Running head: VISUAL ANALOGUE VS. DISCRETE SCALES

This is the final draft of the manuscript accepted for publication at Consciousness and Cognition. Please cite this work as follows:
Rausch, M., & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consicousness and Cognition. 28, 126-140. doi:10.1016/j.concog.2014.06.012*

A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion

Manuel Rausch[ab] and Michael Zehetleitner[a]

[a] Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

[b] Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence should be addressed at
Manuel Rausch
Department of Psychology
Ludwig-Maximilians-Universität München
Leopoldstraße 13, 80802 Munich, Germany
Tel: +49 89 2180 5152, Fax: +49 89 2180 5211
Email: manuel.rausch@psy.lmu.de

Abstract

Can participants make use of the large number of response alternatives of visual analogue scales (VAS) when reporting their subjective experience of motion? In a new paradigm, participants adjusted a comparison according to random dot kinematograms with the direction of motion varying between 0 and 360°. After each discrimination response, they reported how clearly they experienced the global motion either using a VAS or a discrete scale with four scale steps. We observed that both scales were internally consistent and were used gradually. The visual analogue scale was more efficient in predicting discrimination error but this effect was mediated by longer report times and was no longer observed when the VAS was discretized into four bins. These observations are consistent with the interpretation that VAS and discrete scales are associated with a comparable degree of metacognitive sensitivity, although the VAS provides a greater amount of information.

Keywords:

Consciousness, visual awareness, subjective report, random dot motion, visual analogue scale, rating scale, experience, information theory

1. Introduction

The lack of an established measurement for conscious experience is a key challenge to the prosperity of an empirical science of consciousness (Chalmers, 1998). The choice of an adequate measure is delicate because different theoretical perspectives on consciousness can imply different measurements. Some theorists are critical about the use of subjective reports because they assume participants might have conscious experiences they are unable to report (Block, 2005) or they do not report because their criterion is too conservative (Hannula, Simons, and Cohen, 2005). In contrast, proponents of higher-order thought theories often argue that subjective reports are more valid than objective measures because unconscious processes might drive objective performance as well (Dienes, 2004; Lau, 2008). However, as subjective experiences cannot be observed from the third-person point of view (Jackson, 1982; Nagel, 1974), it is impossible to test empirically whether subjective measures of consciousness leave out conscious experiences that observers are unable to report, or whether objective measures suggest falsely that performance in a task is conscious. However, some researchers decide a priori to adopt a perspective that requires the use of subjective reports, either because they endorse a higher-order perspective on consciousness (Cleeremans, 2011; Lau & Rosenthal, 2011), or because they consider subjective reports themselves as the subject of their scientific investigations (Dennett, 2003, 2007); if they do so, the empirical question arises how a scale needs to be designed given the metacognitive abilities of humans to obtain as much information from participants as possible.

1.1. The content of subjective scales

Subjective scales designed to measure conscious experience are constituted out of at least two components: (i) the question participants are instructed to answer and (ii) the way participants deliver their subjective report. Concerning the question, we proposed a classification of subjective scales on the event in the world subjective reports refer to, specifically whether subjective reports refer to the stimulus or to the discrimination response (Zehetleitner & Rausch, 2013). Examples for stimulus-related scales would be to ask participants how visible the stimulus was (Sergent & Dehaene, 2004), to rate clarity of the response defining feature (Zehetleitner & Rausch, 2013), or to report both the experience of specific features as well as feelings of something being shown (Ramsøy & Overgaard, 2004, p. 12). Decision-related scales may ask participants to report how confident they are about the preceding objective task response (Peirce & Jastrow, 1884), whether they attribute their objective task response to guessing, intuition, memory, or knowledge (Dienes & Scott, 2005), how much money they would wager on the accuracy of the objective task response (Persaud, McLeod, and Cowey, 2007), or whether they experienced a “feeling-of-warmth” with respect to the previous task response (Wierzchoń, Asanowicz, Paulewicz, and Cleeremans, 2012).

Several studies compared subjective scales with different questions participants were asked to respond to: Dienes and Seth (2010) reported that wagering was biased by the participants’ risk-aversion, but there were no differences between confidence and wagering after the possibility of loss had been eliminated from wagering. Sandberg, Timmermans, Overgaard, and Cleeremans (2010) observed in a masked object identification task that the perceptual awareness scale (PAS) predicted task performance more efficiently than confidence and wagering did. In an artificial grammar task, it was reported that confidence ratings predicted objective performance more efficiently than ratings of awareness of the artificial grammar rule (Wierzchoń et al., 2012). Szczepanowski, Traczyk, Wierzchoń, and Cleeremans (2013) reported that confidence ratings were more closely correlated with performance than ratings of subjective awareness and wagering, although a recent reanalysis of the data found no significant differences between subjective awareness and confidence (Sandberg, Bibby, and Overgaard, 2013). Finally, subjective reports of visual experience were less strongly correlated with objective performance in masked orientation discrimination tasks or random motion discrimination tasks, but no substantial differences were observed in a masked form discrimination task. In addition, confidence ratings were associated with more liberal thresholds than reports of visual experience across all three visual tasks, and confidence and wagering were more strongly correlated with each other than with reports of visual experience (Zehetleitner & Rausch, 2013).

Four different lines of interpretation for empirical differences between subjective scales with different questions have been suggested: First, it has been assumed (at least for the purpose of a comparison between measurements) that different kinds of subjective reports are equal except the sensitivity (Dienes & Seth, 2010) and the exhaustiveness of the scale (Sandberg et al., 2010). The second suggestion was that different scales might encourage participants to access their conscious contents in different ways: In introspective judgments, participants just directly report their conscious experiences a s they have them; in metacognitive judgments however, participant use their conscious experiences to make more complex cognitive judgments about processes engaged in the objective task (Overgaard & Sandberg, 2012). Third, it has been proposed that different subjective scales might alter the quality of conscious experience itself: Some scales such as wagering might be more motivating for the participants, making them more attentive, and thus cause participants to experience the stimulus more distinctively (Szczepanowski et al., 2013). Finally, it was suggested that different questions may relate to different processes during the task: Stimulus-related reports may be informed by processes involved in stimulus representation, and decision-related reports by processes involved in decision making (Zehetleitner & Rausch, 2013).

1.2. Visual analogue vs. discrete scales

The present study investigated the response format as the second component of subjective scales, specifically whether responses to the same question are more conveniently recorded by a discrete scale or a visual analogue scale (VAS). From the viewpoint of information theory (Shannon, 1948), subjective reports should be collected with a maximum number of scale steps because the maximal amount of information recorded by one report is bounded by number of options provided to the participant. Specifically, as the maximum information is computed as the binary logarithm of the number of options, a binary scale records the information of 1 bit in one trial, 4 scale points 2 bits, 8 scale points 3 bits, etc. The information conveyed by a VAS, where the response is selected along a continuum, would theoretically depend on the number of scale positions differentiated by the equipment (between 2^8 and 2^{16} with custom joysticks), but is in practice limited by the number of positions that participants can differentiate on the continuum, which classical studies estimated to be at least 10 positions (Hake & Garner, 1951).

From the viewpoint of signal detection theory (SDT) (Green & Swets, 1966; Macmillan & Creelman, 2005; Wickens, 2002) however, the use of a high number of scale steps is only feasible if two requirements are met: (i) participants need to be able to maintain a sufficient number of criteria, and (ii) participants' type 2 sensitivity (Galvin, Podd, Drga, and Whitmore, 2003), i. e. their degree of access to their own task performance, should not be impaired by a great number of options. The recent literature has raised doubts about both requirements for high-precision usage of VASs:Overgaard, Rote, Mouridsen, and Ramsøy (2006) proposed that VASs tend to be used like binary judgments: As only the extreme ends of the scale are labelled, reports may be dragged towards the extremes, reducing the number of criteria participants effectively use to two. In addition, they argued as there are no definitions for each experience along the continuum of the VAS, VAS could confuse participants and result in less accurate reports.

Only one study so far has empirically compared a VAS and discrete scale: Wierzchoń et al. (2012) compared subjective reports of rule awareness with four scale steps against a VAS of rule awareness in a 2AFC artificial grammar classification task and observed a tendency that the four-point scale predicted performance more efficiently than the VAS (irrespective of whether the VAS was binned into four scale steps or not), although the statistics were not significant. Wierzchoń et al. (2012) also found that rule awareness measured by a VAS was worse than wagering and feeling-of-warmth both measured by a discrete scale, although there was no significant difference between discrete rule awareness and these two scales; however, these findings are hard to interpret because the content of the scales and the response format are confounded in these comparisons. In domains other than awareness, VASs have been demonstrated to be adequate measurements for state anxiety (Davey, Barret, Butow, and Deeks, 2007), vertigo (Dannenbaum, Chilingaryan, and Fung, 2011), quality of live (de Boer et al., 2004), group cohesiveness (Hornsey, Olsen, Barlow, and Oei, 2012), mood (Kontou, Thomas, and Lincoln, 2012), thermal perception (Leon, Koscheyev, and Stone, 2008), and depression (Rampling et al., 2012), indicated by a strong correlation with an established multi-item questionnaire or by a high reliability of VASs, suggesting that participants are in principle able to make meaningful reports using VASs (although it should be noted that these studies did not compare VASs and discrete scales directly). As VASs were shown to be adequate measurements for a considerable number of different psychological constructs, it is reasonable to hypothesize that a VAS might be a convenient measurement of visual experience as well. Apart from that, it was argued that a VAS may induce more careful responses because it signals to the participant that an exact response is important, while a discrete scale might convey the message that a rough answer is sufficient (Funke & Reips, 2012).

In summary, although VASs are in principle suited to record a large amount of information, it is an open empirical question whether participants are able to use a VAS with a sufficient number of criteria and without loss of type 2 sensitivity, so employing a VAS is feasible.

1.3. Continuous vs. binary discrimination task

While the study by Wierzchoń et al. (2012) contrasted subjective reports and objective performance in a 2AFC discrimination task, the recent development of continuous discrimination tasks (Bays & Husain, 2008; Zhang & Luck, 2008; Zokaei, Gorgoraptis, Bahrami, and Bays, 2011) offers the opportunity to conduct a more powerful test of the amount of information recorded by a VAS. For example, in a typical 2AFC task, participants might be instructed to report whether a previously presented bar is tilted towards left or right. The set of possible stimulus features is two (left or right) and so is the set of possible responses. This paradigm can be changed into a continuous discrimination task by allowing the bar to have any of all possible orientation and asking the participant to indicate the orientation of the bar via a response set of the same cardinality. Errors, defined as the deviation of stimulus and response, are binary in a 2AFC paradigm: either the response corresponds to the stimulus (i.e., is "correct"), or it does not (i.e., is "incorrect"). For continuous tasks however, the deviance between stimulus and response is a continuous variable: When for instance the stimulus consists of a vertical bar, the response may deviate from the true orientation by any angle between 0° and 90°.

The number of task response alternatives is relevant for comparing different scales because the information recorded by a scale depends on the entropy of metacognition, which in turn depends on the entropy of discrimination performance: When there are only two levels of accuracy, i. e. "correct" and "incorrect", there will be a comparably small number of metacognitive states, and consequently, a smaller number of scale steps might perform well to categorize these states. In contrast, when participants are required to adjust a comparison continuously according to a specific stimulus feature, there is a large number of different possibilities how accurate discrimination performance can be, and thus a large number of possible metacognitive states. Consequently, a scale with a larger number of response alternatives might perform better than a discrete scale when the number of response alternatives is large.

In general, performance in a continuous adjustment task can be described mathematically by a combination of a von Mises and a uniform distribution (Bays, Catalao, and Husain, 2009; Zokaei et al., 2011): If participants had to rely completely on guessing, their responses should be evenly distributed across the whole range of possible responses. However, if performance is better than chance, their responses would form a bell-shaped distribution centred at the correct response, with the spread of the distribution indicating the precision of the response. A continuous task for the purpose of the current study would be characterized by a continuous relationship between task difficulty and the precision parameter as well as the guessing parameter. Previous studies suggested that subjective reports are associated with both the precision parameter as well as the probability of guessing in working memory tasks (Rademaker, Tredway, and Tong, 2012), but to our knowledge, no study has so far introduced continuous tasks in the study of visual consciousness.

1.4. Criteria to evaluate subjective scales

As the current experiments entails a comparison between scales with a different number of scale steps, special attention should be paid to the choice of operationally defined criteria to evaluate the scales. We propose to employ three criteria of comparison: (i) the correlation with discrimination performance, (ii) the internal consistency, and (iii) the distribution of ratings.

The correlation with discrimination performance as well as internal consistency come with two very different interpretations depending on whether the amount of information collected with one report is controlled or not. When VAS judgements are binned into the same number of scale steps as the discrete scale and thus the amount of information recorded by the two scales is balanced, the correlation of subjective reports with discrimination performance is indicative of type 2 sensitivity (Galvin et al., 2003), the ability to discriminate between correct and incorrect trials. This is the rationale of numerous previous studies (Dienes & Seth, 2010; Sandberg et al., 2010; Szczepanowski et al., 2013; Wierzchoń et al., 2012) and is analogous to the term resolution in the confidence literature (Baranski & Petrusic, 1994). In contrast, under the assumption that the type 2 sensitivity of participants is comparable, a comparison between the association of the full VAS and objective performance on the one hand and the association between the discrete scale and performance shows whether the VAS is able differentiate between levels of performance that fall equally on the same scale step with the discrete scale and is thus indicative of the amount of information recorded by the scale.

The second criterion we took into account was the internal consistency of subjective reports within experimental conditions: A scale should provide maximally stable estimates of averages of the subjective reports across a number of data points. Again, the comparison between the discretized VAS and a discrete scale shows whether one scale is corrupted from noise unrelated to the number of scale steps; while a comparison between the internal consistency of full VAS and discrete scales shows whether participants can make use of the additional resolution provided by the VAS, i. e. it examines whether VAS reports differentiate between trials that fall on the same scale step at the discrete scale.

Third, another characteristic of subjective scales that has been extensively discussed is the distribution of subjective reports when collected with different scales: Are subjective scales of consciousness used gradually or are they used in a binary fashion? While some scales might be designed in a way that all scale steps are used with relatively equal probability, other scales might induce binary responses (Overgaard et al., 2006). This empirical question is related to the theoretical proposals that consciousness is either dichotomous (Dehaene & Changeux, 2011) or a gradual phenomenon (Cleeremans, 2011). If stimulus consciousness varies binarily (i. e. stimuli are always either conscious or unconscious), an observers would only use the ends of the scale, resulting in a U-shaped distribution of ratings. If stimuli however can be more or less conscious, all points of the scale are potentially used, when stimulus strength increases, resulting in a uniform distribution when averaged across stimulus strength. However, in order to investigate the issue whether consciousness varies gradually or binarily, a scale is required where participants in principle use the intermediate scale steps as well; otherwise a U-shaped distribution would be observed no matter whether consciousness in a specific task in fact gradual or dichotomous (Sergent & Dehaene, 2004).

1.5. Rationale of the present study

The aim of the present study was to investigate whether participants can make use of the high resolution offered by VASs when measuring visual experience of motion. To address this issue, we compared a VAS and a discrete scale with respect to the criteria discussed in 1.4. As stimuli, we presented random dot kinematograms (RDKs), because RDKs allow for a fine-grained manipulation of task difficulty on a metric scale (by manipulating the percentage of coherently moving dots). For the objective task, we assessed objective performance as a continuous variable rather than just correct or false, a procedure that ensured a binary use of subjective reports was not due to binary task performance. To obtain a continuous measurement of task performance, we asked participants to report the orientation of motion by adjusting a clock-hand to point into the direction of the perceived motion, and measured the discrimination error as the angle between clock-handle and direction of motion. For the subjective scales, we asked participants always to report their degree of experience of the coherent motion, which was the same instruction as we used in a previous study (Zehetleitner & Rausch, 2013), and different from the established Perceptual Awareness Scale (PAS, Ramsøy & Overgaard, 2004) in that no instruction to report feelings of something being shown was given.

The experiment was designed to investigate the following three hypotheses:

i. If the participants are able to make use of the additional resolution provided by VASs, the full VAS should predict the discrimination error more efficiently than the discrete scale. In addition, the internal consistency of the full VAS should be better, because the larger amount of data transmitted by each single subjective report would allow for more reproducible statistics based on the same number of trials.
ii. If VAS reduced the type 2 sensitivity of subjective reports, we would expect that the discrete scale would be more efficient in predicting discrimination error and would produce more consistent estimates than the discretized VAS.
iii. If participants are biased by the anchors of the VAS in a way that reports are given binarily, the ratings on the VAS but not on the discrete scale should form a U-shaped distribution. In addition, the discrete scale should outperform both the full and the discretized VAS in predicting discrimination error.

2. Material and Methods

2.1. Participants.

20 participants (5 male, 1 left-handed) took part in the experiment. The age of the participants ranged between 19 and 32 years, with a median age of 24. All participants reported to have normal or corrected-to-normal vision, confirmed that that they did not suffer from epilepsy or seizures and gave written-informed consent

2.2. Apparatus and stimuli.

The experiment was performed with a Mac with OS X 10.7 as operating system and a Diamond Pro 2070 SB (Mitsubishi) monitor with 24 inch screen size. Stimuli were presented at a refresh rate of 120 Hz controlled by MATLAB and Psychtoolbox 3.0.10 (Brainard, 1997;

Pelli, 1997; code adapted from http://www.shadlenlab.columbia.edu/Code/VCRDM). The stimuli were random dot kinematograms, consisting of on average 150 small white squares (sized 2 x 2 pixels, luminance 85.0 cd/m^2) in from of a black background (1.3 cd/m^2), which appeared in a circular aperture (diameter: 5°) centred at the fixation. A set of dots was shown for one video frame and then replotted three video frames later. When replotted, a subset of dots was offset from their original location to create apparent motion while the remaining dots were relocated randomly. The proportion of coherently moving dots was randomly chosen among 1.6, 3.1, 6.2, 12.5, 25, and 50 %. The direction of movement was randomly chosen out of each possible direction. To record the orientation judgment, 12 circles (diameter: 0.2°, 2.2 cd/m^2) were displayed on the screen, forming one large circle centred at the screen with a diameter of 10°. Participants indicated the direction of motion and their rating on the VAS by a Cyborg V1 joystick (Cyborg Gaming, UK). The clock-hand consisted of a bar (length: 5°, width: 0.1°, 2.2 cd/m^2) and a circular head (diameter: 0.2°, 2.2 cd/m^2).

2.3. Trial structure.

The trial structure is shown in Fig. 1. Each trial began with the presentation of a fixation cross at screen centre for 1,000 ms. Then a RDK was presented for 2,000 ms. Next, the circle around the screen centre appeared. As the participants started to move the joystick, the clock-hand appeared, pointing to the direction the joystick was moved to. The circle continued to be displayed on the screen until participants had pulled and released the trigger of joystick. Next, the subjective scale appeared, with either the four response categories from the discrete scale or the VAS. If the error of the orientation judgment had been larger than 45°, the trial ended with the display of “please indicate the direction more carefully” for 1,000 ms.

2.4. Procedure.

Experiment lasted 1 hour on average. Participants were instructed perform the motion discrimination task as carefully as possible, with accuracy being more important than speed. For the verbal reports, participants were told that the subjective scale referred to the global motion experience created by the coherently moving dots. Again, participants were instructed to their ratings as carefully and as accurately as possible.

Participants indicated the direction of motion by using the joystick to move a bar that looked like a clock-hand. When the participants had moved the clock-hand in the direction they saw the dots moving, they confirmed their response by pulling the trigger of the joystick. The clock-hand consisted of a bar (length: 5°, width: 0.1°) and a circular head (diameter: 0.2°). To collect the subjective report, the question “how clearly did you see the coherent motion?” was displayed on the screen. In case of VAS, a continuous scale was shown underneath the question, with the ends labelled as “not at all”, and “clear”. Participants moved an index on the continuous scale by moving the joystick horizontally, and confirmed a position on the scale by pulling the trigger. In case of four point scales, the same question was displayed on the screen, but underneath the question, four response categories were shown, which were “not at all”, “weak”, “almost clear”, and “clear”. Participants responded to the discrete scales by pressing the keys 1, 2, 3, and 4 on the keyboard. At the beginning of the experiment, participants performed a training block with 24 trials. The main experiment involved 10 blocks with 45 trials each. During training, VAS and discrete scale trials were randomly intermixed each of the six possible coherences was presented six times in a from-easy-to-difficult order. During the main experiment, the two subjective scales alternated after each block and the levels of coherence varied randomly between trials.

2.5. Analysis

All analysis were performed in R 2.15.2 (R Core Team, 2012). For both the distribution analysis as well as the regression analysis, fast responses (defined as faster than 200 ms) and slow responses (defined as 2.5 standard deviations slower than the individual average) to the discrimination task or to the scale were omitted. Other exclusion criteria such as 2 or 3 standard deviations gave essentially the same results.

2.5.1. Distribution analysis of the discrimination responses

Discrimination responses were analyzed by fitting a combination of a von Mises distribution and a uniform distribution to the data (Bays et al., 2009; Zokaei et al., 2011). The uniform distribution models the distribution of responses in trials when participants relied on guessing, because when participants guessed, each orientation between 0 and 360 ° was equally probable. The von Mises (circular Gaussian) distribution centred at the true motion direction represents the distribution of responses in trials where participants were not guessing. The better participants performed the orientation judgment, the less responses jittered around the true motion direction; therefore, the concentration parameter of the von Mises distribution can be interpreted as the precision of orientation judgments. The model is described by the following equation:

$$\hat{\theta} = (1-\gamma)\phi_K(\hat{\theta}-\theta) + \gamma\frac{1}{2\pi}$$

where θ is the stimulus motion direction, $\hat{\theta}$ is the motion direction indicated by the participant, γ is the proportion of trials when participants were guessing, ϕ denotes the von Mises distribution with mean of zero and the concentration parameter K. Fitting was performed on the aggregated data across all participants and scales but separately for each level of coherence using maximum likelihood estimation and confidence intervals around each parameter were estimated using 10,000 bootstrap samples. Pooling over participants and scales was necessary to obtain a sufficient number of trials for the fitting algorithms to reach convergence. The purpose of this analysis was a manipulation check if performance in the current task was continuous or binary. As the hypotheses tested in the current study equally apply to metacognition of the precision as well as the guessing aspect of performance, it was legitimate to analyze the relationship between subjective reports and performance without differentiating between guessing and precision (see 2.5.2.)

2.5.2. Relationship between scales and discrimination error

The relationship between the two scales and discrimination error was analyzed by means of mixed model regression analysis based on the cumulative proportional odds model as implemented in the R library ordinal (Christensen, 2013), the ordinal equivalent to the analysis in previous studies (Sandberg et al., 2013; Sandberg et al., 2010; Wierzchoń et al., 2012). The dependent variable, the discrimination error, was determined by the absolute difference between the true motion direction and the reported motion and binned into 12 equal bins between 0 and 90° and a thirteenth bins for errors larger than 90° to allow computation of a proportional odds model. Non-parametric statistics were used to account for the fact that the discrimination error was bounded and strongly skewed. Inter-subject variance was modelled by a random effect on the intercept. Scale (VAS vs. discrete scale), coherence (1.6 vs. 3.1 vs. 6.2 vs. 12.5 vs. 25.0 vs. 50.0) and subjective report and all interactions were treated as fixed effects. Significance of each fixed term was assessed by likelihood ratio tests between the full model and a model where the term was dropped. Confidence intervals were obtained from the likelihood root statistic. Subjective reports given by VAS and discrete scales were standardized separately. To investigate the effects of number of scale steps, two separate models were computed, one with the full VAS included as predictor, and one model where the VAS was binned into four equal partitions. We interpret a comparison between the discretized VAS and the discrete scale as indicative of type 2 sensitivity (i. e. the degree to which participants can access to their own performance) because when the VAS reports are binned to four, the amount of information in discretized VAS and discrete scale are the same, although we acknowledge that ordinal statistics do not provide any means of control over the influence of discrimination bias (Masson & Rotello, 2009). Given that type 2 sensitivity of VAS and discrete scale are the same, a comparison between the full VAS and the discrete scale is indicative of whether participants apply more criteria in the VAS than in the discrete scale and thus the full VAS discriminates between levels of performance that fall on the same scale step with the discrete scale. In addition, we analyzed the effects of feedback and reporting time by computing two additional models comparing full VAS and the discretized scale with feedback and report time as additional fixed effect, respectively.

2.5.3. Internal consistency

Internal consistency was assessed by computing Cronbach's alpha (Cronbach, 1951) separately for each level of coherence using the R library ltm (Rizopoulos, 2006). Confidence intervals were estimated around Cronbach's alpha values based on 10,000 Bootstrap samples.

2.5.4. Distribution of subjective reports

To analyze the distribution of subjective reports, the ratings of VAS was again binned into four categories each covering a fourth of the scale range. The frequency of each bin was then compared against frequency of the corresponding response alternative of the discrete scales using an ANOVA with the factors rating category, coherence, and scale type (VAS vs. discrete scale). When sphericity did not hold, we adjusted the degrees of freedom according to the Greenhouse-Geisser correction. To resolve interactions, post-hoc t-tests were conducted comparing the frequency of each VAS bin with the corresponding response category of the discrete scale separately for each level of coherence. P-values were adjusted by the Holm-correction to account for multiple comparisons.

3. Results

3.1. Discrimination performance

The mean discrimination error was 55.6° (SEM = 2.2) when participants were using the VAS and 56.3° (SEM = 2.2) when the discrete scale was used and ranged from 87.7° (SEM = 1.7) for the lowest to 13.7° (SEM = 1.6) for the highest level of coherence. The relative frequencies of orientation responses and the estimated distributions are shown in Fig. 2. The estimated parameters as well as bootstrapped confidence intervals are shown in Fig. 3. The probability of guessing trials ranged between .94 at the lowest and .05 at the highest level of coherence. Confidence intervals indicated the guessing probability continuously decreased across all levels of coherence. The precision ranged between 5.2 at a Coherence of 3.1 % and 28.5 at the maximum level of coherence. Confidence intervals suggested that there was a continuous increase of precision starting at a coherence of 6.2, while the estimation of the precision parameter was not reliable for coherence levels of 1.3 % and 2.6 % (due to the low number of non-guessing trials).

3.2. Relationship between discrimination error and subjective reports

The regression weights and confidence intervals of the ordinal mixed model regression comparing the full VAS against the discrete scale as predictors of discrimination error can be found in Table 1. Likelihood ratio tests suggested significant main effects of subjective report [$\chi^2(1) = 195.0$, $p < .001$] and coherence [$\chi^2(5) = 1522.0$, $p < .001$], no effect of scale [$\chi^2(1) = 2.1$, n. s.], significant interactions between subjective reports and scale [$\chi^2(1) = 4.3$, $p < .05$] and between subjective reports and coherence [$\chi^2(5) = 50.2$, $p < .001$], and no three-way interaction [$\chi^2(5) = 6.8$, n. s.]. A regression model fitted on VAS ratings only revealed a regression coefficient for subjective reports of -.44 with a 95 % confidence interval of [-.52 -.36]. For the discrete scale, the same analysis revealed a coefficient of -.32 within a confidence interval of [-.40 -.24].

Discrimination error as a function of coherence, scale, and subjective report (with discretized VAS ratings) are depicted in Fig. 4. The ordinal regression model comparing the discretized VAS and the discrete scale revealed significant main effects subjective report [$\chi^2(1) = 178.6$, $p < .001$] and coherence [$\chi^2(5) = 1586.2$, $p < .001$], no effect of scale [$\chi^2(1) = 2.4$, n. s.], a significant interaction between subjective reports and coherence [$\chi^2(5) = 47.4$, $p < .05$], but no interaction between subjective report and scale [$\chi^2(1) = 1.5$, n. s.], and no three-way interaction [$\chi^2(5) = 6.8$, n. s.].

The frequency of feedback, which was provided after discrimination responses with an error greater than 45°, did not substantially differ between VAS trials (M = 40.3, SEM = 1.7) and discrete scale trials (M = 41.0, SEM = 2.1) [$t(19) = .7$, n. s.]. Including feedback on the previous trial into the ordinal regression analysis as an additional predictor revealed no effect of feedback [$\chi^2(1) = 0.1$, n. s.], no interaction between subjective reports and feedback, [$\chi^2(5) = 0.8$, n. s.], between scale and feedback [$\chi^2(1) = 2.1$, n. s.], or between scale, subjective reports, and feedback [$\chi^2(1) = 2.1$, n. s.]. Importantly, the interaction between scale and subjective report was still significant when feedback was included into the analysis [$\chi^2(1) = 3.9$, $p < .05$].

For the VAS, the mean report time, i. e. the time between the orientation judgment and the subjective report, was 1329 ms (SEM = 95.6), compared to 944 ms (SEM = 73.3) with the discrete scale. As can be seen from Figure 5, ordinal regression slopes increased with report time for the VAS, while no such a relation was apparent for the discrete scale. The regression model with report time as additional predictor revealed a significant main effect of report time [$\chi^2(1) = 4.0$, $p < .05$], no interaction between report time and scale [$\chi^2(1) = 0.1$, n. s.], and between subjective report and time [$\chi^2(1) = 1.1$, n. s]. There was however a three-way interaction between subjective reports, scale, and report time [$\chi^2(1) = 5.5$, $p < .05$]. When response times were included into the model, the interaction between subjective reports and scale was no longer significant, [χ^2 (1) = 2.7, n. s]. Separate analyses of the impact of the report time on discrete scales and VAS revealed that the predictive efficiency of subjective reports made with the VAS interacted with rating time [$\chi^2(1) = 6.1$, $p < .05$], while subjective reports on the discrete scale were not influenced by rating time [$\chi^2(1) = 0.4$, n. s]. Overall, this pattern indicates that the differences in predictive power for discrimination error between the VAS and the discrete scale are mediated by longer report times.

3.3. Internal Consistency of subjective reports

Cronbach's alphas ranged between .83 and .93 for the discrete scale, between .84 and .93 for the discretized VAS, and .85 and .93 for full VAS (see Table 3). There was a numeric trend that alphas were larger for both the full and the discretized VAS than for the discrete scale at four out of six levels of coherence, but confidence intervals indicated the only substantial difference between the two scales was at a coherence of 6.2 %, where the internal consistency of the VAS was greater. The internal consistency of the discretized VAS was always within the confidence intervals around the full VAS.

3.4. Distribution of subjective reports

The mean subjective experience reported on the VAS was 49.2 % of the scale range (SEM = 2.2) and 2.3 (SEM = 0.1) on the discrete scale ranging between 1 and 4 (which corresponds to a mean of 41.3 % of the scale range and a standard error of 2.2 %). As can be seen from Fig. 5, the second scale step of the discrete scale was the dominant response even at a coherence of 1.3 % when performance was effectively at chance. The ANOVA on response frequencies revealed a significant main effect of rating category [$F(2.0,38.8) = 5.3$, $p < .001$], significant interactions between scale and rating category [$F(3,57) = 17.4$, $p < .001$], and between rating category and coherence [$F(3.6,68.0) = 53.4$, $p < .001$], as well as a three-way interaction between coherence, scale type, and rating category [$F(5.3,99.9) = 7.2$, $p < .001$]. Post-hoc tests assessing whether the frequency of responses was different between the discrete scale and the VAS separately for each level of coherence and each response category are shown in Table 2. While there was no significant difference between reports of no experience on the discrete scale and the corresponding scale part of the VAS at each coherence, reports of weak experiences occurred more often with the discrete scale than with the VAS at 5 out of 6 coherences, reports of almost clear experiences were more frequently reported with the VAS at lower coherences, and reports of clear experiences were more often with the VAS at a coherence of 25 %.

4. Discussion

The present experiment investigated whether participants are able to use the high number of response alternatives provided by visual analogue scales appropriately when reporting visual experience of motion. We hypothesized that if a VAS allowed to retrieve a larger amount of information from participants' reports than discrete scales, the full VAS should be more efficient in predicting the discrimination error, and should be more internally consistent. Second, if a VAS reduced the type 2 sensitivity of subjective reports, we would expect that the discretized VAS should be less efficient in predicting the discrimination error than the discrete scale. Finally, if participants tended to use VASs in a binary way, ratings on the VAS should form a U-shaped distribution, and the discrete scale should correlate more closely with discrimination error no matter whether the VAS is discretized or not.

Concerning the relationship between subjective reports and discrimination error, the full VAS predicted discrimination error more efficiently than the discrete scale, while there were no substantial differences between the discretized VAS and the discrete scale. The difference between the full VAS and the discrete scale was mediated by the response time to the scale. The analysis of internal consistency revealed no substantial differences for five out of six coherences, while both the full and the discretized VAS were more consistent at a coherence of 6.2 %. Concerning the distribution of subjective reports, we observed that the VAS and the four-point scale were both not used in an all-or-nothing fashion, although participants had a tendency to report weak experiences in the discrete scale while they would report almost clear and clear experiences in the VAS.

4.1. The amount of information in VAS and discrete scales

According to a standard interpretation of differences between scales measuring subjective awareness, subjective reports are created by the same mechanisms, and differences between scales occur due to different qualities of the scale. A key aspect of the quality of the scale is the amount of information transmitted by each rating. According to information theory (Shannon, 1948), subjective reports collected by VAS should provide a larger amount of information that discrete scales, because 4 scale steps allow to record 2 bits of information, while the number of bits collected by a continuous scale is limited only by the number of positions participants are able to differentiate, and was estimated to be at least 10 positions (Hake & Garner, 1951), i.e. at least 3.32 bits. Consistent with the predictions from information theory, the full VAS was more closely correlated to the discrimination error than the discrete scale. We did not detect any substantial differences between the discrete scale and the discretized VAS in terms of type 2 sensitivity, suggesting that the additional alternatives participants have to consider when using a VAS did not add substantial amounts of noise to the subjective reports. Concerning internal consistency, there were no substantial differences between the two scales in five out of six coherences, although the VAS was more reliable at a coherence of 6.2 %. Overall, it seems that a VAS indeed provides a larger amount of information than discrete scales, although the amount of information recorded by discrete scales is sufficient to provide reliable estimates as well.

4.2. The impact of report time

The difference between reports on the full VAS and on discrete scale in predicting discrimination error was mediated by the time of rating: While VAS ratings became more and more efficient in predicting trial accuracy with time, we observed no such a relation for the discrete scale. The first interpretation to these results is that a VAS provides a larger number of response alternatives, and selecting one out of this multitude of options could be more difficult and thus require a longer period of time. Second, it should be noted that also the motor response required by a VAS is more time-consuming than a simple button press: The association of the rating-accuracy relationship and report time at the VAS could also reflect the additional time demand of using a joystick and a decrease of rating precision when participants did not invest enough time to operate the joystick carefully. Third, an alternative explanation to these findings may be based on the dynamics of decision making: While standard SDT models of subjective reports assume that the evidence used for subjective reports is fixed at the time when observers respond to the task (Kepecs, Uchida, Zariwala., and Mainen, 2008; Ko & Lau, 2012; Vickers, 1979), others have proposed that subjective reports are based on evidence participants continue accumulating after the objective decision is made (Pleskac & Busemeyer, 2010). Given that VAS judgements were associated with prolonged time participants needed to give a subjective report, post-decisional accumulation of evidence might be an alternative explanation why ordinal regression slopes are higher with VAS than with a discrete scale, because the additional 400 ms that it takes to make a judgement on the VAS might give participants more time to accumulate evidence. However, we observed a large overlap in the report times between the two scales in the current experiment where VAS regression slopes were larger although the time of the report was the same. In addition, while ordinal regression slopes seemed to increase almost linearly for the VAS, we found no indication of post-decisional accumulation for the discrete scale at all. What is possible is that participants keep accumulating sensory evidence after the decision when using the VAS only, either because they need the additional evidence to make fine-grained VAS ratings, or because they might be more motivated when using a VAS (Funke & Reips, 2012). The (cognitive and motor) cost of precise reporting and on-going accumulation accounts cannot be distinguished on grounds of the current data set. Given that a previous study failed to find any association with report time for both the VAS and discrete scales (Wierzchoń et al., 2012), future studies may be necessary to investigate the dynamics of metacognition.

4.3.Are visual analogue scales used binarily?

VAS received criticism because the continuum in combination with the labelled scale ends might result in a bimodal distribution of subjective reports, with scale extremes being chosen more frequently than the centre of the scale (Overgaard et al., 2006). First, we observed that intermediate scale steps were chosen frequently for both scales. Second, there was no difference between the frequency of the smallest scale step of the discrete scale and the lowest quarter of the VAS, indicating that VAS and discrete scales both applied the same minimal criteria for subjective reports. However, the second smallest scale step of the discrete scale (labelled as “weak”) was more often chosen than the corresponding part of the VAS, while stronger experiences were more frequently reported with the VAS than with the discrete scales. There might be several explanations why more distinct experiences are more frequently reported with the VAS: First, participants could be biased by the labelled extremes of the scale (Ramsøy & Overgaard, 2004). Second, participants might suffer more strongly from an error of central tendency when they respond to discrete scales, and therefore the second scale step was the dominant response in the discrete scale. Finally, it is also possible participants are more motivated when using the VAS, being more attentive, and therefore have in fact clearer experiences (Szczepanowski et al., 2013). Concerning the impact of motivation, the two scales were associated with a comparable discrimination error, suggesting that the scale did not alter the way participants performed the task in general. Concerning a potential bias towards extremes, it should be noted that intermediate positions on the VAS were the most frequent responses for medium levels of coherence, suggesting that participants do use the centre of the scale when they consider it to be appropriate. In contrast, the second scale step of the discrete scale was the dominant response even at the lowest level of coherence when discrimination accuracy was effectively at chance, suggesting that the error of central tendency might be a factor in the distribution of discrete scales. The distribution of VAS is more plausible in a way that low ratings are dominant at low levels of coherence, intermediate ratings at medium coherences, and high ratings at high levels of coherences.

4.4. Discussion of methodology

It should be noted that the current experiment differs from previous studies addressing the topic of subjective reports in several ways. As this task is new to the field of metacognition, future studies are desirable to explore whether the findings obtained with this method are corroborated in more standard experiments. Most importantly, we quantified discrimination error as a continuous variable rather than binary in the current study. In general, such an approach seems promising for the field of consciousness research because some theories of consciousness make specific predictions whether consciousness is gradual (Cleeremans, 2011) or dichotomous (Dehaene & Changeux, 2011), and recording performance in a non-binary way ensures that binary task performance does not artificially cause binary metacognition. Unfortunately, up to know, there is no proposal for a SDT-grounded measure of type 2 sensitivity equivalent to the measures applicable for binary tasks, so our analysis of type 2 sensitivity by ordinal regression does not provide the same control of response bias and confidence thresholds than it is possible for binary tasks. For the purpose of the current study, these potential confounds do not change the interpretation of the data because they would either affect the discrete scale, the full VAS, and the discretized VAS in the same way (response bias), or would affect the full VAS and the discretized VAS to the same degree (confidence thresholds), so it cannot be explained why only the full but not the discretized VAS provides more predictive power. Future studies however need to carefully consider the conceptual advantages of continuous tasks against the methodological disadvantages of the analysis methods available.

It may also be objected that the current task was not as continuous as it could have been, since all responses at above chance performance were concentrated between 45 and -45 degrees (where no error feedback was given), and thus the feedback might have motivated participants to perform at least as accurate as +/- 45 degrees. However, the precision of orientation judgements increased almost linearly although participants no longer received feedback, indicating participants did not perform the task in a binary fashion. As feedback might also have altered performance and type 2 sensitivity in the current task, parameters and coefficients estimated from the current experiment should not be naïvely expected to be the same in standard subliminal perception tasks where error feedback is suspended after a training period or is completely missing. Nevertheless, we did not observe any evidence that feedback on the previous trial influenced any contrast of interest for purpose of the current experiment, suggesting that feedback did not have a major impact on performance in the current study.

4.5. Equivalent conscious access?

Another interpretation of differences between various subjective scales is that different scales might encourage participants to use different mechanisms of conscious access to report their conscious experiences (Overgaard & Sandberg, 2012). Indeed, it is plausible to assume that subjective reports in VASs and discrete scales are accomplished in parts by different processes. Discrete scales rely strongly on verbal categorization, because observers need to have a concept of each of the scale steps, while VAS need only an abstract understanding of the dimension as a whole. In contrast, VAS may depend on visuo-motor coordination, because participants need to translate their experience into spatial coordinates and have to move the joystick accordingly. This might be an explanation for the effects in the current study, although a previous study reported that five scale points cannot convey more information about subjective experiences than four scale points (Ramsøy & Overgaard, 2004). The number of scale steps participants can make use of in labelled scales depends on the participants' ability to categorize their percepts' verbally, which might be limited to four. VASs do not depend to the same degree on verbal categorization; therefore, the amount of information transmitted by a VAS can be greater.

4.6. Conceptual reasons to prefer either VASs or discrete scales over the other

Finally, deciding between VASs and discrete scales is not a question that can be addressed entirely by empirical methods, but needs to be informed conceptually as well. First, a VAS is only feasible if the subjective reports can be given along one dimension. However, the study of visual awareness may require the assessment of several qualitative different patterns of subjective experience: For instance, it has been suggested that observers report "feelings that something has been shown" or "experiences without any content" (Ramsøy & Overgaard, 2004) or even to be confident about the discrimination judgement (Zehetleitner & Rausch, 2013) at low levels of stimulation, and report that they had an experience of a specific stimulus quality only at higher levels of stimulation. These discontinuities in the pattern of subjective reports along the unaware/aware continuum cannot be measured by one single VAS, so other measures are required if the full set of experiences during visual perception is of theoretical interest to a specific study. For example, an established measure that captures qualitatively different experiences is the Perceptual Awareness Scale (Ramsøy & Overgaard, 2004), where participants are asked to differentiate between the absence of an experience, experiences without any content, almost clear experiences of a specific stimulus feature, and full clarity of the specific stimulus feature. Alternatively, different dimensions can be assessed by combining two VASs with different content in one trial (Zehetleitner & Rausch, 2013).

Second, some theorists strongly focus the connection between consciousness and language (Vygotsky, 1962), and such a view might imply verbally categorized scale steps to be more valid than a continuous scale. However, other concepts of consciousness endorse a view where perceptual consciousness is not easily verbalized, and such a view may prefer VASs as they rely less heavily on verbal categorization.

5. Conclusion

We present data that both visual analogue scales as well as discrete scales are reliable measures of subjective reports of global motion experience. We found no evidence that the type 2 sensitivity is decreased or the pattern of reports is binary when participants are provided with a large number of scale steps. The data is consistent with the interpretation that participants are able to maintain a sufficient large number of meaningful criteria so that a VAS retrieves a larger amount of information than a discrete scale with four scale steps, provided that participants take their time to make the more subtle judgements. At least when the number of response alternatives of the objective discrimination task is large, subjective reports of motion experience may be recorded more conveniently by a VAS than by a discrete scale with the same content.

Authors Note

This research was supported by the German-Israeli Foundation for Scientific Research and Development (GIF) grant 1130-158 and the Deutsche Forschungsgesellschaft (DFG, i. e. German Research Council) grant ZE 887/3-1 (both to M. Z.). The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript. Correspondence concerning this article can be addressed to Manuel Rausch or to Michael Zehetleitner.

References

Baranski, J. E., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics, 55*, 412-428.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*, 1-11. doi: 10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851-854. doi: 10.1126/science.1158023

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences, 9*, 46-52. doi: 10.1016/j.tics.2004.12.006

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, , 433-436.

Chalmers, D. (1998). On the search of neural correlates of consciousness. In S. Hameroff, A. Kaszniak & A. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*. Cambridge, MA: MIT Press.

Christensen, R. B. (2013). Analysis of ordinal data with cumulative link models - estimation with the R-package ordinal. http://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology, 2*, 1-12. doi: 10.3389/fpsyg.2011.00086

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrica, 16*(297-334).

Dannenbaum, E., Chilingaryan, G., & Fung, J. (2011). Visual vertigo analogue scale: An assessment questionnaire for visual vertigo. *Journal of Vestibular Research, 21*, 153-159. doi: 10.3233/VES20110412

Davey, H. M., Barret, A. L., Butow, P. N., & Deeks, J. J. (2007). A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *Journal of Clinical Epidemiology, 60*(356-360). doi: 10.1016/j.jclinepi.2006.07.015

de Boer, A. G. E. M., van Lanschot, J. J. M., Stalmeier, P. F. M., van Sandick, J. W., Hulscher, J. B. F., de Haes, J. C. J. M., & Sprangers, M. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Quality of Life Research, 13*, 311-332.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*, 200-227. doi: 10.1016/j.neuron.2011.03.018

Dennett, D. C. (2003). Who's on First? Heterophenomenology explained. *Journal of Consciousness Studies, 10*, 19-30.

Dennett, D. C. (2007). Heterophenomenology reconsidered. *Phenomenology and Cognitive Science, 6*, 247-270. doi: 10.1007/s11097-006-9044-9

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies*, *11*(9), 25-45.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*, 338-351. doi: 10.1007/s00426-004-0208-3

Dienes, Z., & Seth, A. K. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition, 19*, 674-681. doi: 10.1016/j.concog.2009.09.009

Funke, F., & Reips, U.-D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-Point Scales. *Field Methods, 24*, 310-327. doi: 10.1177/1525822X12444061

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10*, 843-876.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hake, H. W., & Garner, W. R. (1951). The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology, 42*, 358-366.

Hannula, D. E., Simons, D. J., & Cohen, N. J. (2005). Imaging implicit perception: Promise and pitfalls. *Nature Reviews Neuroscience, 6*, 247-255.

Hornsey, M. J., Olsen, S., Barlow, F. K., & Oei, T. P. S. (2012). Testing a single-item visual analog scale as a proxy for cohesiveness in group psychotherapy. *Group Dynamics: Theory, Research, and Practice, 16*, 80-90.

Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quaterly, 32*, 127-136.

Kepecs, A., Uchida, N., Zariwala., H., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227-231. doi: 0.1038/nature07200

Ko, Y., & Lau, H. (2012). A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philosophical Transactions of the Royal Society B, 367*, 1401-1411. doi: 10.1098/rstb.2011.0380

Kontou, E., Thomas, S., & Lincoln, N. B. (2012). Psychometric properties of revised version of the Visual nalog Mood Scales. *Clinical Rehabilitation, 26*, 1133-1140. doi: 10.1177/0269215512442670

Lau, H. (2008). Are we studying consciousness yet? In L. Weißkrantz & M. Davies (Eds.), *Frontiers of Consciousness: Chichele Lectures*. Oxford: Oxford University Press.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*, 8-16. doi: 10.1016/j.tics.2011.05.009

Leon, G. R., Koscheyev, V. S., & Stone, E. A. (2008). Visual analog scales for assessment of thermal perception in different environments. *Aviation, Space, and Environmental Medicine*, *79*, 784-786.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory. A user's guide*. Mahwah, NY: Lawrence Erlbaum Associates.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacogntive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509-527. doi: 10.1037/a0014876

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*, 435-450.

Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. Z. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Consciousness and Cognition, 15*, 700-708. doi: 10.1016/j.concog.2006.04.002

Overgaard, M., & Sandberg, K. (2012). Kinds of access: Different methods for report reveal different kinds of metacognitive access. *Philosophical Transactions of the Royal Society B, 367*, 1287-1296. doi: 10.1098/rstb.2011.0425

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *Memoirs of the National Academy of Sciences, 3*, 73-83.

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decisional wagering objectively measures awareness. *Nature Neuroscience, 10*, 257-261. doi: 10.1038/nn1840

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*, 864–901. doi: 10.1037/a0019737

R Core Team. (2012). R: A language and environment for statistical computing. (Version 2.15.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision, 12*, 1–13. doi: 10.1167/12.13.21

Rampling, J., Mitchel, A. J., von Oertzen, T., Docker, J., Jackson, J., Cock, H., & Agrawal, N. (2012). Screening for depression in epilepsy clinics. A comparison of conventional and visual-analog methods. *Epilepsia, 53*, 1713-1721. doi: 10.1111/j.1528-1167.2012.03571.x

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*, 1-23.

Rizopoulos, D. (2006). ltm: An R-package for latent variable modeling and item response theory analysis. *Journal of Statistical Software, 17*(5), 1-25.

Sandberg, K., Bibby, M. B., & Overgaard, M. (2013). Measuring and testing awareness of emotional face expressions. *Consciousness and Cognition, 22,* 806-809. doi: 10.1016/j.concog.2012.12.003

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness- Is one measure better than the other? *Consciousness and Cognition, 19,* 1069-1078. doi: 10.1016/j.concog.2009.12.013

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science, 15,* 720-728.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27,* 379-423.

Szczepanowski, R., Traczyk, J., Wierzchoń, M., & Cleeremans, A. (2013). The perception of visual emotion: Comparing different measures of awareness. *Consciousness and Cognition, 22,* 212-220. doi: 10.1016/j.concog.2012.12.003

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.

Wierzchoń, M., Asanowicz, D., Paulewicz, B., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness and Cognition, 21*, 1141–1153. doi: 10.1016/j.concog.2012.05.012

Zehetleitner, M., & Rausch, M. (2013). Being confident without seeing: What subjective measures of consciousness are about. *Attention, Perception, & Psychophysics, 75,* 1406-1426. doi: 10.3758/s13414-013-0505-2

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 233-236. doi: 10.1038/nature06860

Zokaei, N., Gorgoraptis, N., Bahrami, B., & Bays, P. M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision, 11,* 1–18. doi: 10.1167/11.14.2

Table 1

Results of the mixed-effects ordinal regression model with discrimination error as dependent variable

Predictor	**β**	**95 % CI**		**Likelihood ratio**	**df**	**p**
		lower	**upper**			
Subjective report	-0.38	-0.43	-0.32	195.0	1	< .001
Coherence level				1522.0	1	< .001
– 1.6 % vs. 50 %	1.28	1.17	1.40			
– 3.1 % vs. 50 %	1.06	0.95	1.18			
– 6.2 % vs. 50 %	0.68	0.58	0.79			
– 12.5 % vs. 50 %	-0.39	-0.48	-0.29			
– 25 % vs. 50 %	-1.16	-1.27	-1.05			
Scale type	-0.01	-0.06	0.04	2.1	1	n. s.
Subjective report * coherence level				50.2	5	< .001
– Subjective report * 1.6 % vs. 50 %	0.26	0.14	0.38			
– Subjective report * 3.1 % vs. 50 %	0.21	0.09	0.33			
– Subjective report * 6.2 % vs. 50 %	0.01	-0.10	0.13			
– Subjective report * 12.5 % vs. 50 %	-0.26	-0.37	-0.16			
– Subjective report * 25 % vs. 50 %	-0.15	-0.26	-0.04			
Subjective report * scale type	0.06	0.01	0.11	4.3	1	< .05
Coherence level * scale type				10.3	5	n. s.
– 1.6 % vs. 50 % * scale type	0.10	-0.02	0.21			
– 3.1 % vs. 50 % * scale type	0.06	-0.06	0.17			
– 6.2 % vs. 50 % * scale type	0.10	-0.01	0.20			
– 12.5 % vs. 50 % * scale type	-0.05	-0.15	0.04			
– 25 % vs. 50 % * scale type*	-0.03	-0.14	0.07			
Subjective report * Coherence level * scale type	-0.06			6.8	5	n. s.
– Subjective report * 1.6 % vs. 50 % * scale type	-0.02	-0.14	0.10			
– Subjective report * 3.1 % vs. 50 % * scale type	0.01	-0.11	0.12			
– Subjective report * 6.2 % vs. 50 % * scale type	-0.09	-0.21	0.02			
– Subjective report * 12.5 % vs. 50 % * scale type	-0.06	-0.17	0.04			
– Subjective report * 25 % vs. 50 % * scale type	0.08	-0.03	0.19			

Table 2

Post hoc t-tests comparing the frequency of each scale step of the discrete scale with the frequency of the corresponding part of the VAS

	first bin vs. no experience		second bin vs. weak experience		third bin vs. almost clear		fourth bin vs. clear experience	
Coherence	t^a	p_{cor}^b	t^a	p_{cor}^b	t^a	p_{cor}^b	t^a	p_{cor}^b
1.6	2.9	n. s.	-6.1	< .001	3.7	< .05	0.5	n. s.
3.1	2.3	n. s.	-5.7	< .001	3.8	< .05	1.1	n. s.
6.2	1.1	n. s.	-5.0	< .01	5.1	< .01	0.6	n. s.
12.5	1.4	n. s.	-5.2	< .01	2.9	n. s.	2.5	n. s.
25.0	1.5	n. s.	-3.5	< .05	-0.9	n. s.	3.7	< .05
50.0	0.4	n. s.	-2.4	n. s.	-2.5	n. s.	3.1	n. s.

[a] degrees of freedom were always 19.

[b] p-values are two-sided.

Table 3

Cronbach's alpha of VAS and discrete scales separately each level of coherence.

	Full VAS			Discretized VAS			discrete scale		
Coherence	alpha	CI 2.5	CI 97.5	alpha	CI 2.5	CI 97.5	alpha	CI 2.5	CI 97.5
1.6	.91	.82	.95	.87	.77	.92	.85	.68	.91
3.1	.91	.81	.95	.89	.78	.93	.86	.73	.91
6.2	.92	.87	.95	.91	.85	.94	.83	.63	.88
12.5	.85	.67	.91	.84	.67	.90	.85	.69	.90
25.0	.93	.80	.96	.92	.79	.96	.90	.76	.95
50.0	.93	.83	.96	.92	.83	.96	.93	.83	.96

Figure Captions

Figure 1. Experimental procedure.

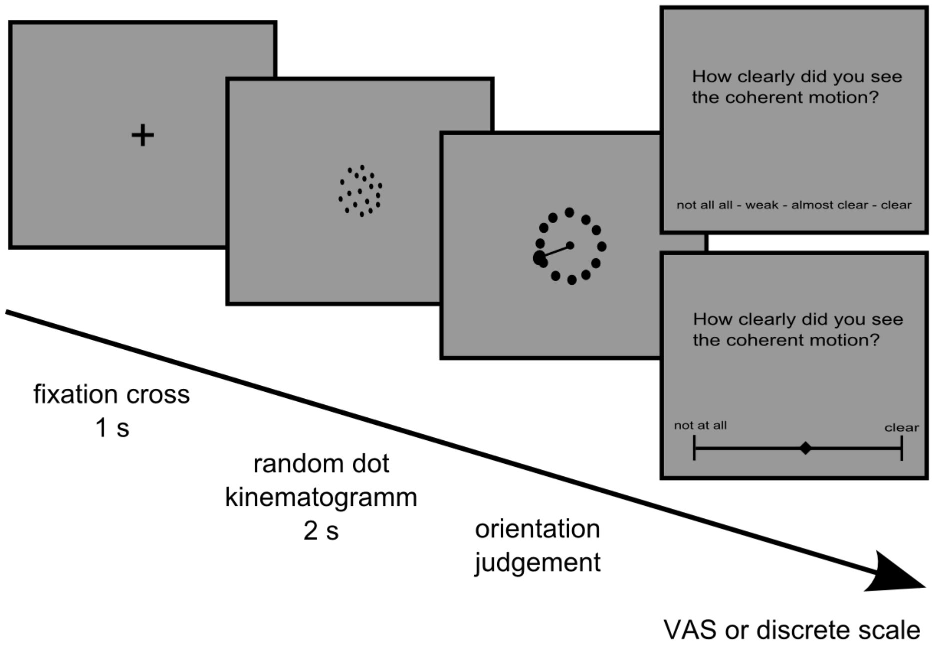
Figure 2. Distribution analysis of discrimination responses. Dots indicate the relative frequency of orientation responses with 0 as the true motion direction with different levels of coherence in each panel. Lines indicate the distribution of responses estimated from the fitted guessing and precision parameters. The grey highlighted area indicates the degree of accuracy between -45° and 45° where no error feedback was given.

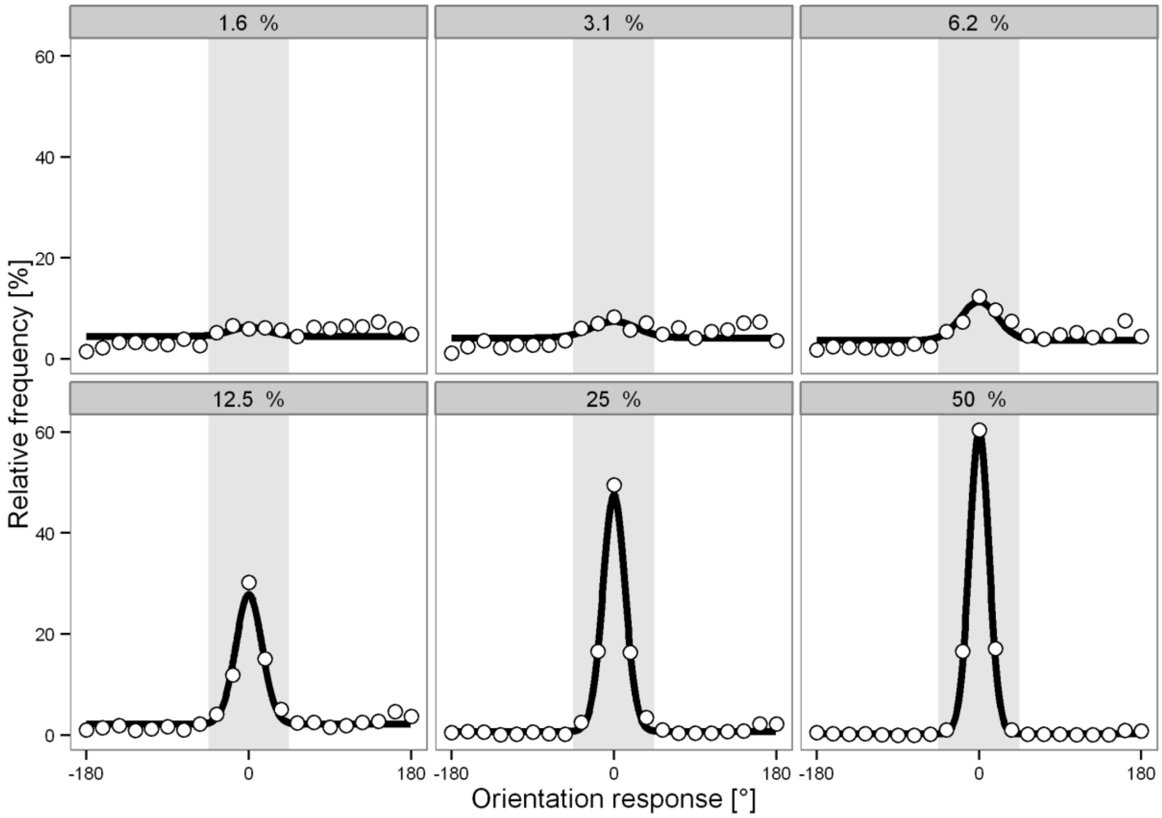
Figure 3. Estimated parameters from the distribution analysis plotted as a function of Coherence. Left Panel: Guessing probability. Right Panel: Precision. The grey areas indicate 95% bootstrapped confidence intervals.

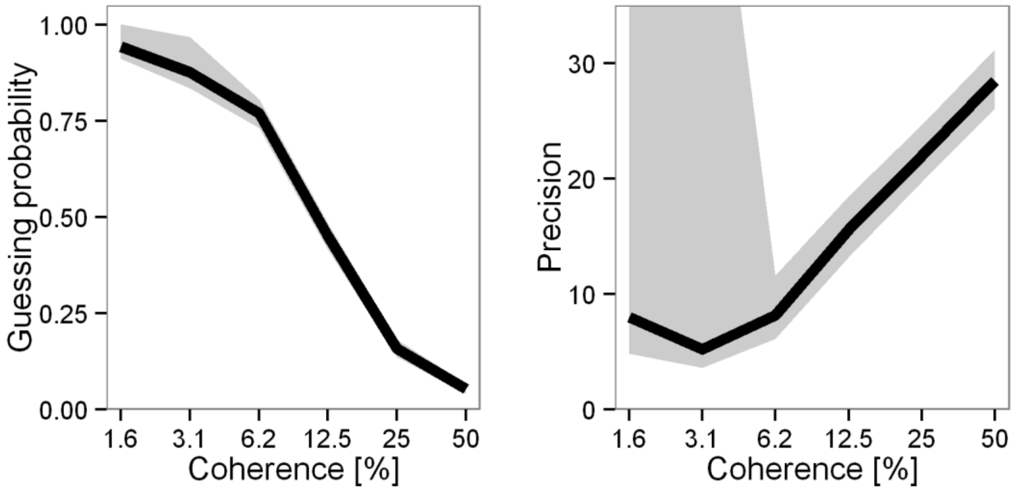
Figure 4. Discrimination error as a function of subjective reports, scale, and levels of coherence. The ratings on the visual analogue scale were discretized into four bins based on individual quartiles. A discrimination error of 90° indicates chance performance.

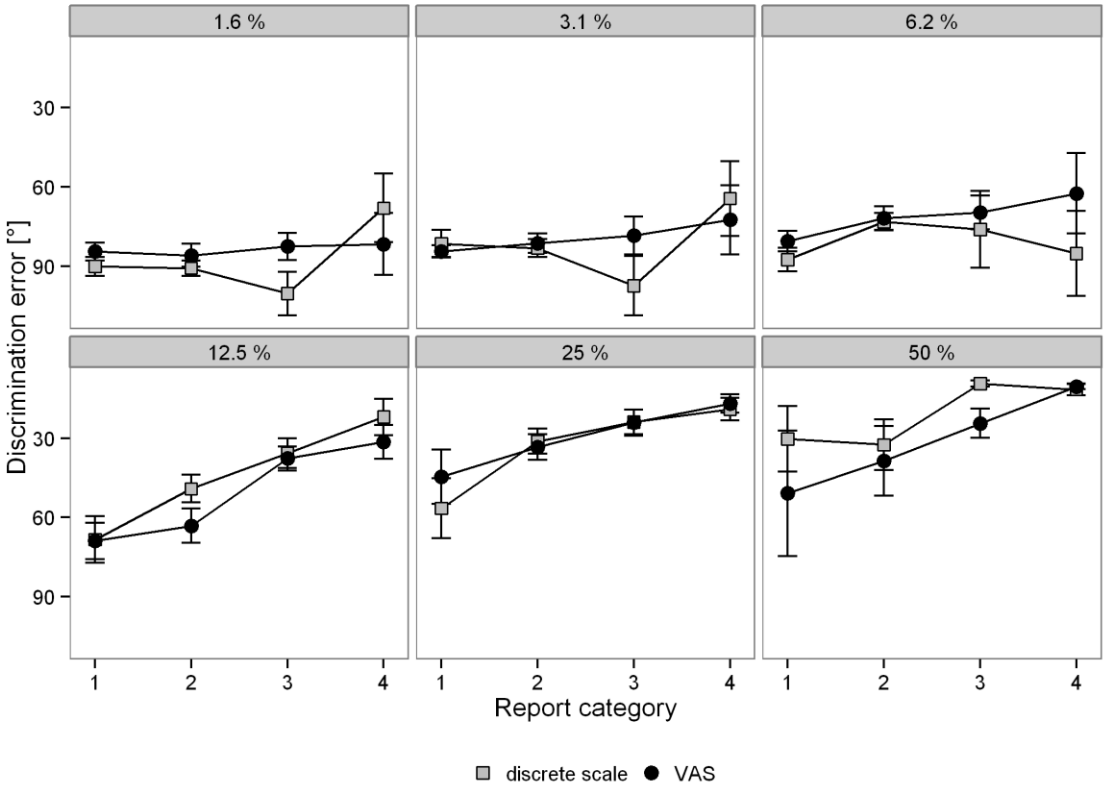
Figure 5. Ordered logistic regression slope of discrimination error predicted by subjective report depending on report time, i. e. time between objective task response and subjective report, and scale. To allow fitting separate regression models, report time is discretized into four bins based on the .25, .5, and .75 quantile.

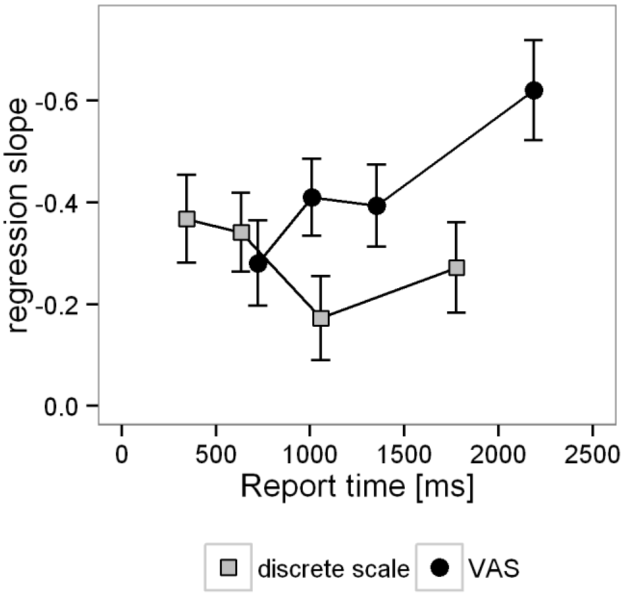
Figure 6. Frequency of each scale step of the discrete scales and the frequency of the corresponding scale parts of the VAS. Black bars indicate the VAS and grey bars the discrete scale. Error bars indicate 1 standard error of the mean.

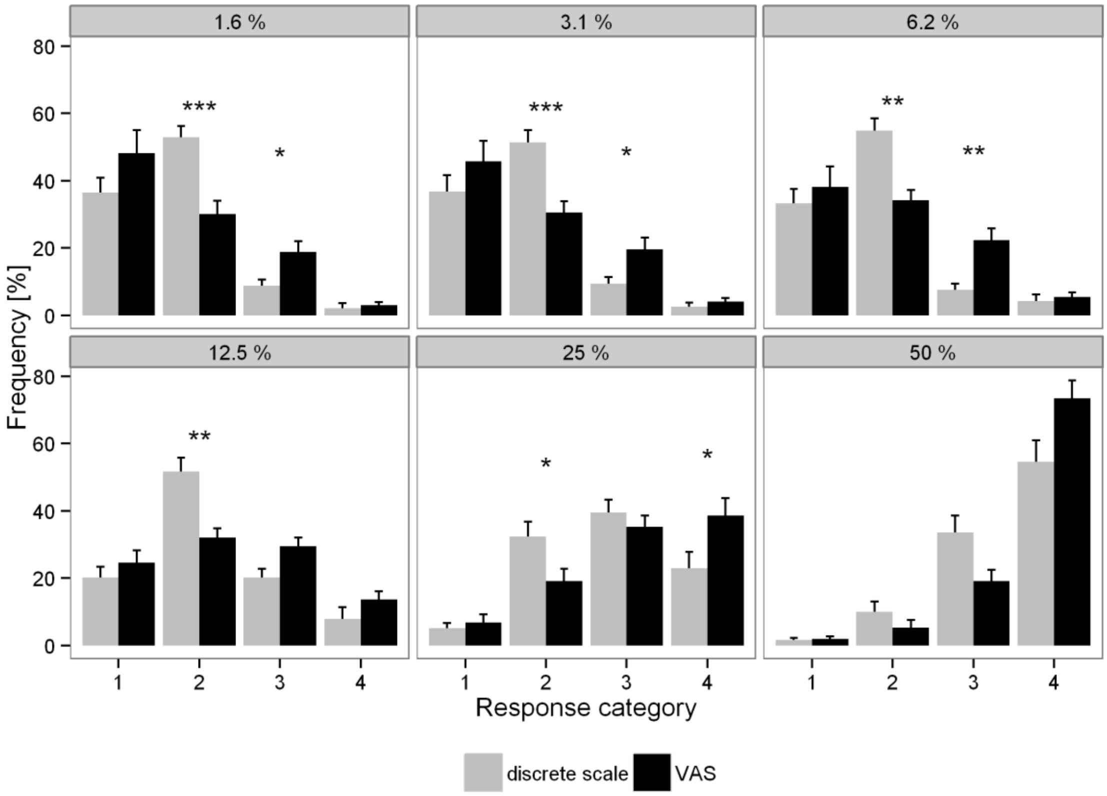

How clearly did you see
the coherent motion?
not all all - weak - almost clear - clear
How clearly did you see
the coherent motion?
not at all
clear
fixation cross
1 s
random dot
kinematogramm
2 s
orientation
judgement
VAS or discrete scale


1.6 %
3.1 %
6.2 %
12.5 %
25 %
50 %
Relative frequency [%]
0
20
40
60
-180
0
180
Orientation response [°]


Guessing probability
1.00
0.75
0.50
0.25
0.00
1.6
3.1
6.2
12.5
25
50
Coherence [%]
Precision
30
20
10
0
1.6
3.1
6.2
12.5
25
50
Coherence [%]


1.6 %
3.1 %
6.2 %
12.5 %
25 %
50 %
Discrimination error [°]
30
60
90
1
2
3
4
Report category
discrete scale
VAS


regression slope
-0.6
-0.4
-0.2
0.0
0
500
1000
1500
2000
2500
Report time [ms]
discrete scale
VAS


1.6 %
3.1 %
6.2 %
12.5 %
25 %
50 %
Frequency [%]
Response category
***
*
**
discrete scale
VAS

1