Revealing criterial vagueness in inconsistencies

Steven Verheyen[a,b]* (steven.verheyen@kuleuven.be)

Anne White[b] (anne.white@kuleuven.be)

Paul Égré[c] (paul.egre@ens.fr)

[a] Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, Paris France.

[b] Laboratory for Experimental Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven Belgium.

[c] Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, Paris France.

* Address for correspondence: Tiensestraat 102 box 3711, 3000 Leuven, Belgium.

Revealing criterial vagueness in inconsistencies

Sixty undergraduate students made category membership decisions for each of 132 candidate exemplar-category name pairs (e.g., *chess –* SPORTS) in each of two separate sessions. They were frequently inconsistent from one session to the next, both for nominal categories such as SPORTS and FISH, and ad hoc categories such as THINGS YOU RESCUE FROM A BURNING HOUSE. A mixture model analysis revealed that several of these inconsistencies could be attributed to criterial vagueness: participants adopting different criteria for membership in the two sessions. This finding indicates that categorization is a probabilistic process, whereby the conditions for applying a category label are not invariant. Individuals have various functional meanings of nominal categories at their disposal and entertain competing goals for ad hoc categories.

Keywords: categorization; vagueness; individual differences; semantic memory; ad hoc categories.

Word count: 4497

Revealing criterial vagueness in inconsistencies

1. Introduction

In 2006 the number of planets in our solar system suddenly dropped from nine to eight. This dramatic change was not due to some astronomical catastrophe, but to a change in the criteria for PLANETS adopted by the International Astronomical Union (IAU). Seeing that Pluto has not cleared the neighborhood around its orbit as the new criteria prescribed, the IAU decided that Pluto should no longer be considered a PLANET, but belongs in the category of DWARF PLANETS. In 2015 the High Court of Tarbes (France) overruled the earlier decision by the supreme court of appeal (Cour de Cassation) that involuntary homicide cannot be committed on a fetus, effectively changing what it means to be a PERSON. Both examples serve to show that even in scientific and legal contexts, where precision is arguably of the utmost importance, concepts are vague and the criteria for determining whether an instance belongs in a category or not are subject to change (Egré, 2018). Most of the concepts we use in our daily lives can be argued to be vague.

In psychology, the vague rather than well-defined nature of categories was convincingly demonstrated by McCloskey and Glucksberg (1978), who showed that participants not only differed in opinion as to whether items should be considered category members or not, but also changed their answer when asked the same question one month later. Participants presented with a list of candidate instances for nominal categories such as FISH and SPORTS, responded with a non-modal answer (a response that is different from the majority response) on 17% of the membership questions and provided inconsistent answers (a change in response after a one-month interval) on 12%. These results have been replicated by Hampton, Dubois, and Yeh (2006), who reported values for these inter- and intra-individual variability measures of 19% and 10%, respectively. Since the abandonment of the classic view of concepts as

involving singly necessary and jointly sufficient membership conditions (Rosch, 1973; Rosch & Mervis, 1975; Ryle, 1949; Wittgenstein, 1953), these differences are not recognized as mistakes, but as a manifestation of faultless disagreement (Kölbel, 2004; Wright, 1995), or permissible variation (Raffman, 2014), indicating that there are multiple, equally competent ways of applying a vague concept.

Although the idea that non-modal responses characterize vague categories was not new at the time (see Borel, 1907, and Black, 1937, for predecessors of the idea), McCloskey and Glucksberg's work contributed to inter-individual variability becoming a hallmark of vague categories. Nowadays, the existence of borderline items for which individuals can faultlessly disagree regarding category membership is considered to be central to what it means for a category to be vague (Kennedy, 2013; Smith, 2008; Wright, 1995).

The inter-individual variability observed in categorization tasks is generally thought to result from both indeterminacy with respect to the conditions for application, and indeterminacy with respect to the extent of application given fixed conditions (Verheyen & Storms, 2013, 2018). Three people may disagree as to whether *chess* and *hiking* are SPORTS, because one believes SPORTS should have competitive and game-like properties, while the other two only label activities that require physical effort SPORTS. On the basis of whether they consider *hiking* sufficiently effortful or not, the latter two could still disagree as to whether to call it a SPORT. The former indeterminacy is commonly referred to as *criterial vagueness*, while the latter is known as *degree vagueness* (Devos, 1995, 2003; see Alston, 1964; Burks, 1946; Kennedy, 2013; Machina, 1976, for a similar distinction).

In contrast to inter-individual variability, intra-individual variability has not caught on as a hallmark of vague categories. Although it has been acknowledged that vague categories have borderline cases for which an individual might feel equally inclined to apply and to deny the

category label (Schiffer, 2003) – evidenced by increased categorization reaction times and lower confidence ratings (Koriat & Sorka, 2015) as well as competing responses to the same stimulus at a given time (Malt, 1990) – within-subject inconsistencies in categorization rarely constitute the topic of investigation themselves (see Hampton, Aina, Andersson, Mirza, & Parmar, 2012, for a notable exception). Intra-individual categorization differences tend to be accounted for in terms of shifting thresholds. What is believed to change from one occasion to the other, is the extent of the evidence the individual requires to apply the category label, not the conditions for application (Hampton, 1995; McCloskey & Glucksberg, 1978). Inconsistent answers are thus thought to reflect degree rather than criterial vagueness. The implicit assumption here seems to be that qualitatively different conceptions of a category might be entertained by different people, but not by an individual. It is this hypothesis that we put to the test in this paper.

2. Outline

The observation by McCloskey and Glucksberg (1978) that people provide inconsistent answers when asked to repeat a categorization task, indicates that the information that is retrieved from semantic memory is not invariant. The probabilistic nature of the semantic retrieval process is corroborated by the modest reliability of repeated exemplar generation (Bellezza, 1984a; White, Voorspoels, Storms, & Verheyen, 2014), category definitions (Barsalou, 1989; Bellezza, 1984b), feature importance ratings (Hampton & Passanisi, 2016) and typicality judgments (Barsalou, 1987, 1989; Hampton & Passanisi, 2016). While these studies allow one to establish *how much* change to expect from one occasion to the next, they do not indicate *what* it is that changes over time. The purpose of this paper is to elucidate whether the criteria that are used to establish category membership may change.

Criterial vagueness has not yet been demonstrated within individuals. Different individuals have been shown to use distinct criteria for categorization, however (Verheyen & Storms, 2013; Verheyen, Voorspoels, & Storms, 2015; White, Storms, Malt, & Verheyen, 2018). This has been achieved using a mixture model that identifies subgroups of categorizers depending on the latent conditions they adhere to for categorization (criterial vagueness). Within each of the identified subgroups, the participants were also found to differ on the extent to which they required instances to demonstrate these conditions to be eligible for categorization (degree vagueness). The rationale behind the mixture model is that the use of distinct criteria will show in the relative frequency with which items are endorsed in subgroups. The items *chess* and *darts* will be more often categorized as SPORTS in a group emphasizing competitive and game-like properties than in a group looking to physical exertion to establish category membership. The use of distinct thresholds will show in the proportion of categorized items. Participants who require little evidence for category membership, will also include items that have relatively lower categorization frequencies in the subgroup, whereas very demanding participants will only include items that are frequently endorsed, as this indicates that these items score high on the subgroup's categorization criterion.

Our design involves having participants complete a categorization task twice. We will apply the mixture model described above to the repeated data to determine whether any of the participants are assigned to different subgroups on the two occasions. This would indicate that their inconsistent answering reflects criterial vagueness. We will investigate both nominal categories like FISH and SPORTS, and ad hoc categories like THINGS YOU RESCUE FROM A BURNING HOUSE. Because unlike nominal categories, ad hoc categories violate the correlational structure of the environment and are not well established in memory (Barsalou, 1983), we expect more intra-individual categorization differences and more criterial accounts of inconsistent answers in ad hoc categories.

3. Design and Procedure

All materials, data, and analysis scripts are available on the Open Science Framework (https://osf.io/pnm9a/).

3.1. Ethics statement

This study was conducted with the approval of the Social and Societal Ethics Committee of KU Leuven. Written informed consent was obtained from all participants both at the start of the first and second categorization session.

3.2. Participants

We invited 65 first-year psychology students at KU Leuven to take a categorization task twice, in exchange for course credit. Sixty of them completed both sessions (92%). Of these 60 participants, 5 were male (8.33%). The participants' age ranged between 17 and 20 years (M=18.08, SD=.65).

3.3. Materials

Verheyen and Storms (2013) and Verheyen, Voorspoels, and Storms (2015) investigated whether degree and criterial vagueness could account for inter-individual categorization differences in nominal and ad hoc categories, respectively. We selected three nominal categories and three ad hoc categories from among the categories in these papers that showed evidence of criterial vagueness in the form of two subgroups of participants identified by the mixture analysis. Among the five qualifying ad hoc categories, we did not include the two categories with a very uneven distribution of participants over subgroups since we expected hardly any participants in our sample to subscribe to the categorization criteria of the smaller

subgroup (comprising less than 10% of the participants in the original paper). In order to have an equal number of nominal categories, we randomly selected three among the four qualifying nominal categories.

The nominal categories FISH, SPORTS, and TOOLS had 24 items each. The ad hoc categories THINGS YOU RESCUE FROM A BURNING HOUSE, MEANS OF TRANSPORT BETWEEN BRUSSELS AND LONDON, and WEAPONS USED FOR HUNTING had 20 items each. These items comprised the full range of category membership, including several clear members and clear non-members, but mainly borderline cases. All the materials were presented in Dutch.

3.4. Procedure

Participants were administered a computerized categorization task in which the materials were presented in two blocks (nominal vs. ad hoc) of three categories each. The presentation order of blocks, categories within a block, and items within a category was randomized for every participant. Separate screens for each category would display the categorization instructions on top, indicating that participants could answer *yes* or *no* to the question of whether the items that followed belong to the category or not. A third response option, labeled *unknown*, was meant to be used when participants did not know a particular item or felt an item was ambiguous and did not know which meaning was intended.

Approximately one month after completing the categorization task, participants were presented the same task again. Following McCloskey and Glucksberg (1978), they were informed that some instances of the first session could appear again.

4. Results

We report the results in two separate sections. In the first, we use linear mixed effects models to investigate whether the prevalence of inconsistencies across categorization sessions differs between nominal and ad hoc categories. In the second section, we apply the mixture model from the studies that informed the stimulus selection to the repeated categorization data in order to determine to what extent differences between sessions represent criterial vagueness.

4.1. Prevalence of inconsistencies

Seventy-three percent of the participants provided at least one inconsistent response (i.e., a change in response across sessions: *yes*/*no*, *yes*/*unknown*, or *no*/*unknown*) for each of the six categories. Sixteen participants (27%) answered consistently on one category, but not on the other five (FISH: $N=7$; RESCUE: $N=6$; SPORTS, TRANSPORT, and WEAPONS: $N=1$). No one answered consistently for the nominal category TOOLS. A parallel pattern was observed for the items. Inconsistent answers were observed for all items ($n=132$) except for eight (94%). The items that yielded perfectly consistent answers were all clear members of the target categories (FISH: *goldfish*, *trout*; SPORTS: *skiing*, *swimming*, *tennis*; TOOLS: *axe*, *hammer*; RESCUE: *people*). None of the items for the other two ad hoc categories yielded perfectly consistent answers.

Figure 1 depicts the proportion of inconsistent responses for each of the participants (left panel) and each of the target items (right panel). On average, the participants provided an inconsistent answer on 16.06% of the membership questions for the nominal categories and on 19.00% of the membership questions for the ad hoc categories.
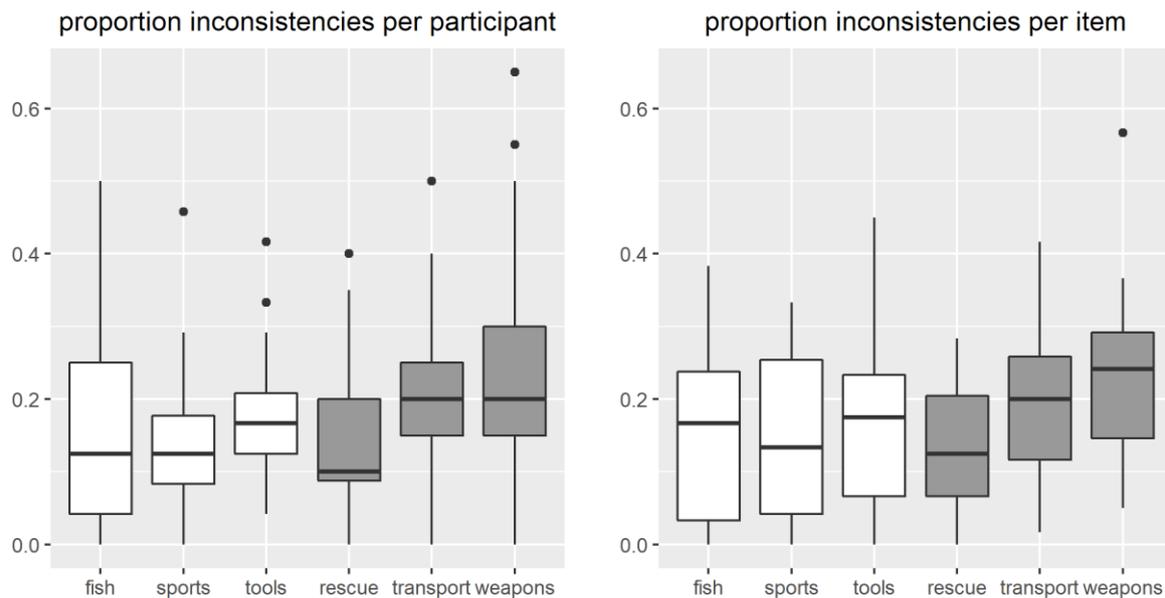
Figure 1: Boxplots of the proportion of inconsistencies across sessions for each participant (left panel) and each target item (right panel). Boxplots for nominal categories are depicted in white; boxplots for ad hoc categories are depicted in gray. The bands in the boxes represent the median values.

To establish whether ad hoc and nominal categories differed with respect to intra-individual differences, we determined whether participants' repeated responses were inconsistent or not and fitted a binomial mixed effects model to the resulting variable, using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.4.3 (R Core Team 2017). The fixed part of the model contained the main effect of the binary variable block, indicating whether the answers pertained to a nominal (1) or ad hoc (0) category. The random part of the model included random category, item, and participant intercepts, and an interaction between the block and participant variables. The main effect of block (β=-.35, SE=.19, z=-1.88, p = .06) was not significant at α = .05. This result was supported by comparing the BIC of the above model to that of an alternative model from which the main effect of block was removed

(BIC$_{simple}$=6894.60 vs. BIC$_{full}$=6900.80)[1]. In other words, we did not reject the null hypothesis that the prevalence of intra-individual differences differs between nominal and ad hoc categories.

4.2. Prevalence of criterial vagueness

Figure 2 holds a graphical depiction of the mixture model (Lee & Wagenmakers, 2014). It considers each categorization decision x_{ip} the outcome of a Bernoulli trial (1 for *yes*, 0 for *no*) with the probability of a membership response to item i by participant p expressed by r_{ip}. It assumes the data result from a mixture of participants who adhere to different criteria for categorization. Depending on a participant's latent group membership z_p, different estimates are obtained for the item parameters β_i, which express the extent to which the items display the group's categorization criterion. The β_i are compared against the participant's internal threshold θ_p to establish the items' category membership. Whereas differences in β_i signal vagueness in criteria, differences in θ_p capture degree vagueness or the amount of evidence participants require for category membership. The parameter α determines for each group the shape of the function that relates the extent to which an item surpasses/falls short of the threshold to the probability of categorization. The function is S-shaped: it starts off at a zero when the $\beta_i - \theta_p$ difference is large and negative, demonstrates an increase for small differences between β_i and θ_p, and asymptotes to one when the difference grows large and

---

[1]The smaller BIC value for the simple model indicates that it provides the more parsimonious account of the data in terms of fit and complexity (Schwarz, 1978). The increase in fit that is obtained by discerning between nominal and ad hoc categories in the full model does not warrant its added complexity/flexibility compared to the simpler model.

positive. The value of α reflects the steepness of the function at the point of subjective equality (the point for which the categorization probability equals .50, when $\beta_i=\theta_p$).
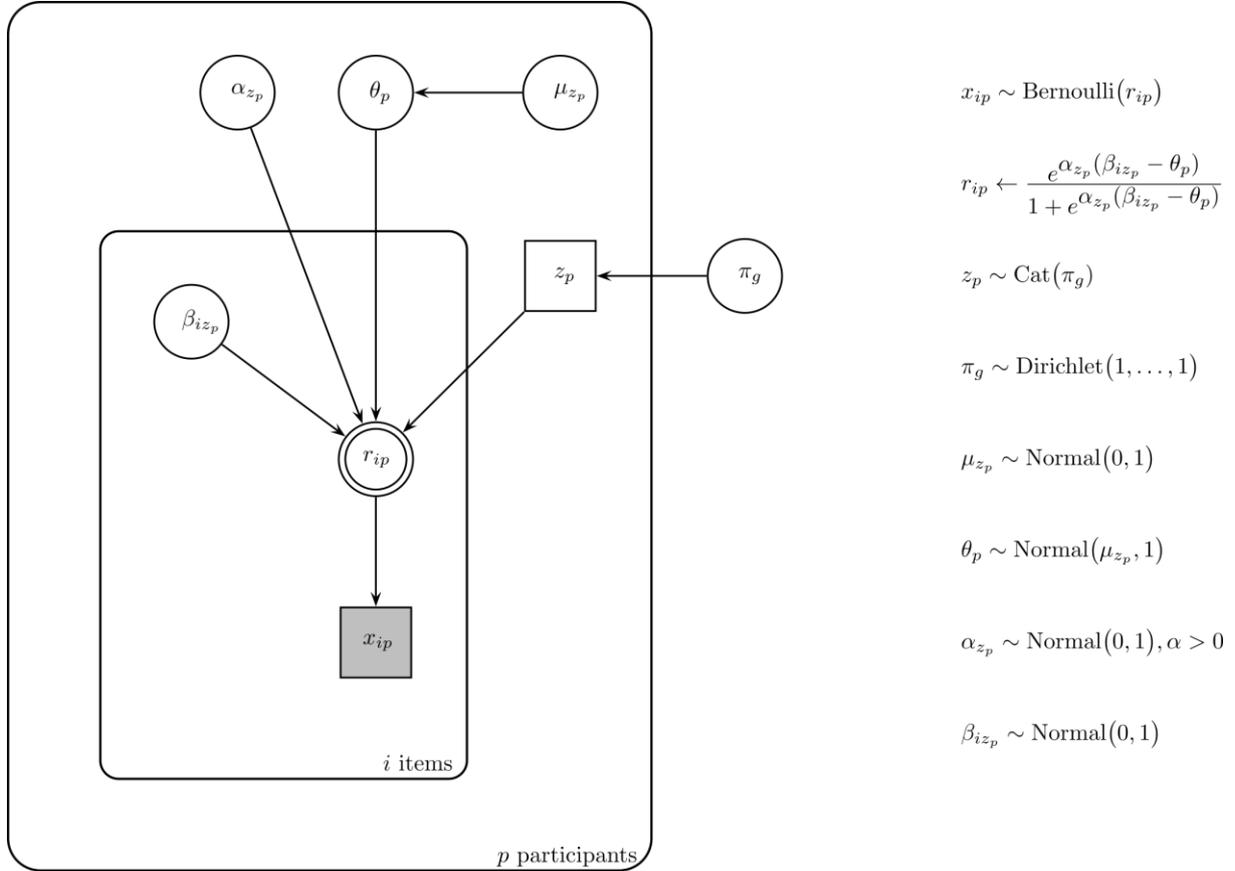

Figure 2: Graphical model representation of the mixture model.

$$x_{ip} \sim \text{Bernoulli}(r_{ip})$$

$$r_{ip} \leftarrow \frac{e^{\alpha_{z_p}(\beta_{iz_p} - \theta_p)}}{1 + e^{\alpha_{z_p}(\beta_{iz_p} - \theta_p)}}$$

$$z_p \sim \text{Cat}(\pi_g)$$

$$\pi_g \sim \text{Dirichlet}(1, \dots, 1)$$

$$\mu_{z_p} \sim \text{Normal}(0, 1)$$

$$\theta_p \sim \text{Normal}(\mu_{z_p}, 1)$$

$$\alpha_{z_p} \sim \text{Normal}(0, 1), \alpha > 0$$

$$\beta_{iz_p} \sim \text{Normal}(0, 1)$$

Latent group membership z_p is parameterized in the model as a categorically distributed random variable with π_g reflecting the probability of belonging to group g. The threshold parameters θ_p are drawn from normal hyper-distributions, parameterized by group-specific means and precision 1. We employed a uniform Dirichlet prior for the membership probabilities π_g, a half-normal distribution centered at 0 with precision 1 for each α, and normal priors centered at 0 with the precision set to 1 for the remaining model parameters.

The repeated categorization data from the current study were merged with the categorization data that were available for the same materials from earlier work (Verheyen & Storms, 2013; Verheyen, Voorspoels, & Storms, 2015). The merging ensures that we have enough data to obtain reliable parameter estimates for the different subgroups. The two sets of categorization responses by the participants who took the categorization task twice were included as independent entries. *Unknown* responses were treated as missing values. For each of the nominal categories, the merged data thus comprised 370 categorization responses to each of 24 items (2 x 60 new responses + 250 responses from Verheyen & Storms, 2013). For the ad hoc categories, the merged data comprised 374 x 20 categorization responses (2 x 60 new + 254 from Verheyen et al., 2015). The mixture model was applied to these merged data sets.

Separate model estimates were obtained for each of the categories using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) by running 3 chains of 10,000 samples each, with a burn-in of 4,000 samples. The chains were checked for convergence and label switching. For every category a 2-group solution was obtained since for all six selected categories two subgroups of participants were identified by the original mixture analyses (Verheyen & Storms, 2013; Verheyen et al., 2015). The original groups were recovered in the analysis of the merged categorization data, as evidenced by the correlations between the posterior means of the old and new group-specific β_i estimates (all $\rho > .95$).[2]

The focus here will be whether the 60 participants who completed the categorization task twice are assigned to a different group upon repetition. This would indicate that they relied on distinct criteria for categorization in the two sessions. Group membership was determined based on the posterior mode of z_p. We observed numerous group changes from Session 1 to

---

[2] For a substantive interpretation of the categorization criteria, see section 1 of the Supplementary Materials.

Session 2.[3] For only 27% of the participants no change in group membership was observed. These participants were placed in the same group on both occurrences of the categorization task for all six categories. Seventy-three percent of participants thus demonstrated a group change for at least one category. Thirty-two percent of participants changed group for two categories. Three percent changed group for three categories. There were no participants for which a group change was established for more than three categories. These percentages indicate that criterial vagueness is present *within* individuals.[4]

For the nominal category FISH, 9 out of 53 participants (17%) who demonstrated at least one inconsistency were placed in different groups on the two occasions. For SPORTS and TOOLS these percentages equaled 24% (14/59) and 28% (17/60), respectively. Fewer group changes were observed for the ad hoc categories: 11/54 (20%) for RESCUE, 7/59 (12%) for TRANSPORT, and 9/59 (15%) for WEAPONS. We constructed a new variable indicating whether the mixture analysis placed participants in different groups on the two repetitions or not, and fitted a binomial mixed effects model to it. The fixed part of the model contained the main effect of the binary variable block, indicating whether the answers pertained to a nominal (1) or ad hoc (0) category. The random part of the model included random category and participant intercepts, and an interaction between the block and participant variables. The effect of block (β=.48, *SE*=.28, *z*=1.75, *p* = .08) was not significant at α = .05. This result was supported by

---

[3] Participants changing groups tended to have a high probability of being assigned to a subgroup in one session and a low probability of being assigned to the same group in the other session, rather than having similar assignment probabilities in both sessions. See section 2 of the Supplementary Materials for details.

[4] The percentages demonstrate that intra-individual variability can be due to criterial vagueness, but presumably overestimate the overall prevalence of criterial vagueness. We observed slightly more intra-individual categorization differences than earlier studies did (16% compared to 12% in McCloskey and Glucksberg, 1978, and 10% in Hampton, Dubois, and Yeh, 2006). This discrepancy might be due to our selection of categories with known criterial vagueness (established between rather than within individuals). The earlier studies included a broader range of nominal categories than we did, which need not all display criterial vagueness. According to Verheyen and Storms (2013) 5 out of 8 categories in Hampton et al. (2006) demonstrated criterial vagueness; 6 out of 10 ad hoc categories in Verheyen, Voorspoels, and Storms (2015) displayed criterial vagueness. The fact that we also counted an *unknown* response on one occasion and a *yes* or *no* response on the other occasion as inconsistencies, might contribute to the discrepancy as well.

comparing the BIC of the above model to that of an alternative model from which the main effect of block was removed (BIC_{simple}=369.35 vs. BIC_{full}=372.30). While group changes thus appeared less frequent for the ad hoc categories than for the nominal categories, this difference was not significant.

5. General Discussion

For three nominal and three ad hoc categories, participants decided on the category membership of various target items. They completed the task twice, separated by a one-month interval. Inconsistent answers were ubiquitous. Participants rarely provided identical responses on both occasions. We established that this intra-individual variability was not exclusively the result of degree vagueness (participants changing the amount of evidence required for membership, given constant conditions across sessions), as was assumed up until now. Several of these inconsistencies could be attributed to criterial vagueness: participants adopting different conditions for application in the two sessions. Each of these participants was placed in distinct groups on the two sessions by a mixture model that identifies latent groups of participants who employ different categorization criteria.

For nominal categories the existence of criteria differences within individuals indicates that people have various "meanings" at their disposal, which are probabilistically retrieved from semantic memory. McCloskey and Glucksberg (1978) refer to these meanings as *functional categories*, suggesting that they can be relied upon to serve different purposes (e.g., FISH in the zoological vs. the seafood sense; see also Hampton, Dubois, & Yeh, 2006, and Verheyen & Storms, 2013). The possibility to recruit different subsets of category knowledge allows for efficient processing in that information that is most relevant to accomplish particular tasks can be focused on (Yeh & Barsalou, 2006). It might make memories and truth judgments less

reliable, however, as information recall and property verification might differ depending on the functional meaning that is accessed (Hampton et al., 2012). The challenge for future work is to determine how particular meanings are likely to become activated on a given occasion and to establish whether it is tenable to argue for context- and task-independent category representations if people are highlighting a particular conceptual content whenever they use a category label (Braisby, 1993).

The observation that inconsistent categorization responses can result from criterial vagueness, holds for both nominal and ad hoc categories. We found no significant difference regarding the prevalence of inconsistencies or of criterial changes in nominal vs. ad hoc categories. This might strike some as surprising given that ad hoc categories are thought to be less rooted in the environment and in semantic memory than nominal categories are (Barsalou, 1983) and therefore might be expected to show less stability. The lack of a stability difference might be an indication that the ad hoc categories we selected should be considered goal-derived categories: ad hoc categories that have become well-established in memory, e.g. through frequent use (Barsalou, 1985). The observation that one and the same individual may use different criteria for recruiting items that fulfill the category's goal, would then be an indication that people sometimes entertain competing goals - such as travelling comfortably or fast between Brussels and London - the prominence of which might change from one occasion to the next (see also Voorspoels, Storms, & Vanpaemel, 2013, who showed that individuals can provide *multiple* ideal characteristics of goal-derived categories).

There is no reason to assume that the occurrence of criterial vagueness is specifically related to particular word classes (Verheyen & Storms, 2018). Our findings pertain to nominal and ad hoc categories, but are likely to generalize to other paradigmatic examples of vague categories, such as gradable adjectives like INTELLIGENT and HEALTHY. The individual-level symptoms of vagueness that were discussed in the introduction for nouns, have also been

shown to exist for gradable adjectives. They too show competing responses to borderline items (borderline contradictions; see Alxatib & Pelletier, 2011; Egré & Zehr, 2018; Hersh & Caramazza, 1976; Ripley, 2011), increased reaction times and decreased confidence ratings for borderline items (Brownell & Caramazza, 1978; Hersh & Caramazza, 1976), and inconsistent responding across categorization sessions (Egré, de Gardelle, & Ripley, 2013; Hersh & Caramazza, 1976). Solt (2018) offers a treatment of how degree and criterial vagueness can account for inter- and intra-individual differences in the application of gradable adjectives. Much like Verheyen and Storms (2013) argued for nouns, she suggests that the judge- and context-dependent weighting of the multiple dimensions that underlie many gradable adjectives, is responsible for the observed variability in their use.

Our examples of vague concepts pertain to higher level categories, which tend to be comprised of heterogeneous instances that share similar functions rather than appearances. For perceptual categories it remains to be seen whether inconsistent answers can be attributed to criterial vagueness. Whether it can, might depend on the frequency with which individuals categorize instances differently. We know that children as young as 14 months old can flexibly shift the criteria they use for categorizing objects in response to tasks requirements or instructions (for instance, from using shape to relying on material; Ellis & Oakes, 2006). We believe that the more this occurs, the more likely it becomes that individuals will develop multiple representations that remain accessible for later (functional) use (Schyns & Rodet, 1997).

Finally, this paper advocates the study of intra-individual differences in vagueness research. Although inter-individual variability is generally considered a hallmark of vague categories, its manifestation is not necessarily due to vagueness, but can be an indication of stable differences between subgroups of categorizers. For example, the same light stimulus may be categorized as one color by a color-normal perceiver, but stably (without uncertainty or

unclarity being experienced) as another color by a person with protanopia (a form of colour blindness characterized by a tendency to confuse reds and greens and by a loss of sensitivity to red light; Paramei, Bimler, & Cavonius, 1998). In addition, inter-individual application differences can often be systematically related to properties of the individuals (tall people imposing higher height requirements than short people to name others tall; Verheyen, Dewil, & Egré, 2018; higher educated people applying nominal categories more conservatively; Verheyen & Storms, 2018; older people looking at traditional rather than modern materials to apply container labels; White, Storms, Malt, & Verheyen, 2018). Intra-individual differences cannot be attributed to participants' background differences and therefore provide a more direct window into the probabilistic nature of categories.

Acknowledgments

References

Alston, W. P. (1964). *Philosophy of language*. Englewood Cliffs, NJ: Prentice-Hall.

Alxatib, S., & Pelletier, F. J. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language, 26*, 287–326. https://doi.org/10.1111/j.1468-0017.2011.01419.x

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211–227. https://doi.org/10.3758/bf03196968

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654. https://doi.org/10.1037//0278-7393.11.1-4.629

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.

Barsalou, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511529863.006

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48. https://doi.org/10.18637/jss.v067.i01

Bellezza, F.S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society, 22*, 324–326. https://doi.org/10.3758/bf03333832

Bellezza, F.S. (1984b). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society, 22*, 377–380. https://doi.org/10.3758/bf03333850

Black, M. (1937). Vagueness: An exercise in logical analysis. *Philosophy of Science, 4*, 427-455. https://doi.org/10.1086/286476

Borel, E. (1907). Un paradoxe économique: Le sophisme du tas de blé et les vérités statistiques. *La Revue du Mois, 4*, 688–699. [English translation by Égré, P., & Gray, E. (2014). An economic paradox: The sophism of the heap of wheat and statistical truths. *Erkenntnis, 79*, 1081–1088. https://doi.org/10.1007/s10670-014-9615-z]

Braisby, N. R. (1993). Stable concepts and context-sensitive classification. *Irish Journal of Psychology, 14*, 426-441. https://doi.org/10.1080/03033910.1993.10557948

Brownell, H. H., & Caramazza, A. (1978). Categorizing with overlapping categories. *Memory & Cognition, 6*, 481-490. https://doi.org/10.3758/bf03198235

Burks, A. W. (1946). Empiricism and vagueness. *Journal of Philosophy, 43*, 477–486. https://doi.org/10.2307/2018954

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

Devos, F. (2003). Semantic vagueness and lexical polyvalence. *Studia Linguistica, 57*, 121-141. https://doi.org/10.1111/j.0039-3193.2003.00101.x

Egré, P. (2017). Vague judgment: A probabilistic account. *Synthese, 194*, 3837-3865. https://doi.org/10.1007/s11229-016-1092-2

Egré, P. (2018). *Qu'est-ce que le vague?* [What is vagueness?]. Paris, France: Collection Chemins Philosophiques, Vrin.

Egré, P., de Gardelle, V., & Ripley, D. (2013). Vagueness and order effects in color categorization. *Journal of Logic, Language and Information, 22*, 391–420. https://doi.org/10.1007/s10849-013-9183-7

Egré, P., & Zehr, J. (2018). Are gaps preferred to gluts? A closer look at borderline contradictions. In G. W. Sassoon, L. McNally, & E. Castroviejo (Eds.), *The semantics of gradability, scale structure, and vagueness: Experimental perspectives* (pp. 25-58). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-319-77791-7_2

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology, 42*, 1000-1011. https://doi.org/10.1037/0012-1649.42.6.1000

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34*, 686−708. https://doi.org/10.1006/jmla.1995.1031

Hampton, J. A., Aina, B, Andersson, J. M., Mirza, H. Z. & Parmar, S. (2012). The Rumsfeld effect: The unknown unknown. *Journal of Experimental Psychology: Learning, Memory & Cognition, 38*, 340-355. https://doi.org/10.1037/a0025376

Hampton, J. A., Dubois, D., & Yeh, W. (2006). The effects of pragmatic context on classification in natural categories. *Memory & Cognition, 34*, 1431-1443. https://doi.org/10.3758/bf03195908

Hampton, J. A., & Passanisi, A. (2016). When intensions don't map onto extensions: Individual differences in conceptualisation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 505-523. https://doi.org/10.1037/xlm0000198

Hersh, H. M., & Caramazza, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General, 105*, 254-276. https://doi.org/10.1037//0096-3445.105.3.254

Kennedy, C. (2013). Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry, 56*, 258-277. https://doi.org/10.1080/0020174x.2013.784483

Kölbel, M. (2004). Faultless disagreement. *Proceedings of the Aristotelian Society, 104*, 53–73. https://doi.org/10.1111/j.0309-7013.2004.00081.x

Koriat, A., & Sorka, H. (2015). The construction of categorization judgments: Using subjective confidence and response latency to test a distributed model. *Cognition, 134*, 21-38. https://doi.org/10.1016/j.cognition.2014.09.009

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337. https://doi.org/10.1023/a:1008929526011

Machina, K. F. (1976). Truth, belief and vagueness. *Journal of Philosophical Logic, 5*, 47-78. https://doi.org/10.1007/bf00263657

Malt, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language, 29*, 289-315. https://doi.org/10.1016/0749-596x(90)90002-h

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6*, 462−472. https://doi.org/10.3758/bf03197480

Paramei, G. V., Bimler, D. L., & Cavonius, C. R. (1998). Effect of luminance on color perception of protanopes. *Vision Research, 38*, 3397-3401. https://doi.org/10.1016/s0042-6989(97)00454-9

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raffman, D. (2014). *Unruly words*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199915101.001.0001

Ripley, D. (2011). Contradictions at the borders. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (eds.), *Vagueness in communication* (pp. 169–188). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-18446-8_10

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York, NY: Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50010-4

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605. https://doi.org/10.1016/0010-0285(75)90024-9

Ryle, G. (1949). *The concept of mind*. London, UK: Hutchinson. https://doi.org/10.4324/9780203875858

Schiffer, S. (2003). *The things we mean.* Oxford, UK: Oxford University Press.

https://doi.org/10.1093/0199257760.001.0001

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

https://doi.org/10.1214/aos/1176344136

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 23, 681-696.

https://doi.org/10.1037//0278-7393.23.3.681

Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford, UK: Oxford University Press.

https://doi.org/10.1093/acprof:oso/9780199233007.001.0001

Solt, S. (2018). Multidimensionality, subjectivity and scales: Experimental evidence. In E.

Castroviejo, L. McNally, & G. W. Sassoon (Eds.), *The semantics of gradability, vagueness,*

and scale structure: Experimental perspectives (pp. 59-91). Berlin, Germany: Springer.

https://doi.org/10.1007/978-3-319-77791-7_3

Verheyen, S., Dewil, S., & Egré, P. (2018). Subjectivity in gradable adjectives: The case of

tall and *heavy*. *Mind & Language, 33,* 460-479. https://doi.org/10.1111/mila.12184

Verheyen, S., Hampton, J. A., & Storms, G. (2010). A probabilistic threshold model:

Analyzing semantic categorization data with the Rasch model. *Acta Psychologica, 135*, 216-

225. https://doi.org/10.1016/j.actpsy.2010.07.002

Verheyen, S., & Storms, G. (2013). A mixture approach to vagueness and ambiguity. *PLoS*

ONE, 8(5), e63507. https://doi.org/10.1371/journal.pone.0063507

Verheyen, S., & Storms, G. (2018). Education as a source of vagueness in criteria and degree. In E. Castroviejo, L. McNally, & G. W. Sassoon (Eds.), *The semantics of gradability, vagueness, and scale structure: Experimental perspectives* (pp. 149-167). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-319-77791-7_6

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*, 97-114.

Voorspoels, W., Storms, G., & Vanpaemel, W. (2013). Similarity and idealness in goal-derived categories. *Memory & Cognition, 41*, 312-327. https://doi.org/10.3758/s13421-012-0252-y

White, A., Storms, G., Malt, B. C., & Verheyen, S. (2018). Mind the generation gap: Differences between young and old in the representation of everyday lexical categories. *Journal of Memory and Language, 98*, 12-25. https://doi.org/10.1016/j.jml.2017.09.001

White, A., Voorspoels, W., Storms, G., & Verheyen, S. (2014). Problems of reliability and validity with similarity derived from category fluency. *Psychiatry Research, 220*, 1125-1130. https://doi.org/10.1016/j.psychres.2014.10.001

Wittgenstein, L. (1953). *Philosophical investigations* (2nd ed.). Oxford, UK: Blackwell.

Wright, C. (1995). The epistemic conception of vagueness. *The Southern Journal of Philosophy, 33*, 133–159. https://doi.org/10.1111/j.2041-6962.1995.tb00767.x

Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *The American Journal of Psychology, 119*, 349-384. https://doi.org/10.2307/20445349