Consciousness and Cognition 21 (2012) 1141–1153

Contents lists available at SciVerse ScienceDirect

Consciousness and Cognition

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

Subjective measures of consciousness in artificial grammar learning task

Michał Wierzchoń [a,c,*], Dariusz Asanowicz [a], Borysław Paulewicz [b], Axel Cleeremans [c]

[a] *Institute of Psychology, Jagiellonian University, Krakow, Poland*
[b] *Warsaw School of Social Science and Humanities, Faculty in Katowice, Katowice, Poland*
[c] *Consciousness, Cognition and Computation Group, Université Libre de Bruxelles, Brussels, Belgium*

A R T I C L E I N F O

Article history:
Received 10 September 2011
Available online 22 June 2012

Keywords:
Subjective measures
Wagering
Confidence
Feeling of warmth
Artificial grammar learning
Implicit learning
Consciousness
Higher order thoughts

A B S T R A C T

Consciousness can be measured in various ways, but different measures often yield different conclusions about the extent to which awareness relates to performance. Here, we compare five different subjective measures of awareness in the context of an artificial grammar learning task. Participants ($N = 217$) expressed their subjective awareness of rules using one of five different scales: confidence ratings (CRs), post-decision wagering (PDW), feeling of warmth (FOW), rule awareness (RAS), and continuous scale (SDS). All scales were equally sensitive to conscious knowledge. PDW, however, was affected by risk aversion, and both RAS and SDS applied different minimal criteria for rule awareness. CR seems to capture the largest range of consciousness, but failed to indicate unconscious knowledge with the guessing criterion. We close by discussing the theoretical implications of scale sensitivity and propose that CR's unique features enable (in conjunction with RAS and FOW) a finer assessment of subjective states of awareness.

1. Introduction

How "implicit" is implicit learning? How "implicit" is implicit perception? The extent to which implicit knowledge is conscious continues to be the central issue in the broad domain of implicit cognition (see e.g. Berry, 1995; Destrebecqz & Cleeremans, 2001; Pothos, 2007). In contrast to initial accounts (Reber, 1967, 1989; Seger, 1994), multiple studies now suggest that participants are at least partially aware of knowledge acquired through implicit learning procedures (Dienes & Seth, 2010; Higham, Vokey, & Pritchard, 2000; Perruchet & Pacteau, 1990; Shanks & St. John, 1994). It has also been suggested that subliminal priming could be driven, at least in part, by conscious perception (Holender, 1986; Kouider & Dupoux, 2004; Pessoa, Japee, & Ungerleider, 2005). However, the extent of the contribution of explicit knowledge or perception observed in those studies seems to depend on a measure of awareness that was used. The challenge of identifying the best measure of consciousness is thus critically important to any study of implicit cognition, and more generally, to investigate differences between conscious and unconscious processing. The main goal of this paper is to investigate the measures of awareness that are currently used, and to propose how they should be used to increase their sensitivity.

In the following section, we briefly describe objective and subjective measures of consciousness. Then, we present and compare different scales of subjective consciousness, discussing why different measures may yield different conclusions about the extent to which awareness relates to performance.

* Corresponding author at: Institute of Psychology, Jagiellonian University, Al. Mickiewicza 3, 31-120 Krakow, Poland.
E-mail address: michal.wierzchon@uj.edu.pl (M. Wierzchoń).

1053-8100/$ - see front matter
http://dx.doi.org/10.1016/j.concog.2012.05.012

1.1. Subjective and objective measures of consciousness

Estimating level of consciousness requires sensitive measures that make it possible to determine to what extent a person is aware of what was acquired or perceived. Multiple measures of consciousness have been proposed in that context. Most often, participants are just asked whether they are aware of being in a certain mental state and their unassisted verbal reports are then collected. Such introspective measures are broadly used in implicit cognition studies,[1] often revealing dissociations between awareness and performance, e.g. in the artificial grammar learning or subliminal priming studies. In the artificial grammar learning paradigm (Reber, 1989), participants are asked to memorize strings of letters which, unknown to them, are based on a set of rules that determine the sequence of successive letters. In a second phase of the task, participants are informed about the occurrence of the rules, and then asked to classify new strings as a function of whether or not they are "grammatical". Interestingly, participants usually report neither awareness of the learning process nor of the acquired knowledge. Nevertheless, their performance on indirect tests of knowledge indicates that learning occurred because their classification accuracy exceeds chance. Thus, dissociation between awareness and performance is observed. Similarly, in the case of subliminal priming, rapid and masked presentation of stimuli prior to targets influences participants' behavior (e.g. a facilitation of processing for congruent visible targets is observed), even though participants claim not to have seen the stimuli (Merikle, Smilek, & Eastwood, 2001). Do such results imply that participants are indeed unaware what was acquired or perceived? Not necessarily—it could be the case, for instance, that participants are partially aware, yet unable or unwilling to express their knowledge verbally, for instance, because of a reluctance to express knowledge held with low subjective certainty, or because the aspect of awareness they base their responses on is difficult to verbalize. Verbal reports have thus been often criticized as being inherently subjective and insensitive measures of consciousness, both in implicit learning (Kinder, Shanks, Cock, & Tunney, 2003; Shanks & St. John, 1994) and subliminal perception studies (Holender, 1986; Pessoa et al., 2005).

It is essentially for this reason that many authors prefer using objective measures to assess awareness. With such measures, participants are usually asked to perform some detection or discrimination task, in which consciousness is indicated by better-than-chance performance. This is in line with the theoretical assumption that awareness might be investigated by means of accessibility in a manner that is not necessarily related to a phenomenological experience (see A-consciousness vs. P-consciousness distinction, Block, 1995). For example, in artificial grammar learning studies, fragment-recognition or fragment-completion tasks are usually used as objective measures of awareness (Jamieson & Mewhort, 2009; Perruchet & Pacteau, 1990; Vokey & Brooks, 1992). Thus, in contrast with the dissociations typically found with subjective measures, objective methods often suggest that the knowledge acquired in artificial grammar learning or other paradigms investigating implicit learning is (at least partially) consciously accessible (Perruchet & Pacteau, 1990; Shanks & Channon, 2002; Shanks, Rowland, & Ranger, 2005; Shanks & St. John, 1994). Signal detection theory (Macmillan & Creelman, 1991) is usually applied to analyze the results of objective measures, and stimulus detection/discrimination sensitivity (d') is computed as an index of conscious access (Gaillard, Vandenberghe, Destrebecqz, & Cleeremans, 2006; Holender, 1986; Snodgrass, Bernat, & Shevrin, 2004).

The higher level of consciousness indicated by objective measures suggests that these measures are more sensitive. However, objective measures may be criticized for failing to capture the very central aspect of consciousness, namely subjective experience (or, in other words, measuring A-consciousness without P-consciousness). From this perspective, the fact that more knowledge is revealed through objective measures does not indicate greater sensitivity to the contents of awareness, but merely the effectiveness of information processing (Lau, 2008). Of course, it may be assumed that subjective experience is not essential to study conscious access at all (see e.g. worldly discrimination theories as described by Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008), or that accessibility of knowledge and subjective experience related to this knowledge should be somehow separated (Block, 1995, 2011). However, it has recently been argued that subjective experience is the critical feature of consciousness that should be always taken into account (Dienes & Perner, 2004; Lau, 2008; O'Regan, Myin, & Noë, 2005). The importance of this assumption is obvious in the case of implicit learning studies, where accuracy of recognition is considered as an objective measure. It may be that the recognition is driven by a feeling of familiarity or other, possibly unconscious process (see Pothos, 2007). If that is so, participants may simultaneously exhibit some knowledge with an objective test and declare being subjectively unaware of that knowledge. Therefore, identifying a measure that is both sensitive and reflective of participants' subjective experience appears to be essential, especially if one assumes that consciousness should be investigated by means of accessibility associated with a direct subjective experience. This is in line with recent theoretical proposals that those two aspects of consciousness cannot be separated (see e.g. Cohen & Dennett, 2011; Kouider, de Gardelle, Sackur, & Dupoux, 2010). Following this reasoning, we assumed that any measure of consciousness should capture both aspects. Such requirement seems to be fulfilled by subjective scales of awareness (Dienes & Perner, 2004; Dienes & Seth, 2010; Persaud, McLeod, & Cowey, 2007; Ramsøy & Overgaard, 2004). In the following section, we provide an overview of such methods.

[1] And in other studies on consciousness, e.g. on neural correlates of consciousness (see: Rees et al., 2002).

1.2. Subjective measures of consciousness

The most typical subjective measure is probably *confidence ratings* (CRs), whereby participants are asked to indicate the extent to which they are confident in their responses or decisions. Different variations of this scale have been proposed, ranging from binary yes–no confidence judgments to graded scales that comprise several intermediate steps between extremes such as ''guessing'' and ''fully certain'' (Cheesman & Merikle, 1986; Dienes & Perner, 2004; Dienes & Seth, 2010). CR might be used, at least in the case of perception studies, to examine the objective accessibility of the stimuli, for instance reflecting the strength of the stimulation (Lau & Maniscalco, 2010). However, in the majority of recent papers, especially those that used CR to study implicit learning (e.g. Dienes & Perner, 2004; Dienes & Seth, 2010), authors assumed that the scale reflects participants' metacognitive judgements on how certain they are of the responses or decisions they had expressed. If participants can indeed access information concerning their subjective certainty, higher confidence ratings should imply consciousness of the process and/or the knowledge that underlies the decisions. This might be analyzed through correlation between performance and awareness ratings. It is assumed that high correlation indicates participants' awareness of knowing (zero-correlation criterion – see Dienes, Altmann, Kwan, & Goode, 1995). Such rationale has been widely applied, for instance, in artificial grammar learning studies, in which participants are asked about their confidence on forced-choice decisions in the test phase. Significant correlation between confidence ratings and accuracy of classifications observed in most of the studies indicates that explicit knowledge influences task performance (cf. Dienes & Seth, 2010).

It seems worth noting that confidence ratings themselves express participants' subjective experience of accessibility (i.e., judgment knowledge) in a way that is not necessarily accompanied by full access to the knowledge itself (i.e. structural knowledge; see Dienes & Scott, 2005 for details on judgment/structural knowledge distinction). In other words, at least theoretically, participants may express high confidence in their decisions although the structure of the grammar underlying the judgments is only partially accessible. This is to some extent unavoidable, if one wants to investigate subjective accessibility of knowledge. Nevertheless, a correlation between confidence and task performance accuracy suggests that judgment knowledge is in fact associated with structural knowledge (i.e., participants may not be fully aware of the abstract rule, but they need to know at least fragmentary rules to judge their confidence level in line with classification accuracy – cf. Dienes & Scott, 2005).

In this paper, we focus our investigations on subjective judgment knowledge, assessing the extent to which it also reflects objective access to knowledge, as investigated with classification accuracy. Confidence ratings seem to fulfill such criteria, being both more sensitive than verbal reports and, contrary to objective measures, rooted in subjective experience. However, the scale has been criticized as being counter-intuitive for participants (Persaud et al., 2007). It seems likely that participants underrate their confidence rating (Tunney & Shanks, 2003) or just fail to indicate conscious knowledge because they have no motivation to do so (Persaud et al., 2007). The scale has also limited use in the case of groups with low introspection abilities (Ruffman, Garnham, Import, & Connolly, 2001). Finally, it seems rather difficult to rate the confidence level in case of tasks characterized by a low level of certainty, like the implicit learning procedures. This is because participants are not confident at all of any decision made (thus a subjective criterium to judge high on an awareness scale may be difficult to reach).

To address these problems, *post-decision wagering* (PDW) scale has been proposed (Persaud et al., 2007; Ruffman et al., 2001). With this scale, participants wager a certain amount of money (or artificial tokens) on their responses or decisions, trying to maximize cash earnings. Analogically to the confidence ratings scale, it is assumed that higher bets reflect higher level of conscious accessibility. That is because participants who are constantly trying to avoid losses should use higher stakes only when being aware of which response or decision is appropriate. An advantage of the PDW scale is that participants do not need to introspect on their certainty. They simply perform the task which by definition requires consciousness, therefore, judgment knowledge should be closely associated with structural knowledge if one wants to avoid losses. That may be a reason why the scale is assumed to measure awareness more directly than e.g. CR. Importantly, because the wagering is both more intuitive and more interesting, participants' motivation and engagement in the task seems to be higher than in case of confidence ratings. Post-decision wagering results were also analyzed in the context of implicit learning. Similarly to confidence ratings, a correlation between accuracy and awareness was observed (Dienes & Seth, 2010).

However, despite being more intuitive and interesting for participants, post-decision wagering is vividly criticized, mainly because its results seem to be strategy dependent, i.e., ratings are influenced by the loss-aversion strategy (Fleming & Dolan, 2010). It was suggested that the scale might be less sensitive than confidence ratings, as participants may in fact acquire some explicit knowledge, but do not express its availability due to the risk associated with the wagering procedure (Dienes & Seth, 2010). Finally, the scale seems to be used in quite dichotomous manner, as compared to confidence ratings, i.e., participants tend to use only extremes of the scale but not all possible ratings (Sandberg, Timmermans, Overgaard, & Cleeremans, 2010), possibly because this strategy allows to maximize cash earnings. This seems to be a particularly strong disadvantage, if one assumes (or would like to test) the graded access to consciousness (Cleeremans, 2011; Overgaard, Rote, Mouridsen, & Ramsøy, 2006; Sergent & Dehaene, 2004).

Attempting to find a method that would share the positive sides of both confidence ratings and post-decision wagering, we have proposed to use a new measure of subjective consciousness namely the *feeling of warmth* scale (FOW; Metcalfe, 1986). The scale was originally proposed as a measure of intuition. It was used to assess participants' intuitions reflecting the proximity of an insight in problem solving (Metcalfe, 1986). The extremes of the scale are usually described as ''cold'' and ''warm'' (like in the well-known children's game), usually with some intermediate points (e.g. ''chilly''). However, it seems that subjective feeling of solution proximity should be based, at least to some extent, on awareness of knowledge (although the feeling may also be the effect of partial access to emotions supporting the correct decisions; Damasio, 2000; see also Bierman, Destrebecqz, & Cleeremans, 2005). Thus, we assumed that the feeling of warmth may in fact represent partial accessibility of knowledge at low levels of information processing (cf. partial awareness hypothesis, Kouider et al., 2010). In the case of artificial grammar learning, we propose to ask participants about their feeling of classification accuracy. Importantly, it does not seem to be necessarily mediated by participants' ability to describe the structural knowledge verbally, as participants are only being asked about the feeling of knowledge accessibility. This seems important, since participants may be (at least partially) aware of regularities even though they are not able to verbalize that knowledge (cf. Kouider et al., 2010). In other words, we assume that the feeling of warmth scale may be a more direct, and thus a more sensitive measure of judgment knowledge than confidence ratings and post-decision wagering scales, because it is focused on participants' direct experience of accuracy related feeling rather than on the accessibility of the knowledge that underlie this feeling. Feeling of warmth thus allows us to concentrate on judgment knowledge to a greater extent.

It has been also proposed that the dissociation between structural and judgment knowledge may be investigated with a few other subjective measures already applied in studies on perceptual awareness. Interestingly, they have not yet been used in the context of implicit learning. The *perceptual awareness scale* (Ramsøy & Overgaard, 2004, PAS) and the *continuous scale* (*Sergent–Dehaene scale*, SDS; Sergent & Dehaene, 2004) seem to be the most prominent examples. Both scales are purely introspective measures by which participants are asked to indicate the level of object visibility. The possible ratings on both scales vary between "no experience" and "absolutely clear image". The difference is that in the perceptual awareness scale, two intermediate ratings are defined, whereas in the continuous scale, participants can choose any point on the continuum in between the scale extremes. Interestingly, the perceptual awareness scale was proposed based on participants' unassisted responses, so the scale should be quite intuitive for participants. Both scales revealed significant correlation between the awareness and perceptual identification task performance (Sandberg et al., 2010; Sergent & Dehaene, 2004). However, the detailed results differ significantly. The perceptual awareness scale ratings seemed to be linearly correlated with the identification task performance suggesting the gradual character of conscious perception (Overgaard et al., 2006; Sandberg et al., 2010), whereas the continuous scale revealed dichotomous results pattern (at least in case of attentional blink experiments), interpreted as an argument for all-or-none character of conscious perception (Sergent & Dehaene, 2004). To analyze those differences, the results of both scales should be directly compared in one study.

The content of consciousness investigated with both perceptual awareness and continuous scales seems to be slightly different than in case of confidence ratings or post-decision wagering scales. This is due to the fact that on both scales, participants report their conscious experience, but not their metacognitive judgement on that experience (hence the first two scales report the direct experience whereas the second two scales report judgment related to that experience).

It seems interesting to ask whether the modified versions of such scales may be used in implicit learning studies, and to compare them with confidence ratings and post-decision wagering. We propose that "visibility" could be replaced with "rule accessibility" for that purpose (*rule awareness scale*, RAS). Thus, we suggest to ask participants to assess the accessibility of knowledge about rule structure underlying the artificial grammar learning task. This, however, opens a question of what exactly both scales measure. The original versions of both scales aimed to measure (at least as it is usually suggested) participant's knowledge about a stimulus (which seems to be equivalent to structural knowledge in implicit learning). In the context of artificial grammar learning, they seem to assess judgment knowledge or, in other words, the subjective accessibility of structural knowledge, because the structure of the knowledge is never explicitly presented, and thus is not directly accessible.

1.3. Why do the results differ?

In the previous paragraph, we have outlined important differences between subjective measures of consciousness. They seem to differ with respect to the level of metacognitive judgment associated with the responses, the gradient of consciousness experience observed with the scale, the influence of response strategies, their intuitiveness and participants' motivation associated with the task performance. However, despite the differences between the scales, it seems that all of them attempt to measure the very same thing – i.e. whether the person is aware of knowing. This is especially the case in the artificial grammar learning task, in which the rule awareness scale and the continuous scale should measure judgment knowledge or subjective accessibility, but not structural knowledge itself. It seems that the rule awareness scale and continuous scale differ with regard to the criterion allowing to rate high on both scales, because both scales ask not only about fragmentary rules influencing the subjective certainty of a certain decision (partial awareness of rules) but also about the complex rules.

Thus, if all scales measure judgment knowledge, albeit not exclusively, the level of conscious access observed in a certain task (e.g. artificial grammar learning) should be comparable, regardless of which subjective measure we would apply. Interestingly, the performance in the very same task was rarely analyzed by comparing results from different scales. More importantly, it seems that when two or more scales were compared in context of one task, slightly different results were observed depending on the measure applied (Dienes & Seth, 2010; Sandberg et al., 2010). As it seems unlikely that the level of consciousness in the particular task differs depending on the scale used, there must be another explanation of those differences. They seem to be caused by differences in scale sensitivity and exhaustiveness. The performance in case of all scales might be analyzed with regard to differences in the range of consciousness assessed by each scale (scale sensitivity) and in the level of consciousness reflected with points on the scale (scale exhaustiveness). We propose that the correlation criterion and the guessing criterion analyses (Dienes et al., 1995) might be adopted for this purpose. The guessing criterion analysis investigates the level of performance associated with the lowest scale point. In other words, it indicates to what extent participants are able to control their behavior, even when they claim not to have any conscious knowledge. It is usually assumed that such above chance performance level for the lowest scale rating is a result of implicit processing. However, it seems plausible that the lowest point of each scale may in fact represent different levels of conscious access. This may be for example due to the risk aversion strategy, which may motivate participants to issue low ratings even if they are partially aware of the acquired knowledge. Another reason may be that participants are sticking to a subjective criterion content that is useful for lower-level experience judgment (in artificial grammar learning this might be partial knowledge of rules underlying subjective certainty), but not useful at the higher-level of subjective experience (complex rule awareness).

The aim of the present study is to compare the results of five subjective scales of consciousness in the context of artificial grammar learning task. We will investigate the differences in scale parameters and discuss the differences in scale sensitivity.

2. Method

2.1. Participants

Two hundred and seventeen undergraduate students from Jagiellonian University voluntarily participated in the study in exchange for course credits. Participants were randomly assigned to one of ten conditions: 5 scales, administered either before or after string classification.

2.2. Materials and procedure

2.2.1. Artificial grammar learning task

The classical Artificial Grammar Learning (AGL) procedure was used (Reber, 1967), in a two grammar design (Dienes & Altmann, 1997). Training and test letter strings were adopted from Dienes and Scott (2005). Design of the AGL task was identical to the one used by Dienes and Seth (2010). In the first phase of the task, participants were asked to memorize letter strings from one of two grammars (half of the participants memorized the strings from grammar 'A' and the other half memorized strings from grammar 'B'). Fifteen strings were presented three times during the acquisition phase. Afterwards, participants were informed that the order of letters followed a complex set of rules. In the second phase of the task, participants were asked to classify sixty new strings (equal mix of 'A' and 'B' grammar strings), as regular (grammatical) or irregular (ungrammatical). Participants were asked to base their decisions on the rules that organized the material in the first part of the experiment (i.e. participants who had been exposed to grammar 'A' in the first phase should classify as regular the strings from grammar 'A', whereas participants who had been exposed to grammar 'B' should classify as regular the strings from grammar 'B'). Half of the test strings were built according to the same set of rules that defined the material in the acquisition phase, thus chance level equals .5. Similarly to Dienes and Scott (2005) study, we used a fixed presentation order both in the acquisition and in the test phases (half of the participants were presented with the strings in one order; the other half in the reverse order). The duration of string presentation in the acquisition phase was fixed to 5 s. In the test phase, participants saw each string for 2.5 s[2]. Then, the screen with decision options (regular vs. non-regular) followed (or preceded) by the screen with one of the subjective scales (depending on the conditions as described below) were presented without further time limits (so in fact participants could classify strings at their own pace).

The experiment was run on PC computers with standard 17″ monitors using DMDX software, except for the continuous scale (SDS) condition, which was, due to technical reasons, carried out with software specifically designed for this study. Parameters of presentation in this condition (except SDS scale itself) were identical to other experimental conditions.

2.2.2. Measures of consciousness

With each decision on classification in the test phase, participants were asked to rate their awareness of rules that organized the material using one of five subjective measures of consciousness:

- Confidence rating scale (CR) – participants were asked to rate their confidence in classification decision with 4 point scale (where '1' is "I am guessing"; '2' – "I am not confident"; '3' – "I am quite confident"; and '4' – "I am very confident") with each decision on string regularity.
- Post-decision wagering scale (PDW) – participants were asked to wager one of four possible stakes (20, 40, 60 or 80 PLN[3]) on each classification decision, expressing their certainty in this decision. They were informed that when they bet on the correct response, they will gain the amount on stake (and it will be added to their account). However, if the wager is inaccurate they will lose the stake (and it will be subtracted from their account). Participants started the classification phase with 200 PLN on their account.

[2] We introduced fixed presentation time in the classification phase in order to ensure participants' engagement in the task.

[3] 1 PLN equals around 0.25 EUR, we recalculated the stakes from Sandberg et al. (2010) study for PLN.

- Feeling of warmth scale (FOW) – in this condition participants were informed that they should base their scale ratings on subjective feelings or intuition related to each classification decision (thus a certainty that bases rater on a subjective feeling than the rational explanation that may be associated with the confidence ratings or decision wagering). Participants rated their feelings about classification accuracy with 4 point scale (where '1' is "cold"; '2' – "chilly"; '3' – "warm"; and '4' – "hot").
- Rule awareness scale (RAS) – the scale was proposed analogously to the perceptual awareness scale (PAS, see Ramsøy & Overgaard, 2004; Sandberg et al., 2010). Because we used the artificial grammar learning task that measures participants' knowledge of the rules, we modified the scale to investigate subjective awareness of knowledge (judgment knowledge on rule accessibility – see also the introduction for the discussion on differences in types of experience associated with different types of the content measured by the scales). Participants were asked to indicate whether they are aware of the rule underlying their classification decision on 4 point scale (where '1' is "I do not have any clue about the rule"; '2' – "I have a glimpse of the rule"; '3' – "I think that I know what is the rule"; '4' – "I know the rule").
- Continuous scale (*Sergant–Dehaene Scale*, SDS) – this scale was built analogously to the continuous scale proposed by Sergent and Dehaene (2004). Similarly to the RAS scale described above, we changed the scale description to adapt it to the measurement of rule awareness. Participants were asked to indicate whether they are aware of the rule underlying their classification decision, rating their rule awareness on a continuous scale represented by a horizontal bar. The bar's extremes were labeled: "I do not have any clue about the rule" on the left side and "I know the rule" on the right side of the screen. Participants were asked to move a cursor on the scale to mark their response. The initial position of the cursor was random.

Because the feeling of warmth scale responses have been previously usually collected before participants responded (cf. Metcalfe, 1986), we have additionally counterbalanced the response – awareness rating order for all scales to control for the influence of this factor (i.e. decision on classification of each subsequent string was either followed or preceded by the awareness rating expressed with CR, PDW, FOW, RAS or SDS). Because we did not observed any significant effects of Order in the initial analyses (see Section 3), these results are not discussed further in the discussion section.

3. Results

3.1. Artificial grammar learning task accuracy

In the first step of the analysis, we compared AGL accuracy depending on which awareness scale was used in the classification task. The results were analyzed with one-way ANOVA with Scale (five levels: CR, PDW, FOW, RAS or SDS) as a between-subjects factor. The main effect of Scale was not significant ($F(4,213) < 1$), hence there is no evidence that accuracy differs between conditions. Participants were able to discriminate grammars accurately to the same extent in all conditions. Classification accuracy in each condition differed significantly from chance level (i.e. .5). Overall accuracy for all conditions was .67 (SE = .008). Accuracy for all conditions assessed through *t* tests to the chance level equals respectively: .67 (SE = .021) for CR scale ($t(33) = 8.03$; $p < .001$); .67 (SE = .018) for PDW scale ($t(42) = 9.02$; $p < .001$); .69 (SE = .019) for FOW scale ($t(38) = 10.12$; $p < .001$); .67 (SE = .019) for RAS scale ($t(39) = 8.82$; $p < .001$); and .66 (SE = .015) for SDS scale ($t(61) = 10.64$; $p < .001$).

3.1.1. Awareness ratings distribution

In the second step of the analysis, we compared the distribution of awareness rating responses for different scales in the classification task. Thus, these results do not take into account the accuracy, but the number of responses recorded for each point of each scale. This analysis shows how participants used the scales and whether each of the scales was used in a graded or all-or-none manner. To compare the results of the SDS scale with other scales, we categorized continuous responses into four bins (by splitting the continuous scale into the 4-score-based bins).

To estimate the effect of the Order factor, we compared two linear regression models with number of observations as the dependent variables, namely a full model with Order (2), Scale (5) and Awareness Rating (4), and a simpler model with two factors (Scale x Awareness Rating), excluding the Order factor. The inclusion of the Order factor resulted in a non-significant improvement in fit ($F(20) = 1.076$, $p = .37$), and it was thus excluded from further analyses. The results were analyzed by means of a linear regression with Scale (5) and Awareness Rating factor (4) as predictors (see Table 1). Number of responses was calculated for each subject x rating. The overall model fit was good ($\chi^2(19) = 171.13$, $p < .0001$).

We used standard, treatment contrast coding, and so coefficients Awareness Rating 2, 3 and 4 estimate CR rating effects, i.e., the differences in response frequency between the respective ratings and rating 1 within CR scale. Scale coefficients estimate the differences in rating 1 frequency between CR and each of the remaining scales. Interactive effects (e.g., Awareness Rating 2: PDW) represent the differences between rating effects for the respective scales and the same rating effects for the CR scale, providing a concise representation of the differences in response distribution patterns across the scales. As can be seen in Fig. 1 and in Table 1, the distribution of responses was different for CR and FOW as compared to the other scales. Participants seemed to avoid the lowest ratings on CR and FOW scales, but used them in the case of the PDW, RAS and SDS scales. The regression coefficients show that participants tend to use middle ratings ('2' and '3') for CR and FOW scales.

Table 1
Regression coefficients for the linear regression model for response distribution.

N = 218 # observations 788	Coefficient	SE	*t*	*p*
Intercept (scale = CR, rating = 1)	8.32	1.99	4.19	<.001***
Awareness Rating 2	13.82	2.75	5.03	<.001***
Awareness Rating 3	13.59	2.77	4.91	<.001***
Awareness Rating 4	2.18	2.89	0.75	.451
Scale PDW	14.05	2.61	5.39	<.001***
Scale FOW	−1.73	2.75	−0.63	0.528
Scale RAS	10.59	2.71	3.91	<.001***
Scale SDS	14.71	2.49	5.92	<.001***
Awareness Rating 2: PDW	−19.32	3.65	−5.29	<.001***
Awareness Rating 3: PDW	−24.67	3.67	−6.71	<.001***
Awareness Rating 4: PDW	−14.48	3.75	−3.86	<.001***
Awareness Rating 2: FOW	−0.16	3.78	−0.04	.967
Awareness Rating 3: FOW	3.06	3.80	0.81	.421
Awareness Rating 4: FOW	2.92	3.93	0.74	.458
Awareness Rating 2: RAS	−8.82	3.74	−2.35	.019*
Awareness Rating 3: RAS	−17.10	3.79	−4.51	<.001***
Awareness Rating 4: RAS	−13.27	4.13	−3.21	.001**
Awareness Rating 2: SDS	−20.14	3.47	−5.80	<.001***
Awareness Rating 3: SDS	−20.17	3.49	−5.77	<.001***
Awareness Rating 4: SDS	−11.38	3.61	−3.15	.002**
Likelihood ratio χ^2 (19) = 171.13, $p < .0001$				

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

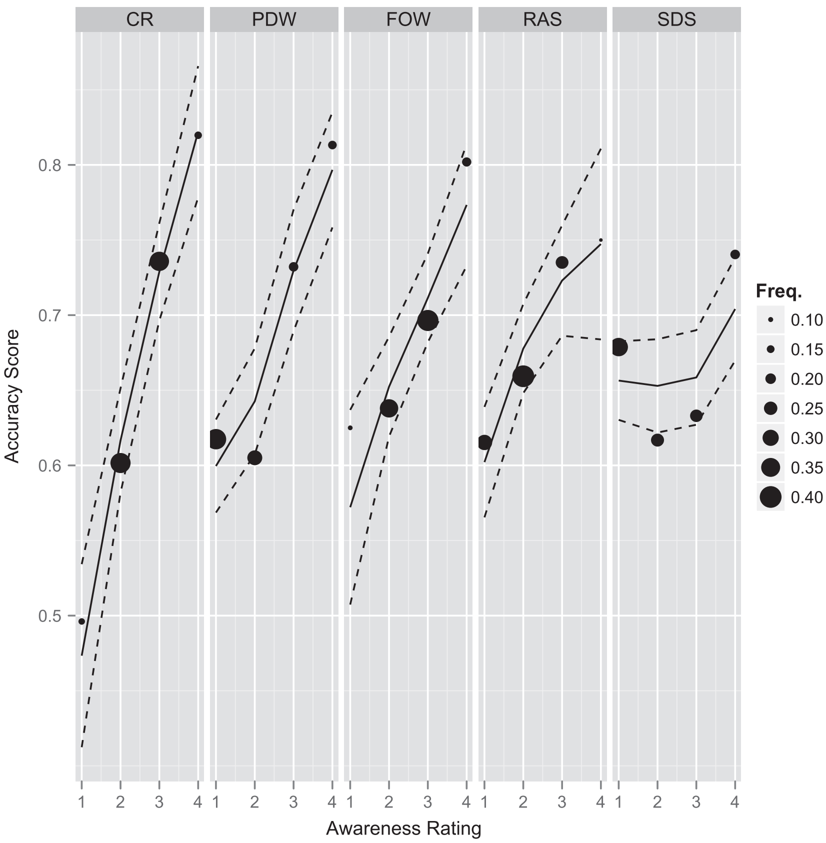

Fig. 1. The average accuracy predicted by Awareness Rating for each Scale. Solid lines represents fitted values based both on fixed and random effects. Dotted lines represents the 95% predictive intervals. Size of dots represents the distribution of responses (frequency within each condition).

A different pattern, a tendency to use the lowest ratings, was observed in case of the RAS scale (where participants tended to use the rating '2'), and was even more pronounced for PDW and SDS scales (where participants tended to use the lowest

Table 2
Regression coefficients for the logistic regression mixed model for accuracy.

$N = 218$ # observations 13,022	Coefficient	SE	z	p
Scale CR	−0.033	0.133	−0.250	.803
Scale PDW	0.301	0.101	2.989	.003**
Scale FOW	0.352	0.128	2.748	.006**
Scale RAS	0.479	0.110	4.360	<.001***
Scale SDS	0.558	0.088	6.347	<.001***
Scale CR * Awareness Rating	0.530	0.062	8.485	<.001***
Scale PDW * Awareness Rating	0.405	0.048	8.505	<.001***
Scale FOW * Awareness Rating	0.304	0.056	5.431	<.001***
Scale RAS * Awareness Rating	0.263	0.057	4.587	<.001***
Scale SDS * Awareness Rating	0.132	0.039	3.370	<.001***
Random effect of subject (intercept): variance = 0.26 (SD = 0.51)				
Likelihood ratio χ^2 (9) = 213.83, $p < .0001$				

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

Table 3
Differences between intercepts (row variables subtracted from the column variables).

	PDW	FOW	RAS	SDS
CR	0.33*	0.39*	0.51**	0.59***
PDW		0.05	0.18	0.26
FOW			0.13	0.21
RAS				0.08

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

ratings). These results are important for further analysis of scale sensitivity by means of the guessing criterion (see below), and also suggest that response strategies for various scales can be different. For example, depending on scale, participants might tend to use a strategy that promotes either the lowest or middle ratings.

3.1.2. Awareness and accuracy

We now turn to accuracy scores, using the percentage of correct classifications for each Awareness Rating in each Scale condition as a dependent variable. This analysis allows us to describe the relation between awareness and accuracy by means of guessing and zero-correlation criteria. To compare the results of the SDS scale with other scales, we again categorized continuous responses into four bins.

Similarly to response distribution analysis, in order to estimate the influence of Order factor we compared two logistic mixed models, namely a full model with factors Order (2), Scale (5) and Awareness Rating (4), this time as a numerical predictor, and their interactions with a simpler model with two factors (Scale × Awareness Rating) excluding the Order factor. Again, the inclusion of the Order factor resulted in a non-significant improvement in fit (χ^2 (10) = 12.34, $p = .26$) thus it was excluded from the analyses and the results were analyzed by means of a logistic mixed model with Scale (5) and Awareness Rating (4) as predictors, and random subject specific intercepts, using separate intercepts and slopes parametrization to improve readability of regression coefficients. The Awareness Rating variable was centered on the lowest value (1), thus the first five coefficients in Table 2 estimate the guessing criterion for each scale. The overall model fit was good (χ^2 (9) = 213.83, $p < .0001$). As can be seen on Fig. 1 most of the aggregated data points fall nicely between the 95% predictive, implying that overdispersion is not an issue here.

3.1.2.1. Guessing criterion results. Accuracy scores, plotted separately for each point of each scale, are presented in Fig. 1 (as the dots' position on the *Y*-axis). To assess the scale sensitivity by means of the guessing criterion we centered Awareness Rating variable on the lowest possible value. Since logit of .5 equals 0, a significant positive regression intercept within a given scale condition implies that accuracy was above chance when the lowest rating was used. As can be seen from Table 2, accuracy scores for the lowest scale point differ significantly from the chance level for all the scales, with the exception of the CR scale.

Mean accuracy for the lowest point in each condition equaled respectively: .5, SE = .03 for CR scale; .62 (SE = .02) for PDW scale; .63 (SE = .03) for FOW scale; .62 (SE = .02) for RAS scale; and .68 (SE = .01) for SDS scale. To investigate if the scales differed with regard to the guessing criterion (estimated by the regression intercepts) we compared the intercepts by means of contrasts analysis (see Table 3).

Table 4
Differences between slopes (row variables subtracted from the column variables).

	PDW * Awareness Rating	FOW * Awareness Rating	RAS * Awareness Rating	SDS * Awareness Rating
CR * Awareness Rating	−0.12	−0.23**	−0.27**	−0.4***
PDW * Awareness Rating		−0.1	−0.14	−0.27***
FOW * Awareness Rating			−0.04	−0.17*
RAS * Awareness Rating				−0.13

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

The results show that CR anchor point differs from the anchor points of all the other scales, suggesting both that the guessing criterion was fulfilled for this scale only and that the general sensitivity (the anchor point) is higher for CR than for the other scales.

3.1.2.2. Zero-correlation criterion results. As expected, the classification accuracy scores increased with ratings for all scales (see Fig. 1). The regression analysis revealed significant Awareness Rating effects (i.e. slopes) for all the scales (see Table 2). To assess the exhaustiveness of the scales in more detail we compared the regression slopes by means of contrasts analysis (see Table 4).

The results indicate that CR slope is higher than the FOW, the RAS and the SDS slopes, suggesting higher exhaustiveness of this scale. We also observed a very small, although still significant slope for the SDS scale that was significantly lower than the slopes for the PDW and the FOW scales. These results suggest a somewhat lower exhaustiveness of the SDS scale.[4]

4. Discussion

The main goal of the present paper was to compare five subjective measures of awareness in the context of an artificial grammar learning task. As expected, all groups performed the AGL task above chance level, showing that participants acquired some knowledge over the time course of the task. Importantly for our scale comparisons, all participants appeared to have acquired similar knowledge in all conditions. Before we discuss the differences for the relation between awareness and accuracy for each subjective scale in details, let us take a closer look at the response distribution analysis.

4.1. Response strategies and the level of consciousness

The results of response distribution analysis are important, as they reflect the response strategies applied by participants. It is well known that subjective measures of consciousness may be influenced by loss aversion (Fleming & Dolan, 2010; Schurger & Sher, 2008) or risk aversion strategies (Dienes & Seth, 2010). When participants try to avoid the uncertainty associated with high ratings (e.g. high bets in PDW), they may choose lower ratings even if they in fact are aware of knowledge (just because they are not confident enough whether their responses will be correct). Such strategy might be observed in the response distribution analysis, i.e. in that case participants should tend to use the lowest scales ratings. However, our results revealed that participants tended to use the lowest ratings in case of three scales: PDW, SDS, and to some extent also RAS. This finding confirmed the previous concerns about PDW scale, showing that the measure might not be sensitive enough due to the effect of loss aversion strategy (Sandberg et al., 2010; Schurger & Sher, 2008. Interestingly, a similar pattern of results was observed for the SDS scale[5] and also, to some extent, for the RAS scale. Neither, to the best of our knowledge, have been used in the context of AGL, so there is no clear explanation for these results. However, it seems unlikely that loss aversion constitutes a reason for an overrepresentation of the lowest scale ratings in the case of those two scales, as participants did not risk anything in this case. Thus, we propose that the interpretation of the results might be a different decision criterion applied in the case of those two scales. As noted in the introduction, in SDS and RAS, participants are asked directly about their knowledge of the rules, whereas CR, FOW and PDW ratings may be rooted in the fragmentary knowledge that allows participants to classify a certain string. Notwithstanding, the criterion for high ratings (over "1") seems to be elevated in case of RAS and SDS, just because accessibility of the rules needs to be higher than in case of the other scales. This seems to resemble the partial awareness theory (Kouider et al., 2010), suggesting that participants may have only a partial access to a multileveled representation of

[4] One may wonder whether the binning did not introduce any artifacts, especially as the lowest rating on 4 score-based bins may not reflect the pure guessing criterion (1/4 of the scale was categorized as the lowest rating). To exclude this possibility we calculated another regression model with Scale and Awareness rating as factors with the SDS responses grouped into 20 bins (this was the way in which the program stored the data). Coefficients for SDS responses were very similar to the previous model, i.e., Intercept = 0.495, Awareness Rating = 0.179, both significant at the $p < 0.001$ level. Qualitatively results look the same – SDS is still the scale with the lowest slope and the highest intercept.

[5] It seems worth noting, that the results of SDS scale were aggregated to be comparable with the other scales results (i.e. the lowest rate reflects in fact 25% of possible judgements). More detailed analysis for the SDS scale only using 20-point categorization (similar to the one applied by Sergent & Dehaene, 2004), revealed that participants tend to use the lowest possible rate in 15.19% cases, which was the most frequent response (the second highest frequency was observed for middle rate (7.42%) and the third, for highest possible rate (6.56%)).

knowledge. In our case, CR, FOW and PDW could thus apply a lower criterion that reflects more partial access than both RAS and SDS. This interpretation is in line with CR and FOW results, for which different pattern of results was observed. Here, participants tended to use the intermediate values, which seems to indicate that the knowledge acquired in AGL is not fully accessible but also that the knowledge acquired is at least partially conscious. The accuracy analysis seems to confirm this interpretation. Regression coefficients reflecting the guessing criterion also suggest that the scales differ by means of the criterion that allows participants to rate high, e.g. the accuracy scores observed for the lowest SDS rating is almost the same that the one observed for the third FOW value (see Fig. 1).

Concluding, CR and FOW results scales seem to be more reliable than PDW when one would like to analyze the level of knowledge awareness in AGL. Although it assumes the same rating criterion, PDW seems to be influenced by loss aversion. Comparing RAS and SDS with the other scales, one needs to consider the difference in rating criterion between the scales.

The second important issue, which may be clarified by the response distribution analysis is the theoretical debate addressing the question whether the consciousness is a graded or dichotomous phenomenon (Cleeremans, 2011; Overgaard et al., 2006; Sergent & Dehaene, 2004). Following the partial awareness theory (Kouider et al., 2010) interpretation of graded access (i.e. that graded access results from partial access to the different level of knowledge representation), our results seem to confirm the graded account. Participants tend to use all possible ratings and not merely the scale extremes. Different scales seem to assume different rating criteria that also might be related with different levels of accessibility. Generally, it seems that participants experienced the AGL task as related to the partial awareness of rules, hence they tended to use low or intermediate scale ratings (depending on the rating criterion applied). The accuracy analysis for the awareness rating seems to confirm this interpretation.

4.2. Awareness and accuracy

The results revealed the significant and strong correlation between awareness (as measured with all subjective scales) and AGL task accuracy. Following zero-correlation criterion rationale (Dienes & Seth, 2010; Dienes et al., 1995), these results suggest that participants were at least partially aware of the knowledge acquired in the AGL. Participants classified strings more correctly when they subjectively experienced higher access of the knowledge representation. Interestingly, the interaction between the scale type and awareness rating was also observed, suggesting that the level of awareness assessed by the scale ratings differs depending on the measure used. As it seems improbable that participants experienced different level of artificial grammar learning task awareness depending on the scale type, we could assume that the most accurate method is the most sensitive one (cf. Dienes & Seth, 2010; Sandberg et al., 2010). Moreover, following our interpretations of judgment criterion and graded access, it seems also probable that each scale measures a different part of the spectrum (i.e. one scale measures e.g. the differences in case of high levels of consciousness, whereas another – in case of low levels). If that is so, we should analyze, which part of the consciousness spectrum is measured by which scale.

Guessing criterion analysis revealed above chance level accuracy for the lowest scale ratings in case of all subjective measures, except for the CR scale. This result is usually interpreted as an indicator of the influence of implicit knowledge. However, it seems unlikely that the same version of the AGL task would reveal and unreveal the implicit knowledge depending on the scale used. Rather, participants should experience the same level of access, but the scales may differ with respect to sensitivity for the low level of consciousness (scale exhaustiveness, cf. Sandberg et al., 2010). Coefficients of the regression model also suggest that the anchor points differ for each scale. Participants generally expressed higher level of consciousness with CR, FOW and PDW, as compared to RAS and SDS. This suggests that participants' subjective experience related with the lowest, highest and intermediate scale points differs significantly. Importantly, the lowest scale point for the CR scale seems to reflect different level of conscious access than the lowest points for other scales. Finally, the spectrum of consciousness probed with CR and with the other scales differs. In the next paragraph, we analyze this idea in more detail, proposing the theoretical analysis of scale sensitivity. To explain the rationale more clearly, we will focus on three scales that seem to differ by means of conscious access level, namely CR, FOW and RAS. We will not go into details for PDW because of the loss

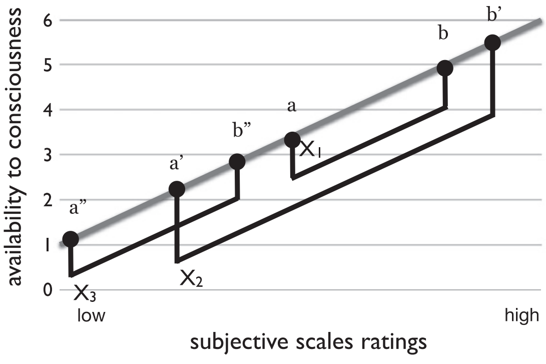

Fig. 2. Sensitivity analysis for the subjective measures of consciousness. Values on *X* axis represents the ratings for each scales (lowest (a) for each scale on the left, and the highest (b) on the right). Axis *Y* represents the level of consciousness (a theoretical variable).

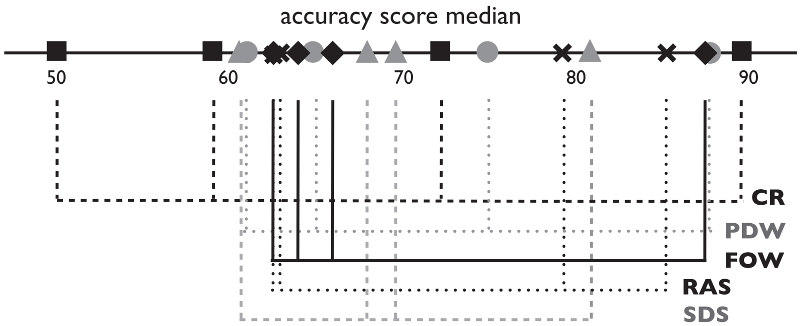

Fig. 3. Scale sensitivity analysis. Values on *X* axis represents the level of consciousness (as measured with the accuracy score median). Bars below the graph represents the range of consciousness that is assessed with the certain subjective scale. Note that CR has broad range with each point equally distant to the other (dashed black bar), whereas e.g. FOW scale measures mainly the variability of the lower stages of consciousness (3 out of 4 points are between 2nd and 3rd point of CR – see solid black bar).

aversion bias. SDS was excluded as it seems comparable, but less exhaustive then RAS (in fact the contrasts analysis for the slopes suggest that it was the worst scale by means of sensitivity, but this might be an effect of splitting the results of continuous scale into 4 bins).

4.3. Scale sensitivity analysis

Significantly different correlations between awareness ratings and accuracy scores, as observed with different subjective measures of consciousness, as well as significant differences for the guessing criterion analysis, suggest that the scales differ with respect to sensitivity. This difference seems to stem from two features characterizing how a scale rating relates to actual level of access. The first feature is the range of consciousness measured by a particular scale. Thus, a given scale may have a wider spread than another, allowing people to express a wider range of degrees of awareness. This will of course influence the observed relationship between awareness and accuracy. The second feature is the anchor point of a scale, that is, the accuracy level associated with its lowest point. The anchor points may vary depending on the scale. For instance, people who report to be "guessing" may also bet somewhat higher than zero. To illustrate our analysis, consider the hypothetical examples shown in Fig. 2.

The X_1 scale measures only a small part of the consciousness spectrum (*Y* axis). Nevertheless, the lowest point on that scale (a) reflects significantly lower level of awareness than the highest (b). In contrast, the X_2 scale measures a higher range of consciousness spectrum – the awareness level associated with the lowest rating (a′) differs from the one equated with the highest value of the scale (b′) to larger extent than X_1 scale extremes. Importantly, the lowest points for both scales do not necessarily indicate implicit knowledge exclusively. The X_3 scale measures a similar spectrum as X_1, but the lowest point (a″) reflects the lowest level of access as compared to lowest points on other scales (a and a′), which are thus still partially conscious. The differences between the lowest scale points reflect the second type of the sensitivity (sensitivity$_2$). The more sensitive$_2$ scale is, the lower level of access is observed in case of lowest scale points. The most sensitive scale may possibly (but not necessarily) reflect the access to implicit knowledge.

In other words, we suggest that the guessing criterion may not necessarily provide the best evidence for the existence of implicit knowledge. Such differences in sensitivity$_2$ were also observed in our study. The accuracy scores on CR scale for the lowest rating suggest that the implicit knowledge was not observed. Assuming that the general level of awareness is actually similar for all experimental groups, it seems that the lowest point on other scales measures higher level of consciousness than the lowest point on CR scale.

To illustrate our theoretical proposal we propose an additional theoretical analysis that is based on our experimental data. Following the assumptions that: (1) the level of awareness is similar in case of all groups, and (2) that consciousness (a theoretical variable that awareness ratings are supposed to represent) should be correlated with the classification accuracy to the same extent, we hypothesized the level of access assessed with the particular scale point might be referenced to the median accuracy score for the particular scale point. In other words, we treat median accuracy as an independent measure of awareness that reflects the most frequent level of the access associated with the certain scale point, assuming that the higher the median accuracy score for a particular scale point, the higher level of consciousness it reflects.[6] If this assumption is well grounded, we may compare and range all points on all scales by plotting medians for each point on each scale on the continuum of consciousness (see Fig. 3).

Here, the medians reflect the distances between the points on consciousness spectrum as measured with a certain scale point. As proposed before, the lowest CR rating seems to reflect the lowest level of access. It also seems to capture the highest

[6] We are aware that the accuracy is not the best measure of awareness. It seems to be proven that accuracy and awareness can be sometimes dissociated (see e.g. Lau, 2008). We nevertheless propose that the accuracy should reflect the awareness to the same extent in case of all groups. Our analyses of the scales should be replicated with another independent measure of subjective accessibility, but to the best of our knowledge no such measure has been yet developed.

range of consciousness as compared to other scales. However, if one would like to take a closer look at the variability of consciousness between the 2nd and 3rd point of CR, the FOW scale seems to be appropriate (as 3 out of 4 points of that scale are in between 2nd and 3rd point of CR). Analogically, the RAS scale, which is anchored on the higher part of the consciousness spectrum because of the rating criterion (see above), seems to capture (with points 3rd and 4th) the variability of consciousness observed in between the 3rd and 4th point of CR scale. Thus, taking into account the three scales we decided to compare, it seems that CRs may be used in conjunction with FOW or RAS to enable finer assessment of subjective states of awareness.

5. General conclusions

To conclude, we have shown that the knowledge acquired in implicit learning, as measured through the artificial grammar learning task is, at least partially, accessible to conscious awareness (in fact because of the chance accuracy for the lowest CR rating, our results suggest that no influence of implicit knowledge was observed in case of our study). However, differences between the scales suggest that the scales differ by means of the consciousness spectrum measured. We propose that those differences are related to two types of sensitivity. The first type (exhaustiveness) reflects the range of conscious variability measured by the scale. With the second, the sensitivity to low levels of consciousness (and possibly implicit knowledge) is measured. Our analyses suggest that each scale evaluates a slightly different range of consciousness. PDW, RAS and SDS seem not to be sensitive by means of $sensitivity_2$, since the lowest points for the scales reflect relatively high level of consciousness (due to the risk aversion strategy or high rating criterion applied). The CR scale's lower extreme seems to reflect the lowest accessibility as compared to other measures. This scale also seems to capture a broader spectrum of consciousness (as higher scale extreme is associated with high accuracy). However, to assess the whole spectrum of consciousness, one might want to use other scales in order to take a closer look into variability in the certain parts of consciousness spectrum. If so, the FOW scale seems to be related with lower levels of subjective experience. On the other hand, the RAS scale seems to probe the higher part of the spectrum with higher precision (also because of different rating criterion applied). Contrary to previous theoretical proposals, our analyses suggest that in order to assess the spectrum of consciousness in details, we should not only select the most sensitive measure, but we should attempt to combine the results obtained with a few measures. Following this argumentation, our data suggest that CR should be used in conjunction with RAS and FOW to enable a finer assessment of subjective states of awareness in AGL.

Acknowledgments

M.W. was supported with "Support for Scientists' International Mobility" Program granted by Polish Ministry of Science and Higher Education for the research described in this paper. M.W. and D.A. are supported by Polish Ministry of Science and Higher Education Grant N106 2548 37. A.C. is a Research Director with the F.R.S.-FNRS (Belgium). The authors would like to thank Anna Marzecová for her valuable comments on the previous versions of this manuscript, and Marta Siedlecka for her help with data collection.

References

Berry, D. C. (1995). How implicit is implicit learning. In G. Underwood (Ed.), *Implicit cognition* (pp. 203–225). Oxford: Oxford University Press.
Bierman, D. J., Destrebecqz, A., & Cleeremans, A. (2005). Intuitive decision making in complex situations: Somatic markers in an artificial grammar learning task. *Cognitive, Affective & Behavioral Neuroscience, 5*(3), 297–305.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–247.
Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences, 15*(12), 567–575.
Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40*(4), 343–367.
Cleeremans, A. (2011). Frontiers: The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology, 2*(86), 1–12.
Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences, 15*(8), 358–364.
Damasio, A. R. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Wilmington, Massachusetts: Mariner Books.
Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review, 8*(2), 343–350.
Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. Berry (Ed.), *How implicit is implicit learning?* Oxford: Oxford University Press.
Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(5), 1322–1338.
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.
Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69*(5–6), 338–351.
Dienes, Z., & Seth, A. (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*(2), 674–681.
Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition, 19*(1), 352–363.
Gaillard, V., Vandenberghe, M., Destrebecqz, A., & Cleeremans, A. (2006). First- and third-person approaches in implicit learning research. *Consciousness and Cognition, 15*(4), 709–722.
Higham, P. A., Vokey, J. R., & Pritchard, J. L. (2000). Beyond dissociation logic: Evidence for controlled and automatic influences in artificial grammar learning. *Journal of Experimental Psychology: General, 129*(4), 457–470.

Holender, D. (1986). Semantic activation without conscious awareness in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9*, 1–23.
Jamieson, R. K., & Mewhort, D. J. (2009). Applying an exemplar model to the serial reaction-time task: Anticipating from experience. *Quarterly Journal of Experimental Psychology, 62*(9), 1757–1783.
Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology: General, 132*(4), 551–565.
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.
Kouider, S., & Dupoux, E. (2004). Partial awareness creates the "illusion" of subliminal semantic priming. *Psychological Science: A Journal of the American Psychological Society/APS, 15*(2), 75–81.
Lau, H. C., & Maniscalco, B. (2010). Neuroscience. Should confidence be trusted? *Science, 329*(5998), 1478–1479.
Lau, H. C. (2008). Are we studying consciousness yet? In L. Weiskrantz & M. David (Eds.), *Frontiers of consciousness: Cichele lectures* (pp. 245–258). Oxford: Oxford University Press.
Macmillan, N. A., & Creelman, C. D. (1991). *Signal detection theory: A user's guide*. Cambridge: Cambridge University Press.
Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*(1–2), 115–134.
Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*(2), 288–294.
O'Regan, J. K., Myin, E., & Noë, A. (2005). Sensory consciousness explained (better) in terms of 'corporality' and 'alerting capacity'. *Phenomenology and the Cognitive Sciences, 4*(4), 369–387.
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.
Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119*(3), 264–275.
Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience, 10*(2), 257–261.
Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion, 5*(2), 243–247.
Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*(2), 227–244.
Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*(1), 1–23.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855–863.
Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219–235.
Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*(3), 201–224.
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.
Schurger, A., & Sher, S. (2008). Awareness, loss aversion, and post-decision wagering. *Trends in Cognitive Sciences, 12*(6), 209–210.
Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115*(2), 163–196.
Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? *Psychological Science: A Journal of the American Psychological Society/APS, 15*(11), 720–728.
Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: Relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences, 12*(8), 314–321.
Shanks, D. R., & Channon, S. (2002). Effects of a secondary task on "implicit" sequence learning: Learning or performance? *Psychological Research, 66*(2), 99–109.
Shanks, D. R., Rowland, L. A., & Ranger, M. S. (2005). Attentional load and implicit sequence learning. *Psychological Research, 69*(5–6), 369–382.
Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(3), 367–395.
Snodgrass, M., Bernat, E., & Shevrin, H. (2004). Unconscious perception at the objective detection threshold exists. *Attention, Perception, & Psychophysics, 66*(5), 888–895.
Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31*(7), 1060–1071.
Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(2), 328–344.