Personality and Individual Differences 141 (2019) 188–195

Contents lists available at ScienceDirect

Personality and Individual Differences

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

The first glance is the weakest: "Tasteful" individuals are slower to judge visual art

Nils Myszkowski

Department of Psychology, Pace University, United States of America

ARTICLE INFO

Keywords:
Aesthetic sensitivity
Good taste
Processing speed
Speed-accuracy tradeoff

ABSTRACT

Recent research (Myszkowski, Çelik, & Storme, 2018) has suggested that the ability to form accurate visual aesthetic judgments – an ability referred to as *aesthetic sensitivity* (Child, 1964), or "good taste" (Eysenck, 1983) – could be explained by the extent to which one taps into extensive processing strategies when facing aesthetic objects. Because individual differences in processing extensiveness may lead to different processing speed, we hypothesize that individuals with high visual aesthetic sensitivity present slower responding to visual aesthetic sensitivity tasks – even without time constraints. 201 adults took the Visual Aesthetic Sensitivity Test-Revised (Myszkowski & Storme, 2017), and their responses and response times were analyzed through joint hierarchical item-response theory modeling (van der Linden & Fox, 2016). As hypothesized, latent speed was negatively correlated at −0.47 (95% HPD [−0.61, −0.32]) with latent accuracy. Similar findings were obtained, accounting for guessing or not, including aberrant response patterns or not. In addition, more difficult items were also more time intensive. These findings are discussed as a substantiation that an important explanation for individual differences in visual aesthetic sensitivity lies in how much individuals are disposed to extensively process aesthetic objects.

1. Introduction

Empirical aesthetics are traditionally focused on finding an empirical definition of beauty, by identifying which aesthetic features are consensually preferred. Yet, early on, a number of notable individual differences psychology researchers (Binet, 1908; Eysenck, 1940; Thorndike, 1916) became interested in how individuals differ in their ability to form accurate art judgments – an ability that is often referred to as *aesthetic sensitivity* (Child, 1964), or more provocatively (Myszkowski, Storme, & Zenasni, 2016) as "good taste" (Eysenck, 1940, 1983).

Even though instruments to understand visual art perception are in constant evolution – especially since the introduction of eye-tracking – the methods used to quantify aesthetic sensitivity and to investigate its relations with other variables generally rely on psychometric testing. In the visual domain, aesthetic sensitivity measures are typically constructed through the *controlled alteration* technique (Meier, 1928), which consists in building pairs of aesthetic stimuli, in which one stimulus is an aesthetically deteriorated version of a base stimulus; the examinee's task is to identify, for each pair, which of the stimuli is aesthetically superior. Although the specific criteria used for the alteration of each item are not described and qualified as "intuitive rather than formally explicit" (Iwawaki, Eysenck, & Götz, 1979, p. 862) – it was inferred (Myszkowski & Storme, 2017) that the alterations are traditionally based on modifying the balance of the stimulus – by, for example, changing the position of the different elements – and quality of execution – by adding breaks in the contours of the elements or disturbances. To illustrate this point, we provide, in Fig. 1, example items of a pre-version of the Visual Aesthetic Sensitivity Test (VAST; Götz, 1985) – correct answers, from top to bottom, are right, left and left. In addition to the stimuli being created with the intention of creating versions of varying aesthetic quality, for some tests – including the one used in this research – the items have been further selected based on the responses of a panel of experts.

Recent research on the relations between visual aesthetic sensitivity and both reasoning abilities and artistic interests (Chamorro-Premuzic & Furnham, 2004; Myszkowski, Çelik, & Storme, 2018; Myszkowski, Storme, Zenasni, & Lubart, 2014), as well as recent psychometric investigations (Myszkowski & Storme, 2017), have suggested that a central factor for the formation of accurate aesthetic judgments may lie in how *extensively* individuals process aesthetic stimuli. Although, originally, visual aesthetic sensitivity measures were constructed to be judged intuitively, it has been advanced that individuals with high aesthetic sensitivity could engage in costly and time-consuming processing activities (Myszkowski et al., 2018) – such as, reflective processing (aiming at extending knowledge rather than seeking emotional

E-mail address: nmyszkowski@pace.edu.

https://doi.org/10.1016/j.paid.2019.01.010
Received 22 September 2018; Received in revised form 29 December 2018; Accepted 3 January 2019
0191-8869/

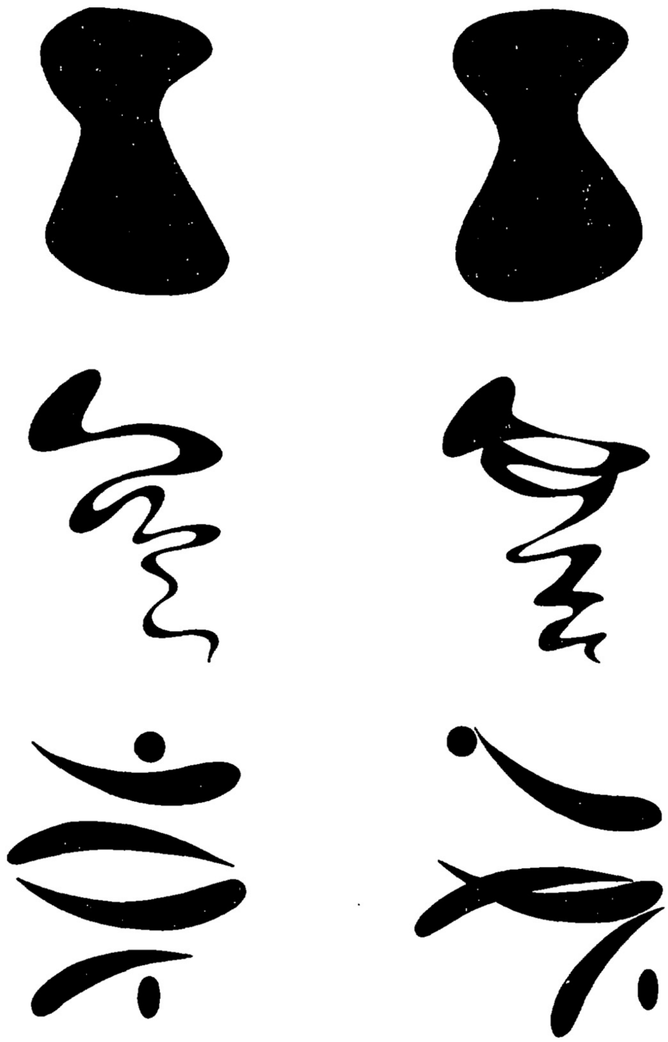

Fig. 1. Example items of the VAST.
(From Götz, Borisy, Lynn, & Eysenck, 1979).

arousal), goal management (spawning goals and subgoals when analyzing the stimuli), abstracting (extracting the structural elements of the stimuli and focusing less on elements readily available in the stimuli) – while individuals with low aesthetic sensitivity would have a quicker surface-level analysis of the stimulus.

In this study, we investigate how differences in "good taste" may be explained by how much effort is put into processing aesthetic stimuli. To investigate this question, we examine the trade-off between processing speed and accuracy in visual aesthetic sensitivity tasks with unlimited time, hypothesizing that slower judges, because they process aesthetic stimuli more extensively, have higher visual aesthetic sensitivity.

1.1. Why speed may impact visual "good taste"

In cognitive tasks with limited time conditions, individuals have to allocate limited cognitive resources between speed and accuracy – a phenomenon referred to as the Speed-Accuracy Tradeoff (SAT; see Heitz, 2014). While visual aesthetic sensitivity measures are not speeded tests – in that the examinee generally has either an unlimited amount of time, or ample time (meaning that a time limit is only meant to regulate the total testing time) to respond – a number of results in fact indicate that, even without any time pressure, an individual's speed could be detrimental to their accuracy.

First, recent research (Goldhammer & Klein Entink, 2011) on figurative reasoning tasks has found that, in these tasks, an individual's processing speed was negatively correlated with their accuracy. Further, the authors explained this relation by suggesting that individuals who are more accurate are the ones who monitor and validate more their responses – in other words the ones who "care more". Figurative reasoning tasks have been found to be consistently related to visual aesthetic sensitivity measures (Myszkowski et al., 2018, 2014), and thus similar responding phenomena may arise in such measures.

In addition, research has indicated that art expertise is characterized by attenuated emotional responses to art (Leder, Gerger, Brieber, & Schwarz, 2014), which favors a more reflective mode of processing, and allows experts to focus their attention on extending their knowledge of the stimulus. This increase in attention could encourage them to engage in a larger number of time consuming cognitive strategies, such as exploring visually the stimulus, extracting its structural skeleton, or generating subgoals to solve the task with more accuracy.

Further and in relation to this, visual aesthetic sensitivity was found to be correlated with openness to aesthetics and art interests (Chamorro-Premuzic & Furnham, 2004; Myszkowski et al., 2014), which suggests that one's motivation to judge art could boost their performance in such tasks. Interest for art may trigger a mode of responding that is both more extensive (thus time consuming) and accurate.

Finally, previous IRT analyses of visual aesthetic sensitivity measures have indicated that individuals may engage in guessing behaviors when responding (Myszkowski & Storme, 2017). This finding may also suggest that individuals engage in strategies that differ in depth, and thus in time-intensity.

1.2. Why the Speed-Accuracy Tradeoff may not apply

In spite of previous research suggesting that individuals with higher aesthetic sensitivity could engage in more time-consuming processing of the presented stimuli, there are a number of reasons why the SAT may not apply in these tasks. First, the SAT is largely induced by the presence of response deadlines (Heitz, 2014). In contrast, visual aesthetic sensitivity measures usually provide generous or unlimited time, and therefore, examinees have no incentive to respond fast. Because of the absence of a time pressure, one may not expect the SAT to apply to aesthetic sensitivity tasks.

Further, the SAT is generally considered to be a *within*-individual level effect (Goldhammer & Klein Entink, 2011), meaning that, for one given individual, responding faster leads to more errors. But this *within*-individual effect does not imply a *between*-individual level speed-accuracy relation. In other words, the SAT does not imply that individuals with higher response speed necessarily present lower accuracy (Fox & Marianti, 2016).

Finally, as previously mentioned, tests of visual aesthetic sensitivity often use visual stimuli that consist of simple abstract forms, which were destined to be judged without needing any interpretation (Eysenck, 1983). This would imply that the stimuli are to be judged with one's intuition and sense of "good gestalt" (Eysenck, Götz, Long, Nias, & Ross, 1984, p. 599), rather than by using complex processing strategies. Thus, if extensive and time-consuming processing strategies are not useful to solve aesthetic sensitivity tasks, one would not expect accuracy and speed to be related.

Perhaps for these various reasons, until now, response times in visual aesthetic sensitivity measures may have appeared irrelevant objects of study, and thus have never (to this day and to the best of our knowledge) been investigated. In this study, we aim at bridging this gap. Because recent research indicates that individuals with high aesthetic sensitivity could engage in more extensive processing strategies (Myszkowski et al., 2018), we hypothesized that individuals with higher visual aesthetic sensitivity – more "tasteful" individuals – are slower judges. We hypothesized that this phenomenon would be observed, even when guessing behavior is accounted for.

2. Method

2.1. Participants

The participants of this study were 201 undergraduate students (69% male and 30% female) from introductory psychology classes, with no prior exposure to visual aesthetic sensitivity measures. The age ranged from 18 to 28 years old, with a mean age of 19.3 years and a standard deviation of 1.8. The participants who volunteered received extra credit for participating in the study. They were told that the study concerned aesthetic judgment.

2.2. Instrument

Visual aesthetic sensitivity was measured with the Revised Visual Aesthetic Sensitivity Test (VAST–R; Myszkowski & Storme, 2017), which is based on the controlled alteration method. After 3 example items, participants are presented with 25 items, which are pairs of black and white paintings. In each pair, one stimulus has been altered to present a lower aesthetic quality. Participants are instructed to identify the painting that is objectively the better designed one – "the more balanced, better formulated" (Götz, 1985, p. 1) – which may not necessarily be the one that the examinee prefers.

To date, this test is the only visual aesthetic sensitivity measure whose items have evidence of content validity – with the unanimity of 8 expert judges in deciding of the aesthetically superior stimuli – of cultural invariance, and, with the revised version, of reliability and structural validity (Myszkowski & Storme, 2017). As with its previous investigation (Myszkowski & Storme, 2017), the VAST–R showed satisfactory internal consistency ($\alpha = 0.85$).

The participants took a computerized version of the test, allowing to record their response times for each item.

2.3. Statistical analyses

To analyze the relation between an individual's speed and accuracy, collapsing per individual both all responses and all response times before computing a correlation coefficient is an insufficient approach. Indeed, collapsing responses and response times leads to confounding item and person effects (van der Linden & Fox, 2016). An example of the insufficiency of this approach is that, if individual accuracy and speed are indeed negatively related, but the most succeeded items are the ones that are the most quickly processed, one could observe no correlation between responses and response times, and thus falsely conclude that speed and accuracy are not related. These issues are generally likened to instances of Simpson's paradox (van der Linden, 2007). Because of this, we used to a procedure both more consistent with the data generating process and more appropriate to test our hypothesis, which we will now describe.

2.3.1. Joint hierarchical IRT modeling

A recently developed approach to disentangle the different effects of individuals and items on both responses and response times consists in modeling responses and response times through a joint hierarchical Item Response Theory model (van der Linden, 2007). Although it has been already been extensively discussed (Fox & Marianti, 2016; Goldhammer & Klein Entink, 2011; Klein Entink, Fox, & van der Linden, 2009; van der Linden, 2007; van der Linden & Fox, 2016), we here discuss the main characteristics of this hierarchical approach.

On a first level, responses and response times are modelled separately using distinct person and item characteristics. The responses are modelled using a person's latent accuracy (noted θ), item discrimination – the strength of the items' relation with accuracy – item difficulty – the ability level where the item best differentiates individuals – and item guessing – the probability that an individual with no ability makes a correct guess. In contrast, response times are modelled using a person's latent speed (noted ζ), item time discrimination – the strength of the relation between speed and the item's response time – and item time intensity – the average response time for the item. A 2 (without guessing) or a 3 (with guessing) parameter normal ogive model is used for the responses, while a 2 parameter log-normal model is used for the response times.

On the second level, relations between the parameters of the level 1 model are specified. Specifically, speed and accuracy are assumed to arise from a bivariate normal distribution, with a covariance structure that allows for their correlation. Similarly, the item parameters are assumed to come from a multivariate normal distribution, with a covariance structure that allows for their correlations.

Such an approach allows to account for the structure of the data and the dependencies within items and within persons, and to account for relations between item characteristics and person characteristics respectively. It thus allows the study of the relation between individual speed and accuracy – our objective. Although this approach is fairly new, it has already been successfully used, notably in the investigation of response times in cognitive ability tasks (e.g., Goldhammer & Klein Entink, 2011; Klein Entink, Kuhn, Hornke, & Fox, 2009).

2.3.2. Model estimation

The model was estimated with the package 'LNIRT' (Fox, Klein Entink, & van der Linden, 2007; Fox, Klotzke, & Klein Entink, 2018) for R. The main purpose of this package is the application of the joint modeling approach previously described. Because previous studies (Myszkowski & Storme, 2017) indicated that guessing should be accounted for, a 3-Parameter Normal Ogive (3PNO) model for the responses, while a 2-Parameter Log-Normal model (2PLN) was used for the response times: We later refer to this model as the 3PNO-2PLN model. 'LNIRT' estimates parameters using a Bayesian Markov Chain Monte Carlo (MCMC) algorithm with Gibbs sampling. Like in similar research (Goldhammer & Klein Entink, 2011), the default non-informative priors of the package were used. Six MCMC chains of 10,000 iterations were used (with the 1000 first iterations discarded as burn-in iterations). As recommended (Klein Entink, Kuhn, et al., 2009), convergence was ensured by examining the trace plots, autocorrelations and posterior densities of the estimates. In addition, the Gelman-Rubin statistic (Gelman & Rubin, 1992) was computed across the chains – a value close to 1 indicating convergence. Currently, tests of model fit for the joint modeling approach have not been fully established (Fox & Marianti, 2017), but person/item fit analyses can be performed. Fox and Marianti (2017) described that the extremeness of response patterns can be analyzed through their posterior probability, which may be used as Bayesian p values – probabilities smaller than .05 indicating an aberrant person/item. The correlations between parameters were studied by examining the posterior MCMC distribution of that correlation across chains and iterations. Point estimates were obtained by averaging across chains and iterations, and 95% High Posterior Density (HPD) intervals were computed.

3. Results

3.1. Model convergence and fit

Across all model parameters of the 3PNO-2PLN model, the observed Gelman-Rubin statistics ranged from 1.00 to 1.02, with effective sample sizes ranging from 2241 to 60,445. Along with the examination of the trace plots, autocorrelation and posterior densities of the estimated parameters, it was concluded that the model had converged successfully.

Regarding model fit, across the chains, no misfitting item was detected in any chain, leading to the general conclusion that the model had adequately fit the data. In all the chains, no participant had an aberrant response pattern, and less than 10% participants (20 participants out of the 201) had aberrant response time patterns – they were

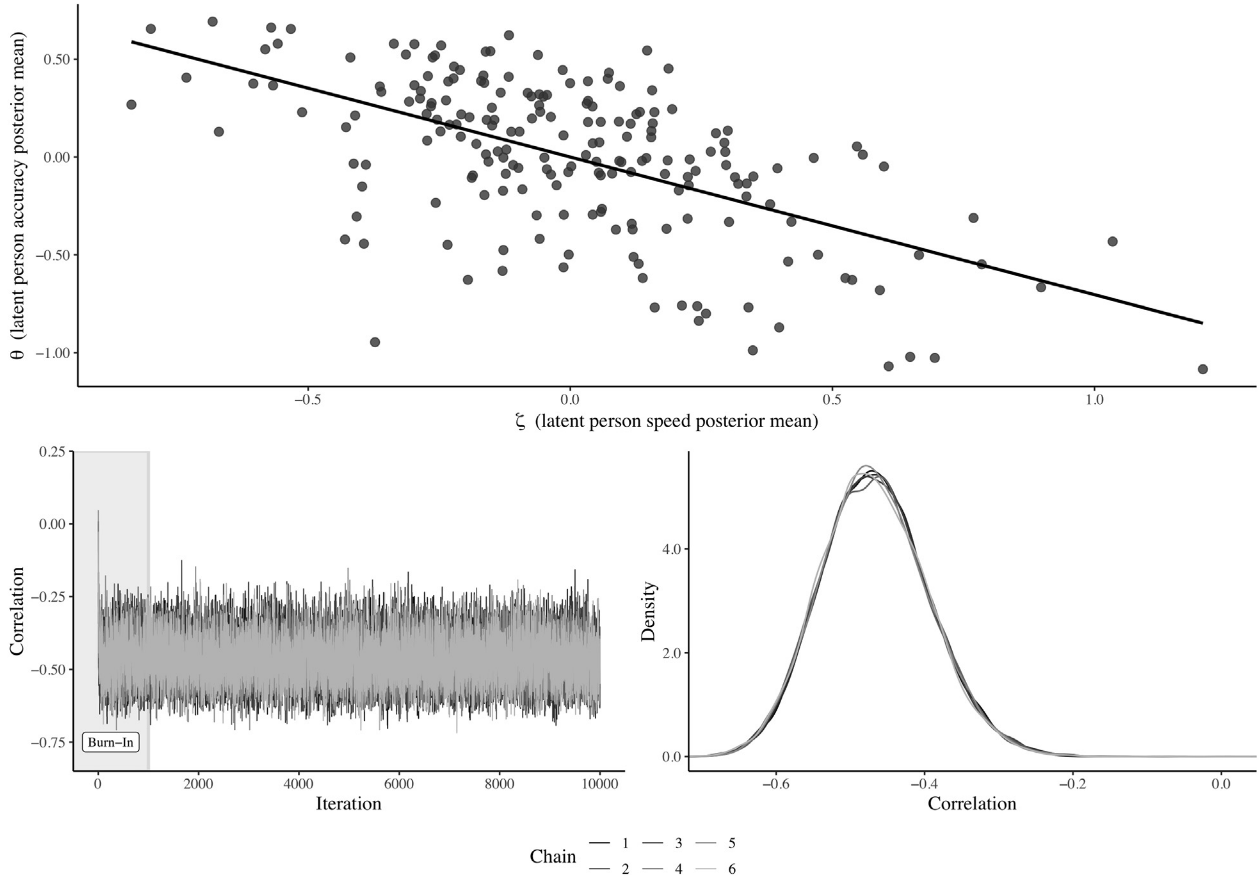

Fig. 2. Correlation between person speed and accuracy for the 3PNO-2PLN model.

the same participants in all chains. We decided to investigate their impact on the speed-accuracy relation in a supplementary analysis (later presented).

3.2. Correlation between person speed and accuracy

As hypothesized, a consistent pattern of negative correlations was observed between individual latent speed and accuracy estimates – with a posterior mean correlation of −0.47 and a 95% HPD of [−0.61, −0.32]. In Fig. 2, we present the scatterplot of the correlation between the mean posterior estimates of speed and accuracy in the sample, along with the trace plot of the MCMC iterations and the posterior density of the correlation.

3.3. Supplementary analyses

3.3.1. Are difficult items more demanding in time?

The main focus of this study is the relation between an individual's speed and ability in aesthetic judgment. Yet, the modeling approach used here also allows to study within-subject (or between-items) effects (Fox et al., 2007). This implies that it allows to study relations between item characteristics, controlling for person effects. Since we hypothesized that, at the between-subjects level, individuals with higher speed have lower accuracy, we formulated a similar hypothesis at the within-subjects level, which is that, the more difficult an item – in other words, the higher the accuracy required to succeed it – the more time intensive it is – in other words, the more time it takes to respond to. As hypothesized, difficulty and time intensity were positively correlated, with a posterior mean of the correlation estimate of 0.49, and a 95% HPD interval of [0.34, 0.63]. In Fig. 3, we present the scatterplot of the correlation between the mean posterior estimates of speed and accuracy in the sample, along with the trace plot of the MCMC iterations and the posterior density of the correlation between item difficulty and time intensity. Time intensity represents the expected time spent to respond an item on a logarithmic scale.

3.3.2. Is the same result obtained without modeling guessing?

Previous results (Myszkowski & Storme, 2017) suggest that the VAST–R may be prone to guessing for some items but not all. Here, accounting for guessing in our modeling approach was important, as guessing could have biased the relation between speed and ability. Yet, since this previous study was not completely conclusive on this point, we decided to replicate our analysis, but fixing guessing to zero, and thus using a 2-Parameter Normal Ogive (for the responses) and 2-Parameter Log-Normal (for the response times) model (2PNO-2PLN).

Across the model parameters, all observed Gelman-Rubin statistics rounded to 1.00, with effective sample sizes ranging from 5277 to 61,445, indicating successful convergence. The item response model fit the data slightly worse than the 3PNO-2PLN, as 3 participants had misfitting response patterns. Like for the 3PNO-2PLN, the same 20 cases showed aberrant response time patterns in all chains. Negative correlations were observed between individual latent speed and accuracy estimates – with a posterior mean correlation of −0.46 and a 95% HPD of [−0.60, −0.31], which is very similar to the result obtained for the 3PNO-2PLN. In Fig. 4, we present the scatterplot of the correlation between the mean posterior estimates of speed and accuracy in the sample, along with the trace plot of the MCMC iterations and the posterior density of the correlation between speed and accuracy.

3.3.3. Did cases with aberrant response time patterns bias the correlation?

As previously noted, 20 out of 201 participants were identified as having some aberrant response time patterns. Since it appeared possible

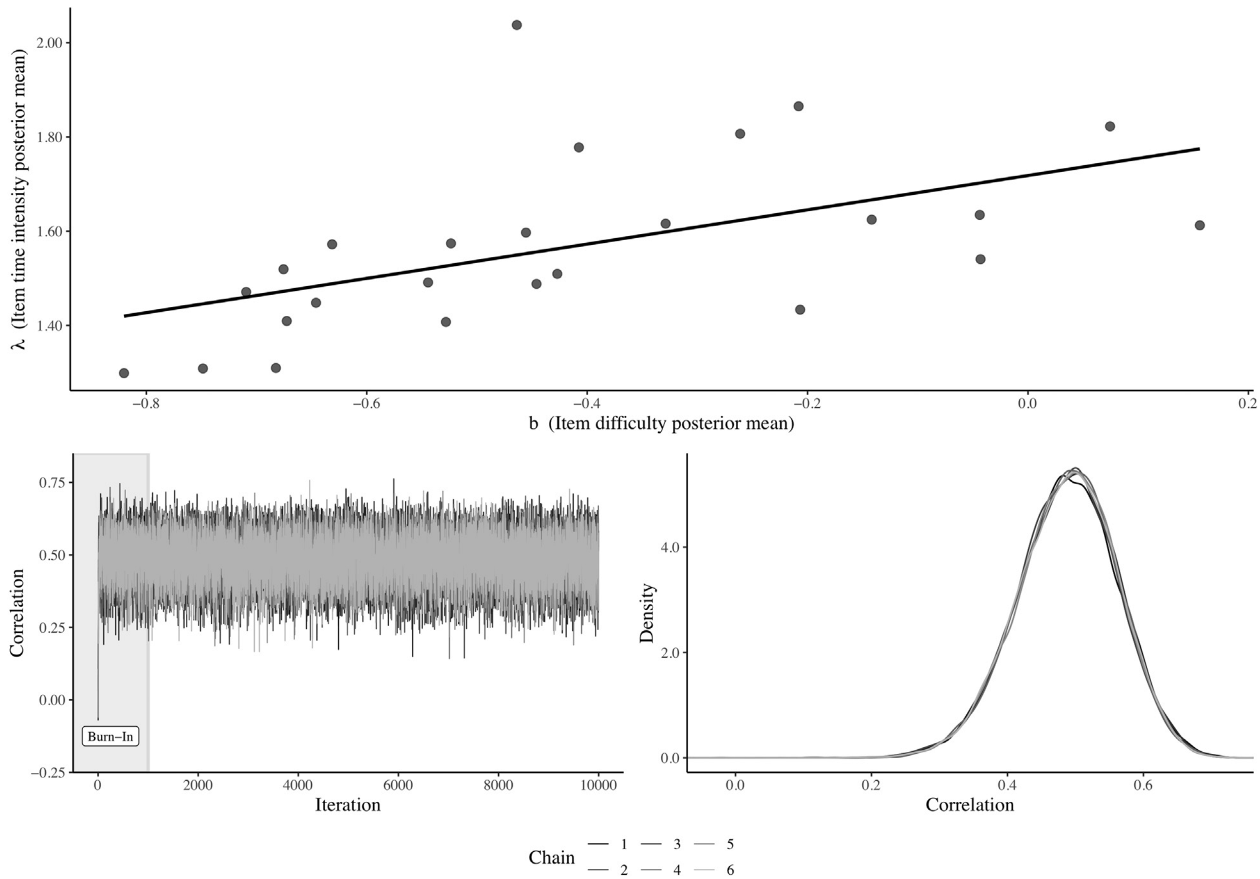

Fig. 3. Correlation between item difficulty and time intensity for the 3PNO-2PLN model.

that these participants may have biased the correlation between the individual speed and accuracy, we re-estimated the 3PNO-2PLN model without these subjects.

The Gelman-Rubin statistic was 1.00, indicating that the iterations successfully converged. Like previously, speed and accuracy were negatively correlated, with a posterior mean of the correlation estimate of −0.46, and a 95% HPD interval of [−0.61, −0.30]. This result is very similar to the one observed with all cases, and indicates that the negative correlation between speed and accuracy observed overall was not due to the presence of cases with aberrant response times. In Fig. 5, we present the scatterplot of the correlation between the mean posterior estimates of speed and accuracy in the sample without the cases with aberrant response time patterns, along with the trace plot of the MCMC iterations and the posterior density of the correlation.

4. Discussion

By the past, research endeavors on visual aesthetic sensitivity have solely focused on the ability to form *accurate* judgments of "taste". After all, aesthetic judgment is rarely affected by time pressure, in reality or in testing situations. Thus without any time limit, why would speed impact "good taste"?

Yet, research on the topic has suggested that a differentiating factor between individuals with high and low visual good taste could be the fact that individuals with high good taste use more demanding processing strategies when facing such tests. Although we did not here study directly the very processes engaged in visual aesthetic sensitivity tasks, our results are in line with these suggestions, as they indicate that one's visual aesthetic sensitivity is strongly impacted by one's responding speed.

Indeed, through the joint hierarchical modeling of responses and response times, we were able to separate item effects and person effects on both responses and response times. We found consistent negative correlations between individual speed and accuracy – accounting for guessing or not, including or excluding cases with aberrant patterns. Although we were here primarily focused on differences between individuals, it was also noted that, at the item level, the most difficult items were also the most time intensive.

4.1. Implications

While our study does not allow to further explain the speed-accuracy tradeoff observed, recent research in visual aesthetic sensitivity allows to advance a few concurrent explanations. First, as previous research suggested (Myszkowski et al., 2014), it could be that visual aesthetic sensitivity integrates an important motivational element. In other words, this research could suggest that individuals may differ in the accuracy of their judgments because, to some extent, they were more *motivated* to use more (or deeper) processing strategies. Another explanation, which correlations between visual aesthetic sensitivity and cognitive ability tests could suggest (Myszkowski et al., 2018), may be that individuals who process more slowly and more accurately are in fact more *able* to access such extensive processing strategies.

In addition, this study has psychometric implications. Indeed, the results suggest that estimates of visual aesthetic sensitivity may in fact be contaminated by response speed – even when guessing is accounted for. Although one may argue that visual aesthetic sensitivity could comprise slow processing as a component – since, after all, it may indicate more extensive processing – we would suggest that researchers question if they are interested in accuracy *in general*, or in accuracy *controlling for speed*. It may notably be that, through statistically controlling for speed, one is allowed to, at least partially, rule out the

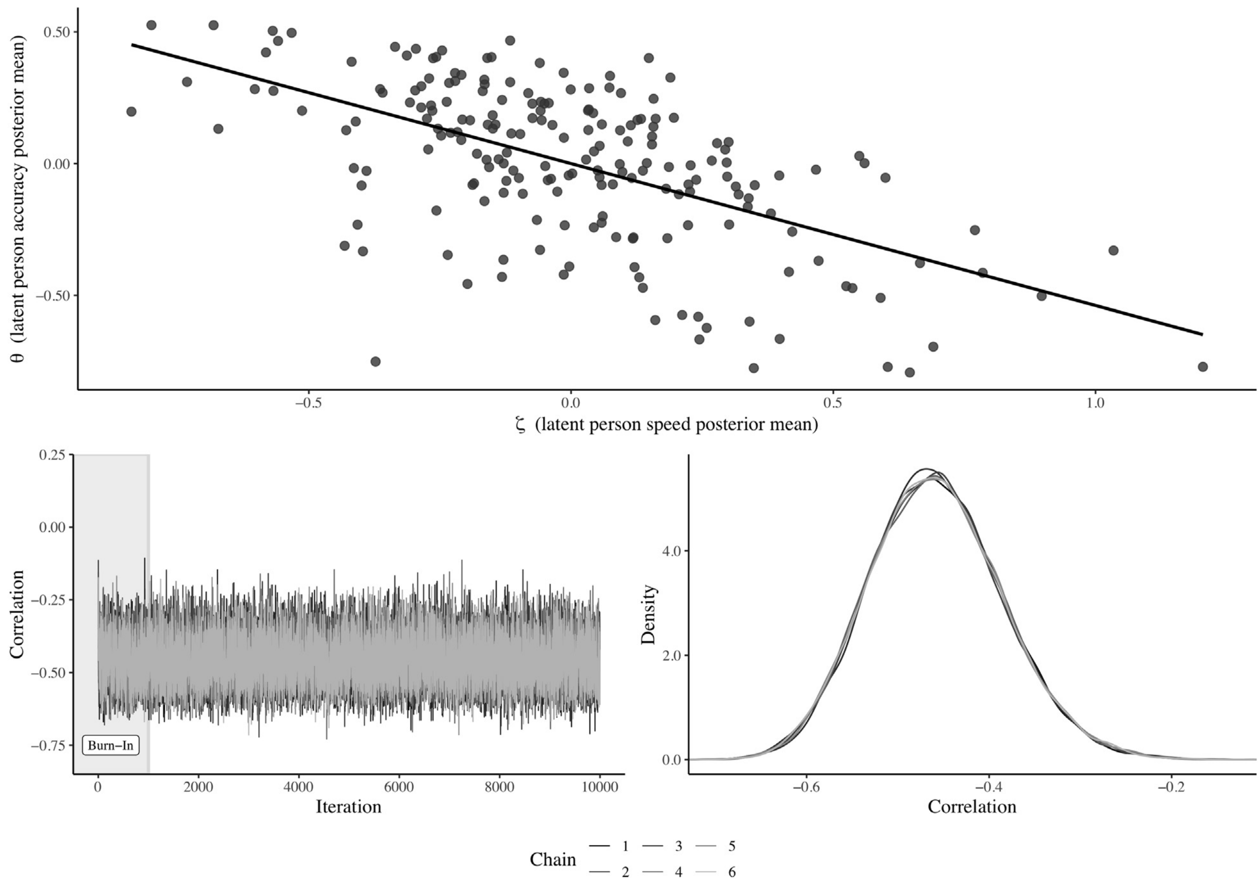

Fig. 4. Correlation between person speed and accuracy for the 2PNO-2PLN model.

motivational component of visual aesthetic sensitivity, and thus study it as a purer ability. Provided that response times are collected – which has become fairly easy when using computerized testing – the joint modeling approach used in the present study (van der Linden & Fox, 2016) appears both ideal – as it allows to disentangle individual and item effects on both responses and response times – and available (Fox et al., 2018). In addition, although the study of cases with aberrant patterns did not lead to different conclusions regarding our hypothesis, we see the methodology used to identify these aberrant response times patterns as essential, notably in situations where phenomena like careless responding is susceptible to occur.

4.2. Limitations and further directions

While this research effort is the first to study response times in visual aesthetic sensitivity tasks, it is limited in several aspects.

First, it is limited by the fact that we used only one measure of visual aesthetic sensitivity – albeit the one with (currently) the most evidence of psychometrical robustness (Myszkowski & Storme, 2017). For better generalizability, we would suggest that this result be replicated with other visual aesthetic sensitivity measures. Another limitation is the use of a convenience sample of students. Such samples have recurrently and successfully been used by the past to study visual aesthetic sensitivity, and we do not anticipate that the relation between speed and accuracy would differ in the general population of adults, but replications using other sampling methods, and perhaps larger samples, are probably called for.

Moreover, it is important to note that we here studied *response* speed, not *processing* speed. In other words, it may be that individuals who are more fluent *processors* of visual aesthetic stimuli do have a higher accuracy in visual aesthetic sensitivity tasks, but that, at the same time, fast respondents tend to not fully (or properly, or deep enough) get to processing visual aesthetic stimuli. In other words, accuracy could be both positively related to *processing* speed, but negatively related to *response* speed. One may for example speculate that some individuals may process the stimuli fast, but then hesitate and reflect more extensively to cross-validate their response, leading to increased response times – what appears to be slow processing. Such individuals may exhibit both slow response speed and high processing speed. Methods such as verbalization during task and eye-tracking may help address this question.

Another limitation of this study is that all items were presented here without time limit for all participants, and thus, it remains unclear whether the negative effect of speed on accuracy could be induced experimentally. It may be that, for example, when participants are speeded to respond, one's speed of processing becomes an advantage rather than a weakness in responding accurately. Further research may experiment on how time constraints may limit one's accuracy.

Finally, we did not study here how various sources of individual differences may affect speed, accuracy, or both. One could for example suspect here that personality traits, especially those related to openness to aesthetics and interest in art, would lead to slower and more accurate responses, for the reason that individuals that are more interested in art may be more interested in forming accurate aesthetic judgments. Likewise, we may suggest that art expertise may have an effect on speed (and thus also perhaps on accuracy), as experts may be have developed skills to access more extensive aesthetic processing strategies. Similarly, as intelligence has been showed to be related to visual aesthetic sensitivity (Myszkowski et al., 2018), this relation may be explained by intelligent individuals being more able to use extensive strategies to form aesthetic judgments.

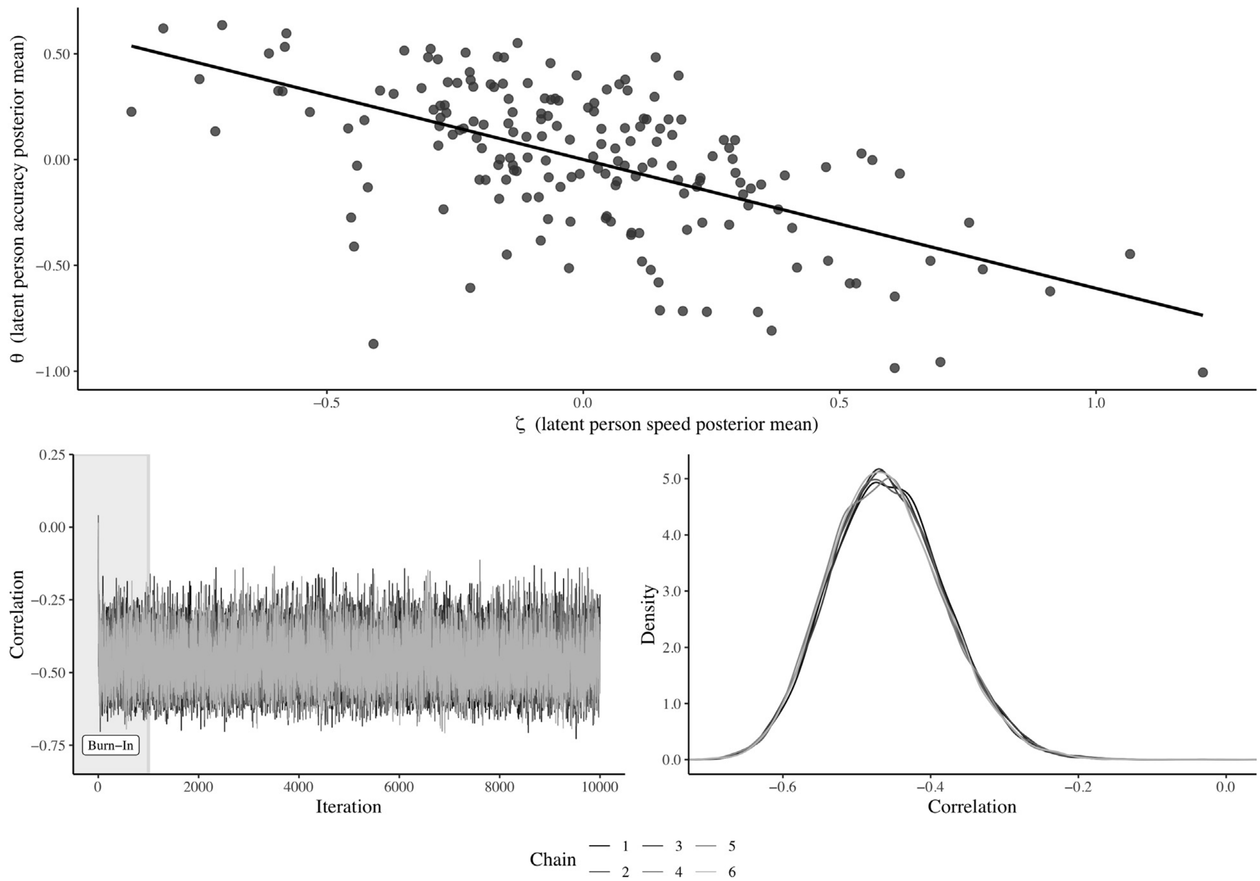

Fig. 5. Correlation between person speed and accuracy for the 3PNO-2PLN model without cases with aberrant response times.

5. Conclusion

When one refers to reasoning, speed is often thought of as a sign of high performance. Faster individuals are considered as being able to mobilize more resources, allowing them in return to solve problems more efficiently. Yet, when it comes to art, individuals may use various processing strategies – from promptly scanning to exhaustively investigating the stimulus.

Although the study of individual differences in aesthetic ability is receiving increased interest, the phenomena that underlie accurately judging the quality of art – as opposed to the formation of individual preferences – remain largely unknown. Through recent advances in joint response and response time modeling, this study provides evidence that indicates that individuals who judge art with more discernment – individuals with "good taste" – actually judge art slower, and thus probably with more care.

References

Binet, A. (1908). La psychologie artistique de Tade Styka. *L'Année Psychologique, 15*(1), 316–356. https://doi.org/10.3406/psy.1908.3760.

Chamorro-Premuzic, T., & Furnham, A. (2004). Art judgment: A measure related to both personality and intelligence? *Imagination, Cognition and Personality, 24*(1), 3–24. https://doi.org/10.2190/U4LW-TH9X-80M3-NJ54.

Child, I. L. (1964). Observations on the meaning of some measures of esthetic sensitivity. *The Journal of Psychology, 57*(1), 49–64. https://doi.org/10.1080/00223980.1964.9916671.

Eysenck, H. J. (1940). The general factor in aesthetic judgements. *The British Journal of Psychology. General Section, 31*(1), 94–102. https://doi.org/10.1111/j.2044-8295.1940.tb00977.x.

Eysenck, H. J. (1983). A new measure of "good taste" in visual art. *Leonardo, 16*(3), 229. https://doi.org/10.2307/1574921.

Eysenck, H. J., Götz, K. O., Long, H. Y., Nias, D. K. B., & Ross, M. (1984). A new Visual Aesthetic Sensitivity Test: IV. Cross-cultural comparisons between a Chinese sample from Singapore and an English sample. *Personality and Individual Differences, 5*(5), 599–600. https://doi.org/10.1016/0191-8869(84)90036-9.

Fox, J.-P., Klein Entink, R. H., & van der Linden, W. J. (2007). Modeling of responses and response times with the package CIRT. *Journal of Statistical Software, 20*(7), 1–14.

Fox, J.-P., Klotzke, K., & Klein Entink, R. H. (2018). LNIRT: LogNormal response time item response theory models. Retrieved from https://CRAN.R-project.org/package=LNIRT.

Fox, J.-P., & Marianti, S. (2016). Joint modeling of ability and differential speed using responses and response times. *Multivariate Behavioral Research, 51*(4), 540–553. https://doi.org/10.1080/00273171.2016.1171128.

Fox, J.-P., & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed: Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement, 54*(2), 243–262. https://doi.org/10.1111/jedm.12143.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472. https://doi.org/10.1214/ss/1177011136.

Goldhammer, F., & Klein Entink, R. H. (2011). Speed of reasoning and its relation to reasoning ability. *Intelligence, 39*(2), 108–119. https://doi.org/10.1016/j.intell.2011.02.001.

Götz, K. O. (1985). *VAST: Visual Aesthetic Sensitivity Test* (4th ed.). Dusseldorf, Germany: Concept Verlag.

Götz, K. O., Borisy, A. R., Lynn, R., & Eysenck, H. J. (1979). A new Visual Aesthetic Sensitivity Test: I. Construction and psychometric properties. *Perceptual and Motor Skills, 49*(3), 795–802. https://doi.org/10.2466/pms.1979.49.3.795.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*. https://doi.org/10.3389/fnins.2014.00150.

Iwawaki, S., Eysenck, H. J., & Götz, K. O. (1979). A new Visual Aesthetic Sensitivity Test: II. Cross-cultural comparison between England and Japan. *Perceptual and Motor Skills, 49*(3), 859–862. https://doi.org/10.2466/pms.1979.49.3.859.

Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika, 74*(1), 21–48.

Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods, 14*(1), 54–75. https://doi.org/10.1037/a0014877.

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition & Emotion, 28*(6), 1137–1147. https://doi.org/10.1080/02699931.2013.870132.

Meier, N. C. (1928). A measure of art talent. *Psychological Monographs, 39*(2), 184–199. https://doi.org/10.1037/h0093346.

Myszkowski, N., Çelik, P., & Storme, M. (2018). A meta-analysis of the relationship between intelligence and visual "taste" measures. *Psychology of Aesthetics, Creativity, and*

the Arts, 12(1), 24–33. https://doi.org/10.1037/aca0000099.
Myszkowski, N., & Storme, M. (2017). Measuring "good taste" with the Visual Aesthetic Sensitivity Test-Revised (VAST-R). *Personality and Individual Differences, 117*, 91–100. https://doi.org/10.1016/j.paid.2017.05.041.
Myszkowski, N., Storme, M., & Zenasni, F. (2016). Order in complexity: How Hans Eysenck brought differential psychology and aesthetics together. *Personality and Individual Differences, 103*, 156–162. https://doi.org/10.1016/j.paid.2016.04.034.
Myszkowski, N., Storme, M., Zenasni, F., & Lubart, T. (2014). Is visual aesthetic sensitivity independent from intelligence, personality and creativity? *Personality and Individual Differences, 59*, 16–20. https://doi.org/10.1016/j.paid.2013.10.021.
Thorndike, E. L. (1916). Tests of esthetic appreciation. *Journal of Educational Psychology, 7*(9), 509–522.
van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*(3), 287–308. https://doi.org/10.1007/s11336-006-1478-z.
van der Linden, W. J., & Fox, G. J. A. (2016). Joint hierarchical modeling of responses and response times. *Handbook of item response theory, volume one: Models*. Retrieved from https://research.utwente.nl/en/publications/joint-hierarchical-modeling-of-responses-and-response-times.