Education as a Source of Vagueness in Criteria and Degree

Steven Verheyen and Gert Storms

Abstract Individual differences in application are considered a hallmark of vague terms. When a term is truly vague there exists a range of applications that are considered permissible by competent users of the language. The divergence in application may be the result of indeterminacy with respect to the conditions for application (vagueness in criteria) and indeterminacy with respect to the extent of application given fixed conditions (vagueness in degree). We propose a formal procedure to determine whether individual application differences result from vagueness in criteria and/or vagueness in degree. The procedure provides an experimental perspective on vagueness in that it involves the comparison of two groups of participants that differ on a variable of interest. The procedure establishes whether the variable systematically affects application of a term. We present a case study in which we compare categorization data from participants who went on to higher education after completing compulsory education and participants who did not. Application of the proposed procedure shows that education systematically affects categorization. Higher education participants tend to apply common terms like VEGETABLES, FURNITURE, and TOOLS more conservatively than compulsory education participants do (vagueness in degree). For terms they are arguably more familiar with, like SCIENCES, they are found to employ different conditions for application (vagueness in criteria). The results demonstrate that part of the permissible variation that is deemed characteristic of vagueness reflects sociolinguistic variation.

S. Verheyen (✉)
Institut Jean Nicod, Département d'études Cognitives, ENS, EHESS,
PSL Research University, CNRS, Pavillon Jardin, 29, rue d'Ulm, 75005 Paris, France
e-mail: steven.verheyen@ens.fr

G. Storms
Laboratory of Experimental Psychology, Faculty of Psychology and Educational Sciences,
KU Leuven, Tiensestraat 102, 3000 Leuven, Belgium
e-mail: gert.storms@kuleuven.be

E. Castroviejo et al. (eds.), *The Semantics of Gradability, Vagueness, and Scale Structure*,
Language, Cognition, and Mind 4, https://doi.org/10.1007/978-3-319-77791-7_6

Keywords Categorization · Individual differences · Threshold theory Permissible variation · Sociolinguistics · Vagueness in criteria · Vagueness in degree

1 Vagueness and Individual Differences

Whether a woman of 1m75 is TALL or chess is a SPORT are questions without single, matter-of-fact answers. TALL and SPORT are vague words, meaning that there is no established demarcation of the instances to which they apply and the instances they do not apply to. Individuals can use these words in different ways without committing an error (Kölbel 2004; Raffman 2014; Wright 1995). This "permissible variation" becomes readily apparent in categorization tasks in which the participants diverge widely regarding the instances they feel these words apply to (Black 1937; McCloskey and Glucksberg 1978). Among psychology undergraduates, the odds in favour of calling a woman of 1m75 TALL are 65:35 (Verheyen et al. 2018) and the odds in favour of considering chess a SPORT are about 50:50 (Verheyen et al. 2010), for instance.

The idea that some of these individual differences in categorization are systematic and can be brought back to properties of the participants, has often been entertained, but seldom demonstrated. Both Barsalou (1993) and Smith and Samuelson (1997) have suggested that in addition to the context language users find themselves in, their individual learning histories, aptitudes, and dispositions influence their categorization behavior. The idea also finds support in the work of Gardner (1953) and Verheyen et al. (2010), who showed that participants display relatively stable categorization patterns across tasks. Which properties of the individuals are responsible for the observed stability? And to what extent can individual differences in categorization be attributed to different participant properties?

Questions like these naturally fit an experimental perspective on vagueness. A comparison of the categorization behaviour of two groups of participants who differ in a property of interest, allows one to determine if the property under investigation systematically affects the participants' word use. The work described in this chapter is an illustration of how one might effectively do that. Inspired by a quote from Chomsky: "*Word meaning is intimately bound up with matters of knowledge and belief*" (1980: 225), we chose to study the effect of education on vagueness[1].

The effect of education on semantic tasks has rarely been explicitly investigated (see Rosenzweig 1964, for a notable exception), unless educational level is used as a proxy for verbal skill (Kuperman et al. 2013). Participants' level of education is regularly included as a control variable in semantic norming studies, though (e.g., De Witte et al. 2015; Loonstra et al. 2001). In these studies the transition from compulsory to higher education regularly yields pronounced effects (Charchat Fichman et al.

[1]To be fair, although Chomsky recognizes that variation is an important part of language, he considers the study of the *shared* knowledge of language users paramount.

2009; Keuleers et al. 2015). For illustrative purposes, we too therefore decided to focus on the categorization differences between participants who went on to complete higher education after compulsory education and those who did not.

In what follows we describe how a property such as the level of education of participants can be brought into a formal account of vagueness. This account takes the form of a statistical model that makes a number of assumptions regarding the manner participants arrive at a categorization decision *and* the kind of differences one might expect to see herein. Indeed, while the vagueness of words comes with variation in their application by different individuals, it is not the case that anything goes. The latter would render communication among language users impossible. An explanation of the permissible individual variation should thus be an intrinsic part of any formal account of vagueness (Black 1937). The proposed statistical model works by characterizing categorization differences between groups as vagueness in criteria or vagueness in degree.

2 Vagueness in Criteria and Degree

Devos (1995, 2003) distinguishes vagueness in criteria and vagueness in degree (see also Alston 1964; Kennedy 2013; Machina 1976). He defines vagueness in criteria as the indeterminacy with respect to (the combination of) the conditions for application of a term. Individuals might employ different criteria for establishing whether an activity is a SPORT or not. While some might emphasize that SPORTS require physical activity, others might require an element of competition. According to the first criterion *hiking*, but not *chess*, is likely to be considered a SPORT, while the reverse holds when the second criterion is employed. Devos (1995, 2003) defines vagueness in degree as the extent to which a term can be applied given that the conditions have been determined. Even when individuals agree on which criterion to employ to establish whether an activity is a SPORT or not, they might still disagree as to whether a particular activity sufficiently meets that criterion. That is, while some might deem both *hiking* and *running* sufficiently demanding, others might feel that only *running* requires sufficient physical activity to be considered a SPORT.

Devos (1995, 2003) argues that vagueness in criteria coincides primarily with nouns and vagueness in degree with adjectives. The rationale behind this is that while for many adjectives there exists a unique criterion that determines application (e.g., height for TALL, price for EXPENSIVE), many criteria can be considered for applying a noun like SPORT (competitiveness, physical activity, …). This line of reasoning ignores the fact, however, that (a) many adjectives too are multifaceted (Kamp 1975; Klein 1980; Sassoon 2012) and (b) the application of many multifaceted nouns is effectively governed by one-dimensional constructs such as typicality or similarity to the target category (Hampton 1998, 2007; McCloskey and Glucksberg 1978). We therefore see no reason to expect a principled relationship between vagueness in criteria and nouns on the one hand, and vagueness in degree and adjectives on the other.

What Devos (1995, 2003) terms vagueness in degree is addressed in the so-called threshold theory (Hampton 1995, 1998, 2007; Raffman 1994, 1996). The threshold theory provides a psychological account of individual differences in categorization. It does so by positing that categorizers position a threshold that separates members from non-members on a dimension along which the candidate items are organized. The threshold theory thus assumes that categorization is governed by a single latent dimension, which can be composed of one or more (weighted) substantial criteria (see Égré 2017; Keefe 2000, for similar assumptions)[2]. In the work of Raffman (1994, 1996), who is mainly interested in adjectives such as RICH and RED, the dimension reflects a single substantial criterion such as dollar amounts or wavelengths. In the work of Hampton (1995, 1998, 2007), who is primarily concerned with noun categories such as FRUITS, VEHICLES, and SPORTS, the dimension is thought to reflect the items' typicality or similarity to the target category, which can be shown to reflect a weighted combination of several substantial criteria (Dry and Storms 2010; Hampton 1979; Rosch and Mervis 1975; see De Deyne et al. 2014, for evidence that the same holds for multifaceted adjectives). Crucially, however, in both accounts all categorizers are assumed to rely on the same dimension for their categorization and only to differ with respect to the threshold they employ. When categorizers use different thresholds on the same dimension, they are essentially diverging on the *degree* to which the target term applies, given fixed conditions (be it wavelength for RED, or a weighted combination of competitiveness and activity level for SPORTS).

Although vagueness in criteria lay outside its original scope, the threshold theory can easily be extended to encompass it, by allowing not only the threshold, but also the dimension that is relied upon for categorization, to be subject to individual variation. Hampton (2006) already alluded to this possibility when he showed that the weighting of substantial criteria may differ from person to person (see Hampton and Passanisi, 2016; Verheyen and Storms 2013; Zee et al. 2014, for additional empirical support). The resulting dimensions of categorization could differ rather subtly when the same criteria are merely accentuated differently, or more profoundly when they reflect the use of distinct criteria as in the SPORTS example at the beginning of this section. Both are potential instantiations of vagueness in *criteria*; the latter being a special case which amounts to setting to zero the weights of all potential criteria, except the one that is relied on.

2.1 A Formalization of Threshold Theory

Verheyen et al. (2010) provided a formalization of the threshold theory that can be applied to binary categorization tasks, in which categorizers go through a set of candidate items, indicating whether (1) or not (0) these belong to the target category. Like the threshold theory, the model assumes a single dimension along which the

[2]This assumption does preclude the possibility that individuals employ a disjunctive set of criteria, such as "SPORTS should be competitive *or* involve intense physical activity".

candidate items and the categorizers' thresholds are positioned and it is their relative position that determines the answers. Unlike the original theory, the thresholds in the model do not deterministically separate members from non-members.

The model encompasses a free parameter β for each one of the items *i* and a free parameter θ for each one of the categorizers *c*. The β_i values reflect a rank ordering of the items according to the propensity with which they are endorsed (the more often *i* is endorsed as a category member, the higher β_i is). The θ_c values reflect a rank ordering of the categorizers according to the number of items they endorsed (the fewer items *c* endorses, the higher θ_c is). As such, the values of β_i and θ_c can respectively be interpreted as the extent to which item *i* meets the categorization criteria and categorizer *c*'s threshold (Verheyen et al. 2010). The better an item meets the categorization criterion, the further it is positioned on the dimension. The higher the requirements for category membership categorizers' impose, the further their thresholds θ_c are positioned along the dimension.

Both the item positions (β_i) and the thresholds (θ_c) are estimated from the categorization data. As such, the model does not presume any a priori knowledge about the underlying dimension, although it can be subsequently interpreted by inspecting the relative positions of the items indicated by β_i. Items are positioned further along the dimension the better they meet the categorization criteria. If physical activity thus governed the categorization decisions for SPORT, *hiking* would be found to the right of *chess*. If competition governed the decisions, the relative positions of *hiking* and *chess* would be reversed.

According to the model formula in Eq. (1), the more the position of the item (β_i) surpasses the position of the categorizer's threshold (θ_c) along the dimension, the higher the probability that categorizer *c* considers item *i* a category member will be (and vice versa):

$$\Pr\left(Y_{ci} = 1\right) = \frac{e^{\beta_i - \theta_c}}{1 + e^{\beta_i - \theta_c}}. \tag{1}$$

The membership function thus starts of at 0 (clear non-member) for items that fall short of the categorizer's threshold ($\beta_i \lll \theta_c$) and steadily increases until 1 (clear member) for items that clearly surpass it ($\beta_i \ggg \theta_c$). The threshold θ_c thus does not rigorously separate members from non-members into, but rather reflects the point at which the categorizer is indifferent with respect to the category membership decision. The probability that the model assigns to the categorization answer for an item that coincides with the threshold θ_c is 0.5, making the decision effectively a coin toss. By casting them as stochastic variables, the model leaves room for some statistical variation in the categorization answers, which is particularly interesting for explaining intra-individual differences in categorization (see Verheyen et al. 2010, for details); but this does not mean that categorizers are completely free to fill in the meaning of a word. The categorization patterns are constrained by the positions of the items (β_i), which reflect how the items score on the criteria that are relevant for category membership. Like the threshold theory, the model assumes that the grounds for making the categorization decisions are shared by all categorizers. In Eq. (1) the agreement on which items score high/low on the relevant criteria is reflected in a single set of β_i estimates. That is, there is one dimension that governs categorization. As such, the model is only concerned with degree differences in categorization. It can be naturally extended to encompass criteria differences, though.

2.2 A Formalization of Criteria and Degree Differences

The model in Eq. (1) makes establishing whether a participant's group membership (e.g., compulsory vs. higher education) affects vagueness in criteria and degree tangible. Vagueness in degree would show in a meaningful difference in the threshold positioning θ_c of the members of the two groups. This would indicate that one group applies the term under investigation more conservatively than the other group does. Vagueness in criteria would show in a meaningful difference in the item positioning β_i of the two groups. This would indicate that different dimensions govern categorization in the groups. In order to establish whether two groups employ different criteria for categorization, the modeling framework again requires no intuitions about potential criteria. It suffices that the estimated item positions are different for the two groups, as this constitutes evidence that the used criteria are not the same. The nature of these criteria can later be investigated through interpretation of the items' relative positions in the manner described above.

To make statistical inferences regarding the existence of both types of vagueness feasible, the model in Eq. (1) needs to be extended as neither the threshold positions, nor the item positions are group dependent. In addition to indices *c* and *i* to indicate individual categorizers and items, respectively, an index *g* is introduced to make a distinction between groups. In order to establish whether group membership is a source of vagueness in degree, we assume that the thresholds follow a normal distribution with a group-specific mean and variance: $\theta_{cg} \sim N\left(\mu_{\theta_g}, \sigma^2_{\theta_g}\right)$. By constraining μ_{θ_1} to equal 0, μ_{θ_2} can be thought of as the mean threshold difference between the groups (compulsory vs. higher education). If μ_{θ_2} is reliably different from zero, this constitutes evidence for vagueness in degree resulting from education.

To establish whether group membership is a source of vagueness in criteria, we introduce binary latent indicators D_i that signal whether individuals who have the same threshold but are from different groups have a different probability of categorizing item *i* as a category member and therefore require a separate β_i estimate. If D_i equals 0 the model reads:

$$\Pr\left(Y_{cig} = 1 \mid D_i = 0\right) = \frac{e^{\beta_i - \theta_{cg}}}{1 + e^{\beta_i - \theta_{cg}}}. \tag{2}$$

β_i has no index *g* here, signaling that the position of item *i* is the same in both groups.

If D_i equals 1 the model reads:

$$\Pr\left(Y_{cig} = 1 \,|D_i = 1\right) = \frac{e^{\beta_{ig}-\theta_{cg}}}{1 + e^{\beta_{ig}-\theta_{cg}}}. \qquad (3)$$

β_i does receive an additional index *g* here, signaling that item *i* is positioned differently in the two groups. When the modeling procedure uncovers items with different positions in the two groups, this constitutes evidence for vagueness in criteria resulting from education.

3 Method

In order to demonstrate the modeling framework, we will re-analyze part of the data from Verheyen et al. (2018) in which adult participants who were recruited online completed a categorization task. The re-analysis is restricted to the data from female participants, who make up the majority of the participants sample (55%), since previous research has shown that gender produces both vagueness in degree and vagueness in criteria (Stukken et al. 2013). We want to avoid mistaking gender differences for education differences.

3.1 Participants

The selection of participants is comprised of 1036 adult female participants aged 18—92 ($M = 56.82$, $SD = 16.10$) from Flanders (Belgium). Since not all participants completed all categories, the actual number of participants per category ranges between 1004 and 1011. Sixty-four percent of the participants indicated to have obtained a diploma beyond secondary education, which is the compulsory level in Flanders. They make up the higher education group.[3]

3.2 Materials

The materials were Dutch translations of the categories and items in Hampton et al. (2006). In what follows we will use the original English terms to refer to them. The materials included eight categories with 24 items each. The majority of these items were borderline items for the target categories, but clear members and non-members

[3]Because the higher educated participants tended to be younger, we also conducted an additional analysis in which we equated the two groups in terms of age. We biased this analysis against the educational effect found in the main analysis by choosing for the higher education group participants with a more practical university college education over participants with a scientific university education whenever possible. This additional analysis yielded similar results, indicating that the results of the main analysis are not due to a confounding of education with age.

were included as well to make the task more natural. The categories will be printed in small capitals (FRUITS, VEGETABLES, FISH, INSECTS, SPORTS, SCIENCES, TOOLS, and FURNITURE) and the items in italic (*avocado*, *garlic*, *shrimp*, *maggot*, *chess*, *economics*, *funnel*, and *piano*, are examples of borderline items for the respective categories). A list of the original materials along with their Dutch translation can be found in the Appendix.

3.3 Procedure

The data were gathered through a web survey. The participants completed a categorization task in which they were asked to indicate for the eight categories whether the 24 candidate items belonged to the category or not. In addition to "yes" and "no" participants could also answer "I don't know the item". It was emphasized that we were interested in participants' personal opinions rather than the answers considered appropriate by the general public or official authorities. The categories were presented on separate pages in a random order. The corresponding items were presented in randomized lists. Participants could proceed at their own pace. The majority of participants completed the survey in less than ten minutes.

3.4 Model Analysis

The models that have been discussed in this paper are all existing models that have been developed in the Item Response Theory (IRT) literature. IRT models tend to be used to infer latent traits from individuals' manifest responses to the items in a questionnaire or test (Hambleton et al. 1991). Verheyen et al. (2010) recognized the potential of IRT models for the study of vagueness when they introduced the model in Eq. (1) as a formalization of the threshold theory. In the IRT literature this model is known as the Rasch model (Rasch 1960). The model we describe in this chapter was introduced in the IRT literature by Frederickx et al. (2010) to investigate group bias in high stakes testing situations.

We analyzed each category's categorization data separately using the extended model. This was done using WinBUGS (Lunn et al. 2000) following the procedures for the Bayesian estimation of the model outlined in Frederickx et al. (2010). These include the specification of the priors for the model parameters. For every analysis five chains were run of 10,000 iterations each, with a burn-in sample of 1,000. To determine whether compulsory education (group 1) and higher education (group 2) participants employ different criteria for categorization, we investigate whether there are items for which the posterior probability of indicator D_i exceeds .5, indicating that the item is positioned differently in the two groups. To determine whether the two education groups differ regarding the degree they feel the target categories apply, we investigate the mean threshold difference between them. We deem a difference in degree reliable if the 95% credibility interval for μ_{θ_2} does not include 0.

3.5 Predictions

This is an exploratory study that is first and foremost intended to be an illustration of a modeling framework to characterize group differences in categorization as degree or criteria differences. The study was not designed with the intent of investigating the effect of education level on categorization. The choice for education level as the group variable was a matter of convenience, as we had this information available from a large scale study where this was not the variable of interest. We did not entertain any a priori hypotheses as to the extent to which we would observe criteria and degree differences resulting from education differences, as to our knowledge, the effect of education level on semantic categorization has not been investigated yet. We did deem education level a promising group variable for an illustration, though, as it seemed probable that individual differences in knowledge would affect how terms are applied.

4 Results

4.1 Vagueness in Criteria

We hardly found any evidence for vagueness in criteria. For FRUITS, VEGETABLES, INSECTS, SPORTS, and TOOLS there were no items for which the indicator D_i exceeded .5. That is, in the higher education group the candidate items were positioned the same as in the compulsory education group. For FISH and FURNITURE one item was positioned differently. Higher educated participants considered *sea horses* more representative FISH than compulsory education participants did (higher β_i in the higher education group). They also considered *shelves* more representative FURNITURE than compulsory education participants did.

For SCIENCES the picture looked completely different. There were twelve items (50%) for which D_i exceeded 0.5. *Astrology*, *philosophy*, *palm reading*, *literature*, and *psychology* were considered less representative of SCIENCES in the higher education group than in the compulsory education group (lower β_i in the higher education group). The items *geography*, *geometry*, *meteorology*, *mineralogy*, *chemistry*, *dentistry*, and *nutrition* were considered more representative by higher education participants than by compulsory education participants (higher β_i in the higher education group).

4.2 Vagueness in Degree

For FRUITS and SPORTS the mean threshold difference could not reliably be discerned from zero, indicating that there is no reliable difference in the number of items endorsed by the two groups. A reliable positive threshold difference was found for VEGETABLES, INSECTS, FURNITURE, FISH, and TOOLS. Higher education participants used a higher threshold than the compulsory education participants did, resulting in the endorsement of fewer items as category members by the former group. For the category of SCIENCES, we found the opposite pattern: the mean threshold difference was reliably negative, indicating that higher education participants tended to endorse more items as SCIENCES than compulsory education participants did.

4.3 Criteria-Degree Interplay

The findings regarding group differences in degree should be interpreted in light of the findings regarding vagueness in criteria. To aid this interpretation we have included a figure that depicts categorization patterns for the three combinations of vagueness in criteria and degree we identified through the model analyses: (i) absence of vagueness in criteria and degree, (ii) vagueness in degree but not criteria, (iii) vagueness in both criteria and degree. The three combinations are exemplified by the results for FRUITS, INSECTS, and SCIENCES, respectively.

All three panels in Fig. 1 show for 24 candidate items the proportion of participants endorsing the items as a category member. Each panel contains two graphs, one for the compulsory education participants (black circles) and one for the higher education participants (gray squares). The items are organized along the horizontal axis in increasing order of endorsement according to the compulsory education group. The item on the far left is thus the one least endorsed by the compulsory education participants, while the item on the right is the one most endorsed by these participants.

For FRUITS (left panel) the analyses yielded no reliable differences between the two education groups. The absence of vagueness in degree and criteria clearly shows in the categorization proportions as well: the black and gray curves almost completely overlap. This pattern of categorization is exemplary for SPORTS as well.

For INSECTS (middle panel) the analyses yielded a reliable threshold difference, but no differently positioned items. The absence of vagueness in criteria shows in that the categorization curves of the education groups take the same shape: the items are ordered in the same manner in both groups. The vagueness in degree shows in the displacement of the two curves: the gray squares are systematically lower than the black circles, indicating that the higher educated participants were more strict in categorizing items as category members. This pattern of categorization is exemplary for the majority of the studied categories (VEGETABLES, FURNITURE, FISH, TOOLS).

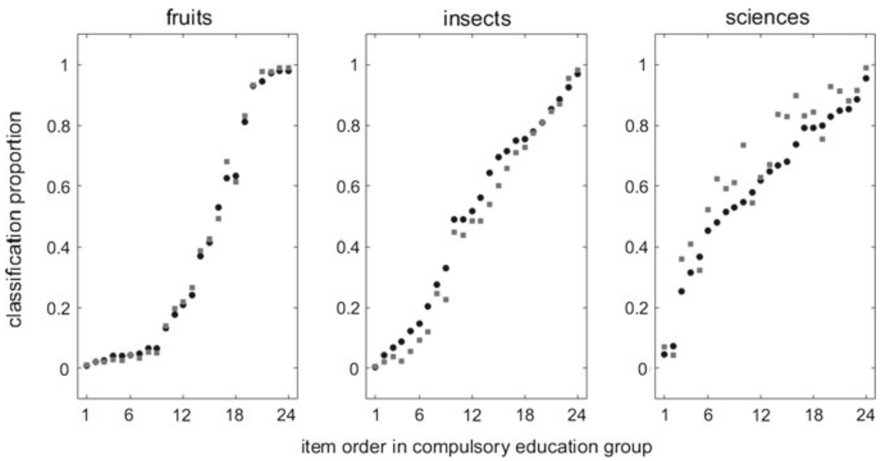

Fig. 1 Categorization proportions for FRUITS (left), INSECTS (middle), and SCIENCES (right) of the compulsory education group (black circles) and the higher education group (gray squares). Items are ordered along the horizontal axis according to their categorization rank in the compulsory education group

For SCIENCES (right panel), we observe that the categorization proportions of the higher education group tend to be higher (instead of lower) with respect to those of the compulsory education group. This is an indication of the vagueness in degree the analyses established. The additional vagueness in criteria shows in that the two curves no longer have the same shape: The order of the categorization proportions in the two groups is not the same. It is not just the case that the curve of one group is shifted with respect to the other group, as we saw for INSECTS. The nature of the group difference is different for individual items (those for which D_i exceeded .5) with the compulsory – higher education categorization divide being greater for some than for others and in some instances even showing the reverse pattern. This was the case for items 2 (*palm reading*), 5 (*literature*), 11 (*philosophy*), and 19 (*astrology*). They were considered less representative category members by the higher education group, along with *psychology* (item 12) for which the categorization proportion difference is smaller than one would expect in light of the group threshold difference. For other items such as item 7 (*nutrition*), 10 (*dentistry*), and 16 (*mineralogy*) the categorization difference is larger than expected based on the threshold difference solely.

4.4 Discussion

We observed pronounced differences in semantic categorization between the group of compulsory education participants and the group of higher education participants. The model analyses qualified the categorization differences more often as degree differences (with higher education participants endorsing fewer items than compulsory education participants), than as criteria differences. From this observation one should conclude that education can give rise to both these types of vagueness, but not that degree differences are more prevalent than criteria differences. Which combination of degree and criteria differences emerges, appears to be dependent upon the stimulus materials, and the employed materials were not selected with the education difference in mind. The current study does offer a number of interesting hypotheses regarding the origin of the degree and criteria differences between the education groups, which can be tested in future research with materials tailored to these questions (see Conclusions section for details).

5 Conclusions

The purpose of this chapter was to introduce a procedure that allows group differences in categorization to be identified as criteria and/or degree differences. As such, the procedure yields a number of insights regarding the nature and the sources of vagueness. We established that for noun categories, both vagueness in criteria and vagueness in degree are likely in play at any given moment. From the early work by McCloskey and Glucksberg (1978) and the subsequent replication of their work by Verheyen et al. (2010) we already knew that individuals diverge in the degree to which they consider items members of noun categories. Verheyen and Storms (2013), on the other hand, identified latent groups of participants who use different criteria for categorizing items in categories such as FISH, SPORTS, and SCIENCES. The novelty of the current procedure lies in its focus on vagueness in degree and vagueness in criteria *simultaneously*, and in its ability to relate them to external information about the participants, such as their education. In the current application it allowed us to show that higher education participants tend to apply common terms like VEGETABLE, FURNITURE, and TOOL more conservatively than compulsory education participants do (vagueness in degree) and that for terms they are arguably more familiar with, like SCIENCES, they employ different conditions for application (vagueness in criteria).

The procedure opens up the possibility to investigate which other properties of the participants systematically affect vagueness. In other work along these lines, we have so far identified a number of factors with ties to degree and criteria differences. We found degree differences in the application of adjectives that one can also apply to oneself to be related to one's own standing on the relevant dimension. Subjects' height and weight requirements for applying TALL and HEAVY, for instance, correlate positively with their personal measurements (Verheyen et al. 2018). Gender results in categorization patterns that are opposite to those we observed for education. Using a selection of categories that were expected to yield gender differences in categorization (e.g., CLOTHING, PROFESSIONS, SPORTS, TOYS) few degree differences were observed, while most categories yielded criteria differences (Stukken et al. 2013). Age gives rise to very intricate patterns of categorization. Both young children (<13 years old) and older adults (>62 years old) are found to overextend common noun categories compared to young adults (degree difference; Verheyen et al. 2011a; Verheyen et al. 2018). Older and young adults also differ in the criteria they use for categorization (White et al. 2018). For the categorization of storage containers, for instance, older adults rely more on "classic" materials such as glass or cardboard, whereas younger adults emphasize relatively "new" materials such as plastics. Using the proposed modeling framework, it is straightforward to extend this line of study to the comparison of different contexts, cultures, language groups, and even regional varieties, and/or to look into categorization differences at the level of individual items instead of entire categories, as we have done.

Ultimately, this research program will allow us to determine to what extent vagueness, as shown in inter-individual application differences, can be accounted for in terms of a limited number of external participant properties. As such, the program could be characterized as researching vagueness from a sociological or psychological perspective, depending on the nature of the property of interest that is investigated. The complete resolution of vagueness is not within reach of this program, however, as intra-individual application differences (Hampton and Passanisi 2016; McCloskey and Glucksberg 1978) cannot be explained in these terms. Our findings do raise the fundamental question of whether sociolinguistic variation of the kind uncovered in this study should be part of a theory of vagueness proper, or whether such a theory should only address the variation that is left after these external influences have been partialled out[4]. A choice for the latter might lead to inter-individual application differences being struck as hallmarks of vagueness, as it is impossible to ever ascertain whether all relevant participant properties have been taken into account.

The current project falls short when it comes to explaining where the effects of education on categorization come from. The purpose of this project was to establish *whether* education gives rise to particular categorization differences. Establishing *how* education gives rise to these differences, is more of an endeavor for sociolinguists and differential psychologists and is out of the scope of this chapter. Below we will nevertheless offer some suggestions as to the origins of the education effects we established, with the primary purpose of indicating how they could be formally tested within the framework we have proposed.

One explanation for the observation that compulsory education participants have broader categories than higher education participants do, might be the latter group's higher lexical familiarity (familiarity with printed words). Individuals that score high on lexical familiarity, have been found to display a higher rejection rate of non-category members that are semantically related to the category (Lewellen et al. 1993), reminiscent of the difference in degree we established. According to this explanation, educational level would be a proxy for verbal skill (see Kuperman and Van Dyke 2013, for support of this argument). An alternative explanation for the difference in degree could also be attempted in terms of personality characteristics that correlate with education. Education level correlates positively with both

[4]Conversely, one might ask whether we are not construing sociolinguistics too broadly if that might encompass the possession of specific forms of non-linguistic knowledge, which might explain the differences we found (see below).

openness and conscientiousness (Denissen et al. 2008), but academic performance is more strongly related to conscientiousness than to openness (Rocklin 1994; see also Paunonen and Ashton 2001). The observation that higher educated participants are more conscientious than compulsory education participants are, could be used to explain the degree differences observed in the majority of categories. According to this reasoning, higher education participants would ultimately reject more semantic foils because their deliberations are more thorough and deliberate than those of the compulsory education participants are. However, based on the relationship between education and openness, one could have predicted the opposite pattern as well. Higher education participants would then include more items in their categories than compulsory education participants do because they are more imaginative and creative and can therefore more easily come up with conditions in which an item fulfills the requirements for category membership.

For the pronounced differences in criteria for the category of SCIENCES, one can similarly come up with several explanations: (i) higher education participants differ in familiarity with the different disciplines on the basis of information they acquired directly in higher education courses or indirectly through contact with fellow students, (ii) higher education participants have been explicitly instructed about the nature of SCIENCES, (iii) higher education participants have been exposed to the manner in which higher education institutes are organized. Evidence for the second explanation can be found in our data in that higher education participants make a stronger distinction between pseudosciences like *astrology* and *palm reading* and prototypical natural sciences like *chemistry*, *meteorology*, and *mineralogy*. Evidence for the third explanation can be found in the clustering of *dentistry* and *nutrition* with medicine among higher education participants. All three explanations offered for the vagueness in criteria with respect to the SCIENCES category are reminiscent of categorization differences between experts and novices (e.g., Chi et al. 1981).

We did not entertain any a priori predictions as to how education might give rise to categorization differences. For demonstrative purposes we re-analyzed an existing data set, which was not gathered with the purpose of explaining education differences. If one were interested in the effects of lexical familiarity, conscientiousness, or openness on degree differences in categorization, it would be straightforward to collect this information from the participants and apply the procedure outlined in this chapter after dichotomizing these external variables (as is commonly done in experimental approaches). Alternatively, if one would like to honor the continuous nature of these variables, one could extend the model hierarchically to evaluate their relationship with the estimated thresholds (for a demonstration see Verheyen et al. 2011a). To interpret the criteria differences, one could regress the item positions onto the criteria under consideration and see which of these impact the item positions in the two groups differently. This is the approach taken by Verheyen and Storms (2013) and Verheyen et al. (2015). The regression could also be made part of the modeling by extending the model hierarchically, as in Verheyen et al. (2011b). Ideally, if one were to have specific hypotheses about potential criteria, one would compile the item set in a way that optimally allows one to disentangle the various criteria under consideration. Manipulations like these would make the proposed procedure an even more valuable experimental perspective on vagueness.

Acknowledgements Part of this research was conducted while SV was a postdoctoral fellow at the Research Foundation-Flanders (FWO). At the time of writing, SV was funded by ANR project *TriLogMean* (ANR-14-CE30-0010). SV also acknowledges grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL* for research conducted at the Department of Cognitive Studies of ENS in Paris. SV and GS thank Elisabeth Droeshout for her help with the data collection.

Appendix, Part 1: Categories and Items from Hampton, Dubois, and Yeh (2006) Along with Their Dutch Translation Used in the Current Study

FRUITS	FRUIT	VEGETABLES	GROENTEN	FISH	VISSEN	INSECTS	INSECTEN
acorn	eikel	apple	appel	alligator	krokodil	amoeba	amoebe
almond	amandel	artichoke	artisjok	catfish	zeewolf	ant	mier
avocado	avocado	asparagus	asperge	clam	mossel	bacterium	bacterie
banana	banaan	bamboo shoot	bamboescheut	crab	krab	bat	vleermuis
carrot	wortel	bread	brood	eel	paling	caterpillar	rups
coconut	kokosnoot	celery	selder	frog	kikker	centipede	duizendpoot
cucumber	komkommer	cereal	graan	goldfish	goudvis	dust mite	mijt
date	dadel	chili pepper	peper	gull	meeuw	earthworm	regenworm
eggplant	aubergine	cloves	kruidnagel	jellyfish	kwal	grashopper	sprinkhaan
ginger	gember	dandelion	paardenbloem	lobster	kreeft	hamster	hamster
mint	munt	garlic	look	oyster	oester	head lice	luis
mushroom	champignon	lettuce	sla	plankton	plankton	leech	bloedzuiger
olive	olijf	milk	melk	salmon	zalm	lizard	hagedis
onion	ajuin	parsley	peterselie	sardine	sardine	maggot	made
orange	sinaasappel	peanut	pinda	sea horse	zeepaardje	mosquito	mug
pine cone	denneappel	pineapple	ananas	seal	zeerob	moth	mot
pomegranate	granaatappel	potato	aardappel	shark	haai	scorpion	schorpioen
pumpkin	pompoen	rice	rijst	shrimp	garnaal	silkworm	zijderups
rhubarb	rabarber	sage	salie	sponge	spons	snail	slak
strawberry	aardbei	seaweed	zeewier	squid	inktvis	spider	spin
sugar beet	suikerbiet	soybean	soja	starfish	zeester	tapeworm	lintworm
tomato	tomaat	spinach	spinazie	tadpole	kikkervisje	tarantula	tarantula
walnut	walnoot	turnip	raap	trout	forel	termite	termiet
watermelon	watermeloen	watercress	waterkers	whale	walvis	wasp	wesp

Appendix, Part 2: Categories and Items from Hampton et al. (2006) Along with Their Dutch Translation Used in the Current Study

SPORTS	SPORTEN	SCIENCES	WETENSCHAPPEN	TOOLS	WERKTUIGEN	FURNITURE	MEUBELS
aerobics	aerobics	advertising	advertising	axe	bijl	ashtray	asbak
ballroom dancing	salondansen	agriculture	landbouw	broom	bezem	bed	bed
billiards	biljart	archaeology	archeologie	calculator	rekenmachine	book	boek
bridge	bridgen	architecture	architectuur	dictionary	woordenboek	bookends	boekensteun
bullfighting	stierengevecht	astrology	astrologie	funnel	trechter	bucket	emmer
chess	schaken	astronomy	sterrenkunde	hammer	hamer	chair	stoel
conversation	praten	chemistry	scheikunde	key	sleutel	curtains	gordijnen
croquet	croquet	criminology	criminologie	pen	balpen	cushion	poef
crosswords	kruiswoordpuzzelen	dentistry	tandheelkunde	photograph	foto	desk	bureau
darts	darts	economics	economie	pitchfork	hooivork	dishwasher	vaatwasmachine
fishing	vissen	geography	geografie	rake	hark	door mat	mat
frisbee	frisbee	geometry	geometrie	scalpel	scalpel	lamp	lamp
hiking	trektocht	literature	literatuur	scissors	schaar	painting	schilderij
hunting	jagen	mathematics	wiskunde	screw	schroef	piano	piano
jogging	joggen	medicine	geneeskunde	screwdriver	schroevendraaier	pillow	kussen
kite flying	vliegeren	meteorology	meteorologie	sewing needle	naald	plate	schotel
mountaineering	bergbeklimmen	mineralogy	mineralogie	shovel	schop	refrigerator	koelkast
picnicking	picknicken	nutrition	voedingsleer	stone	steen	rug	vloerkleed
skiing	skiën	palm reading	handlezen	string	koord	shelf	schap
surfing	surfen	pharmacy	farmacie	toothbrush	tandenborstel	suitcase	aktentas
swimming	zwemmen	philosophy	filosofie	tractor	tractor	table	tafel
tennis	tennis	psychology	psychologie	trunk	koffer	telephone	telefoon
weightlifting	gewichtheffen	religious studies	godsdienstleer	umbrella	paraplu	television	televisie
wrestling	worstelen	sociology	sociologie	varnish	vernis	waste basket	vuilbak

References

Alston, W. P. (1964). *Philosophy of language*. Englewood Cliffs: Prentice-Hall.
Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memories* (pp. 29–101). East Sussex, UK: Lawrence Erlbaum Associates.
Black, M. (1937). Vagueness: An exercise in logical analysis. *Philosophy of Science, 4*(4), 427–455.
Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.
Chomsky, N. (1980). *Rules and representations*. Oxford, UK: Basil Blackwell.

De Deyne, S., Voorspoels, W., Verheyen, S., Navarro, D. J., & Storms, G. (2014). Accounting for graded structure in adjective categories with valence-based opposition relationships. *Language and Cognitive Processes, 29*(5), 568–583.

Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90*(2), 152–157.

Devos, F. (1995). Still fuzzy after all these years. A linguistic evaluation of the fuzzy set approach to semantic vagueness. *Quaderni di Semantica, 16*(1), 47–82.

Devos, F. (2003). Semantic vagueness and lexical polyvalence. *Studia Linguistica, 57*(3), 121–141.

De Witte, E., Satoer, D., Robert, E., Colle, H., Verheyen, S., Visch-Brink, E., et al. (2015). The Dutch linguistic intraoperative protocol: A valid linguistic approach to awake brain surgery. *Brain and Language, 140,* 35–48.

Dry, M. J., & Storms, G. (2010). Features of graded category structure. *Acta Psychologica, 133*(3), 244–255.

Égré, P. (2017). Vague judgment: A probabilistic account. *Synthese, 194*(10), 3837–3865.

Charchat Fichman, H., Santos Fernandez, C., Alves Lourenço, R., Martins de Paiva Paradela, E., Carthery-Goulart, M. T., & Caramelli, P. (2009). Age and educational level effects on the performance of normal elderly on category verbal fluency tasks. *Dementia & Neuropsychologia, 3*(1), 49–54.

Frederickx, S., Tuerlinckx, F., De Boeck, P., & Magis, D. (2010). RIM: A random item mixture model to detect differential item functioning. *Journal of Educational Measurement, 47*(4), 432–457.

Gardner, R. W. (1953). Cognitive styles in categorizing behavior. *Journal of Personality, 22*(2), 214–233.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*(4), 441–461.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34*(5), 686–708.

Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition, 65*(2–3), 137–165.

Hampton, J. A. (2006). Concepts as prototypes. *The Psychology of Learning and Motivation: Advances in Research and Theory, 46,* 79–113.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*(3), 355–384.

Hampton, J. A., Dubois, D., & Yeh, W. (2006). The effects of pragmatic context on classification in natural categories. *Memory & Cognition, 34*(7), 1431–1443.

Hampton, J. A., & Passanisi, A. (2016). When intensions don't map onto extensions: Individual differences in conceptualization. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*(4), 505–523.

Kamp, H. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal semantics of natural language* (pp. 123–155). Cambridge, UK: Cambridge University Press.

Keefe, R. (2000). *Theories of vagueness*. Cambridge, UK: University Press.

Kennedy, C. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry, 56*(2–3), 258–277.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*(8), 1665–1692.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy, 4*(1), 1–45.

Kölbel, M. (2004). Faultless disagreement. *Proceedings of the Aristotelian Society* (Vol. 104, pp. 53–73).

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of work recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802–823.
Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General, 122,* 316–330.
Loonstra, A. S., Tarlow, A. R., & Sellers, A. H. (2001). COWAT metanorms across age, education, and gender. *Applied Neuropsychology, 8*(3), 161–166.
Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modeling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*(4), 325–337.
Machina, K. F. (1976). Truth, belief and vagueness. *Journal of Philosophical Logic, 5*(1), 47–78.
McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*(4), 462–472.
Paunonen, S. V., & Ashton, M. C. (2001). Big five predictors of academic achievement. *Journal of Research in Personality, 35*(1), 78–90.
Raffman, D. (1994). Vagueness without paradox. *Philosophical Review, 103*(1), 41–74.
Raffman, D. (1996). Vagueness and context-relativity. *Philosophical Studies, 81*(2–3) 175–192.
Raffman, D. (2014). *Unruly words*. Oxford, UK: Oxford University Press.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
Rocklin, T. (1994). Relation between typical intellectual engagement and openness: Comment on Goff and Ackerman (1992). *Journal of Educational Psychology, 86*(1), 145–149.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.
Rosenzweig, M. R. (1964). Word associations of French workmen: Comparisons with associations of French students and American workmen and students. *Journal of Verbal Learning and Verbal Behavior, 3*(1), 57–69.
Sassoon, G. W. (2012). A typology of multidimensional adjectives. *Journal of Semantics, 30*(3), 335–380.
Smith, L. B., & Samuelson, L. K. (1997). Perceiving and remembering: Category stability, variability and development. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 161–195). East Sussex, UK: Psychology Press.
Stukken, L., Verheyen, S., & Storms, G. (2013). Representation and criterion differences between men and women in semantic categorization. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3474–3479). Austin, TX: Cognitive Science Society.
Verheyen, S., Ameel, E., & Storms, G. (2011a). Overextensions that extend into adolescence: Insights from a threshold model of categorization. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2000–2005). Austin, TX: Cognitive Science Society.
Verheyen, S., De Deyne, S., Dry, M. J., & Storms, G. (2011b). Uncovering contrast categories in categorization with a probabilistic threshold model. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(6), 1515–1531.
Verheyen, S., Dewil, S., & Égré, P. (2018). Subjectivity in gradable adjectives: The case of *tall* and *heavy*. *Mind & Language*. In Press.
Verheyen, S., Droeshout, E., & Storms, G. (2018). *Age-related degree and criteria differences in semantic categorization*. Ms. University of Leuven and École Normale Supérieure.
Verheyen, S., Hampton, J. A., & Storms, G. (2010). A probabilistic threshold model: Analyzing semantic categorization data with the threshold model. *Acta Psychologica, 135*(2), 216–225.
Verheyen, S., & Storms, G. (2013). A mixture approach to vagueness and ambiguity. *PLoS ONE, 8*(5), e63507.

Verheyen, S., Voorspoels, W., & Storms, G. (2015). Inferring choice criteria with mixture IRT models: A demonstration using ad hoc and goal-derived categories. *Judgment and Decision Making, 10*(1), 97–114.
White, A., Storms, G., Malt, B. C., & Verheyen, S. (2018). Mind the generation gap: Differences between young and old in everyday lexical categories. *Journal of Memory and Language, 98,* 12–25.
Wright, C. (1995). The epistemic conception of vagueness. *The Southern Journal of Philosophy, 33*(S1), 133–159.
Zee, J., Storms, G., & Verheyen, S. (2014). Violations of the local independence assumption in categorization. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1670–1675). Austin, TX: Cognitive Science Society.