmatized and non-stigmatized variation so as to look for putative differences in the influence of the experimental instructions as a consequence of the semi-conscious application of linguistic rules in the former case (cf. Schütze 1996).

## 2 Background

Evidence that a manipulation of the scale dimension affects the output of linguistic judgment experiments is limited. Cowart (1997) presents an experiment in which two participant groups took part in the same task but under different instructions: 'intuitive instructions' based on personal judgment criteria and 'prescriptive instructions' eliciting judgments of well-formedness. This experiment did not yield any differences which are particularly relevant to linguistic theory. Langsford et al. (2019) investigate potential differences between judgments of acceptability and (confidence of) grammaticality for various grammatical illusion phenomena, as well as a set of judgment contrasts from *Linguistic Inquiry* (2001–2010). They find that reported judgments may differ somewhat between the two dimensions, but the instructions at least do not impinge on the relative acceptability between conditions. Turning now to cases of stigmatized variation, Bennis and Hinskens (2014) investigate judgments about ten different prescriptive norm violations in Dutch using a large scale questionnaire. Participants rated the norm violations on four scales (*good–bad, ugly–beautiful, sloppy-diligent, dialect–standard language*). The judgments were remarkably similar across the board, with only the last scale as a notable exception. Participants were thus at least sufficiently invested in the experiment to engage with its instructions, yet the scale manipulation did not yield linguistically relevant differences. Vogel (2019) investigates three German norm violations, eliciting judgments in terms of their

normativity, possibility, and aesthetic quality. Although the judgment scores of aesthetic quality were only slightly lower than those of normativity (18.7% and 24.5% respectively), the scores of probability were much higher (36.1%). Thus, prescriptive norm violations are not particularly good or pretty, but they do exist in the linguistic reality and speakers are aware of this. Crucially, here the instructions do influence the relative judgments between conditions.

It thus seems as though the dimension of the scale may impact the outcome of a linguistic judgment experiment, yet evidence is scarce and the reported effects so far are limited to cases of stigmatized variation. The present study replicates Vogel's (2019) experiment of prescriptive norm violations for Dutch, and expands on it by including an item set of non-stigmatized variation, *viz.* scrambling (see Schoenmakers 2022). More specifically, definite objects in the Dutch middle-field may appear on the left or right side of a clause adverb, a type of word order variation which has been argued to be driven by topic-focus structure and, crucially, which is not associated with sociolinguistic stigmatization.

## 3 The experiment

The experiment was an online questionnaire and contained two distinct item sets: one item set with prescriptive norm violations (stigmatized variation) and another with scrambling constructions (non-stigmatized variation). Three types of prescriptive norm violations were included in the experiment: subject *hun* 'them', comparative *als* 'as', and auxiliary *doen* 'do'. Items from this set either did or did not contain a norm violation and were additionally grammatical or ungrammatical (following Vogel 2019). The scrambling items contained a definite object on either side of a clause adverb. A sample target sentence from the scrambling item set is given in (1).

(1) *Nora gaat (het museum) absoluut (het museum) bezoeken.*
Nora goes the museum absolutely the museum visit

All experimental items were preceded by three-sentence preambles, which in the scrambling items licensed the object as the topic or focus, in order to test for the 'discourse template' (see Schoenmakers 2020). Both item sets thus had a 2×2 design (±*violation* × ±*grammatical* in the stigmatized item set; *object position* × ±*topicality* in the scrambling set). 153 participants ($M_{age}$ 48.51, range 18–91, SD = 20.89) rated 108 sentences (36 norm violations, 24 scrambling, and 48 fillers) in one of three dimensions, illustrated in (2) (i.e. *dimension* was a between-subjects factor). Judgments were given on a slider scale from 0–100%.

(2) a. **Aesthetic judgment:**
*Hoe mooi vind je de formulering van de bovenstaande zin?*
How pretty do you find the wording of the above sentence?

b. **Acceptability judgment:**
*Hoe goed vind je de bovenstaande zin als Nederlandse constructie?*
How good do you find the above sentence as a construction of Dutch?

c. **Probability judgment:**
*Hoe waarschijnlijk vind je het dat de bovenstaande zin is uitgesproken door een moedertaalspreker van het Nederlands?*
How likely do you think it is that the above sentence has been uttered by a native speaker of Dutch?

## 4 Results

The judgment patterns of the raw scores are visually presented in Figure 1 for both item sets. The standardized scores were entered into two LMER models, with the above-mentioned factors and the scale dimension as fixed effects (with *acceptability* set as the reference category). The random structure of the models contained by-participant and by-item intercepts and slopes for the effects of both fixed factors and the by-participant (stigmatized) or by-item (scrambling) interaction. The full model specifications are given in Tables 1 and 2.

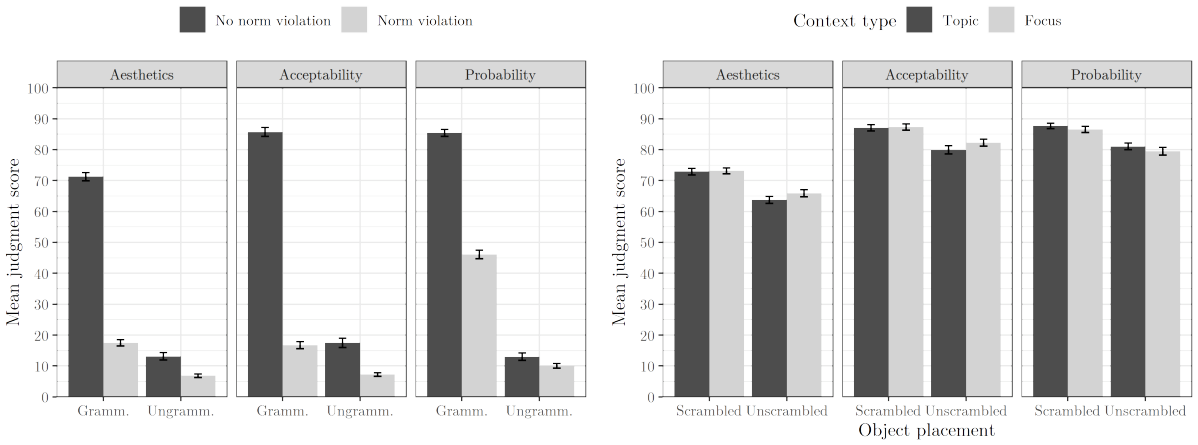


**Figure 1:** Mean judgment scores per condition for the two item set in three dimensions (error bars indicate within-subject standard errors from the mean)

| | **Fixed effects** | | | | **Random effects** (SDs) | |
|---|---|---|---|---|---|---|
| **Parameters** | β | Std. Error | *t*-value | *p* | *by-participant* | *by-item* |
| (Intercept) | -0.475 | 0.023 | -20.452 | $< .001$ | 0.082 | 0.095 |
| norm violation | -0.996 | 0.055 | -18.248 | $< .001$ | 0.210 | 0.230 |
| grammaticality | 0.994 | 0.071 | 14.029 | $< .001$ | 0.220 | 0.351 |
| dimension (aesthetics) | -0.011 | 0.023 | -0.492 | .623 | - | - |
| dimension (probability) | 0.086 | 0.023 | 3.719 | $< .001$ | - | - |
| dimension (aesthetics) * norm violation | 0.097 | 0.053 | 1.823 | .068 | - | - |
| dimension (probability) * norm violation | 0.439 | 0.053 | 8.328 | $< .001$ | - | - |
| dimension (aesthetics) * grammaticality | 0.066 | 0.055 | 1.203 | .229 | - | - |
| dimension (probability) * grammaticality | 0.440 | 0.054 | 8.077 | $< .001$ | - | - |
| norm violation * grammaticality | -1.467 | 0.074 | -19.752 | $< .001$ | 0.388 | - |
| dimension (aesthetics) * norm violation * grammaticality | 0.009 | 0.102 | 0.086 | .931 | - | - |
| dimension (probability) * norm violation * grammaticality | 0.544 | 0.101 | 5.392 | $< .001$ | - | - |

**Table 1:** Model specifications of the linear mixed-effects model for the stigmatized item set (number of observations: 5506, groups: participant, 153; item, 36)

| | **Fixed effects** | | | | **Random effects** (SDs) | |
|---|---|---|---|---|---|---|
| **Parameters** | β | Std. Error | *t*-value | *p* | *by-participant* | *by-item* |
| (Intercept) | 0.842 | 0.036 | 23.281 | $< .001$ | 0.126 | 0.139 |
| topicality | 0.026 | 0.029 | 0.888 | .374 | 0.054 | 0.058 |
| object position | 0.164 | 0.048 | 3.429 | $< .001$ | 0.227 | 0.113 |
| dimension (aesthetics) | -0.079 | 0.031 | -2.564 | .010 | - | - |
| dimension (probability) | -0.046 | 0.030 | -1.519 | .129 | - | - |
| dimension (aesthetics) * topicality | -0.015 | 0.036 | -0.416 | .677 | - | - |
| dimension (probability) * topicality | -0.048 | 0.036 | -1.327 | .185 | - | - |
| dimension (aesthetics) * object position | 0.089 | 0.058 | 1.549 | .121 | - | - |
| dimension (probability) * object position | 0.004 | 0.057 | 0.064 | .949 | - | - |
| topicality * object position | -0.038 | 0.061 | -0.624 | .533 | - | 0.168 |
| dimension (aesthetics) * topicality * object position | -0.009 | 0.069 | -0.132 | .895 | - | - |
| dimension (probability) * topicality * object position | 0.027 | 0.068 | 0.393 | .694 | - | - |

**Table 2:** Model specifications of the linear mixed-effects model for the scrambling item set (number of observations: 3671, groups: participant, 153; item, 24)

Statistical analysis led to the following conclusions:

i. Prescriptive norm violations were rated higher on the scale of probability (46.1%) than on the scales of acceptability (16.7%) and aesthetic quality (17.5%), but lower than unmarked sentences on all three scales (71–86%). The difference between judgments of probability and acceptability was significant, and the effect was moderated by $\pm$*violation* and $\pm$*grammatical*. Further, the three-way interaction was also significant. The experimental instructions thus influenced judgment behavior in the case of stigmatized variation.

ii. The manipulation of scale dimension did not impinge on the relative acceptability between conditions in items with a scrambling configuration. Thus, the present experiment does not provide evidence for an effect of the experimental instructions in non-stigmatized variation or for the idea that scrambling adheres to a 'discourse template', which is commonly assumed in the literature (see Schoenmakers 2020).
iii. Unmarked items (grammatical fillers, non-violations, scrambling configurations) were rated considerably lower on the scale of aesthetic quality than on the two other scales (by at least a 10-point margin numerically). The difference between the dimensions of acceptability and aesthetic quality was significant in the scrambling set but not in the stigmatized item set.

Taken together, the results indicate that participants take into account the scale dimension, in both stigmatized and non-stigmatized variation, but the effects depend on the type of variation.

## 5 Theoretical implications

A crucial difference between the norm violations and the scrambling items is that participants have conscious access to the prescriptive rules of their language. Schütze (1996: 83) notes that "we could imagine that expected judgment causes people to revert to conscious reasoning *about* sentences, rather than processing *of* them." In case the norm is violated, the judgments of acceptability and aesthetic quality may thus reflect a binary opposition, in that the sentence either does or does not match a prescriptively correct form, whereas frequency estimations are much more open-ended (cf. Featherston 2021). Although work on norm violations is currently only limited, this type of rationale can explain findings of previous experimental studies as well (e.g. Hubers et al. 2020). The new results therefore call for much needed future research on norm violations and the way in which they are judged and processed.

Regarding the non-stigmatized (scrambling) item set, the findings do not provide much new evidence, but they do not reject Schütze's (1996) theory of judgments either. Only main effects of the instructions were found (cf. Cowart 1997). One could argue that the new findings imply that the reported reactions do not reflect technical introspection in the Wundtian sense; rather, they consist of an acceptability core with additional effects from other cognitive processes that influence judgments of aesthetic quality and presumably other dimensions of judgment.

That is to say, judgments may reflect an amalgamation of (partially non-instructed) considerations on the part of the participant, and these can be vastly different conceptually. The data may thus serve as a proxy for grammaticality, but the judgment scale(s) used may have put an additional coat of paint on them. This must be considered when an attempt is made to answer the question what linguistic judgments quantify.

## References


Bennis, H., & Hinskens, F. (2014). Goed of fout: Niet-standaard inflectie in het hedendaags Standaardnederlands. *Nederlandse Taalkunde*, *19*(2), 131–184.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. SAGE.

Featherston, S. (2021). Response methods in acceptability experiments. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 39–61). Cambridge University Press.

Goodall, G. (2021). Sentence acceptability experiments: What, how, and why. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 7–38). Cambridge University Press.

Hubers, F., Redl, T., de Vos, H., Reinarz, L., & de Hoop, H. (2020). Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking [Article 186]. *Frontiers in Psychology*, *11*.

Langsford, S., Stephens, R., Dunn, J., & Lewis, R. (2019). In search of the factors behind naive sentence judgments: A State Trace Analysis of grammaticality and acceptability ratings [Article 2886]. *Frontiers in Psychology*, *10*.

Schindler, S., Drożdżowicz, A., & Brøcker, K. (2020). *Linguistic intuitions: Evidence and method*. Oxford University Press.

Schoenmakers, G. (2020). Freedom in the Dutch middle-field: Deriving discourse structure at the syntax-pragmatics interface [Article 114]. *Glossa: A journal of general linguistics*, *5*(1).

Schoenmakers, G. (2022). *Definite objects in the wild: A converging evidence approach to scrambling in the Dutch middle-field* (Doctoral dissertation). Radboud University. Nijmegen.

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology* [Reprinted in 2016 by Language Science Press]. University of Chicago Press.

Sprouse, J. (2020). A user's view of the validity of acceptability judgments as evidence for syntactic theories. In S. Schindler, A. Drożdżowicz, & K. Brøcker (Eds.), *Linguistic intuitions: Evidence and method* (pp. 215–232). Oxford University Press.

Vogel, R. (2019). Grammatical taboos. *Zeitschrift für Sprachwissenschaft*, *38*(1), 37–79.