RESEARCH ARTICLE

REVISED Pupil dilation prediction of random events [version 2; peer review: 2 approved with reservations]

Patrizio E Tressoldi , Massimiliano Martinelli, Luca Semenzato

Dipartimento di Psicologia Generale, Università di Padova, Padova, 35131, Italy

v2 First published: 02 Dec 2013, 2:262
<https://doi.org/10.12688/f1000research.2-262.v1>
Latest published: 09 May 2014, 2:262
<https://doi.org/10.12688/f1000research.2-262.v2>

Abstract

We report the results of a conceptual replication of a study that reported that pupil dilation can predict potentially threatening random events above chance level. In this study, participants' pupil dilation was used to predict the appearance of a threatening or a neutral stimulus, presented randomly in a double sequence of ten trials with replacement, i.e. replacing the chosen trial for the future extractions.

In the first experiment, with a sample of 100 participants, the average correct prediction was 55.9%, with a small difference between the two stimuli.

This effect was further tested in an exact pre-registered study where the average correct prediction was 58.7%. The reliability of these findings was checked utilizing both a frequentist and a Bayesian statistical parameters estimate approach.

These findings collectively support the hypothesis that pupil dilation can be used to anticipate random and therefore theoretically "unpredictable" events in an implicit unconscious way that is without a conscious awareness, and that this ability is another characteristic of the powerful anticipatory adaptive systems of our psychophysiological system.

Keywords

anticipation; random events; pupil dilation; emotion;

Open Peer Review**Approval Status** ? ?

	1	2
version 2 (revision) 09 May 2014	 view ↑	 view ↑
version 1 02 Dec 2013	 view	 view

1. **Chris Baker** , National Institute of Mental Health, Bethesda, MD, USA
2. **Leo McHugh**, Immunexpress Inc., Seattle, WA, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Patrizio E Tressoldi (Patrizio.tressoldi@unipd.it)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2014 Tressoldi PE *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Tressoldi PE, Martinelli M and Semenzato L. Pupil dilation prediction of random events [version 2; peer review: 2 approved with reservations] F1000Research 2014, 2:262 <https://doi.org/10.12688/f1000research.2-262.v2>

First published: 02 Dec 2013, 2:262 <https://doi.org/10.12688/f1000research.2-262.v1>

REVISED Amendments from Version 1

The main revisions to Version 1 are:

- A clarification about which PD data are used for the calculation of the prediction accuracy on the paragraph “Formal experimental phase” and on Figure 2.
- The presentation of the confidence intervals related to the hits percentages in Figure S1 and Figure S2.

A comment to each of the reviewers is posted in the comments section of the online version.

See referee reports

Introduction

The anticipation, or prediction, of future events is a fundamental activity of our conscious and implicit, (i.e. unconscious) cognitive abilities. To be able to predict, even approximately, future movements, perceptions and events, reduces the risks and harms of a future threatening situation and more importantly, optimizes the use of our limited cognitive and energetic resources (e.g. Friston¹).

It is therefore not surprising that the study of anticipation has become an interdisciplinary line of research encompassing psychophysiology and neurophysiology (e.g. Van Boxtel and Böcker²), cognitive processes (Barcelo, Bestmann and Yu³) and artificial intelligence (Butz, Sigaud and Baldassarre⁴).

At the core of the anticipatory activity is an innate sensitivity to the structure and statistics of the environment in order to build up correct representations of each event in order to prepare the organism to perceive and act in a way that minimizes errors and reduces unnecessary adjustments (e.g. Clark⁵). But this raises the question of how it is possible for the organism to prepare for future events if they are unpredictable or equally probable?

Our interest is focused on this more extreme form of anticipation, that of random events. Even if the study of predicting unpredictable events seems a paradox, there is cumulative evidence related to almost 15 years of investigation, that humans, and perhaps every other living organism, can predict some classes of events above the level expected by chance (see the meta-analysis by Mossbridge, Tressoldi, and Utts⁶). If proven, this phenomenon will further enhance our knowledge about innate predictive survival abilities.

In this study, we demonstrate that pupil dilation (PD) reactions differ before the random presentation of a neutral or a potentially threatening stimulus.

Similar to other psychophysiological variables (e.g. heart rate, skin conductance, etc.), PD reacts to different emotional states that are correlated with the arousal of the autonomic sensory system (e.g. Bradley *et al.*⁷). Because the pupil dilates with sympathetic activity and constricts with parasympathetic activity (Beatty and Lucero-Wagoner⁸; Steinhauer *et al.*⁹), it represents an accurate and unobtrusive measure of the cognitive resources invested in a task, for example in doing arithmetic tasks.

An interesting application of PD is shown in tasks where participants are not consciously aware of some information but their PD

reveals that it is unconsciously available and is being processed by the cognitive system. For example Bijleveld, Custers and Aarts¹⁰, used PD to reveal the strategic recruitment of resources upon presentation of subliminal reward cues.

However PD not only gives information about unconscious cognitive activity but can even anticipate future perceptual or cognitive activities. Einhäuser, Stout, Koch, and Carter¹¹ used PD to reveal perceptual selection and its prediction of subsequent stability in perceptual rivalry and Einhäuser, Koch, and Carter¹² employed PD to reveal decision making before a person voluntarily reported it.

Starting from these studies and the cumulative evidence that the human psychophysiological and electrophysiological systems react differently before random presentation of two categories of emotional stimuli such as pictures, sounds, etc. (Mossbridge, Tressoldi, and Utts⁶), we aimed to further investigate whether PD can actually predict random events at the level of each single trial of each participant (see Procedure). An increase of approximately 20% of alerting sounds above the mean chance expected (MCE) has already been observed by Tressoldi *et al.*¹³ in the prediction of auditory alerting and neutral sounds presented randomly at the level of single trials (a summary of results obtained in that study is presented in [Table S1](#) in the Supplementary Material of this paper).

In this paper, we will present the results of a conceptual replication using visual information contrasting a potentially “threatening” stimulus with a neutral one using an experimental procedure that simulate an ecological condition where a dangerous or a neutral event could happen randomly.

Methods

Participants

Participants were recruited by advertisements mainly among students of Padova University. Exclusion criteria included uncorrected vision and use of drugs that could affect pupil size and pupil dilation. These characteristics were ascertained by asking each participant.

Sample definition

Estimating an effect size of approximately 0.30, to achieve a statistical power above 0.80, setting $\alpha=0.05$, an opportunity sample of 100 students and personnel from Padova University (Faul *et al.*¹⁴) were recruited by a research assistant to participate in an experiment on a gambling task. The final sample comprised 32 males and 68 females with a mean age of 29.3 and with a standard deviation of 3.8.

Ethics statement

Participation inclusion followed the ethics guidelines in accordance with the Helsinki Declaration and the study was approved by the Ethics Committee of Dipartimento di Psicologia Generale, the hosting institution. Before taking part in the experiment, each participant provided written consent after reading a brief description of the experiment.

Procedure

The experiment consisted of two different phases, a preliminary and a formal one. The preliminary phase aimed at familiarizing each

participant with the procedure and testing if they reacted differently to the two stimuli.

Preliminary phase

Each participant was seated in front of a 19 inch monitor in a sound and light attenuated lab of approximately 120 cd/m² measured with a Minolta light meter.

Before the formal sessions, each participant was told: *“Before the formal experiment, we must record your personal pupil dilation reactivity to the two types of stimuli you will see behind the door. You must simply watch what will happen on the screen without doing anything. When the door opens, you will see a gun shooting at you, hearing a shot, or you will see a smile. You will see the shooting gun and the smile ten times each, in random order”.*

We chose to limit the test to 10 trials per stimulus to avoid boredom and reduce the possibility of using controlled strategies to predict the target.

If there was no need for further clarification, the task started with the calibration of the eye position for the Eye Tracker apparatus. This consisted of participants following a dot moving slowly in different positions on the monitor in a natural way without the need to fix the head position.

Once the calibration was completed, the task started with the stimuli presentation. The sequence of events is presented in **Figure 1**. The inter-item interval was randomly chosen between 2 and 4 sec.

The two target stimuli, the gun and the smile, and the door, were calibrated for luminance (300 × 471 pixel; 72 horizontal and vertical dpi). The door was colored in black similar to the video background to avoid PD modification consequent to differences in luminosity. Their luminance measured using cd/m² with a Minolta® photometer was, for the gun: 15 center, 90 periphery, Smile: 73 center, 4 periphery, door: 48 center, 8 periphery.. taken at fifty centimeters from the monitor. After the preliminary phase, the formal phase started.

Formal experimental phase

The research assistant's instruction to each participant was: *“Now your task is to let your pupil dilation predict what you will see behind a closed door that will be shown in the center of the monitor. Behind the door you can see a gun shooting at you or a smile. The computer will monitor your pupil dilation and will predict for you what you will see. Remember that the choice of the shooting gun and the smile, is completely random and hence it is not possible to find an underlying rule to predict their sequence. The task consists in two sessions of 10 trials each. For each correct hit you will earn 0.5 euros”.*

Figure 1. Sequence of events related to the preliminary phase.

The sequence of events is presented in **Figure 2**. In this case pupil dilation was measured for 5 seconds during the fixation of the door and used to calculate the prediction accuracy (see Individual prediction accuracy paragraph).

Software and apparatus

Eye-Tracker Apparatus: The eye-tracker model Tobii T120®, Tobii, Stockholm, has the following technical characteristics: data rate, 120 Hz; accuracy, 0.5 degrees; freedom of head movements, 30 × 22 × 30 cm; monitor, 17 inch; 1280 × 1024 pixels; automatic optimization of bright-dark pupil tracking. PD is measured automatically in millimeters by the apparatus using the incorporated near infrared detectors and software. These data were fed to an original software for their storage. This program, created using E-Prime™ v2.0, written by two of the authors (MM and LS) and interfaced with the eye tracker, controlled events presentation and pupil size automatic recording. The source code is available at: [dx.doi.org/10.6084/m9.figshare.848604](https://doi.org/10.6084/m9.figshare.848604).

The sampling with replacement of the two stimuli in the two series of ten trials was randomized using E-Prime™ v2.0 randomized statement and Random function which was reset after every trial. This procedure guarantee against the possibility to guess the incoming stimulus by learning implicit and explicit rules.

The light in the laboratory was constantly dim, approximately 30 cd/m² to avoid undesired or unrelated changes to the participants' pupils.

The time necessary to complete the calibration, 2 min on average and give the instructions to participants was long enough to accommodate their pupils to the ambient light before starting the experiment.

Statistical methods

We used both a frequentist parameters estimation and a Bayesian model comparison approach, according to the American Psychology Association (APA)¹⁵, Kruschke¹⁶ and Wagenmakers, Wetzels, Borsboom and van der Maas's¹⁷ statistical recommendations.

This statistical approach is recommended to limit the shortcomings of the classical Null Hypothesis Significant Testing (e.g. Tressoldi *et al.*¹⁸). Basically, each parameter of interest (mean, correlation, etc.) is estimated for its precision by the confidence intervals, and its effect size or Bayes Factor. For those interested in the classical statistical significance with this approach, it sufficient to check if the confidence intervals include (not significant) or exclude (significant) zero.

Inferential frequentist estimates were applied both to the sum and the average of correct guesses (hits) using a binomial and a one-sample t-test statistical test respectively to take in account the sum and the percentages of hit responses. Confidence intervals were estimated using a bootstrap procedure based on 5000 samples.

Bayesian statistics

We adopted a model comparison approach contrasting the alternative hypothesis of a higher difference with respect to MCE (H1)

Figure 2. Sequence of events of the prediction phase.

with the Null Hypothesis (H_0) of a nil difference with respect to the MCE. We calculated the Bayes Factor ($BF_{H1/H0}$) using the software implemented by Morey and Rouder¹⁹ for the comparison with the one-tailed one-sample t-test, applying Jeffreys, Zellner, Siow (JZS) prior (see Jeffreys²⁰) setting an effect size of 0.3, as suggested by Rouder *et al.*²¹.

Preliminary data analysis

Before proceeding with the statistical analyses, the data for each participant were screened for artifacts. All artifacts, i.e. missing or anomalous (PD values close to/below 1, or above 10) data recordings related to PD easily detected by inspecting the raw scores saved in the individual files, were eliminated. If they exceeded the threshold of 60%, that is 12 out of 20 trials, the entire participant was excluded and substituted to keep the total sample equal to 100. The overall percentage of artifacts was 4%. The full raw data and corrected for anomalous data are available at <http://dx.doi.org/10.6084/m9.figshare.818978> (Tressoldi)²².

Individual prediction accuracy

In order to take into account individual differences, we standardized the PD values related to the 20 trials measured in the anticipation phase to z scores for each participant. Next, the means associated with the two stimuli chosen by the software were calculated. In this way a mean was always above zero and the second one below zero except in the case of identical mean between the two stimuli. The prediction for each trial was obtained simply by defining whether the value of PD, above or below zero corresponded to the stimulus that was chosen randomly. For example, if the PD standardized means associated with the smile and the gun were respectively 0.25 and -0.15, each PD value above zero predicted a smile and vice versa, each value below zero predicted a gun. At the end of the trial, the sum and the percentage of hits (correct predictions) were calculated for each participant.

Results

Prediction accuracy

In Table 1 we report the descriptive statistics, in Figure 3 the hits percentages with their 95% Confidence Intervals (CIs), and in Table 2

Table 1. Descriptive statistics of hits percentage for the two stimuli and overall. (n=100).

Hits	Smile	Gun	Overall
Mean %	0.549	0.569	0.559
SD	0.28	0.16	0.94
SUM	508/931	554/986	1062/1917

the effect sizes estimation and the $BF_{H1/H0}$ of the two stimuli and overall with respect the MCE with the corresponding 95% CIs.

The means estimate related to both stimuli and overall, show clearly that the prediction accuracy is above the mean chance expected of 50%.

The estimates of all effect sizes, both those referred to the binomial test and to the one-sample t-test, are above zero and in the range of medium effects. Furthermore the BF values range from 6 for the smile to 284394 for the overall accuracy.

Predictors of hits percentages

It is plausible to expect a correlation in the difference between the anticipatory PD associated with the two stimuli and their prediction accuracy. The variance explained by this correlation is $R^2=0.348$, 95%CI: 0.20 to 0.49. This moderate correlation suggests that anticipatory PD differences between the two stimuli explain only a part of the hits or correct predictions. This finding will be commented on further after the results of the exact replication.

Expectation bias control

Expectation bias, arises when a random sequence including multiple repetitions of the same stimulus type (e.g., five non-arousing stimuli) produces an expectation in the participant that the next stimulus should be of another type (e.g., an arousing stimulus) and the contrary (the Gambler's Fallacy). In the Figure S1 in the Supplementary Material we report the trend observed for the two types of targets. Ninety-eight percent of all series of identical stimuli

Figure 3. Hits Percentages with corresponding Confidence Intervals.

Table 2. Parameters estimation. Effect sizes with 95% CIs and $BF_{H1/H0}$ values of hits percentage for the two stimuli and overall with respect the MCE = 50%.

	Smile	Gun	Overall
Binomial z	2.48	3.61	4.34
$ES(z/\sqrt{n})$	0.24 [0.04,0.44]	0.36 [0.16,0.56]	0.43 [0.22,0.63]
One Sample t	2.78	3.94	5.98
$ES d (t/\sqrt{n-1})$	0.28 [0.08,0.48]	0.39 [0.19,0.59]	0.60 [0.39,0.81]
$BF_{H1/H0}$	6±0.02%	154±0.07%	284394±0.01%

ES=effect size; BF=Bayes Factor

were comprised between 1 and 5. The visual inspection supported by the estimate of the linear trend, -0.0071 for the smile and -0.0128 for the gun, excludes an expectation bias for both the stimuli.

General comment

The overall prediction accuracy turned out above 50%, the chance expected. Even if of small magnitude in absolute terms, approximately 5%, the parameter estimates suggest that this is quite substantial in term of effect size and BF.

Before commenting further on these findings we wanted to test their reliability in an exact replication of the experiment.

Exact replication method

Experiment registration

Following the suggestions of Wagenmakers *et al.*²³ and of the Open Science Collaboration²⁴, the experiment was registered on the site <http://www.openscienceframework.org> before data collection.

Participants

We preplanned to recruit the same number of participants as in the original study assuming a similar effect size and setting the statistical power to 0.80. The final sample recruited as in the first experiment, comprised 26 males and 74 females with a mean age of 23.02 with an associated standard deviation of 2.7.

Procedure

Identical to the original study. The overall percentage of artifacts was 7.3%.

Results

Overall prediction accuracy

In Table 3 we report the descriptive statistics, in Figure 4 the hits percentages with their 95% CIs and in Table 4 the effect sizes estimation and the $BF_{H1/H0}$ of the two stimuli and overall with respect the MCE with the corresponding 95% CIs.

The means estimates related to both stimuli and overall, show clearly that the prediction accuracy is above the mean chance expected of 50%.

The estimates of all effect sizes, both those referred to the binomial test and to the one-sample t-test, are above zero and in the range

Table 3. Descriptive statistics of hits percentage for the two stimuli and overall (n=100).

Hits	Smile	Gun	Overall
Mean %	0.578	0.596	0.587
SD	0.16	0.14	0.09
SUM	533/943	536/911	1069/1854

Figure 4. Hits Percentages with corresponding Confidence Intervals.

Table 4. Inferential statistics. Effect sizes with 95% CIs and $BF_{H1/H0}$ values of hits percentage for the two stimuli with respect the mean chance expected, 50%.

	Smile	Gun	Overall
Binomial z	3.97	5.3	6.57
$ES(z/\sqrt{n})$	0.39 [0.19,0.59]	0.53 [0.32,0.74]	0.65 [0.43,0.86]
One Sample t	4.7	6.5	9.8
$ES d (t/\sqrt{n-1})$	0.48 [0.27,0.69]	0.65 [0.43,0.86]	0.97 [0.73,1.21]
$BF_{H1/H0}$	2317 $\pm$ 0.13%	3159414 $\pm$ 0.02%	1.5×10^{13}

of medium to large effects. Furthermore the BF values range from 2317 for the smile to 1.5×10^{13} for the overall accuracy.

Correlation between PD and hit percentages

The correlations between the difference between the anticipatory PD associated with the overall prediction accuracy was $R^2=0.42$, 95%CI: 0.27,0.56, overlapping that observed in the original experiment.

Expectation bias

The same analysis used in the original experiment yielded similar results (see Figure S2 in the Supplementary Material), showing no sign of expectation bias.

Comparison with the original study

In this replication, the hit percentages of the two stimuli and the overall hit percentage are slightly larger, as well as the effect sizes and BFs estimates, confirming the results of the original study. The BFs in this case are superior to those observed in the original study and in the range of extreme evidence according to Jeffreys²⁰ criteria.

The difference between the anticipatory PD associated with the two stimuli predicts approximately one third of the variance related to the overall accuracy. At present we do not have hypotheses about which other predictors can contribute to the remaining variance.

General discussion

The results of the two experiments support the idea that PD can predict future random stimuli therefore adding more evidence to the findings reported by Tressoldi *et al.*¹³ using auditory stimuli.

Even if these results were due to unpredicted methodological or statistical artifacts, as in all experiments, we can rule out that in our case they are due to an improper randomization algorithm, the characteristics of our participants or a fault detection of PD by our apparatus.

The observed estimated prediction accuracy is between 5 to 10% above the chance level expected. It remains to be explored whether the 5 to 10% above chance represents an upper limit of this prediction or whether it can be enhanced.

It seems then that PD can be implicitly (unconsciously) modulated to predict random and hence statistically “unpredictable” events. Even if of small magnitude, this predictive ability could have important adaptive consequences, for example in cases of serious threats to life, suggesting that this characteristic is another expression of the powerful adaptive functions of our psychophysiological system that can anticipate future events, for example by promoting advantageous decision-making (e.g. Denburg *et al.*²⁵), anticipating a reward (e.g. Hackley *et al.*²⁶) and the pain of others (e.g. Caes *et al.*²⁷).

It seems that PD, like possibly all other apparatuses regulated by the psychophysiological system (i.e. heart rate, skin conductance level), has innate characteristics specifically dedicated to the anticipation of future events, no matter how predictable, extending the potentialities of survival mechanisms.

The investigation of the mechanisms at the base of this capacity is an open question. If random events cannot be predicted using previous experiences and information, we can argue that some “guessing” mechanisms based on probabilistic estimations are adopted. For example, we are currently testing if our results can be modeled using the Bayesian hierarchical generative model suggested by Mathys, Daunizeau, Friston and Stephan²⁸ to investigate individual learning under uncertainty.

Another hypothesis is the possibility that our psychophysiological system can manifest a sort of temporal quantum-like entanglement, restoring brief time-symmetry situations, apparently violating the past to future flow of time, allowing a connection between present and immediately future events like those observed in quantum physics (e.g. Ma *et al.*²⁹) and recently studied in the perception of ambiguous images by Atmanspacher and Filk³⁰.

The fact that it is possible to study this characteristic at the level of a single trial taking into account individual differences, opens up the possibility of devising proof of concept experiments for potential possible applications to be adopted in real life. For example, it is not very complicated to devise technical devices to add to glasses or a smartphone for instance, that can amplify the subtle variations

of PD at the level of a conscious overt detection in a way to be used intentionally by every person or to use these variations to activate automatically an alarm that could enhance personal safety when driving or walking.

However only independent exact or conceptual replications can support our findings, for example changing the type of stimuli and/or the randomization algorithm.

Author contributions

Authorship: P. Tressoldi developed the study concept. M. Martinelli and L. Semenzato, devised the software. All authors contributed to the study design. Testing and data collection were performed by two research assistants. P. Tressoldi performed the data analysis and

interpretation. All authors approved the final version of the paper for submission.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

We acknowledge the relevant advices offered by Dick Bierman, Eva Lobach, Dean Radin and Julia Mossbridge in particular on how to analyse the anticipatory signals and Maaïke Pols for revising the English.

Supplementary materials

Table S1. Descriptive and inferential statistics obtained in Tressoldi *et al.* 2011.

n.80	Alerting sounds	Neutral sounds
Mean [95%CI]	60.3 [53.6,67]	44.6 [38.2,51]
SD	3.06	2.9
SUM	482	357
Binomial z	5.76	3.01
ES ($z/\sqrt{n}$)	0.20 [-0.02,0.42]	0.11 [-0.11,0.33]
One Sample t	3.05	1.67
ES d ($t/\sqrt{n-1}$)	0.33 [0.10,0.55]	0.18 [-0.4,0.40]
BF _{H1/H0} *	3225	7.7

*Using Rouder *et al.* (2009) method, setting the scale of effect size $r=0.3$.

Figure S1. Expectation Bias control of Experiment 1. Error bars represent the 95% confidence intervals.

Figure S2. Expectation Bias of Exact replication of Experiment 1. Error bars represent the 95% confidence intervals.

References

1. Friston K: **The free-energy principle: a unified brain theory?** *Nat Rev Neurosci.* 2010; **11**(2): 127–138.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. van Boxtel GJM, Böcker KBE: **Cortical measures of anticipation.** *J Psychophysiol.* 2004; **18**(2–3): 61–76.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Barcelo F, Bestmann S, Yu AJ: **Cognitive control, uncertainty and the brain.** Research Topic, Frontiers in Human Neuroscience. 2010.
[Reference Source](#)
4. Butz MV, Sigaud O, Baldassarre G: **Anticipatory Behavior in Adaptive Learning Systems.** From Brain to Individual and Social Behavior. *Berlin Springer.* 2007; **4520**: 379.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Clark A: **Whatever next? Predictive brains, situated agents, and the future of cognitive science.** *Behav Brain Sci.* 2013; **36**(3): 181–204.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Mossbridge J, Tressoldi P, Utts J: **Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis.** *Front Psychol.* 2012; **3**: 390.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Bradley MM, Miccoli L, Escrig MA, *et al.*: **The pupil as a measure of emotional arousal and autonomic activation.** *Psychophysiology.* 2008; **45**(4): 602–607.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Beatty J, Lucero-Wagoner B: **The pupillary system.** In J.T. Cacioppo, L.G. Tassinari, & G.G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed.). New York: Cambridge University Press. 2000; 142–162.
[Reference Source](#)
9. Steinhauer SR, Siegle GJ, Condray R, *et al.*: **Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing.** *Int J Psychophysiol.* 2004; **52**(1): 77–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Bijleveld E, Custers R, Aarts H: **The unconscious eye opener: pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues.** *Psychol Sci.* 2009; **20**(11): 1313–1315.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Einhäuser W, Stout J, Koch C, *et al.*: **Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry.** *Proc Natl Acad Sci U S A.* 2008; **105**(5): 1704–1709.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Einhäuser W, Koch C, Carter O, *et al.*: **Pupil dilation betrays the timing of decisions.** *Front Hum Neurosci.* 2010; **4**: 18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Tressoldi PE, Martinelli M, Semenzato L, *et al.*: **Let Your Eyes Predict: Prediction Accuracy of Pupillary Responses to Random Alerting and Neutral Sounds.** *SageOpen.* 2011; **1**(2): 1–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Faul F, Erdfelder E, Lang AG, *et al.*: **G*Power 3: a flexible statistical power analysis for the social, behavioral, and biomedical sciences.** *Behav Res Methods.* 2007; **39**(2): 175–191.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. **Reporting standards for research in psychology: why do we need them? What might they be?** APA Publications and Communications Board Working Group on Journal Article Reporting Standards. *Am Psychol.* 2008; **63**(9): 839–851.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Kruschke J: **Introduction to Special Section on Bayesian Data Analysis.** *Perspect Psychol Sci.* 2011; **6**(3): 272–273.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Wagenmakers EJ, Wetzels R, Borsboom D, *et al.*: **Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011).** *J Pers Soc Psychol.* 2011; **100**(3): 426–432.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Tressoldi PE, Giofrè D, Sella F, *et al.*: **High impact = high statistical standards? Not necessarily so.** *PLoS One.* 2013; **8**(2): e56180.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Morey RD, Rouder JN: **Package BayesFactor.** 2013.
[Reference Source](#)
20. Jeffreys H: **Theory of probability.** Oxford, England. Oxford University Press. 1961.
[Reference Source](#)
21. Rouder JN, Speckman PL, Sun D, *et al.*: **Bayesian t-tests for accepting and rejecting the null hypothesis.** *Psychon Bull Rev.* 2009; **16**(2): 225–237.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Tressoldi P: **PDAnticipationDatabase.** Figshare. 2013.
[Publisher Full Text](#)
23. Wagenmakers EJ, Wetzels R, Borsboom D, *et al.*: **An agenda for purely confirmatory research.** *Perspect Psychol Sci.* 2012; **7**(6): 632–638.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Open Science Collaboration. **An open, large-scale, collaborative effort to estimate the reproducibility of psychological science.** *Perspect Psychol Sci.* 2012; **7**(6): 657–660.
[Publisher Full Text](#)
25. Denburg NL, Recknor EC, Bechara A, *et al.*: **Psychophysiological anticipation of positive outcomes promotes advantageous decision-making in normal older persons.** *Int J Psychophysiol.* 2006; **61**(1): 19–25.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Hackley SA, Muñoz MÁ, Hebert K, *et al.*: **Reciprocal modulation of eye-blink and pinna-flexion components of startle during reward anticipation.** *Psychophysiology.* 2009; **46**(6): 1154–1159.
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Caes L, Uzieblo K, Crombez G, *et al.*: **Negative emotional responses elicited by the anticipation of pain in others: psychophysiological evidence.** *J Pain.* 2012; **13**(5): 467–476.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Mathys C, Daunizeau J, Friston KJ, *et al.*: **A bayesian foundation for individual learning under uncertainty.** *Front Hum Neurosci.* 2011; **5**: 39.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Ma XS, Zotter S, Kofler J, *et al.*: **Experimental delayed-choice entanglement swapping.** *Nat Phys.* 2012; **8**: 479–484.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Atmanspacher H, Filk T: **A proposed test of temporal nonlocality in bistable perception.** *J Mathematical Psychol.* 2010; **54**(3): 314–321.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ?

Version 2

Reviewer Report 13 June 2014

<https://doi.org/10.5256/f1000research.4388.r5116>

© 2014 Baker C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chris Baker

Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA

In their response to my previous comments, the authors have clarified that only the data from the “Experimental phase” were used to calculate prediction accuracy. However, if I now understand the analysis procedure correctly, there are serious concerns with the approach adopted.

First, let me state what I now understand the analysis procedure to be:

1. For each subject the PD values across the 20 trials were converted to z-scores.
2. For each stimulus, the mean z-score was calculated.
3. The sign of the mean z-score for each stimulus was used to make predictions.
4. For each of the 20 trials, if the sign of the z-score on that trial was the same as for the mean z-score for that stimulus, a hit (correct prediction) was assigned. In contrast, if the sign of the z-score on that trial was the opposite as for the mean z-score for that stimulus, a miss (incorrect prediction) was assigned.
5. For each stimulus the total hits and misses were calculated.
6. Average hits (correct prediction) for each stimulus was calculated across subjects.

If this is a correct description of the procedure, the problem is that the same data were used to determine the sign of the z-score that would be associated with a correct prediction and to determine the actual correct predictions. This will effectively guarantee a correct prediction rate above chance.

To check if this is true, I quickly generated random data and used the analysis procedure as laid out above (see MATLAB code below). Across 10,000 iterations of 100 random subjects, the average

“prediction” accuracy was ~57% for each stimulus (standard deviation, 1.1%), remarkably similar to the values reported by the authors in their two studies. In this simulation, I assumed that all subjects contributed 20 trials, but in the actual data analyzed in the study, some subjects contributed fewer than 20 trials due to artifacts in the pupil measurements.

If the above description of the analysis procedure is correct, then I think the authors have provided no evidence to support pupil dilation prediction of random events, with the results reflecting circularity in the analysis procedure.

However, if the above description of the procedure is incorrect, the authors need to clarify exactly what the analysis procedure was, perhaps by providing their analysis scripts.

MATLAB code:

```
nTrials = 10; % 10 trials for each stimulus/condition

for boot = 1:10000 % 10,000 iterations for bootstrapping

    for i = 1:100 % 100 subjects

        data = randn(nTrials,2); % generate random values for each trial

        meandata = squeeze(mean(data(:))); % calculate mean
        stddata = std(data(:)); % calculate standard deviation

        zdata = (data - meandata)/(stddata); % convert to z-scores

        meancond1 = mean(zdata(:,1)); % calculate mean for each stimulus/condition
        meancond2 = mean(zdata(:,2));

        if meancond1 > 0 % evaluate sign of the mean values
            conscore = 1; % conscore indicates for which condition, positive z-values will indicate
correctness
            conscoreB = 2; % conscoreB indicates for which condition, negative z-values will indicate
correctness
        elseif meancond2 > 0
            conscore = 2;
            conscoreB = 1;
        else
            error = 'They are equal' % if mean z-values are equal, arbitrarily assign correctness
            conscore = 1;
            conscoreB = 2;
        end

        accScores(i) = sum(squeeze(zdata(:,conscore))>0)./nTrials; %calculate average correct for each
condition for each subject
        accScoresB(i) = sum(squeeze(zdata(:,conscoreB))<0)./nTrials;
```

```
end

mAcc(boot) = mean(accScores); % calculate average correct for each condition across subjects
for each iteration
    mAccB(boot) = mean(accScoresB);

end

meanBoot = mean(mAcc)% calculate mean correct for each condition across iterations
meanBootB = mean(mAccB)

stdBoot = std(mAcc)% calculate standard deviation for each condition across iterations
stdBootB = std(mAccB)
```

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 13 Jun 2014

Patrizio Tressoldi, Università di Padova, Padova, Italy

Dear Chris,

Many thanks indeed for the time you have been dedicating to our study. We first are interested to know if what we found is a statistical artifact or a real phenomenon. Your description of the prediction procedure is correct but it is necessary to consider PD was measured before the presentation of stimuli. How is it possible for the PD differences to be associated with future stimuli above chance?

I was surprised at the results obtained with your script. I also ran your script and apart from the multiple simulations results which were almost identical with minor differences in the 3rd and 4th decimals, I expected a theoretical mean of 0.50. Am I missing anything?

If I change lines 17-22 as follows:

```
if meancond1 > 0 % evaluate sign of the mean values
    conscore = 2;
    conscoreB = 1;
elseif meancond2 > 0
    conscore = 1;
    conscoreB = 2;
```

I obtain mean values of 0.42, complementary to 0.57.

Competing Interests: No competing interests were disclosed.

Author Response 19 Jun 2014

Patrizio Tressoldi, Università di Padova, Padova, Italy

I found the error in the MatLab code that yielded the above or below chance results.

Old version:

```
if meancond1 > 0 % evaluate sign of the mean values
    conscore = 1; % conscore indicates for which condition, positive z-values will indicate
correctness
    conscoreB = 2; % conscoreB indicates for which condition, negative z-values will
indicate correctness
elseif meancond2 > 0
    conscore = 2;
    conscoreB = 1;
```

Correct version:

```
if meancond1 > 0 % evaluate sign of the mean values
    conscore = 2;
    conscoreB = 1;
elseif meancond2 < 0
    conscore = 1;
    conscoreB = 2;
```

Competing Interests: I am the corresponding author

Reviewer Report 13 June 2014

<https://doi.org/10.5256/f1000research.4388.r5109>

© 2014 McHugh L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Leo McHugh

Immunexpress Inc., Seattle, WA, USA

1. **Artifact removal:** Unfortunately the authors have not updated the paper with a 2x2 table showing guns and smiles by removed data points. This could dispel criticism that an asymmetrical expectation bias that has been shown to exist in similar experiments is not driving a bias leading to inappropriate conclusions. This is my strongest criticism of the paper and should be easily addressed as per my previous review comment. The fact that this simple data presentation was not performed to remove a clear potential source of spurious results is disappointing.
2. The authors have added 95% CIs to figures S1 and S2. This clarifies the scope for

expectation bias in these data. The addition of error bars permits the authors' assumption of a linear trend, indicating that the effect of sequences of either guns or smiles may not skew results. Equally, there could be either a downwards or upwards trend fitting within the confidence intervals that could be indicative of a cognitive bias that may violate the assumptions of the authors, leading to spurious results. One way to remove these doubts could be to stratify the analyses by the length of sequences of identical symbols. If the results hold up in each of the strata, this potential bias could be shown to not be present in the data. If the bias is strong, particularly in longer runs, this could indicate that the positive result was due to small numbers of longer identical runs combined with a cognitive bias rather than an ability to predict future events.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 16 Jun 2014

Patrizio Tressoldi, Università di Padova, Padova, Italy

1. Artifact removal. In the previous reply we clarified that "artifacts are completely random and due to anomalous or missing values, and not to participants' bias toward one specific stimulus". We remind that the randomisation of the two stimuli presentation was with replacement and hence for each participant the number of the two stimuli was completely different. We do not see any reason why technical artifacts could be associated with a particular stimulus type.

2. *"..there could be either a downwards or upwards trend fitting within the confidence intervals that could be indicative of a cognitive bias that may violate the assumptions of the authors, leading to spurious results. One way to remove these doubts could be to stratify the analyses by the length of sequences of identical symbols."*

Here we add the n, means and sds of the sequences above five:

Exp1

Smile:

length 6: n=10; m = 0.40; sd = 0.51

length 7: n=5; m = 0.40 ; sd = 0.54

length 8: n = 2; m = 0 ; sd = 0

Gun:

length 6: n=12; m = 0.25; sd = 0.45

length 7: n=6; m = 0.50 ; sd = 0.55.

Exp2

Smile:

length 6: n=9; m = 0.67; sd = 0.50

Gun:

length 6: $n=9$; $m = 0.78$; $sd = 0.44$

length 7: $n=7$; $m = 0.43$; $sd = 0.53$

length 8: $n=3$; $m = 0.67$; $sd = 0.57$

length 9: $n=1$; $m = 1$; $sd =$

length 10: $n=1$; $m = 1$; $sd =$

Even with these more rare sequences, it doesn't emerge that responses are constantly correct to justify the overall results. Is it what you meant as "stratify the analyses"?

Competing Interests: I am the corresponding author

Version 1

Reviewer Report 23 April 2014

<https://doi.org/10.5256/f1000research.2733.r4545>

© 2014 **McHugh L.** This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Leo McHugh

Immunexpress Inc., Seattle, WA, USA

This paper presents evidence that the average person can predict random events better than chance in the context of the experimental setup described by the authors. This is a remarkable result if true. Sociologist Marcello Truzzi coined the aphorism "extraordinary claims require extraordinary proof", from which standpoint a review of any such data must be based. It appears that the sample sizes, study design and statistical methods used are generally appropriate throughout. However I can think of a number of potential artifacts of data collection and processing in this experiment not addressed in this paper that may produce positive results inappropriately. These potential artifacts could be tested statistically within the existing dataset (i.e. without further experimentation), and the paper would be strengthened if each of these potential sources for false positive results could be eliminated.

This paper is supported by a nice dataset and it would be good to see the paper refined through exposure in F1000 to accommodate data excluding other sources of bias as others read and comment on the paper to the authors. Ultimately it would be great to see a well-powered dataset (such as this one), pointing to an uncomfortable conclusion, that has addressed all the criticisms of the community and becomes an outstanding challenge for refutation.

Expectation bias control:

The authors report that 98% of random sequences have identical sequential elements of 5 or less, and show a constant hits percentage over lag to support this argument in Figure S1. This constant rate is an important assumption in the paper because assuming a linear trend removes the requirement to correct for gamblers fallacy, as pointed out by the authors.

We would expect wider confidence intervals around larger lag values shown in S1 because these events are less frequent. It would be useful if the authors could include confidence intervals around these points else it could be argued that the 'linear' trend is merely point values within ever widening confidence intervals that could equally support a non-constant hits percentage. An example of the consequence of a non-constant hits percentage is that because of the smaller numbers of such events at higher lags, chance, especially if combined with a participants' inner bias to select more often a gun or smile, may have a small effect on the overall hits accuracy. Admittedly, this would at most be only a few percentage points, but for a result a few percent above chance, this could be important.

As a caveat, it would also be nice to see second order data, say looking at participant's likelihood of selecting the same class 2 or 3 steps later. This is analogous to the gambler's fallacy in its potential effect but is not discussed by the authors and no data is shown.

Removed artifacts:

The authors report 4% of trials removed due to artifacts. It would be helpful if the authors would report the number of artifacts removed for each corresponding random event. A significant imbalance would immediately be indicative of bias. False conclusions may be reached under these conditions. For example, if the smile produced a PD which was more likely to be removed as an artifact leaving an excess of guns for analysis, and the participants in general had a bias towards expecting a gun over a smile, as is supported by evidence for prospect theory (Khaneman), then the result would be a higher than 'chance' hits percentage merely because participants are naturally biasing towards selection of a more prevalent class. Again, this is unlikely to be more than a few percent but may still be important in this context. One way to easily show this bias as non-existent in this data would be to remove the artifacts, and randomly remove trials so that the number of guns and smiles is the same for analysis, then re-run the numbers.

Minor comments (typos):

- "In this way a mean was always above zero and the second one below zero unless of an identical mean between the two stimuli." Unless? Perhaps should be 'except in the case of an identical mean...'
- "This finding will be commented on further after the results of the exact replication"

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 24 Apr 2014

Patrizio Tressoldi, Università di Padova, Padova, Italy

First of all, many thanks for your interest and useful suggestions to our paper.

As suggested, in our revised article we have added confidence intervals to the expectation bias measures. As you predicted, the decreasing number of data as the lag increases makes the estimates of the average prediction more wide.

As for whether the artefacts could be unbalanced between the two stimuli, they are completely random and due to anomalous or missing values, and not to participants' bias toward one specific stimulus.

Competing Interests: I am the correspondent author of the paper

Reviewer Report 28 March 2014

<https://doi.org/10.5256/f1000research.2733.r4300>

© 2014 Baker C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chris Baker

Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD, USA

At first glance, this appears to be a thorough demonstration that pupil diameter can be predictive of upcoming events. The study is a conceptual replication of an earlier study by the same group (Tressoldi *et al*, 2011) and the manuscript contains two separate experiments, each with 100 participants. Further, the second experiment was registered at <http://www.openscienceframework.org> before data collection (although at the time of writing this, I was unable to access any public record of the registration on the web site). Finally, the authors analyzed their results using both frequentist and Bayesian approaches, producing similar results.

However, the claim being made - that somehow pupil responses can predict upcoming random events - would be incredibly remarkable (if true). Given such a bold, and some would say impossible claim, it is critical to carefully examine the experimental design and the veracity of the results. To their credit, the authors have made much of the raw data available as well as the source code for the experiments.

Here, I'm going to just focus on evaluating the specific methods and analyses in this particular study, ignoring any evidence from prior work.

Briefly, each experiment was composed of two phases:

1. Preliminary Phase - The experimenters first measured individual participants' pupil responses to two visual stimuli, a smiley face and a pointed gun (accompanied by the sound of a shot), which appeared following the presentation of a picture of a door. There were 10 trials of each condition presented in a random order (without replacement) and pupil diameter was measured during the presentation of the second stimulus (smiley or gun).
2. Experimental Phase - In two blocks of ten trials, participants were again presented with a picture of a door followed by a smiley or a gun, but the trials were fully randomized (no replacement) and pupil diameter was now measured during the period when the door was on the screen.

The authors' claim is that in the Experimental Phase, the pupil diameter measured during the presentation of the door is predictive of the upcoming (random) stimulus.

Unfortunately, the method for generating and testing predictions is not entirely clear. My understanding is that for each phase of the experiment, they separately z-scored the data across all (both smiley and gun) trials. Then for the Preliminary Phase data, they determined whether the average z-score for smileys and guns was above or below zero. This was necessary because it appears there were individual differences in the relative pupil size for the two stimuli. The prediction for each Experimental Phase trial was then generated by determining whether the z-score for that trial was above or below zero, and assigning the corresponding stimulus that had a positive or negative mean in the Preliminary Phase as the predicted stimulus.

This method does not generate trial-by-trial predictions since it simply assesses whether the pupil diameter for a given stimulus tended to be in the same half of the data in both the Preliminary and Experimental Phase.

Using this method the authors then generated percentage hits, comparing the predicted and actual stimuli, and determined that they were significantly above chance (for both stimuli considered separately and combined) using both frequentist and Bayesian approaches.

All results were replicated in the second (pre-registered) experiment.

Concerns:

There are many issues that could be raised about this study, questioning whether the result should be taken to reflect unconscious prediction of random events (many, including myself, would be unwilling to accept this interpretation) or whether it is an artifact of some experimental procedure. I will admit that I don't have a ready explanation for what sort of artifact can have produced these results (across two separate experiments), but it's worth making a few points:

1. Replication, in and of itself, does not prove this is a real effect. If there is a methodological or statistical artifact, this would readily replicate too. Replication is much more valuable when conducted by independent research groups.
2. The authors could do much more to convince a skeptic as to the quality of their underlying data. As someone who has recently started to use pupil diameter as an experimental measure, I've become aware of how noisy this measure can be from trial-to-trial and I find it

extremely hard to believe that the signals will be sufficiently robust across such a small number of trials for any kind of subtle prediction, whether it be of a physical stimulus or an upcoming random event (as is claimed here). The authors have provided no information to attest to the quality of their pupil recordings. What is particularly surprising to me is that the pupil diameter data is from completely different time periods of a trial between the Preliminary and Experimental Phases, when there are great differences in the luminance of the stimuli that are on the screen at that time. What I would like to see is much more careful presentation and analysis of the raw pupil data. For example, time course plots of individual trials for both experimental phases, assessments of trial-to-trial variance etc.

3. Aside from the lack of clarity on the method for generating predictions, there are a couple of other details that need to be more fully specified.
 1. It's not clear exactly what luminance measurements were taken of the stimuli and how they were "calibrated for luminance". How were the center and periphery measurements taken?
 2. Were subjects asked to fixate, or were they free to move their eyes?
 3. Over what time period was the pupil diameter measured in each phase? Across the whole 5-second periods?
 4. What accounts for the individual differences in relative pupil diameter to the same exact stimuli with the same luminance properties?

Providing this sort of information is critical if any other group is to try and replicate the current findings.

Ultimately, these results do not convince me that it is possible for participants to unconsciously predict future random events. I view the results as a curiosity and suspect that there is some underlying methodological or analytical artifact that can explain the results – but this is something that may be very hard to pick up from the manuscript, even though the authors have provided much of the raw data. I remember a story a former postdoc of mine told me about a behavioral experiment he ran, in which participants had to indicate whether two sequential stimuli were the same or different. He found that some subjects were incredibly fast and accurate at the task and couldn't work out how it was possible until one of his participants told him that he was listening to the sound of the computer disk spinning – on different trials the computer would have to spin and load up a second image!

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 24 Apr 2014

Patrizio Tressoldi, Università di Padova, Padova, Italy

Many thanks for your interest and useful comments to our paper.

We completely agree that a more "convincing" support to our findings can derive only by independent replications. However we remind that in the cited meta-analysis by [Mossbridge et al. \(2012\)](#), there are independent replications, even if the basic experimental protocol is limited to the averaging of the anticipatory physiological variables.

As to the differences in PD between the Preliminary and the Experimental phases, as explained on page 2 and 3, in the Preliminary phase PD was measured during the two stimuli presentation, whereas in the Experimental phase PD was measured before their presentation with participants fixating the door image. Furthermore, on page 2 we clarified that the Preliminary phase was used to familiarise the participants with the task, and not to obtain data for the prediction. In our revised article we have added "Anticipation phase" in Figure 2, and the sentence *"In this case pupil dilation was measured for 5 seconds during the fixation of the door and used to calculate the prediction accuracy"* to clarify that only these data were used to predict the future stimuli.

The individual differences with the same exact stimuli and the same luminance properties are expected, given the individual differences in the sympathetic and parasympathetic reactivity.

Competing Interests: I am a corresponding author of the paper

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research