Consciousness and Cognition 75 (2019) 102798

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

In search of the optimal measure of awareness: Discrete or continuous?

Michał Wierzchoń[a,*], Anna Anzulewicz[a,b], Justyna Hobot[a], Borysław Paulewicz[c], Jérôme Sackur[d,e]

[a] Consciousness Lab, Institute of Psychology, Jagiellonian University, Krakow, Poland
[b] Jan Matejko Academy of Fine Arts in Krakow, Poland
[c] SWPS University of Social Sciences and Humanities, Katowice Faculty of Psychology, Katowice, Poland
[d] Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), Départment d'Études Cognitives de l'École Normale Supérieure, Centre National de la Recherche Scientifique, École des Hautes Études en Sciences Sociales, Paris Sciences et Lettres Research University, Paris, France
[e] Laboratoire Interdisciplinaire de l'X, École Polytechnique, Palaiseau, France

A B S T R A C T

Studies of perceptual awareness require sensitive measures reflecting subjective judgments of visibility. Two scales have been proposed for this purpose: the Continuous Scale (CS) and the Perceptual Awareness Scale (PAS). Here we compare the scales in the context of the Gabor patch orientation discrimination task and propose a Continuous Perceptual Awareness Scale (C-PAS) that aims to combine their advantages. The results of the study shown no differences in sensitivity between the scales. However, we observed differences between the scales in awareness ratings frequencies and accuracy associated with the lowest ratings. We concluded that visibility ratings are often biased, and thus, the scale sensitivity may not be optimal. Furthermore, based on the additional analyses, we argued that there is an advantage of using C-PAS over CS. The scale allows to use an additional variability of judgment within PAS categories and thus it may enable more fine-grained measurement of visibility at near-threshold conditions.

1. Introduction

The question of how to measure visual awareness has caused a long-standing debate in consciousness research. This is because verbal reports on visibility that are typically used to assess subjective visual experience might not capture all the relevant information due to their low sensitivity (Dienes & Perner, 2004; Dienes, 2004; Nisbett & Wilson, 1997; Shanks & St. John, 1994). On the other hand, the objective tests proposed in the literature (such as forced choice or discrimination tasks) measure to what extent participants were sensitive to visual information, but are frequently criticised as they do not measure subjective awareness accurately (e.g. participants often claim to have no awareness of a stimulus even when exhibiting above-chance performance with these measures). The absence of better alternatives leads to the conclusion that researchers should use awareness measures based on verbal reports, as well as try to improve their validity by combining objective and subjective measures and increase their sensitivity using statistical analyses. Here we take a closer look at the metacognitive sensitivity (the degree of association between behavioral accuracy and visibility ratings) of the two types of scale that are most often used to measure visual awareness: the Perceptual Awareness Scale (PAS, Ramsøy & Overgaard, 2004) and the continuous Visual Analogue Scale (Continuous Scale, CS; Hayes & Peterson, 1921; Sergent & Dehaene, 2004). We also introduce a novel scale, which combines properties of both measures.

The PAS was constructed by allowing participants to create their own response categories to express the level of stimulus visibility. In a study using a visual stimulus feature identification task, participants were asked to report the degree of clarity of visual

* Corresponding author at: Institute of Psychology, Jagiellonian University, Ingardena 6, 30-060 Kraków, Poland.
E-mail address: michal.wierzchon@uj.edu.pl (M. Wierzchoń).

https://doi.org/10.1016/j.concog.2019.102798
Received 12 January 2018; Received in revised form 21 July 2019; Accepted 22 July 2019
1053-8100/

experience after each identification response (Ramsøy & Overgaard, 2004). They were asked to construct a scale embracing levels of clarity ranging from "no experience at all" to "clear image", and were allowed to create their own intermediate categories. In the end, most of the participants agreed that a four-point scale (with the categories "no experience", "vague experience", "almost clear experience" and "clear experience") is optimal for capturing degrees of perceived clarity (at least in the case of the stimuli material that was used in this study, i.e. simple coloured shapes in different positions). It was observed that visibility ratings predict identification task accuracy such that higher visibility ratings are associated with better identification task performance. Furthermore, a positive correlation was found between the duration of stimuli presentation and the reported level of visibility. Proponents of the PAS claimed that such a design makes the scale more intuitive and that the usage of a simple dichotomous "unaware/aware" measure is often perceived as more difficult (Ramsøy & Overgaard, 2004). Subsequently, Overgaard and colleagues demonstrated that the PAS is indeed more metacognitively sensitive than a dichotomous scale. The results of their study, which employed a visual recognition task, revealed that the probability of correct recognition of a visual stimulus (four-alternative choice) was slightly lower when the participants rated it as "unseen" on the PAS (31%) as compared to an "unseen" response on a dichotomous scale (35%). However, the accuracy at the highest PAS level was significantly higher (94%) than the average accuracy observed with the "seen" response on the dichotomous scale (78%, see: Overgaard, Rote, Mouridsen, & Ramsøy, 2006). Thus, it seems that response criteria for awareness differ, depending on the awareness scale applied (Overgaard et al., 2006; Wierzchoń, Paulewicz, Asanowicz, Timmermans, & Cleeremans, 2014; Zehetleitner & Rausch, 2013). It was suggested that the PAS allows participants to express their conscious experience (visibility itself), but does not rely on a metacognitive judgment on the recognition task response (Overgaard & Sandberg, 2012; Ramsøy & Overgaard, 2004, but see: Wierzchoń et al., 2014 for the opposite view). Paradoxically, this also may be considered as a disadvantage of the PAS, as reports depend on participants' introspective skills in reporting their experience. Similarly as in the case of other introspective reports, the scale might be confounded by poor insight and low verbal abilities, both of which could lead to an underestimation of awareness level. It may also be criticised for providing only four categories of visibility and not allowing participants to report levels of awareness between the assumed categories (like CS does).

Although the gradual character of perception has been suggested in early works on consciousness (e.g. Lange, 1893; Werner & Wapner, 1952), there was no consistent method of measuring this characteristic. The CS is another popular method of awareness measurement. The most known version of the continuous scale—employed by Sergent and Dehaene (2004)—was developed specifically to address the question of whether conscious perception is an all-or-none or a gradual phenomenon. In two studies employing the Attentional Blink paradigm (AB, see: Raymond, Shapiro, & Arnell, 1992), Sergent and Dehaene (2004) asked participants to rate the clarity of their visual experience on a scale where only the extremes were described with verbal categories ("not seen" and "maximal visibility"), but participants could select a response in-between the extremes of the scale.[1] It was assumed that if the transition from unconscious to conscious perception is gradual, the participants should use the available scale spectrum, thus revealing different levels of visual awareness. Conversely, if the transition from unconscious to conscious perception is all-or-none, then participants should use the scale in a binary manner: they should report that they either had a clear experience or no experience at all. Participants evaluated the subjective visibility of a target (number word) after it was presented. The results indicated that participants used the scale in all-or-none fashion, suggesting a sharp transition between unconscious and conscious processing (see also: Del Cul, Baillet, & Dehaene, 2007 for similar results). It was, however, suggested that this all-or-none pattern is a result of the design characteristics of the CS. First of all, because only the extremes of the scale were labelled, the space between them was devoid of clear meaning (Overgaard et al., 2006). For this reason, participants might be biased towards the ends of the scale, which resulted in a bimodal distribution of scale ratings. Notably, Sergent and Dehaene (2004) addressed this issue in their paper. They argue that the pattern observed in AB experiments does not reflect a response bias towards the ends of the scale, as participants spontaneously used the scale in a continuous manner in the masking task. In the third experiment employing an AB paradigm with variable duration of the second target, participants used the scale gradually, but the AB phenomenon still yielded all-or-none response patterns. On the one hand, contrary to the results obtained by Sergent and Dehaene (2004) in their masking experiment, some results indicate that the CS leads to a more dichotomous rating pattern regardless the underlying subjective visual experience (see: Windey, 2014). On the other hand, studies of Rauch & Zehetleitner (2014, 2016) show that the distribution of visibility ratings are uniform or multimodal at intermediate levels of stimulation while using a CS scale. The other problem pointed out by Overgaard et al. (2006) was that the continuous scale used by Sergent and Dehaene (2004) had too many intermediate points. It is unlikely that the participants were able to define so many levels of conscious experience (see also: Hake & Garner, 1951), thus they might find the continuous scale confusing.[2] This argument is in line with the aforementioned results obtained by Ramsøy and Overgaard (2004), who demonstrated that the use of a four-point scale is intuitive, and adding more verbally defined categories does not provide any meaningful data. However, when slightly different CS was used to measure conscious experience of motion—the one in which participants provided responses by moving a joystick horizontally and confirming the answer by pulling the trigger—the obtained results showed that the scale was used gradually and was more efficient at predicting discrimination error than the discrete scale with four scale categories (Rausch & Zehetleitner, 2014). Thus, the results seem to depend on the task used (how much perceptual variability is provided in the task).

To sum up, empirical evidence suggests that the level of clarity of conscious visual experience reported by participants depends

[1] Precisely, the procedure allowed one of 21 continuous positions to be chosen; in each trial a cursor was set at a random position on the scale and participants were asked to estimate the level of visibility of stimuli by pressing two keys on the keyboard to move the cursor towards the left or right extreme of the scale.

[2] This is even more probable, as the original Sergent and Dehaene scale was implemented such that participants had to move the cursor indicating their visibility rating by multiple key-presses of the left or the right arrow keys (Sergent & Dehaene, 2004).

not only on actual stimulus visibility, but also on the experimental setup. Importantly, the theoretical assumptions may also lead researchers to prefer some measures over others (e.g. the assumption of a gradual access to the content of awareness may lead to the proposal of a scale containing intermediate steps), and vice versa; thus, methodological details can lead to certain conclusions on the characteristics of consciousness. For example, the application of a dichotomous scale does not rule out the possibility that awareness of a stimulus is gradual, but it precludes the possibility of investigating the effects of different levels of visual awareness. This is however possible with the PAS, as it distinguishes additional levels of awareness between "unseen" and "seen" categories. However, as mentioned above, it has also been demonstrated that even when participants are required to use the continuous scale (which should allow intermediate levels of awareness to be tested), depending on the procedure they use it in a more dichotomous (Del Cul et al., 2007) or in a more gradual manner (2016; Rausch & Zehetleitner, 2014; Sergent & Dehaene, 2004). Therefore, the number of scale points available to describe the level of stimulus awareness may influence visibility judgements. Similarly it is likely that the manner in which the categories are described influences scale sensitivity. However, it may be also the case that every measure containing verbal categories is affected by problems of arbitrary assessment criteria and difficulties associated with the subject-specific or language-based differences in language comprehension. On the other hand, the lack of any descriptions may render such a scale difficult to use for subjects and difficult to interpret for experimenters.

Following the three theoretical assumptions that (1) there is a direct relation between the task performance accuracy and the level of stimulus awareness, (2) verbal categories may limit the reports on subjective experience, and (3) the nature of visual experience might be investigated as a gradual phenomenon, we may conclude that the most valid measure of subjective awareness should enable an adequate amount of response options, but also make some additional response variability possible. Thus, it should allow evaluation of subjective experience by proposing a certain number of verbal categories, but at the same time, it should not force participants to express their visual experience with these categories. Such an approach should lead to the best match between ratings of subjective visual experience and task performance. To address these assumptions, we propose a Continuous Perceptual Awareness Scale (C-PAS) that aims to combine the advantages of the CS and the PAS. By containing continuous spaces in four response categories, it allows the states "in-between" to be reported, which may not be captured by verbalization. The descriptions of the four scale steps are intended to make the choice more intuitive than on the CS and require only minor cognitive effort. The C-PAS is also meant to be less susceptible to the influence of verbalization abilities and subjects' strategies than the aforementioned scales. This paper aims to compare the PAS and the CS by means of scale sensitivity and the extent to which they support the measurement of the intermediate levels of awareness. We also analysed the properties of the new scale we proposed, namely a C-PAS scale that investigates whether and when it may be preferred over the previously proposed scales.

2. Method

2.1. Participants

71 healthy participants (48 females, mean age = 22.3 years, SD = 2.7) took part in the experiment in exchange for a small financial gratification. All participants reported normal or corrected-to-normal vision and no history of neurological disorders. The participants were randomly assigned to one of three experimental groups: the PAS (25 participants), the CS (17 participants), or the C-PAS (29 participants).

2.2. Materials

The experiment was run on computers using software developed specifically for the purpose of this study (time-critical portion of the software was written in C+ +.). The LCD monitors had a screen resolution of 1280 × 800 pixels and a refresh rate of 60 Hz. A visual orientation discrimination task was employed with Gabor patches masked by a checkerboard matched in frequency serving as target stimuli. The stimuli and the mask were displayed on a light grey background and were viewed from a distance of 60 cm. The diameter of the stimuli subtended 3.7° of visual angle. The parameters of the Gabor patches were constant (contrast = 30%, frequency = 20 cycles per screen width, standard deviation of the Gaussian envelope = 0.1 screen width) and the patches were tilted either 45° to the left or right from the vertical position. An awareness scale was presented in the lower part of the screen in a frame sized c.a. 35.14 by 1.3° of visual angle. The scales were labelled differently depending on the experimental condition: (1) in the PAS condition four response categories were used ("no experience", "a vague experience", "an almost clear experience", and "a clear experience"); (2) in the CS condition only the scale extremes were labelled "no experience" and "a clear experience"; and (3) in the C-PAS condition four categories with PAS tag names were embedded inside a continuous scale.

The participants were asked to indicate their response by moving the mouse and pressing the left mouse button. The three types of scales are presented in Fig. 1 (panel B).

2.3. Procedure

The experimental procedure is depicted in Fig. 1 (panel A). Each trial began with a fixation cross presented at the centre of the screen for 1500 ms, followed by a Gabor patch randomly varying between two orientations. The stimulus duration was randomly intermixed across trials (16, 32, 64 and 128 ms). The target was immediately followed by the 200 ms mask and the orientation discrimination task. In each trial, participants were asked to identify the Gabor patch orientation as quickly as possible with maximal accuracy. Participants responded by pressing one of two keys (the left or right arrow on a standard QWERTY computer keyboard).

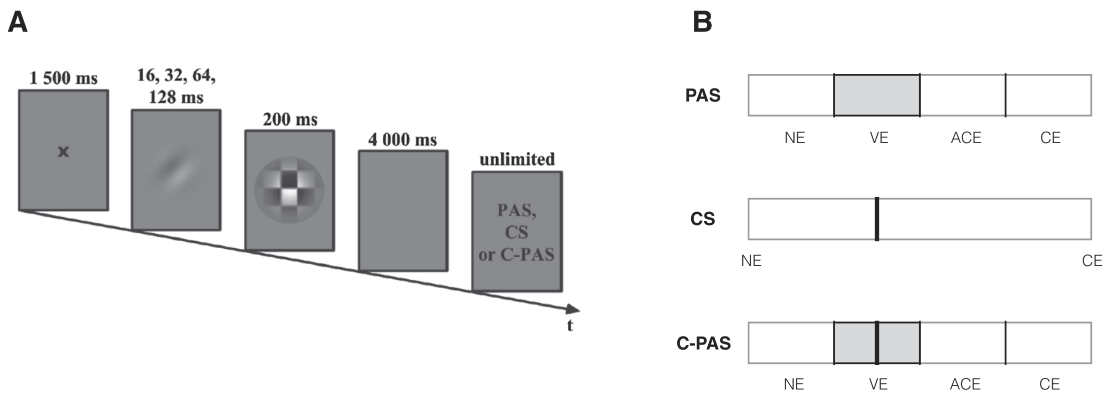

Fig. 1. The experiment procedure (panel A) and a visualisation of the awareness scales applied in the experiment (panel B: NE = no experience, VE = vague experience, ACE = almost clear experience and CE = clear experience).

The maximum time allowed for the response was 4000 ms. After the orientation discrimination response, participants were asked to rate the visibility of the target stimulus using one of the scales. A graphical representation of one of the three scales of visual awareness (PAS, CS, or C-PAS) was presented. Each scale was displayed in a one-row table containing descriptions of scale categories, and/or presenting the area in which participants could choose their response within a continuous range (i.e. CS and C-PAS). Participants marked their response using a standard computer mouse, and depending on the condition, they: (1) choose one of four possible categories in which a selected category was highlighted (PAS), (2) specify the cursor position on a continuous horizontal line that did not contain intermediate categories, so only the scale extremes were described as “no experience” and “clear experience” (CS) and, (3) choose one of four possible categories and then decided on the specific position of the cursor within a selected PAS category (C-PAS). The response time was unlimited.

The experiment consisted of a practice session and a main task. In each practice session the first phase of the training consisted of 8 trials (target duration: 500 ms); the second phase of the training consisted of 16 trials with target durations of 16, 32, 64, 128 ms. The main task consisted of 320 experimental trials. Thus, each of the four target durations (16, 32, 64, 128 ms) was repeated 80 times over the course of the experiment. The order of trials was randomized across participants. The scales were described in the following manner:

- CS - The left end means the lack of pattern visibility. The next parts on the scale refer to intermediate experiences between the lack of pattern detection and its full visibility. In other words: the closer to the right end of the scale, the stronger the experience of pattern perception was. The right edge means that the pattern was clearly seen. We ask for the most precise assessment of how clearly you have seen the pattern.
- PAS - The category “I have not seen anything” means the lack of pattern visibility. The next categories of scale refer to intermediate experiences between the lack of pattern detection and its full visibility. In other words: the closer the category is to the category “I have seen very clearly”, the stronger the experience of the pattern perception was. The last category of the scale means clearly seeing the pattern. We ask for the most precise assessment of how clearly you have seen the pattern.
- C-PAS - The left end means the lack of pattern visibility. The next parts on the scale refer to intermediate experiences between the lack of pattern detection and its full visibility. In other words: the closer to the right end of the scale, the stronger the experience of pattern perception was. The right edge means that the pattern was clearly seen. The scale within the categories is continuous to allow the most accurate assessment of the pattern's visibility. This means that it is possible to differentiate the experience within the following categories (the more on the right within a given category the indicator is placed the better the pattern was visible). We ask for the most precise assessment of how clearly you have seen the pattern.

3. Results

Participants whose accuracy was lower than 0.6 (n = 5) were excluded from the analysis, because they were not performing the task adequately, especially at longer presentation times.[3] First, we compared the Scale conditions response criteria, d′, accuracy and meta-d′/d′ ratio (the so-called the M ratio; see Maniscalco & Lau, 2014). The response criteria differed only between CS and PAS. In the PAS condition participants tended to respond left more often ($p < .05$). Sensitivity was higher in the PAS compared to the CS condition ($p < .001$) and the C-PAS condition ($p < .05$) and was higher in the C-PAS condition than in the CS condition ($p < .05$). The same pattern of results was observed when average accuracy was compared instead of sensitivity. Finally, the comparison of the M ratio indices did not reveal any systematic differences between groups. Participants provided comparable awareness ratings on average in all Scale conditions. For the average task accuracy, average ratings, d’ and M ratio across Scale conditions, see Table 1.

[3] The datasets and analyses script are publicly available via the Open Science Framework website: osf.io/w46xv.

Table 1
Averaged task accuracy, average ratings, d' and M ratio across Scale conditions.

	Average accuracy	Average rating (0–3)	d′	M ratio
CS	0.76	1.27	1.56	1.10
PAS	0.85	1.46	2.22	1.05
C-PAS	0.81	1.35	1.91	1.09

Accuracy data were analysed using a logistic mixed model with Presentation Time (16 ms, 32 ms, 64 ms and 128 ms), Scale (CS, PAS, and C-PAS), Awareness Rating (0–3, either continuous or four-point depending on the condition) and their interaction as fixed effects and a subject-specific random intercept and slope. In order to increase the readability of the results, we have used a nested parametrization with separate intercepts and slopes for each presentation time × scale condition within the same model. Thus, the regression intercepts depict whether accuracy in a given condition deviates from chance at the lowest rating, whereas the regression slopes represent metacognitive sensitivity, i.e., how well visibility ratings predict discrimination accuracy (see: Table 2), which is conceptually equivalent to the presented M ratio indices compared within each Presentation Time condition (see e.g., Wierzchoń, Asanowicz, Paulewicz, & Cleeremans, 2012).

The intercepts and slopes of the regression lines were then compared between scales for each Presentation Time condition. The results are presented in Table 3.

The results show that the only significant difference between slopes was observed for the 64 ms condition, where C-PAS predicts accuracy better than CS. Given the number of comparisons, this single difference should be treated with caution. We further analysed the metacognitive sensitivity as measured with M ratio within each Presentation Time and found no significant differences between the Scale conditions. Thus, we concluded that the scales have comparable metacognitive sensitivity.

Despite the lack of differences in metacognitive sensitivity, it seems worth taking a closer look at the effects we observed when we compared the regression intercepts (which could be interpreted as the scale's sensitivity to low levels of consciousness; see Wierzchoń et al., 2012; Timmermans & Cleeremans, 2015, or just as a measure of conservative/liberal rating bias). The results show that C-PAS accuracy at 16 ms was significantly lower than the accuracy observed at the lowest rating with other scales. A similar effect was observed in the 32 ms and 64 ms conditions between PAS and C-PAS. We also observed that CS accuracy was significantly lower than PAS in the 64 ms condition. Finally, we also observed a significant difference in the opposite direction between C-PAS and CS in the 128 ms condition, but it seems that with this presentation time the accuracy at the intercept is not informative as participants mainly used higher ratings.

To sum up, we did not observe differences in metacognitive sensitivity as measured with M ratio and regression slopes. Furthermore, the accuracy observed at the lowest rating was lower in C-PAS compared to the PAS conditions at most of the Presentation Times (and CS at the lowest presentation time), which suggests that the C-PAS scale enables better reporting of the

Table 2
Regression coefficients of the logistic regression mixed model for the accuracy of awareness ratings (intercepts and slopes nested within Presentation Time and Scale Type).

N66 #observations 19,640	Estimate	SE	z	p
16: CS	0.28	0.12	2.33	< .05
16: PAS	0.34	0.11	3.04	< .01
16: C-PAS	−0.05	0.11	−0.46	ns
32: CS	0.21	0.12	1.80	ns
32: PAS	0.51	0.11	4.49	< .001
32: C-PAS	−0.04	0.11	−0.37	ns
64: CS	0.88	0.16	5.29	< .001
64: PAS	2.03	0.39	5.24	< .001
64: C-PAS	0.61	0.23	2.65	< .01
128: CS	0.52	0.26	0.99	< .05
128: PAS	0.61	0.50	1.23	ns
128: C-PAS	1.88	0.53	3.54	< .001
16: CS: Awareness Rating	0.75	0.16	4.69	< .001
16: PAS: Awareness Rating	1.07	0.15	7.08	< .001
16: C-PAS: Awareness Rating	1.02	0.15	6.88	< .001
32: CS: Awareness Rating	0.82	0.16	5.14	< .001
32: PAS: Awareness Rating	1.06	0.15	7.07	< .001
32: C-PAS: Awareness Rating	1.09	0.15	7.42	< .001
64: CS: Awareness Rating	0.99	0.17	5.94	< .001
64: PAS: Awareness Rating	1.20	0.26	4.56	< .001
64: C-PAS: Awareness Rating	1.81	0.19	9.26	< .001
128: CS: Awareness Rating	1.47	0.19	7.80	< .001
128: PAS: Awareness Rating	1.70	0.26	6.46	< .001
128: C-PAS: Awareness Rating	1.20	0.27	4.44	< .001

Likelihood ratio: $\chi^2(25) = 3800$ $p < .001$.

Table 3
Differences between the intercepts and slopes (the column variables subtracted from the row variables). Contrasts were calculated based on the fit summary statistics for the full model of mixed regression, parameterized for the analysis presented in Table 2.

Intercepts		CS	PAS
16 ms	C − PAS	−0.33*	−0.39*
	CS		−0.06
32 ms	C − PAS	−0.26	−0.55***
	CS		−0.30
64 ms	C − PAS	−0.27	−1.42**
	CS		−1.15**
128 ms	C − PAS	1.36*	1.27
	CS		−0.09
Slopes		CS	PAS
16 ms	C − PAS	0.27	−0.05
	CS		−0.32
32 ms	C − PAS	0.27	0.03
	CS		−0.25
64 ms	C − PAS	0.82**	0.62
	CS		−0.21
128 ms	C − PAS	−0.29	−0.50
	CS		−0.21

* $p < .05$.
** $p < .01$.
*** $p < .001$.

experience in near-threshold conditions (i.e. it has higher scale sensitivity to low levels of consciousness; see Wierzchoń et al., 2012), or that we observed stronger rating bias on one of the conditions.

The main aim of this study was to compare the sensitivity of the different measures of perceptual awareness. The initial hypothesis was that the C-PAS combines the advantages of both the PAS and CS, allowing for more precise judgements of visibility (using continuous judgments), but at the same time organising them in the PAS categories, thus unifying the rating criteria. As reported above, we did not observe any systematic differences between regression slopes. However, this does not mean that continuous ratings within PAS do not provide additional information. When continuous and discrete ratings are on the same scale in terms of minimum and maximum possible rating, then they might not affect logistic regression slopes, even if continuous ratings were used meaningfully by participants. We reason that, if only the discrete ratings matter, the position of continuous ratings within the boundaries of discrete rating categories should only be a source of additional noise. However, if continuous ratings are informative about accuracy even within the boundaries of discrete rating categories, then a model that uses continuous rating data from a given scale should explain more variance in accuracy than a model that uses discrete rating data from the same scale. In order to test if continuous rating data provide some information that is not already present in discrete rating data for the CS or the C-PAS scale, we separately compared the fit of the two models for each of these two scales, one using discrete ratings and one using continuous ratings. Otherwise, the mixed logistic models were the same as described above (i.e. with Presentation Time, Scale Type (only C-PAS and CS), Awareness Rating and their interaction as fixed effects and a subject-specific random intercept and slope). Because these models are not nested, we used Bayesian Information Criterion (BIC) to compare the continuous and discrete versions. Comparing discrete and continuous version of C-PAS we found a very strong evidence (BIC difference = 12) in favour of the discrete version of the C-PAS scale. This means that the C-PAS metacognitive sensitivity relies on the presence of the categories. However, in case of comparison discrete and continuous versions the CS, the BIC index provided very strong evidence in favour of the continuous model (BIC difference = 39). This in turn suggests there is a benefit that comes from allowing participants to use a continuous rating (if we force arbitrary categorisation into four equal categories as we did in CS scale for the sake of this comparison). Taken together, the categorisation is beneficial when the categories are meaningful, and the continuity of the scale allows for providing relevant information about metacognitive accuracy. Thus, this result is evidence that introducing the PAS categories allows better categorization of subjective experience than CS alone and thus that CS can benefit from categorisation; however, on the basis of the presented analysis we cannot conclude whether this effect is observable only in C-PAS, or also in the regular PAS, which could not be compared using this method. In other words, we confirmed that C-PAS should be preferred over CS and that continuous responses within C-PAS do not add meaningful information to the awareness ratings; however, we do not know whether this will also be the case for the regular PAS scale. To decide which version of the scale should be preferred, we should take a closer look at rating distributions.

3.1. Awareness ratings distributions

We compared the response frequencies for each type of scale and each rating in each Presentation Time condition (to increase the readability of Fig. 2, distributions on the continuous scales are plotted in 11 bins). Upon preliminary visual inspection, Fig. 2 shows differences in empirical distributions between the scales. Apart from the obvious differences between four-category and continuous scale frequencies, we observed a change in the frequency of the extreme ratings associated with the use of the different types of scale.

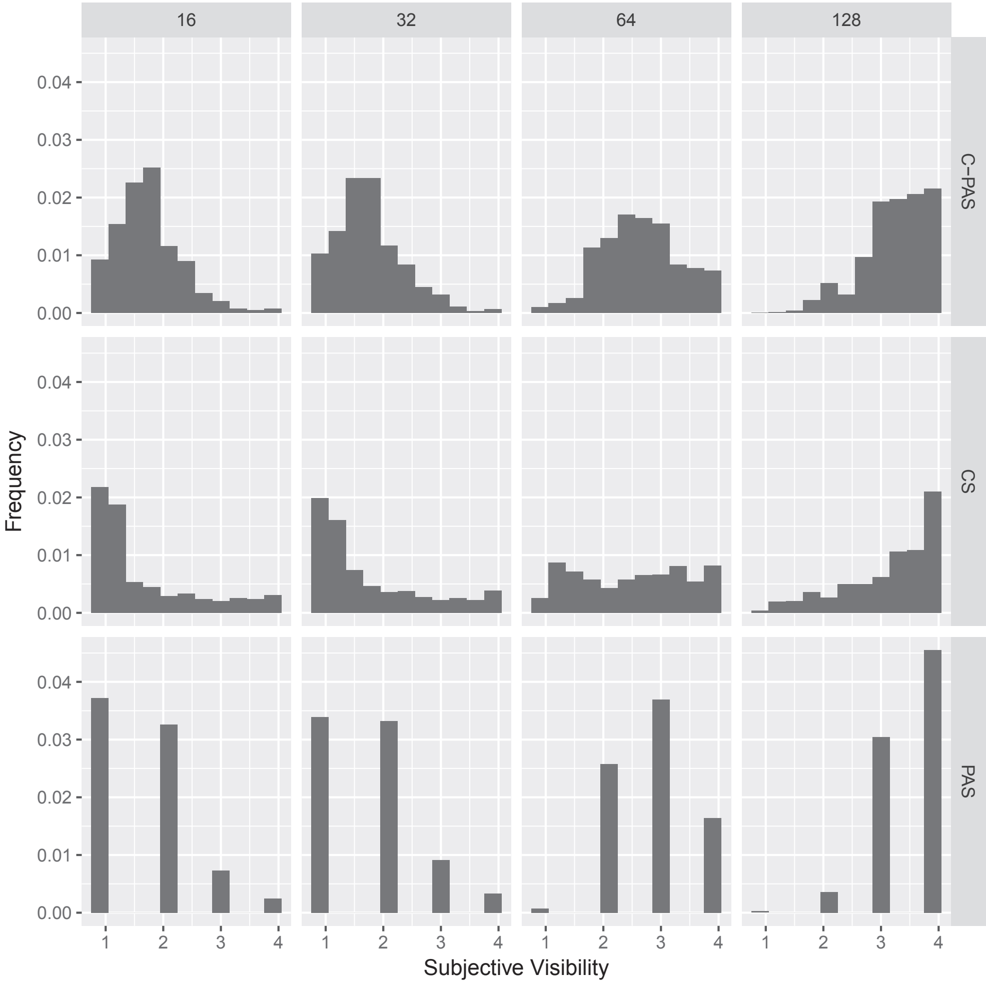

Fig. 2. Scale rating distributions as a function of Presentation Time.

Extreme ratings were given more frequently (compared to all other ratings) in case of the CS. We also noted an increased number of responses of the C-PAS lower-middle ratings (that is not the lowest rating, but low visibility), which may reflect differences in the rating bias.

To test if the rating distributions were indeed different, we fitted a Bayesian hierarchical Beta regression model to ratings normalized to fit in the (0,1) interval. Thus, the shapes of the distributions could be compared even though the number of rating levels differed. It was assumed that the distributions may differ by Scale and Presentation Time and that both Beta parameters (mean and scale) may vary by participant (a random effect) within Presentation Time. The model was fitted using the brms package. There were 6 chains with 4000 iterations, half of which served as a warm-up stage. The chains exhibited good mixing performance, i.e., the $\hat{R}$ statistic was lower or equal to 1.01 and there were at least 1000 effective samples for all the relevant parameters. Analysis by highest posterior density intervals revealed that the shapes of all the rating distributions differed between all the scales within each duration (the 95% HPD intervals computed for all the contrasts in scale parameters did not contain 0). The analysis is available within the OSF script, but the fact that the shapes differ does not yet imply that they differ in specific features.

To test specific effects of response frequencies, we created additional binary indicator variables representing the lowest ratings (1 when rating = 1, otherwise 0), middle (1 when rating = 2 or 3) and highest (1 when rating = 4). We aggregated frequency data in the three categories (low, middle, and high, so as to compare between scales) for all trials in each Scale Type condition (for CS and C-PAS, the middle ratings corresponded to PAS ratings 1 and 2). Next, we tested the differences between the indexes obtained for each scale by means of mixed logistic regression analyses with Scale condition. The contrast analyses show that the frequency of lowest ratings in the CS was significantly higher compared to both the PAS ($p < .001$) and C-PAS ($p < .05$) conditions. The frequency of the lowest PAS ratings was not significantly different than C-PAS. We also observed significant differences in the middle ratings' distributions between the CS and both PAS and C-PAS. In line with the lowest ratings' distribution results, the frequency of CS middle ratings was lower than in the other scales ($p < .001$). We did not observe a significant effect between PAS and C-PAS frequency ratings. Finally, the frequencies of the highest ratings were comparable for all scales. The middle CS ratings were used slightly less often (for most of the Presentation Time conditions), whereas both PAS and C-PAS were used with similar frequencies across all ratings. Thus, the analyses did not allow the differences observed with the Bayesian hierarchical Beta regression model to be confirmed, but note that the lowest category assumed here was quite broad, especially for C-PAS and CS conditions (i.e. the difference reported above may in fact appear within the first quarter of the visibility ratings).

4. General discussion

All the scales used in the experiment enabled us to predict accuracy in the Gabor patch orientation discrimination task. The higher participants rated their awareness, the better they performed the task, regardless of which scale was used. Thus, all scales seem to be sensitive measures of awareness, but this paper aims to compare them. The main goals of the paper were (1) to compare the two most popular measures of subjective visibility described in the literature, i.e. the Perceptual Awareness Scale (PAS) and the Visual Analogue Scale (Continuous Scale, CS) and (2) to introduce a new measure, the Continuous-Perceptual Awareness Scale (C-PAS). We proposed the latter to combine the advantages of both the PAS and the CS scales. We hypothesise that the C-PAS should allow more precise judgements of visibility due to the continuous judgments that will be organised by the main PAS categories to unify rating criteria. However, the results were not straightforward.

We did not observe any systematic differences in metacognitive accuracy between the scales as measured both with regression slopes and the M ratio index. However, we propose that continuous ratings may provide some additional information, even if they do not affect the logistic regression slope or the M ratio index. This is because the continuous and discrete ratings are in fact on the very same scale in terms of scale extremes. Thus, we decided to compare the model fits of the regression slopes for continuous and discrete versions of two scales that involved continuous ratings, i.e. C-PAS and CS. We assumed that if only the discrete ratings mattered, then the continuous ratings within the boundaries of a given discrete rating category would simply be a source of noise; thus, the discrete version of the model should have a better fit. However, if continuous ratings are informative within the boundaries of a given discrete rating category then the continuous version of the model should have a better fit. Given these assumptions, the C-PAS scale is the most interesting as it allows the four PAS categories to be compared with a continuous version of the same scale. The results showed that continuous data do not add much information to what is already present in the discrete PAS categories. Furthermore, the results obtained in the CS condition confirmed that the PAS categories (not just a separation of the continuous scale into the four categories) were informative, because when the PAS categories were not introduced, participants seemed to use the CS scale in a more continuous way. The CS results also suggest there is a benefit that comes from allowing participants to use a continuous rating. Importantly, the analyses of the lowest ratings' distribution showed that the CS lowest scale extreme ratings were often represented, whereas the middle scale ratings were observed less frequently than in other scales. This could mean that participants apply a much more conservative response strategy with CS, i.e. when PAS categories are not provided (see: Sandberg & Overgaard, 2015). However, it is (at least theoretically), possible that the rating bias is related to PAS (than more liberal one as compared to CS[4]). To sum up, PAS and C-PAS seem to involve more ratings reflecting low levels of consciousness than CS, as suggested by both the comparison of regression intercepts and the analysis of the distribution of ratings. The additional analysis comparing the continuous and discrete versions of C-PAS and CS scales seems to confirm that the verbal categories introduced by the PAS help participants categorise their experience. Therefore, they can accurately discriminate between different states of conscious access at near-threshold presentation times.

The response to the question of whether PAS should be preferred over C-PAS is even more difficult, as these scales not only do not differ in terms of sensitivity, but we were not able to compare them directly within the aforementioned model comparison analysis. Our comparison of model fits for the continuous and discrete versions of C-PAS and CS confirmed that PAS categories allow the subjective experience of the participants to be organised, but did not allow us to compare C-PAS and the regular PAS scales. Interestingly, visual inspection of the data for response distribution at near-threshold presentation times suggests that the number of C-PAS ratings does not decrease monotonously with increasing visibility ratings. Contrary to CS, where most of the ratings for short durations (16 ms, 32 ms) indicate no visibility (i.e. the left extreme of the visibility scale), in C-PAS the mode of the ratings is above the extreme low end of the scale. The difference in C-PAS distribution as compared both to CS and PAS was confirmed by our Bayesian hierarchical Beta regression model. This may suggest that C-PAS is more sensitive at near-threshold Presentation Times than PAS and CS (see: Wierzchoń et al., 2012), but we were not able to confirm this interpretation with any additional analysis. The conclusion seems to be in line with the results of the analysis comparing regression intercepts. The intercept for C-PAS at near-threshold Presentation Time (16 ms) was lower than for both the PAS and CS scales (it was the only intercept that was not different from zero for the 16 ms condition). This would mean that the C-PAS scale is more sensitive for the near-threshold Presentation Times (that is, it is least influenced by the conservative strategy) and that the lowest rating indeed represents no visibility only for this scale (guessing criterion – see Dienes & Perner, 2004; Timmermans & Cleeremans, 2015). Given the verbal descriptions of the lower two PAS ratings ("no visibility" and "a brief experience that is not necessarily related to the identification of the stimuli"), the distribution of the C-PAS responses may suggest that it allows capturing of the variability of the visual experience around the visual threshold with greater precision than PAS. However, as stated above, it is also possible that C-PAS and PAS are introducing liberal rating bias. Thus, abovementioned differences may also reflect stronger rating bias in case of those scales.

To sum up, the results described above suggest that C-PAS and PAS use more intermediate ratings than CS. The model comparison analyses show that PAS categories are informative. However, we concluded that continuous ratings may still be useful when general response categories are provided. Therefore, we propose that the C-PAS scale allows participants to further specify their subjective

[4] We thank an anonymous reviewer for suggesting such an alternative interpretation.

visibility. This is not reflected in the metacognitive sensitivity measures, but it influences ratings at the near-threshold conditions, changes response strategy and influences response distribution. In other words, our results suggest that one may effectively make use of an additional variability of judgment within PAS categories. Thus, the combination of PAS and CS features allows the variability of experience around the subjective threshold of awareness to be described more precisely. Therefore, C-PAS could be used whenever the theoretical assumptions of the study suggest using a continuous measure (e.g. in the case of computational modelling).

Our data demonstrate that the effectiveness of continuous ratings should be investigated in more detail. In order to do so, different options of the C-PAS scale may be introduced in future studies to identify the most sensitive version of the scale. In particular, it should be noted that the C-PAS scale that we tested implied fixed boundaries between categories, and allowed participants to introduce nuances *within* a category. It might be that this procedure is not the most effective one for revealing the combination of categorical and continuous information in awareness judgments. One concern might be that our procedure constrains participants to first select one category and that this first step may lead to a collapse of all additional information. In opposition to this “fixed boundaries” interpretation of categories, we might put forward the notion that categories of visual awareness could be defined by their focal prototype and separated by fuzzy boundaries (Rosch, 1973, 1975). Thus defined, PAS categories would serve as prototype levels of visual awareness, represented as reference points on a continuous scale. It remains to be investigated whether such usage of labels for prototypical experiences would help participants organize reports about their perception, while still enabling them to distinguish as many nuances as they have experienced. One may further compare the sensitivity of the scales at near-threshold presentation times by replicating our effects in a similar study with trials in which no stimulus is presented (catch trials).[5] This will allow to test whether the differences in the regression intercepts and response distributions indeed reflect higher sensitivity at near-threshold presentation times or rather reflect the stronger rating bias.

Finally, we have to acknowledge several limitations of our study. First, one may argue that the number of participants in the CS condition is not optimal (N = 17) as it is lower than in the case of the other two conditions. However, the majority of the analysis we applied should not be sensitive to the differences in the group size. More importantly, most of the observed effects are consistent with those observed in the literature (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010; Wierzchoń et al., 2012). Another critical limitation refers to the very issue of the definition of the minimal conditions allowing to choose the appropriate measure of consciousness experience. We have expected differences in scale sensitivity, assuming this will allow selecting the optimal scale. Extensive statistical analyses did not prove that one of the scales is more sensitive than others. We are also not entirely convinced that the sensitivity reflects the actual visual experience (see, e.g. Lau, 2007) and thus allows selecting the optimal subjective scale.

One may argue that at near-threshold conditions participants' ratings may reflect not only stimulus sensitivity, but also expectations, priors and other non-perceptual information that affects perceptual awareness ratings (see: Aru, Tulver, & Bachmann, 2018; Kouider, De Gardelle, Sackur, & Dupoux, 2010; Anzulewicz, Hobot, Siedlecka, & Wierzchoń, 2019). Those expectations, priors or other information are not necessarily in line with the actual stimuli; thus, sensitivity does not provide a proper approximation for such cases. Given such a possibility, it is also challenging to disentangle between rating bias and actual change in the visual experience. A recent development in the computational models of awareness ratings (see. e.g. King & Dehaene, 2014; Fleming, 2019; Rausch & Zehetleitner, 2018) may allow addressing those disadvantages. However, future studies applying those models may require research protocols adjustment (e.g. involving catch trials or detection and identification conditions). Finally, the limitations are also related to the very measure of awareness applied in the study. The PAS response categories are subject to individual interpretations of the participants. They had to decide on which aspect of the stimulus the ratings are to be based – e.g. brightness, contrast, spatial frequency (2015; Bachmann, 2012). Thus, we should ensure that the subjective scale is measuring the availability of the stimuli feature of interest. Importantly, this limitation refers to all versions of the visibility rating scales applied in the study, that is CS, PAS and C-PAS.

To sum up, we have compared three scales of awareness: PAS, C-PAS and CS. None of them shows a clear advantage in scale sensitivity. We observed differences in rating frequencies and the lowest rating accuracy between CS and other scales. We concluded that visibility ratings are often biased, and thus, the scale sensitivity may not be optimal. Based on the literature, we hypothesised that CS is associated with a more conservative response strategy than PAS. Based on the additional analyses, we argued there is an advantage of using C-PAS over CS. The scale allows to use an additional variability of judgment within PAS categories and thus it may enable more fine-grained measurement of visibility at near-threshold conditions. We argue that C-PAS may be useful, especially in cases when a more fine-tuned measure of near-threshold perception is needed.

Acknowledgements

This work was supported by the National Science Centre, Poland HARMONIA grant given to Michal Wierzchon (2014/14/M/HS6/00911) and PRELUDIUM grant given to Anna Anzulewicz (2014/15/N/HS6/04214).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2019.102798.

[5] We thank the reviewer of the paper for suggesting such an experiment.

References

Anzulewicz, A., Hobot, J., Siedlecka, M., & Wierzchoń, M. (2019). Bringing action into the picture. How action influences visual awareness. *Attention, Perception, & Psychophysics,* 1–6.
Aru, J., Tulver, K., & Bachmann, T. (2018). It's all in your head: Expectations create illusory perception in a dual-task setup. *Consciousness and Cognition, 65*, 197–208.
Bachmann, T. (2012). How to begin to overcome the ambiguity present in differentiation between contents and levels of consciousness? *Frontiers in Psychology, 3*, 82.
Bachmann, T. (2015). Unmasking the pitfalls of the masking method in consciousness research. *Behavioral methods in consciousness research* (pp. 49–76). Oxford University Press.
Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(1), 260.
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., & Perner, J. (2004). Assumptions of a subjective measure of consciousness: Three mappings. In R. Gennaro (Ed.). *Higher order theories of consciousness* (pp. 173–199). Amsterdam: John Benjamins Publishers.
Fleming, S. M. (2019). Awareness reports as inference in a higher-order state space. arXiv preprint arXiv:1906.00728.
Hake, H., & Garner (1951). The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology, 42*(5), 358.
Hayes, M. H. J., & Peterson, D. G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin, 18*, 98–99.
King, J. R., & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1641), 20130204.
Kouider, S., De Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*(7), 301–307.
Lange, N. N. (1893). *Psychological investigations: The law of perception and the theory of voluntary attention.* Odessa: District Military Press In Russian.
Lau, H. C. (2007). A higher order Bayesian decision theory of consciousness. *Progress in Brain Research, 168*, 35–48.
Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d', response-specific meta-d', and the unequal variance SDT mode. In S. M. Fleming, & C. D. Frith (Eds.). *The Cognitive Neuroscience of Metacognition* (pp. 25–66). Springer.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
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*(4), 700–708.
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.
Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*, 1–23.
Rausch, M., & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consciousness and Cognition, 28*, 126–140.
Rausch, M., & Zehetleitner, M. (2016). Visibility is not equivalent to confidence in a low contrast orientation discrimination task. *Frontiers in Psychology,* 7.
Rausch, M., & Zehetleitner, M. (2018). Modelling visibility judgments using models of decision confidence. https://doi.org/10.31219/osf.io/7dakz.
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849–860.
Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328–350.
Rosch, E. H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192.
Sandberg, K., & Overgaard, M. (2015). Using the perceptual awareness scale (PAS). *Behavioral methods in consciousness research* (pp. 181–198). Oxford University Press.
Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition, 19*(4), 1069–1078.
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.
Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(3), 367–395.
Werner, H., & Wapner, S. (1952). Toward a general theory of perception. *Psychological Review, 59*(4), 324–338.
Windey, B. (2014). Is consciousness graded or dichotomous? A level of processing account. Unpublished PhD thesis.
Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In M. Overgaard (Ed.). *Behavioural methods in consciousness research* (pp. 21–46). Oxford: Oxford University Press.
Wierzchoń, M., Asanowicz, D., Paulewicz, B.i., & Cleeremans, A. (2012). Subjective measures of consciousness in artificial grammar learning task. *Consciousness & Cognition, 21*(3), 1141–1153. https://doi.org/10.1016/j.concog.2012.05.012.
Wierzchoń, M., Paulewicz, B., Asanowicz, D., Timmermans, B., & Cleeremans, A. (2014). Different subjective awareness measures demonstrate the influence of visual identification on perceptual awareness ratings. *Consciousness and Cognition, 27*, 109–120.
Zehetleitner, M., & Rausch, M. (2013). Being confident without seeing: What subjective measures of visual consciousness are about. *Attention, Perception, & Psychophysics, 75*, 1406–1426.