Look into my eyes: Pupillometry reveals that a post-hypnotic suggestion for word blindness reduces Stroop interference by marshalling greater effortful control

Benjamin A. Parris[1][*], Nabil Hasshim[1] and Zoltan Dienes[2]

1. Department of Psychology, Bournemouth University

2. Department of Psychology, University of Sussex

Dr. Benjamin A. Parris

Department of Psychology

Faculty of Science and Technology

Bournemouth University

Talbot Campus

Fern Barrow

BH12 5BB

Email: bparris@bournemouth.ac.uk

Abstract

The mechanisms underpinning the apparently remarkable levels of cognitive and behavioural control following hypnosis and hypnotic suggestion are poorly understood. Numerous independent studies have reported that Stroop interference can be reduced following a post-hypnotic suggestion that asks participants to perceive words as if made up of characters from a foreign language. This effect indicates that frontal executive functions can be more potent than is generally accepted and has been described as resulting from top-down control not normally voluntarily available. We employed eye tracking and pupillometry to investigate whether the effect results from voluntary visuo-attentional strategies (subtly looking away from the word to prevent optimal word processing), reduced response conflict but not overall conflict, Stroop effects being pushed from response selection to response execution (response durations) or increased proactive effortful control given enhanced contextual motivation (as indexed via pupil dilation). We replicated the reduction of Stroop interference following the suggestion despite removing any trials on which eye movements were not consistent with optimal word processing. Our data were inconclusive with regards to conflict type affected by the suggestion in the latency data, although preserved semantic conflict was evident in the pupil data. There was also no evidence of Stroop effects on response durations. However, we show that baseline corrected pupil sizes were larger following the suggestion indicating the socio-cognitive context and experimental demands motivate participants to marshal greater effortful control.

1. Introduction

Hypnosis usually follows an induction procedure that involves the hypnotist asking the person being hypnotised to concentrate on relaxing until they are eventually counted up (or down) into "hypnosis". By listening to the suggestions delivered by the hypnotist some individuals can produce an apparent form of control over their thoughts and actions not usually possible via more self-directed, effortful approaches. Under the influence of hypnotic suggestion (or post-hypnotic suggestion[1]) the experiences of pain (Derbyshire, Whalley, Stenger, & Oakley, 2004; Tan et al., 2015; Thompson et al., 2019), colour (Kosslyn, Thompson, Costantini-Ferrando, Alpert, & Spiegel, 2000), and delusions (Rahmanovic, Barnier, Cox, Langdon, & Coltheart, 2011) can be produced or extinguished (see Oakley & Halligan, 2009; 2013, for reviews). Hypnosis has been shown to modify frontal lobe activity (Crawford, 1996; Jamieson, Dwivedi, & Gruzelier, 2005; Rainville, Carrier, Hofbauer, Bushnell, & Duncan, 1999) and function (Farvolden & Woody, 2004; Gruzelier & Warren, 1993; Jamieson & Sheehan, 2004; Jamieson & Woody, 2007; Sheehan, Donovan, & Macleod, 1988; Terhune, Cleeremans, Raz & Lynn, 2017; Wagstaff, Cole, & Brunas-Wagstaff, 2007), although there is a debate as to whether responses to suggestion are the result of increased or decreased executive control (see Parris, 2017, for a review).

One of the more remarkable findings reported in the literature is the elimination of the Stroop interference effect following a post-hypnotic suggestion describing the

---

[1] A post-hypnotic suggestion is a suggestion given whilst 'under hypnosis' but not acted upon until the participant is no longer in a 'hypnotized state'. When the participant is in their normal 'non-hypnotized state' a cue is given (e.g. a clap) as a sign to activate the suggestion.

word dimension of the Stroop stimulus as being made up of 'meaningless symbols' and 'characters of a foreign language' (to be referred to as the *word blindness* suggestion; Raz, Shapiro, Fan & Posner, 2002). The Stroop interference effect (Stroop, 1935) refers to the finding that people are slower to name the colour that a word is printed in when the word denotes an incongruent colour (e.g., the word *red* in blue) compared to when naming the print colour of a colour neutral word (e.g., the word *top* in red), and has been referred to as the "gold standard" of attentional measures (MacLeod, 1992). Raz and colleagues (Raz et al., 2002; Raz et al., 2005) reported that both the Stroop interference effect and the activity in the anterior cingulate cortex that represents conflict processing in the Stroop task are substantially reduced following the post-hypnotic suggestion for word blindness. Importantly, the word blindness suggestion effect (WBSE), has been replicated numerous times across different independent labs (e.g. Augustinova & Ferrand, 2012; Parris, Dienes, & Hodgson, 2012; Raz, Fan, & Posner, 2005).

Raz and colleagues (Raz et al., 2002; Raz et al., 2005) have argued that the WBSE reflects a level of top-down control over word reading through a means not *normally voluntarily available*; that is, a level of control conferred by post-hypnotic suggestion (or suggestion without hypnosis; see Parris & Dienes, 2013; Raz, Kirsch, Pollard, & Nitken-Kaner, 2006) that participants are unable or unlikely to marshal outside of the context of hypnosis and/or suggestion. Lifshitz, Aubert-Bonn, Fischer, Kasham and Raz (2013) have argued that this atypical control mechanism can modify automatic cognitive processes. However, the nature of this mechanism has not been elucidated. Given its implications for understanding the efficacy of cognitive and neural mechanisms

of control, the aim of the present study was to explore four alternative accounts of the WBSE based on normal mechanisms of control.

1.1.A motivational account.

Highly suggestible participants might be highly motivated to act as good hypnotic subjects and might therefore fully engage control mechanisms that are not normally fully engaged (De Jong et al., 1999). Highly suggestible participants could be motivated to proactively fully focus attention on the Stroop task following the suggestion; more so than they would under standard experimental conditions. In the Dual-Mechanisms of Control framework (Braver, 2012) proactive control refers to the effortful preparation and then maintenance of control mechanisms for ongoing goal-oriented behaviour (this is in contrast to reactive control which is preset, but is not maintained in memory, and is later triggered by predefined cues). If it were shown that the WBSE were the result of participants marshalling frontal executive functions to engage more effortful control under suggestion, it could be understood as an effect resulting from the motivational context in a similar way to Stroop interference reductions following financial incentives (Krebs, Boehler & Woldorff, 2010) and not from an atypical control mechanism (Lifshitz et al., 2013; Raz et al., 2002). Indeed, involuntary, effortless responding is thought to be a key marker of responding to hypnotic suggestions (Bowers, 1982; Weitzenhoffer, 1980), indicating that responding to suggestions does not involve effortful executive control mechanisms. There are however, reasons to reject the hypothesis that hypnotic responding is involuntary and effortless including evidence of regulation of hypnotic responses, and the demonstration of active attention-demanding attempts to fulfil the requirements of hypnotic suggestions (Lynn, Rhue & Weekes, 1990). Indeed, subjective

reports of involuntary responding could simply be a marker of demand characteristics and their effect on awareness of intentions (Dienes et al., in press).

1.2.A visuo-attentional account

Stroop interference is substantially reduced when attention is focussed at the end letter position (Besner et al., 1997; Parris, Sharma & Weekes, 2007; Sheehan, Donovan & MacLeod, 1988). One account of the WBSE therefore is that participants engage in strategic eye movements to focus their attention on the end letter to impair word processing. In an attempt to counter this explanation Raz et al. (2003) showed that the WBSE is observed even when active blurring of vision is temporarily impaired, and when video recordings show no attempts at moving the head or eyes or squinting to achieve the effect. However, noting that Raz et al.'s (2003) experimental manipulations would not have completely prevented blurring or been able to detect all eye movements, Palfi, Parris, and Dienes (submitted) showed that both looking-away from the irrelevant word and deliberate visual blurring substantially reduced Stroop interference indicating both as candidate strategies for producing the WBSE. However, neither strategy resulted in a reduction of reaction times to incongruent stimuli (indeed the RTs substantially increased), which they described as a key marker of the WBSE.

1.3. A response-selection (not word blindness) account

Augustinova and Ferrand (2012) showed that whilst the WBSE substantially reduced one form of conflict in the Stroop task known as response conflict, semantic conflict was not affected by the suggestion, which was interpreted as showing that semantic processing of the words continues, despite the suggestion for word blindness. Response conflict is typically measured by comparing RTs to Stroop stimuli that involve

response conflict to stimuli that do not, but do involve conflict at the level of semantics. For example, response set stimuli are those in which the irrelevant word denotes a colour that is also part of the response set (e.g. the word RED in blue where both red and blue are possible responses). Non-response set stimuli on the other hand do not involve irrelevant words denoting colours in the response set (e.g. the word ORANGE in blue where orange in not a response option). In non-response set stimuli words and colours are still incongruent colours and would compete at the level of semantic activation; they would not however strongly compete at the level of response selection. Response conflict can be measured by subtracting RTs to non-response set trials to those to response set trials. Semantic conflict can be measured by subtracting RTs to non-colour word neutral trials (e.g. the word TABLE in red) from RTs to non-response set trials (Hasshim & Parris, 2017). Thus, the data from Augustinova and Ferrand (2012) indicate that participants are not engaging the suggested strategy of inducing word meaninglessness but are instead targeting response level conflict. However, Augustinova and Ferrand (2012) did not report evidence for no effect on semantic conflict, merely a non-significant effect. Furthermore, the effect has not yet been replicated.

1.4.A response execution account

One of the interesting findings regarding the WBSE is that response times are substantially reduced in all Stroop conditions following the suggestion (Parris et al., 2012; Parris & Dienes, 2013; Parris, Dienes, Bate & Gothard, 2014; Raz et al., 2002; 2005; Raz & Campbell, 2011; although see Augustinova & Ferrand, 2012; Raz et al., 2003; Raz, Kirsch, Pollard & Nitkin-Kaner, 2006, in which reductions occur for some but not all Stroop conditions) indicating a modified response threshold. Previous work has

indicated that faster responses in the Stroop task can lead to absent Stroop effects at the level of response selection but also that the effects get pushed into response execution (Kello, Plaut & McWhinney, 2000). Kello et al. (2000) initially also observed a Stroop interference effect on response selection (~110ms on vocal naming latencies) but no Stroop effects on response execution (vocal naming durations; i.e., the time it takes to actually vocalise the response "blue"). In a follow up experiment, however, in which they introduced a fast response deadline of 575ms, the Stroop interference effect was substantially reduced to ~70ms and they observed a Stroop interference effect of ~45ms on response durations. Thus, if, in an attempt to perform better following the suggestion, participants set a different response threshold (they speed up their responses) it is possible that the missing Stroop effects (the WBSE) results from the effects being pushed into response durations.

1.5 Testing the accounts

In the present study we employed similar methods to Raz et al. (2002; 2005) to produce the WBSE with a few modifications that permitted us to test the effort-, visuo-attentional-, response level, and response execution accounts of the effect. Instead of the manual Stroop task employed in previous studies (Augustinova & Ferrand, 2012; Parris, Dienes, & Hodgson, 2012; Raz et al., 2002; 2005) we had participants complete an oculomotor version of the Stroop task (Hasshim & Parris, 2015; Hermans & Walker, 2012; Hodgson, Parris, Gregory & Jarvis, 2009; Singh & Mishra, 2012). In this version of the Stroop task, participants respond by looking at the on-screen patch that corresponds to the colour of the Stroop stimulus. Any eye movements made in a direction away from the target patch are counted as errors. Importantly, participants need to fixate

on a central fixation stimulus to trigger the appearance of the Stroop stimulus, meaning that their fixation will be on the centre of the Stroop stimulus when it appears. Moreover, any eye movements that are small enough to represent a saccade from the middle of the word to near its end letter (roughly 1-2° in amplitude) can be excluded from analysis. In other words, if participants achieve the WBSE by strategically looking away from the region of the screen that permits efficient word reading (the word's centre) the WBSE will not be observed since these trials can be removed from the analyses. Furthermore, if Stroop effects are being pushed from response selection to response execution (Kello et al., 2000), the duration of the response saccades (saccades from the Stroop stimulus to the correct patch location) will exhibit Stroop effects. Finally, we employed a block design for the different trial types. This decision was motivated by recent work showing that response conflict is substantially reduced when trial types are mixed, thus reducing the likelihood of observing a reduction in its magnitude (Hasshim & Parris, 2017).

The use of eye tracking also permitted the use of pupillometry. Pupillometry is a robust measure of mental effort with the pupil becoming larger as more cognitive effort is experienced (Hess & Polt, 1960; 1964; Kahneman & Beatty, 1966; see Beatty & Lucero-Wagoner, 2000, and Laeng, Sirois, & Gredebäck, 2012, for reviews). These mental effort-related pupil dilations are thought to be related to activity of the hypothalamus and the locus coeruleus with the latter thought to receive input from the anterior cingulate and orbitofrontal cortices to optimise utility of action (Aston-Jones & Cohen, 2005). Recent work employed pupillometry to measure motivation-cognition interactions in attention tasks (Chiew & Braver, 2013; Massar, Sasmita, Lim & Chee, 2018) and pupillometric Stroop effects have been reported (e.g. Hasshim & Parris, 2015; Hershman & Henik,

2019; 2020; Laeng, Ørbo, Holmlund, & Mizzo, 2011; see Laeng, Sirois, & Gredebäck, 2012, for a review of the use of pupillometry in cognitive studies). If the WBSE results from increased effort and motivation, increased pupil dilation would be expected in the Suggestion Present condition. If the WBSE results from atypical (Raz et al., 2002; Lifshtiz et al., 2013) and perhaps involuntary, effortless control associated with responding to suggestion (Bowers, 1982; Weitzenhoffer, 1980), reduced pupil dilation would be expected in the Suggestion Present condition.

Finally, the present experiment also intended to investigate potential differences in the effect of the word blindness suggestion on different types of conflict in the Stroop task by employing response set trials, non-response set trials, neutral trials, and trials with words that matched the font colour (congruent trials). To measure overall Stroop interference we compared response set and neutral trial performance. The difference between response set and non-response set trials was taken as a measure of response conflict; the difference between non-response set and neutral trials was taken as a measure of semantic conflict; the difference between neutral and congruent trials was taken as a measure of facilitation. We employed a block design to maximise conflict types (Hasshim & Parris, 2017). If the WBSE works via a reduction of only response-level processing and not via inducing word blindness, semantic conflict will be unaffected by the suggestion.

Following Raz and colleagues work, participants completed the Stroop task under two conditions delivered in counterbalanced order. In one condition participants were first induced into hypnosis and given the word blindness suggestion. They were then counted out of hypnosis. The experimenter then clapped once which participants were

told would activate the suggestion. Immediately following the clap, the participants completed the Stroop task. This will be referred to as the Suggestion Present condition. In the other condition, participants were simply asked to complete the Stroop task without the induction, the delivery of the suggestion and the clap. This will be referred to as the Suggestion Absent condition.

2. Method

2.1. Design

The experiment had a 2 (Suggestion: Present, Absent) x 4 (Stroop conditions: congruent, response set, non-response set, neutral) fully within-subjects design. The dependent variables were saccade latencies on the first saccade post-stimulus onset (i.e., the time it took to make a saccade towards the corresponding area of the screen) and pre-trial and intra-trial pupil dilation.

2.2. Participants

Participants were selected from a pool of students on Bournemouth University's Experiment Participation Scheme were pre-screened on the Stanford Hypnotic Susceptibility Scale, Form C (SHSS-C; Weitzenhoffer & Hilgard, 1962) with the age regression suggestion removed. Hypnosis scores from this pool ranged from 1-11. Participants that scored 6 or above (medium-high hypnotizable) were invited to take part. Sixteen (15 females, 1 male) participants agreed to take part in the study of which 6 were classified as mediums and 10 as highs. Previous research has shown the presence of the WBSE even in medium and low suggestible individuals, although to a lesser extent (Parris & Dienes, 2013; Parris et al., 2014; Raz & Campbell, 2011) The average age was

22.24 years (*SD*=6.40). Participants were given course credits for their participation.

All procedures performed in this study were approved by the ethics committee of Bournemouth University. All participants gave their written informed consent to participate in the study and for the associated procedures.

2.3. Eye movement and pupil size recording

Stimuli were presented using a standard PC running Experiment Builder software (SR Research Ltd) and displayed on a colour monitor displaying at 120 Hz. An SR Research Eyelink 1000 (SR Research Ltd) video-based pupil/CR tracker was used to record eye movements. Calibration and validation of eye movements was carried out prior to the commencement of each trial block using a nine-point calibration process. A monocular sampling rate of 1000Hz was used. Saccade parameters were extracted off line using Eyelink DataViewer software (SR Research Ltd). Saccades were detected using a combined velocity and acceleration criteria of 30 °/s and 8000 °/s^2. Saccades with a latency greater than 3 standard deviations from the mean or <80 ms were excluded from analysis as were trials where the saccade amplitude was <2°. The primary measure of interest for reaction time was the latency of onset of the first saccadic response from the onset of the target patches and Stroop stimulus. Saccades which deviated from the correct target direction were classified as response errors.

Pupil size was measured in pixels. After each participant completed the task, a single measurement of a 4-mm dot was recorded from the same camera location (the placement of the camera was adjusted for each participant for comfort), and this was used as a reference point to convert all measurements from the arbitrary pixel units to millimeters and to determine pupil diameter changes. The same stimuli were presented in both

conditions the same number of times and the conditions were run one after the other in the same room with lighting in the lab was controlled, so the conditions did not differ in luminance.

Our study was designed to test for pupil size differences between the Suggestion Absent and Suggestion Present conditions. Pupil sizes were sampled at two phases of the task: 1) The intratrial-response phase: The average pupil size within the period from stimulus onset to response completion; 2) The post-response phase: The average pupil size within a 500ms window from 250ms after a response was made (see Figure 1). These phases were chosen because pupil dilations associated with effort related to Stroop task performance have been reported in pupil data both intratrial (Hasshim & Parris, 2015) and post-response, with post-response Stroop effects peaking around 500-600ms after the response is made (see Hershman & Henik, 2019; 2020; Laeng et al., 2011). A 300ms pre-trial period (just before stimulus onset) acted as the pupil size baseline for both phases and was subtracted from the intra-trial phase to provide a baseline corrected measure of performance as recommended by Mathôt, Fabius, Heusden, & Van der Stigchel (2018), and is a method used to show pupillometric Stroop effects (Laeng et al., 2011). The benefit of having both a intratrial- and post-response phase is that it has been argued that the post-response phase might simply represent residual change due to the response that was made (Simpson, 1969). Moreover, whilst the two response modes have not been directly compared, intratrial-response Stroop effects have been reported with the saccadic response Stroop task that is employed in the present study, whereas post-response pupil Stroop effects have been reported with manual response Stroop tasks (Hershman & Henik, 2019; Laeng et al., 2011).

2.4.   *The Stroop task*

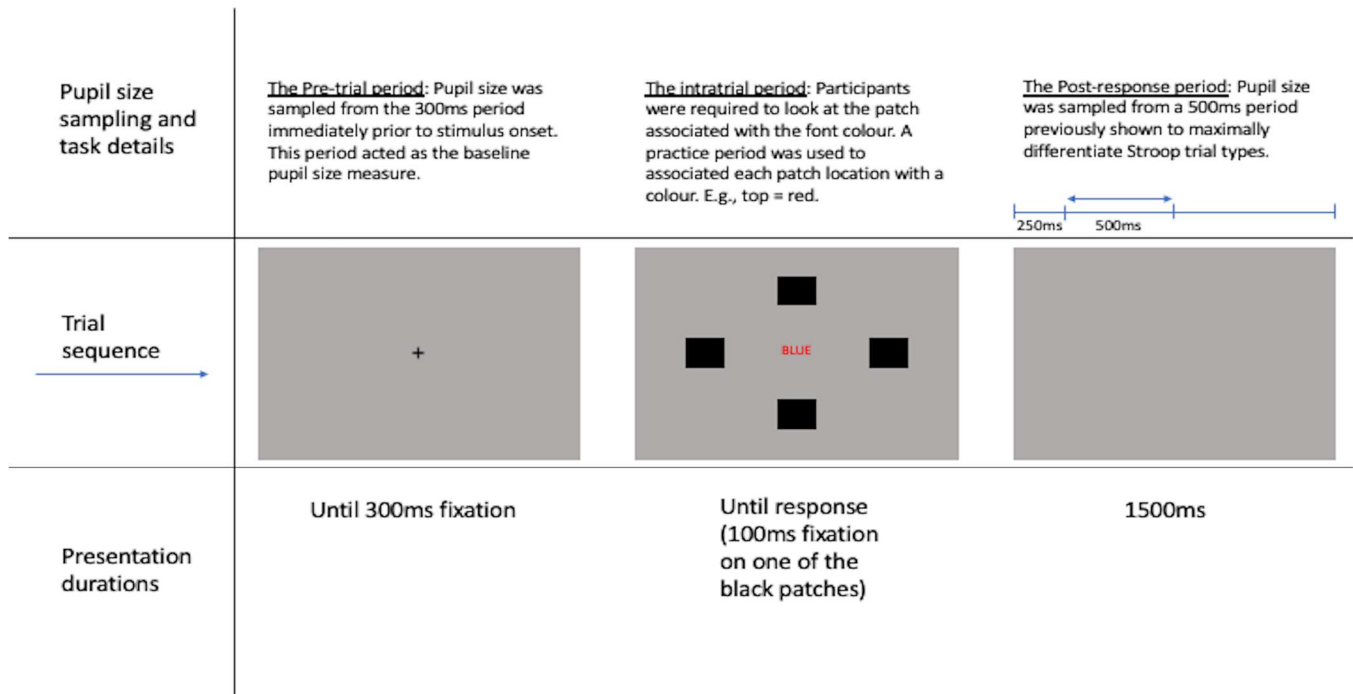| Pupil size sampling and task details	The Pre-trial period: Pupil size was sampled from the 300ms period immediately prior to stimulus onset. This period acted as the baseline pupil size measure.	The intratrial period: Participants were required to look at the patch associated with the font colour. A practice period was used to associated each patch location with a colour. E.g., top = red.	The Post-response period: Pupil size was sampled from a 500ms period previously shown to maximally differentiate Stroop trial types.
Trial sequence			
Presentation durations	Until 300ms fixation	Until response (100ms fixation on one of the black patches)	1500ms

Figure 1 The trial sequence and pupil sampling periods.

There were four trial types, congruent trials (words spelling out a congruent colour that was part of the response set), neutral trials (words not associated with a colour), non-response set trials (words spelling out a colour not part of the response set), and incongruent (response set) trials (words spelling out an incongruent colour that was part of the response set; standard incongruent trials) all presented in pure blocks. Each participant began with 8 practice trials of each trial type and the order of blocks was counterbalanced. There were 48 trials of each trial type[2]. Two versions of the experiment

---

[2] Due to a programming error, there were 64 non-response set trials instead of the

were administered, counterbalanced between participants. The difference between the

versions was the colours making up the response set. Words that spelled out the four

possible colour responses in one version acted as the word stimuli in the other version's

non-response set trials. The first version used the colours yellow (RGB: 255, 255, 0),

pink (RGB: 255, 153, 204), green (RGB: 0, 200, 0) and white (RGB: 255, 255, 255),

while the second version used blue (RGB: 0, 0, 255), purple (RGB: 204, 102, 255),

orange (RGB: 255, 153, 0) and red (RGB: 255, 0, 0). The versions were counterbalanced

between participants. The neutral words were 'due', 'wall', 'story', 'marvel' and were

matched for length and frequency using the English Lexicon Project (Balota et al., 2007).

Words were presented on the centre of the screen and all words were printed in upper-

case, bold, and in size-20 Courier New font against a white background. No word

subtended an angle of >2.5° meaning that to look at the end letter of word a saccade

would have to be smaller than roughly half of that value. Any trial on which the first

saccade was less than 2° was excluded. Four squares ("patches") 200 × 200 pixels in size

appeared to the right, left, above and below the screen's central position. The black

patches subtended approximately 3° of arc at an eccentricity of 7.5° from the fixation

point. Participants placed their heads on a chinrest approximately 60 cm from the screen

---

intended 48. A re-analysis of the latency data including just the first 48 non-response set

trials revealed no differences in the outcomes of the analyses (Bayes Factors for

difference in Stroop interference, response competition, semantic competition and

facilitation were, 14.213, 7.389, 0.606, and 0.688, respectively, which are very similar to

those reported below).

and made saccadic responses towards one of four target colour patches in the periphery.

2.5. *Post-hypnotic suggestion*

In the posthypnotic suggestion present condition the participants were given a standard induction (taken from the SHSS-C) followed by the following suggestion taken from Raz et al. (2002):

Very soon you will be playing the computer game. When I clap my hands, meaningless symbols will appear in the middle of the screen. They will feel like characters of a foreign language that you do not know, and you will not attempt to attribute any meaning to them. This gibberish will be printed in one of 4 inks colours: red, blue, green or yellow. Although you will only be able to attend to the symbols' ink colour, you will look straight at the scrambled signs and crisply see all of them. Your job is to quickly and accurately look to the patch that corresponds to the ink colour shown. You will find that you can play this game easily and effortlessly.

2.6. Procedure

Participants were first given an information sheet describing what they were going to be asked to do and were then invited to sign a consent form. The order in which participants completed the Suggestion Present and Suggestion Absent conditions was counterbalanced across participants. In the Suggestion Present condition participants were administered a hypnotic induction using the procedure in the SHSS-C. The word-blindness suggestion was then delivered. Participants were then counted out of "hypnosis". Following de-induction, and before the Stroop task was started, the posthypnotic suggestion was activated via a single clap. During the Stroop task, participants were presented with a fixation cross in the centre of the screen which had to be fixated on for 300ms before the

Stroop stimulus was presented. Participants' heads rested on a chin support. The Stroop stimulus was then presented until the participant fixated one of the four black peripheral patches for 100ms. To train participants on the locations corresponding with each of the response set colours, participants completed 32 practice trials presented as described above with the exception that the patches were in colour. Hence, the patches would be linked to a specific colour which participants had to remember after the patches turned black during the experimental trials. Whilst this would likely result in increasing task difficulty, its purpose was to prevent colour matching during the experimental trials and was more akin to responding by keypress for which there are no on-screen reminders of colour locations. Once a response was made and one of the patches fixated for 100ms, a blank screen was presented for 1500ms until the next fixation cross was presented, beginning the next trial. The trial types were presented in their own blocks and the order of the blocks was counterbalanced across participants. After completion of all the blocks in the Suggestion Present condition, the suggestion was deactivated. In the Suggestion Absent condition participants performed the Stroop task but without the induction / post-hypnotic suggestion being delivered. Following Raz et al. (2002) there was a break (10-minutes) between the Suggestion Present and Suggestion Absent conditions.

2.7. Analyses of eye movements and pupil size

For saccade latencies, 2 (Suggestion: Present, Absent) x 4 (Word Type: Response Set, Non-response Set, Neutral, Congruent) repeated measures ANOVAs were planned. However, given predictions included potential null effects, Bayes factors (*B*) were also planned to assess strength of evidence for all tests with 1-degree of freedom, H1, over the

null, H0 (see below for further details on Bayes Factors). For analysis of saccadic latencies we applied the strict exclusion criteria employed by Hodgson et al. (2009) since this represents a previously reported method and permits us to exclude any trials on which participants shifted their gaze to anywhere else in the word. Therefore, only the first saccades following stimulus onset whose latency was >80ms (which are known as express saccades and are generally assumed to not result from the onset of the stimulus (Fischer et al., 1993; Fischer & Boch, 1983; Fischer & Ramsperger, 1984; Wenban-Smith & Findlay, 1991)) and whose amplitude was greater than 2° were analysed.

Pupil size was continuously sampled except for when blinks occurred; when blinks did occur pupil sizes 100ms either side of the blink were removed without interpolation and therefore did not contribute to the mean pupil size values.

For pupil size, whilst we report analysis using a 2 (Suggestion: Present, Absent) x 4 (Word Type: Response Set, Non-response Set, Neutral, Congruent) repeated measures ANOVAs for consistency with the saccadic latency analyses, the main prediction concerned the main effect of Suggestion. As noted above, our study was designed to test for pupil size differences between the Suggestion Absent and Suggestion Present conditions with the intra-trial pupil sizes baseline corrected calculated by subtracting the pre-trial pupil sizes (Mathôt et al., 2018). The pupil analyses included on those trials analysed in the latency analysis.

2.8. Bayes Factors

A *B* of above 3 indicates moderate evidence for H1 over H0 and below 1/3 moderate evidence for the H0 over Hl. All Bayes factors, *B*, reported here represent the evidence

for H1 relative to H0; to find the evidence for H0 relative to H1, take $1/B$. Bs between 3

and 1/3 indicate data insensitivity (see Dienes, 2014). Here, $B_{H(0, x)}$ refers to a Bayes

factor in which the predictions of H1 were modelled as a half-normal distribution with an

SD of x (see Dienes, 2014, 2016); the half-normal can be used when a theory makes a

directional prediction where x scales the size of effect that could be expected. All Bayes

Factors were calculated with an adjusted standard error where standard error (SE) =

SE*(1 + 20/df*df) due to the sample size being less than 30 (Dienes, 2014).

To indicate the robustness of Bayesian conclusions, for each B, a robustness

region is reported (Dienes, 2019), giving the range of scales that qualitatively support the

same conclusion (i.e. evidence as insensitive, or as supporting H0, or as supporting H1),

notated as: RR$_{conclusion}$ [x1, x2] where x1 is the smallest SD that gives the same

conclusion and x2 is the largest; the "conclusion" will be notated as "$B < 1/3$"; "$1/3 < B < 3$"; or "$B > 3$".

2.8.1. *Derivation of models of H1*

For saccade latencies relevant to the main effect of the suggestion, we contrasted the

theory that the suggestion had some effect with the null hypothesis that the suggestion

had no effect. Predictions of the theory were represented as a half- normal scaled with an

expected reduction of 50ms which represent the significant main effect observed in Raz

et al. (2002).

For saccade latencies relevant to the effect of the suggestion on each conflict type,

we contrasted the theory that the suggestion had some effect with the null hypothesis that

the suggestion had no effect. Predictions of the theory were represented as a half- normal

scaled with an expected reduction of 70% of the Suggestion Absent interference effect

and 40% of the Suggestion Absent facilitation effect which represent the average effect sizes of the suggestion across numerous studies (see Parris et al., 2013).

For saccade durations relevant to the effect of the suggestion on Stroop interference, we contrasted the theory that the suggestion had some effect (by producing a Stroop effect), with the null hypothesis that the suggestion had no effect (the Stroop effects in the Suggestion Present and Suggestion Absent conditions are equal). Predictions of the theory were represented as a half- normal scaled with an expected value of 1ms which represents the approximate size of an effect of night shift-related tiredness on saccade durations (Skrzypek, Szeligg, Jabloński & Nowakowski, 2017) where saccade durations were similar in magnitude to those observed in the present experiment (~50-60ms).

For pupil size, we based expectations on Laeng, Ørbo, Holmund, and Miozzo (2011) who obtained a baseline corrected change in pupil diameter of about 0.026 mm for the pupil Stroop effect when averaging over mean and peak dilation changes; thus, this was used as the SD for all tests.

With these choices of models of H1, results significant at $p < .05$ in fact corresponded to Bs above 3 (though there is in general no guarantee of such a correspondence between p values and Bayes factors, Dienes, 2014).

3. Results

We excluded 34% of trials due to the first saccade being smaller than 2° in amplitude. Next, we excluded 2% of the remaining trials because the latency was less than 80ms. We next excluded error trials which were defined as any trial on which saccades were

made in the wrong direction, resulting in the exclusion of 20.1% of the remaining trials.

This large proportion of errors attests to the difficulty of having to remember the location

associated with each colour after the patches turned black following the practice block.

However, this also means that errors are not approaching the upper bound and thus likely

avoid underadditive or overadditive artefacts that would prompt the need for logistic

regression analysis (Dixon, 2008). Finally, following all other studies on this effect we

excluded any of the remaining trials whose latency was greater than 3 standard deviations

from the overall mean (e.g. Raz et al., 2002; Parris & Dienes, 2013; Parris, Dienes, Bate

& Gothard, 2014) resulting in the exclusion of 0.06% of the remaining trials. In total,

following the criteria employed by Hodgson et al. (2009) meant that 48.7% of trials were

excluded from analysis of the saccadic latencies. Raw data are available at

https://osf.io/enpdy/.

3.1.   *Saccade latencies*

The data were entered into a 2 (Suggestion: Present, Absent) x 4 (Word Type:

Response Set, Non-response Set, Neutral, Congruent) repeated measures ANOVA. There

was no evidence one way or the other for the main effect of Suggestion, $F(1, 15) = .737$,

$p = .404$, $B_{H(0,50)} = 0.428$, $RR_{1/3 < B < 3}$ [0, 73], where the expected effect size was taken

from Raz et al. (2002). The main effect of Word Type was significant, $F(3, 45) = 8.798$, p

$< .001$, $\eta_p^2 = .370$. The interaction between Suggestion and Word Type was also

significant, $F(3, 45) = 5.572$, $p = .013$ (Greenhouse-Geisser), $\eta_p^2 = .271$. See Table 1 for

latencies in all conditions.

To explore the interaction and to test for the presence of the WBSE and a

differential effect of the WBSE on conflict types, the magnitudes of Stroop interference (response set – neutral), response competition (response set – non-response set), semantic competition (non-response set – neutral) and facilitation (neutral – congruent) in the Suggestion Absent condition were compared to their counterparts in the Suggestion Present condition. For the key effect of interest (the WBSE), the analyses revealed that Stroop interference was greater in the Suggestion Absent condition than in the Suggestion Present condition (72ms vs. -1ms), $t(15) = 2.720$, $p = .016$, $r = .575$ $B_{H(0, 50.4)} = 16.871$, $RR_{B>3}$ [12, 695], replicating the WBSE (see Figure 2). Response competition was also substantially modified, (54ms vs. -53ms), $t(15) = 2.661$, $p = .018$, $r = .567$, $B_{H(0, 37.8)} = 7.478$, $RR_{B>3}$ [19, 923]. However, the Bayes Factors for the remaining comparisons were insensitive but favoured the null: semantic competition (18ms vs. 54ms), $t(15) = -1.419$, $p > .17$, $B_{H(0, 12.6)} = 0.575$, $RR_{1/3 < B < 3}$ [0, 32.5]; facilitation (14ms vs. 46ms), $t(15) = -1.651$, $p > .12$, $B_{H(0, 5.6)} = 0.686$, $RR_{1/3 < B < 3}$ [0, 22]. See Table 2.
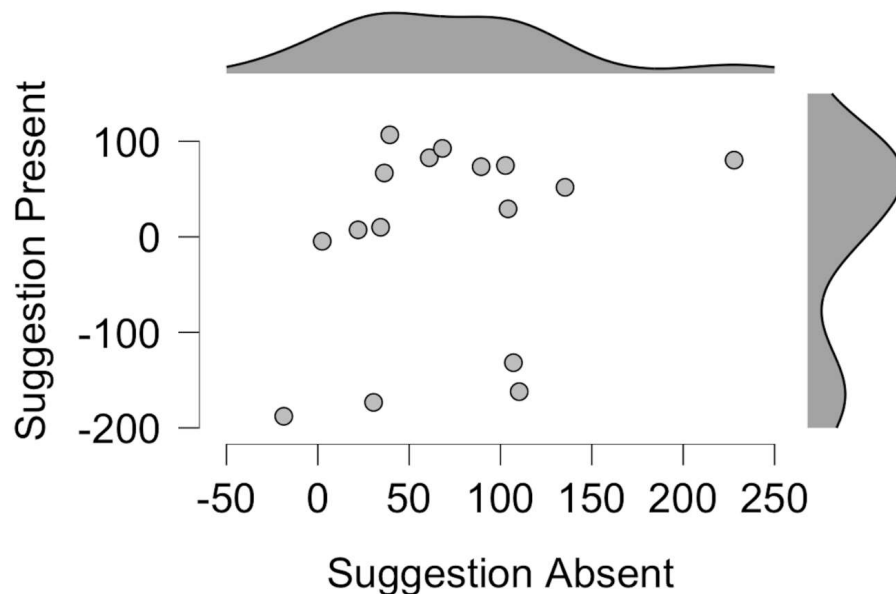

Figure 2 Scatterplot (with density displayed above) of the observed Word Blindness Suggestion Effect showing the Stroop interference effect (Incongruent –

Neutral trials) for saccade latency was smaller in the Suggestion Present condition compared to the Suggestion Absent condition for the majority of the participants.

Given that the application of the criteria set by Hodgson et al. led to the exclusion of almost half the trials, we re-analysed the data but included trials on which second (29%) or third (8%) saccades satisfy the above criteria (fourth saccades only added a negligible number of trials (2%)). Their inclusion meant that only 16.9% of trials were excluded in total. Importantly the WBSE remained. However, the magnitudes of response competition (response set – non-response set trial RTs) in the Suggestion Present vs Suggestion Absent conditions were no longer sensitive.

The analyses of the saccade latency data therefore discount the visuo-attentional account of the WBSE since it was present despite removing unnecessary saccades, but do not permit a conclusion with regards specific effects of the WBSE on conflict types.

3.2.   *Saccade durations*

The analysis of saccade durations was motivated by the finding that speeded responding, often observed following the word blindness suggestion (e.g., Parris et al., 2012; Raz et al., 2002; 2005), can push Stroop effects out of response latencies and into response durations (Kello et al., 2000). However, there was not a clear main effect of suggestion on response latencies in the present data. Nevertheless, for the interested reader we report the outcome of a 2 (Suggestion: Present, Absent) x 4 (Word Type: Response Set, Non-response Set, Neutral, Congruent) repeated measures ANOVA. The

analysis did not reveal substantial evidence for a main effect of Suggestion, $F(1, 15) =$ 4.679, $p = .047$, $\eta_p^2 = .290$ (Greenhouse-Geisser), $B_{H(0, 1)} = 1.8$, $RR_{1/3 < B < 3}$ [0, 200], where sample durations were longer in the Suggestion Present condition (65ms vs. 59ms). There was no significant main effect of Word Type ($p = .290$) nor interaction ($p = .545$). No firm conclusions follow.

3.3. *Proportion errors*

Errors were defined as the first saccade following stimulus onset whose latency was >80ms and whose amplitude was greater than 2° and that was executed in the wrong direction. 20.1% of all trials were errors. The proportion error data were arcsine transformed before being entered into the analysis. A repeated measures ANOVA revealed no evidence one way or the other for a main effect of Suggestion (F (1, 15) = .027, $p = .873$), $B_{H(0, 0.8\%)} = 0.97$, RR [0, 0.14], no significant main effect of Word Type (F (3, 45) = 1.964, $p = .133$, and no significant interaction where F(3, 45) = 1.564, $p = .211$.

3.4. *Pupil Size*
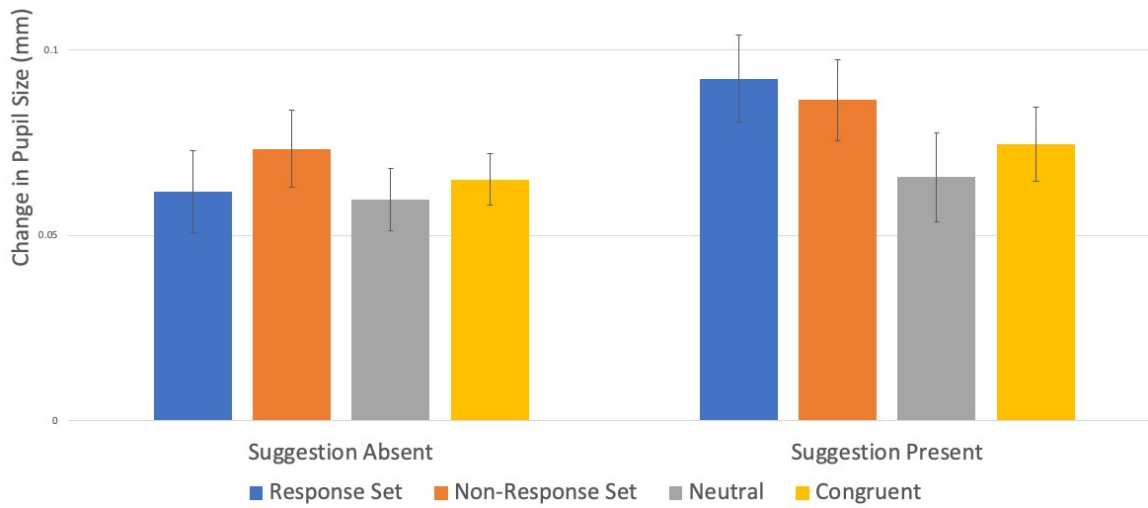
3.3.1 Baseline corrected Intratrial Pupil Sizes

Figure 3 Bar chart showing pupil size changes in both conditions and for each trial type in the intratrial period. Only the main effects (of Suggestion condition (Absent, Present) and Word Type (response set, non-response set, neutral, congruent)) were significant.

For baseline corrected Intratrial pupil sizes the interaction was non-significant, $F(3, 45) = 1.281$, $p = .293$ (Greenhouse-Geisser). However, both the main effects of Word Type, $F(3, 45) = 4.527$, $p = .011$, $\eta_p^2 = .2$ (Greenhouse-Geisser) and Suggestion, $F(1, 15) = 9.223$, $p = .008$, $\eta_p^2 = .381$ were significant with the latter being supported by a Bayes Factor of $B_{H(0,0.026mm)} = 31.927$, $RR_{B>3}$ [0.002, 0.3]. See Figure 3.

The main effect of suggestion was the result of baseline corrected pupil sizes being larger in the Suggestion Present (0.0798mm) compared to the Suggestion Absent (0.0649mm) condition indicating more effort expenditure in the former (see Figure 4). Paired-sample t-tests revealed the main effect of Word Type was the result of a Stroop

interference effect (response set**:** 0.077mm – neutral: 0.063mm), $t(15) = 2.453$, p = .027,

$r = .535$, $B_{H(0,0.026mm)} = 7.666$, $RR_{B > 3}$ [0.075, 0.78], and a semantic Stroop effect

(nonresponse set: 0.08mm – neutral: 0.063mm) t(15) = 3.263, p = .005, $r = .644$,

$B_{H(0,0.026mm)} = 66.953$, RR $_{B>3}$ [0.002 – 0.73]. However, for our measure of response

conflict (response set: 0.077mm - nonresponse set trials: 0.08mm) there was evidence for

the null hypothesis $t(15) = .549$, $p = .591$, $B_{H(0,0.026mm)} = 0.325$, $RR_{B<1/3}$ [0.055, >0.1]. For

Stroop facilitation (neutral: 0.063mm - congruent trials: 0.07mm) the result was

insensitive, $t(15) = 1.497$, $p = .155$, $B_{H(0,0.026mm)} = 0.999$, RR $_{B<1/3}$ [0.092, >0.1].
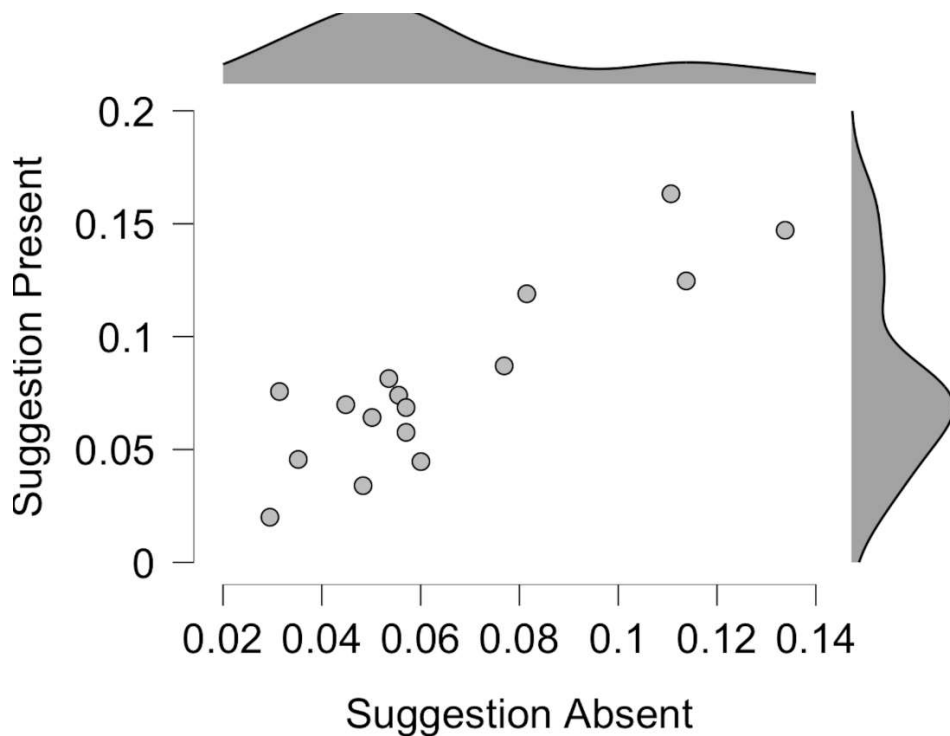

Figure 4 Scatterplot (with density displayed above) showing increased overall

intratrial pupil sizes in the Suggestion Present condition which is taken as evidence for

the recruitment of effortful control to achieve the word blindness suggestion effect.

3.3.2 Baseline corrected Post-response Pupil Sizes

For baseline corrected post-response pupil sizes the interaction was non-significant, $F(3, 45) = 1.191$, $p = .319$ (Greenhouse-Geisser) and neither were the main effects of Word Type, $F(3, 45) = 1.019$, $p = .382$ (Greenhouse-Geisser) and Suggestion, $F(1, 15) = 2.136$, $p = .164$, $B_{H(0,0.026mm)} = 1.703$, $RR_{1/3 < B < 3}$ [0, 0.17].

3.3.3 Uncorrected pretrial pupil sizes

Due to the fact the we used a block design where the Suggestion Present and Suggestion Absent conditions were undertaken in different blocks separated by a break of 10-15 minutes, a reviewer pointed out that it is possible that the larger baseline corrected pupil sizes in the Suggest Present condition reported above could be due to pupil sizes already being different between the two conditions (that is, a change that occurred between blocks that had nothing to do with the presence or absence of the suggestion). To check for this possibility, we analysed the uncorrected pretrial pupil sizes. For pre-trial uncorrected pupil sizes the interaction was non-significant, $F(3, 45) = 2.076$, $p = .122$, as was the main effect of Word Type, $F(3, 45) = 1.882$, $p = .155$, (Greenhouse-Geisser). There was a main effect of Suggestion $F(1, 15) = 5.313$, $p = .036$, $\eta_p^2 = .262$, $B_{H(0,0.1mm)} = 7.364$, $RR_{B>3}$ [0.00325, 0.15] but this was the result of the uncorrected pupil sizes being larger in the Suggestion Absent (3.47mm) compared to the Suggestion Present (3.37mm) condition, which is the opposite of the effect reported above.

4. Discussion

The aim of the present study was to test potential accounts of the reduction in Stroop interference that follows a post-hypnotic suggestion for word blindness that do not rely on atypical processes. Specifically, we tested whether: 1) the WBSE resulted from the visuo-attentional strategy of looking away/at the end letter of the irrelevant word to impair word processing; 2) only response-level processes are affected to achieve the appearance of word blindness; 3) the missing interference is pushed into response execution or; 4) the experimental context provided highly suggestible participants with the extra motivation required to more fully engage effortful control mechanisms. Below we consider the implications of the reported results for each of these accounts.

Our results showed that the WBSE was observed despite eliminating any trial on which the initial eye movement after stimulus onset was made in any direction other than the correct direction or was of a magnitude indicative of a move to either end (i.e. to the initial or final letter positions) or near either end of the irrelevant word. The preservation of the WBSE following the removal of these trials indicates that the WBSE is not the result of a visuo-attentional strategy. Therefore, whilst it has been shown that Stroop interference can be reduced by altering the spatial distribution of attention across the word (Besner et al., 1997; Parris et al., 2007) this does not appear to be underpinning the WBSE. Indeed, our design permitted us to lock participants attention at or near the optimal viewing position for maximum Stroop interference (see Parris et al., 2007). A caveat to this conclusion is that whilst eye tracking permitted us to track where participants had directed their gaze, it is possible that participants' attention was not focussed where they were looking (Rayner, 2009; Shepherd, Findlay & Hockey, 1986) and therefore that, despite controlling for gaze location, participants were actively

refocussing their attention whilst keep their eye position constant to de-optimise word processing. However, Rayner (2009) argued that such a dissociation between eye position and attention in complex tasks like reading, scene perception and visual search is unlikely to be the result of a strategy employed by participants.

The saccade latency data indicated that it was a reduction in response conflict that led to the reduction in overall Stroop interference which is consistent with the results from Augustinova and Ferrand (2012). However, this effect became insensitive when we added some of the previously excluded trials in the analysis rendering our results inconclusive on this matter. The semantic conflict effect numerically increased in the Suggestion Present condition but again Bayes Factors were insensitive indicating that our data cannot be taken as evidence for or against a modification of semantic conflict by the word blindness suggestion. The analysis of the pupil data indicated the presence of a semantic conflict effect in the intratrial period. In contrast, our measures of response conflict and Stroop facilitation were not significant in the pupil data.

Following the relatively consistent finding that the WBSE results in overall reductions in response times (a main effect of suggestion; Parris et al., 2012; Parris & Dienes, 2013; Parris, Dienes, Bate & Gothard, 2014; Raz et al., 2002; 2005; Raz & Campbell, 2011), and the finding showing that speeded responses can push Stroop effects from the level of response selection to response durations (Kello et al., 2000), we sought to identify whether the missing Stroop effects associated with the WBSE were to be found in the saccade duration data. Our data turned out not to be non-evidential on this matter.

Analysis of the pupil size data revealed larger pupil sizes in the Suggestion Present condition. This finding is consistent with the notion that participants were more motivated to more fully engage effortful control in this condition (Chiew & Braver, 2013; De Jong et al., 1999; Massar et al., 2018) in a similar way to Stroop interference reductions following financial incentives (Krebs, Boehler & Woldorff, 2010). This finding contrasts with the notion that involuntary, effortless responding is key marker of responding to hypnotic suggestions (Bowers, 1982; Weitzenhoffer, 1980) and is more consistent with the notion that responses to suggestions are active attention-demanding attempts to fulfil the requirements of goals (Lynn, Rhue & Weekes, 1990). These data do not necessarily indicate that participants are being deceptive in their attempts to achieve a reduction in Stroop interference and could be a marker of how demand characteristics can be turned into genuine experience by their effect on the awareness of a participants intentions to act to achieve a desired outcome (Dienes et al., in press, a, b).

The pupil size data also permit us to rule out explanations based on squinting (Raz et al., 2003) since squinting results in occluding the pupil, rendering pupils smaller, the opposite to the finding reported here. However, a potential explanation of the WBSE is that participants blur their vision to impair word reading. As noted above Raz et al. (2003) attempted to rule this out by showing that, when the muscles used to blur vision were temporarily paralysed, the WBSE was still observed. Noting that Raz et al.'s method (cyclopentolate eye drops) does not fully prevent blurring, Palfi et al. (submitted) asked participants to deliberately blur their vision and reported that whilst doing so substantially reduced Stroop interference, it also substantially slowed responding, which they noted is not a marker of the WBSE. However, the present results contribute evidence

consistent with the blurring account. It is well established that visual accommodation, used to focus gaze on a particular object, results in pupillary constriction (via constriction of the ciliary muscles; Atchison, Smith & Efron, 1979; Campbell & Gregory, 1960; Liang & Williams, 1997; Sheedy et al., 2003; Woodhouse, 1975). Relaxation of accommodation (blurring) would therefore lead to less constriction and larger pupils which is what we observed here. Nevertheless, one cannot discount the results from Palfi et al, (submitted) and Raz et al. (2003), so on balance the evidence favours an account of the larger pupils based on the employment of more effortful control.

Raz et al. (2002; 2005) invoked an account of the WBSE based on a form of top-down control they describe as being not normally voluntarily available; and one that is able to deautomatise cognitive processes, and perhaps one that is special to the context of hypnosis and suggestion (Landry et al., 2013). As noted, this account needs further elucidation. However, one result from the present study that could be interpreted as being broadly consistent with this account is one showing the WBSE to be the result of more fully engaged effortful control mechanisms which have been described as being available but not normally utilised (De Jong et al., 1999; Parris, 2014). Notably, however, this is not limited to the context of hypnosis and suggestion. Chiew & Braver (2013) and Massar et al. (2018) have argued that larger pupils evidence more effortful (proactive) control and thus here we argue that the experimental context provides highly suggestible participants with extra motivation to perform well and resultantly increase the level of effortful control employed.

Notably, De Jong et al. (1999) used an external manipulation (shortened the response-stimulus interval) to draw out this extra level of control. The WBSE is therefore

evidence that manipulations of stimulus presentation are not needed to marshal this extra level of control. It is possible that participants engaged an effortful strategy to achieve the WBSE. For example, Palfi et al. argued that the WBSE could be the result of deliberating imagining a counterfactual world in which words are meaningless. This argument was based on the finding that priming the concept of dyslexia has been shown to reduced Stroop interference in a similar way to the word blindness suggestion (Augustinova & Ferrand, 2014; Goldfarb, Aisenberg, & Henik, 2011), and which would require a similar method to prevent/slow word processing.

Theoretical accounts of responding to suggestion based on the strategic relinquishment of awareness of the intention to act according to the suggestion (Barnier et al., 2008; Dienes & Perner, 2007) are not contradicted by the present results.   Such accounts hold that whatever control is achieved under hypnosis and suggestion should also be achievable outside of the context of hypnosis and suggestion, because the only difference is the modified awareness of the intention to act in accordance with the suggestion. We did not measure awareness however and so do not know whether participants were aware or not of the extra effort applied to achieve the WBSE. A recent paper has shown that in the context of the Stroop task pupil sizes can increase in response to an increase in task difficulty indicating extra effort but that participants are unaware of a change in task difficulty (Diede & Bugg, 2017).

To permit the measurement of response execution (saccade durations), the present study employed a relatively rare mode of response for the Stroop task which might have modified the strategy participants employed to achieve the WBSE. Indeed, research has indicated that there might important differences between response modes in terms of the

nature of Stroop interference and its control (Augustinova, Parris & Ferrand, 2019; Parris et al., 2019; see Parris et al., submitted, for a review). For example, it has been argued that the manual response Stroop task does not result in semantic conflict (e.g. Sharma & McKenna, 1998) or does not result in equally large magnitudes of various types of conflict (e.g. task and phonological conflict; e.g. Augustinova et al., 2019; Parris et al., 2019). To the best of our knowledge, there has not been a comparison of the oculomotor Stroop task with manual and vocal versions and thus it is not possible to comment on whether differences exist or whether there are differences in the strategies adopted or the mechanisms employed to produce the WBSE. Thus, the conclusions from the present work need to be verified with the more common manual response Stroop task. Indeed, a notable difference between most previous studies employing pupillometry as a measure of Stroop task performance and the present study is that previous studies have observed post-response pupillometric Stroop effects (see Hershman & Henik, 2019; 2020; Laeng et al., 2011). In contrast, in the present study we observed intratrial pupillometric Stroop effects, a finding that is consistent with a previous oculomotor Stroop task study (Hasshim & Parris, 2015). This indicates that the oculomotor Stroop task differs in some way from the manual response Stroop task. However, the benefit of the observation of an intratrial effect is that it is does not simply represent residual change due to the response that was made; a criticism levied at the post-response pupil effects (Simpson, 1969). Nevertheless, despite the different response mode, we reported substantial Stroop effects with the oculomotor Stroop task and the WBSE is of an expected magnitude (Parris, Dienes & Hodgson, 2013) rendering it unlikely that the mode of response selected for the

present study would modify strategies employed. Indeed, the present results are consistent with those of Palfi et al. (submitted) who employed a manual response.

A potential further limitation of our study is the make-up of the participant sample. Most of our participants were female. There is evidence indicating women are more suggestible than men (Page & Green, 2007), and that women are suggestible for different reasons (Geiger, Peter, Prade, & Piesbergen, 2014). For example, Geiger et al. (2014) showed that hypnotic suggestibility in women was more strongly related to crystallised intelligence. Notably, in contrast to Page and Green (2007), Geiger et al. did not report a difference in males and females in terms of the levels of suggestibility. Nevertheless, together these studies indicate that males and females might differentially respond to suggestions. In turn, this means that it is possible that females might achieve the WBSE in a different way to males. Our results do not permit us to effectively assess this possibility, but it would be an interesting area of investigation for future research.

We also included medium hypnotisable participants in our sample. As noted above, previous research has shown the presence of the WBSE even in medium and low suggestible individuals, although to a lesser extent (Parris & Dienes, 2013; Parris et al., 2014; Raz & Campbell, 2011). It remains possible however that medium hypnotisable participants use different strategies/mechanisms to highly hypnotisable participants. With only 6 medium participants our data do not permit us to provide a meaningful analysis of this possibility and thus a question for future research is whether level of hypnotic suggestibility also modifies the strategies/mechanisms employed to achieve the effect.

A large proportion (20.1%) of trials in the present study were classified as errors. This is larger than the proportion of errors reported in other studies of the WBSE (<10%)

and larger than that reported in other oculomotor Stroop papers (<6%). In designing the present study we wanted to avoid the potential colour matching that would have occurred in the original oculomotor Stroop study reported by Hodgson et al. (2009). To this end, only during the practice session were the patches coloured. During the experimental phase of the study the patches were rendered black as per Hasshim & Parris (2015). This in fact better mimics the manual response Stroop task in which the coloured labels on the response keys are often not visible (covered by the response fingers) during the task. However, in contrast to Hasshim and Parris (2015) the present task involved four, not three response locations making the demands on working memory all the greater. Whilst it is surprising that the number of errors in the two studies differs by quite so much, the analysis of errors in the present task showed that this extra level of difficulty did not affect any particular condition or trial type.

To conclude, the present study replicated the effect of the word blindness suggestion on Stroop task performance and showed that it is not the result of deliberate visuo-attentional strategies, but is the result of the fuller engagement of control processes which we argue is due to enhanced motivation/experimental demands. Hypnosis has been shown to modify the efficacy of frontal lobe function (see Parris, 2017, and Terhuneet al., 2017, for reviews), although the neural mechanistic consequences of post-hypnotic suggestion are far less clear (Parris, 2017). It has been argued that this modification of frontal lobe executive functions confers benefits upon those able to be affected by hypnotic induction (Crawford, 1996; Jamieson et al., 2005; Farvolden & Woody, 2004; Gruzelier & Warren, 1993; Rainville, et al., 1999) such that highly suggestible participants are able to achieve a rare level of control over cognition (Jamieson &

Sheehan, 2004; Jamieson & Woody, 2007; Sheehan et al., 1988; Terhune et al., 2017;

Wagstaff et al., 2007).   However, Parris (2017) argued that the evidence for modification

of frontal lobe activity and function following hypnotic induction and its relationship to

hypnotic suggestibility is mixed. Furthermore, a recent meta-analysis of imaging studies

of hypnosis did not support the predicted involvement of frontal functions known to be

involved in cognitive control (Landry, Lifshitz, & Raz, 2017). Further research is needed

to clarify these relationships, because the evidence for the modification of the experience

of pain and cognition following hypnotic suggestion is plentiful and, as the results of the

present study attest, replicable and robust. The present results indicate that the complex

socio-cognitive context associated with hypnosis and suggestion, could lead participants

to engage an unusual but normally available level of control.

The data and materials for all experiments are available at: https://osf.io/enpdy/. *The experiment was not preregistered.*

Abbreviations list

WBSE = Word Blindness Suggestion Effect
RT = Reaction Time
H1 = Alternative hypothesis
H0 = Null hypothesis
B = Bayes Factor
$B_{H(0, x)}$ = a Bayes factor in which the predictions of H1 were modelled as a half-normal distribution with an SD of x
RR = Robustness Region
SE = Standard Error
SD = Standard Deviation

References

Atchison, D. A., Smith, G., & Efron, N. (1979). The effect of pupil size is uncorrected and corrected myopia. *American Journal of Optometry and Physiological Optics, 56*, 315-323.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Reviews Neuroscience*, **28**, 403–450

Augustinova, M. & Ferrand, L. (2012). Suggestion does not de-automatize word reading: Evidence from the semantically-based Stroop task. *Psychonomic Bulletin & Rev*iew, **19**, 521-527.

Augustinova, M. & Ferrand, L. (2014). Social priming of dyslexia and reduction of the Stroop effect: what component of the Stroop effect is actually reduced? *Cognition, 130*(3), 442-454.

Augustinova, M., Parris, B. A. and Ferrand, L. (2019). The Loci of Stroop Interference and Facilitation Effects With Manual and Vocal Responses. *Frontiers in Psychology. 10:1786*. doi: 10.3389/fpsyg.2019.01786

Barnier, A. J., Dienes, Z., & Mitchell, C. J. (2008). How hypnosis happens: New cognitive theories of hypnotic responding. In M. Heap., R. J. Brown & D. A. Oakley (Eds.), *The Oxford handbook of hypnosis: Theory, research, and practice* (pp. 141-177). London: Routledge.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In: Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.), *Handbook of Psychophysiology*, 2, 142–162. New York, NY: Cambridge University Press

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review, 4*(2), 221-225.

Braver, T.S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113.

Bowers, P. (1982). The classic suggestion effect: Relationships with scales of hypnotizability, effortless experiencing, and imagery vividness. *International Journal of Clinical and Experimental Hypnosis, 30*(3), 270-279.

Campbell, F.W. & Gregory, A. H. (1960). Effect of size of pupil on visual acuity. *Nature, 187*, 1121-1123.

Chiew, K.S., Braver, T.S. (2013). Temporal dynamics of motivation–cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Psychology, 4*:15.

Crawford, H. J. (1996). Cerebral brain dynamics of mental imagery: Evidence and issues for Hypnosis and imagination (pp. 253-281). In R. G. Kunzendorf, N. P. Spanos, & B. Wallace (Eds.), *Hypnosis and Imagination*. Amityville, New York: Baywood Publishing Company.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: dissociable causes of interference effects in conflict situations. *Acta Psychologica*, **101**(2-3), 379–394.

Derbyshire, S. W. G., Whalley, M. G., Stenger, V. A., & Oakley, D. A. (2004). Cerebral activation during hypnotically induced and imagined pain. *Neuroimage*, **23**(1), 392–401.

Diede, N. T. & Bugg, J. M. (2017). Cognitive effort is modulated outside of the explicit awareness of conflict frequency: Evidence from pupillometry. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **43**(5), 824–835.

Dienes, Z. (2012). Is hypnotic responding the strategic relinquishment of metacognition?. In Beran, M., Brandl, J. L., Perner, J., & Proust, J. (Eds.), *Foundations of metacognition* (pp. 267-277). Oxford University Press.

Dienes, Z. (2014). Using Bayes to get the most out of nonsignificant results. Frontiers in Psychology, **5**, 781.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78-89.

Dienes, Z., Palfi, B., & Lush, P. (in press). Controlling phenomenology by being unaware of intentions. In Weisberg, J. (Ed.),*Qualitative Consciousness: Themes from the Philosophy of David Rosenthal*. Cambridge University Press. https://psyarxiv.com/4zw6g/

Dienes, Z., Lush, P., Palfi, B., Scott., R., Parris, B. A., Seth., A., & Lovell, M. (in press). Phenomenological Control as Cold Control: Hypnosis and beyond. *Psychology of Consciousness.*

Dienes, Z., & Perner, J. (2007). Executive control without conscious awareness: the cold control theory of hypnosis. In Jamieson, G. (Ed.), *Hypnosis and conscious states: The cognitive neuroscience perspective*, (pp. 293-314). Oxford University Press.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*, 447-456.

Farvolden, P., & Woody, E. Z. (2004). Hypnosis, memory, and frontal executive functioning. *The International Journal of Clinical and Experimental Hypnosis*, *52*(1), 3–26. http://doi.org/10.1076/iceh.52.1.3.23926

Fischer B. & Boch R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Research*, **260** (1), 21–26.

Fischer B. & Ramsperger E. (1984). Human express saccades: Extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, **57** (1), 191–195.

Fischer B., Weber H., Biscaldi M., Aiple F., Otto P.,& Stuhr V. (1993). Separate populations of visually guided saccades in humans: Reaction times and amplitudes. *Experimental Brain Research*, **92** (3), 528–541.

Geiger, E., Peter, B., Prade, T., & Piesbergen, C. (2014). Intelligence and hypnotizability: is there a connection? *The International Journal of Clinical and Experimental Hypnosis*,

62(3), 310–329. http://doi.org/10.1080/00207144.2014.901083.

Goldfarb, L., Aisenberg, D., & Henik, A. (2011). Think the thought, walk the walk – social priming reduces the Stroop effect. *Cognition, 118*(2), 193-200.

Gruzelier, J., & Warren, K. (1993). Neuropsychological evidence of reductions on left frontal tests with hypnosis. *Psychological Medicine*, *23*(01), 93–101. http://doi.org/10.1017/S0033291700038885

Hasshim, N., & Parris, B. A. (2015). Assessing stimulus-stimulus (semantic) conflict in the Stroop task using saccadic two-to-one colour response mapping and preresponse pupillary measures. *Attention, Perception and Psychophysics, 77,* 2601-2610.

Hasshim, N. & Parris, B. A. (2018). Trial type mixing substantially reduces the response set effect in the Stroop task. *Acta Psychologica, 189,* 43-53.

Hermans, F. & Walker, R. (2012). The site of interference in the saccadic Stroop effect. *Vision Research, 73*, 10-22.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, **132**(3423), 349–350.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, **143**(3611), 1190–1192.

Hodgson, T. L., Parris, B. A., Gregory, N. J., & Jarvis, T. (2009). The saccadic Stroop effect: Evidence for involuntary programming of eye movements by linguistic cues.

Vision Research, *49*, 569–574.

Jamieson, G. A., Dwivedi, P., & Gruzelier, J. H. (2005). Changes in mismatch negativity across pre-hypnosis, hypnosis and post-hypnosis conditions distinguish high from low hypnotic susceptibility groups. *Brain Research Bulletin*, *67*(4), 298–303. http://doi.org/10.1016/j.brainresbull.2005.06.033.

Jamieson, G. A., & Sheehan, P. W. (2004). An empirical test of Woody and Bowers's dissociated-control theory of hypnosis. *The International Journal of Clinical and Experimental Hypnosis*, *52*(3), 232–249. http://doi.org/10.1080/0020714049052349

Jamieson, G. A., & Woody, E. Z. (2007). Dissociated control as a paradigm for cognitive neuroscience research and theorizing in hypnosis. In G. Jamieson (Ed.), *Hypnosis and Conscious States* (pp. 111-129). Oxford, UK: Oxford University Press.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, **154**(3756), 1583–1585.

Kello, C. T., Plaut, D. C. & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General, 129*(3), 340-360.

Kosslyn, S. M., Thompson, W. L., Costantini-Ferrando, M. F., Alpert, N. M., & Spiegel, D. (2000). Hypnotic visual illusion alters color processing in the brain. *American Journal of Psychiatry*, *157*(8), 1279–1284.

Krebs, R.M., Boehler, C. N. & Woldorff, M. G. (2010). The influence of reward

associations on conflict processing in the Stroop task. *Cognition, 117*(3), 341-347.

Laeng, B., Ørbo, M., Holmlund, T., & Mizzo, M. (2011). Pupillary Stroop effects. *Cognitive Processes, 12*(1), 13-21.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: a window to the preconscious? *Perspectives on Psychological Science, 7*(1), 18-27.

Landry, M., Lifshitz, M., & Raz, A. (2015). Brain correlates of hypnosis: A systematic and meta-analytic exploration. Neuroscience & Biobehavioral Reviews, **57**(3), 285-313.

Liang, J. & Williams, D. R. (1997). Aberrations and retinal image quality of the normal human eye. *Journal of the Ophthalmological Society of America A, 14*(11), 2873-2883.

Lifshitz, M., Aubert-Bonn, N., Fischer, A., Kasham, I. F., and Raz, A. (2013). Using suggestion to modulate automatic processes: From Stroop to McGurk and beyond. *Cortex, 49*(2), 463-473.

Lush, P., Caspar, E. A., Cleermans, A., Haggard, P., Magalhães De Saldanha da Gama, P. A., & Dienes, Z. (2017). The power of suggestion: Post-hypnotically induced changes in the temporal binding of intentional action outcomes. *Psychological Science, 28* (5), 661-669.

Lynn, S. J., Rhue, J. W., & Weekes, J. R. (1990). Hypnotic involuntariness: A social cognitive analysis. *Psychological Review, 97*(2), 169-184.

MacLeod, C. M. (1992). The Stroop task: The" gold standard" of attentional measures. *Journal of Experimental Psychology: General, 121*(1), 12-14.

Mathôt, S., Fabius, J., Heusden, E. V., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods, 50*, 94-106.

Massar, S. A. A., Sasmita, K., Lim, J., & Chee, M. L. W. (2018). Motivation alters implicit temporal attention through sustained and transient mechanisms: A behavioural and pupillometric study. *Psychophysiology, 55*(12), e13275.

Oakley, D. A., & Halligan, P. W. (2009). Hypnotic suggestion and cognitive neuroscience. *Trends in Cognitive Sciences*, *13*(6), 264–270. http://doi.org/10.1016/j.tics.2009.03.004.

Oakley, D. A., & Halligan, P. W. (2013). Hypnotic suggestion: opportunities for cognitive neuroscience. *Nature Reviews Neuroscience* , *14*, 565–576.

Page, R. A. & Green, J. P. (2007). An update on Age, hypnotic suggestibility and gender: A brief report. *The American Journal of Clinical Hypnosis, 49*(4), 283-287.

Palfi, B., Parris, B. A. & Dienes, Z. (submitted). Strategies that reduce Stroop interference.

Parris, B. A. (2014). Task conflict in the Stroop task: When Stroop interference decreases as Stroop facilitation increases in a low task conflict context. *Frontiers in Psychology, 5*:1182.

Parris, B. A. (2017). The Role of Frontal Executive Functions in Hypnosis and Hypnotic Suggestibility. *Psychology of Consciousness: Theory, Research and Practice, 4*(2), 211-229.

Parris, B. A., Dienes, Z., Bate, S. & Gothard, S. (2014). Oxytocin impedes the effect of the word blindness posthypnotic suggestion on Stroop task performance. *Social, Cognitive and Affective Neuroscience, 9*(7), 895-9.

Parris, B. A., Dienes, Z., & Hodgson, T.L. (2012). Temporal constraints of the word-blindness post-hypnotic suggestion on Stroop task performance. *Journal of Experimental Psychology Human Perception and Performance, 38*, 833–837.

Parris, B. A., Dienes, Z., & Hodgson, T.L. (2013). Application of the ex-Gaussian function to the effect of the word blindness suggestion on Stroop task performance suggests no word blindness. *Frontiers in Psychology, 4*:647.

Parris, B. A., Sharma, D., & Weekes, B. S. (2007). An optimal viewing position effect in the Stroop task when only one letter the colour carrier. *Experimental Psychology, 54*(4), 273-280.

Parris, B. A., Sharma, D., Hackett Weekes, B. S., Momenian, M., Augustinova, M. & Ferrand, L. (2019). Response modality and the Stroop task: Are there phonological Stroop effects with manual responses? *Experimental Psychology, 66*(5), 361-367.

Parris, B. A., Hasshim, N., Wadsley, M. G., Augustinova, M. & Ferrand, L. (submitted). The loci of Stroop effects: A critical review of methods and evidence for levels of

processing contributing to colour-word Stroop effects and the implications for the loci of attentional selection.

Rahmanovic, A., Barnier, A. J., Cox, R. E., Langdon, R. A., & Coltheart, M. (2011). "That's not my arm": A hypnotic analogue of somatoparaphrenia. *Cognitive Neuropsychiatry*, *17*(1), 36–63.

Rainville, P., Carrier, B., Hofbauer, R. K., Bushnell, C. M., & Duncan, G. H. (1999). Dissociation of sensory and affective dimensions of pain using hypnotic modulation. *Pain*, *82*(2), 159–171. http://doi.org/10.1016/S0304-3959(99)00048-2

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Raz, A. & Campbell, N. K. J. (2011). Can suggestion obviate reading? Supplementing primary Stroop evidence with exploratory negative priming analysis. *Consciousness and Cognition, 20*(2), 312-320.

Raz, A., Fan, J., & Posner, M. (2005). Hypnotic suggestion reduces conflict in the human brain. *Proceedings of the National Academy of Sciences of the U.S.A, 102*, 9978–9983.

Raz, A., Kirsch, I., Pollard, J., & Nitkin-Kaner, Y. (2006). Suggestion reduces the stroop effect. *Psychological Science, 17* (2), 91-5.

Raz, A., Landzberg, K. S., Schweizer, H. R., Zephrani, Z. R., Shapiro, T., Fan, J., & Posner, M. I. (2003). Posthypnotic suggestion and the modulation of Stroop interference under cycloplegia. *Consciousness and Cognition, 12*, 332-346.

Raz, A., Shapiro, T., Fan, J., & Posner, M. (2002). Hypnotic suggestion and the modulation of Stroop interference. *Archives of General Psychiatry, 59* (12), 1155-61.

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*(5), 1033-1040.

Sheedy, J. E., Truong, S. D., & Hayes, J. R. (2003). What are the visual benefits of eyelid squinting? *Optometry and Vision Science, 80*(11), 740-744.

Sheehan, P. W., Donovan, P. & MacLeod, C. M. (1988). Strategy manipulation and the Stroop effect in hypnosis. *Journal of Abnormal Psychology, 97*(4), 455–460.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experiment Psychology, 38*(3), 475-491.

Simpson, H. M. (1969). Effects of a task-relevant response on pupil size. *Psychophysiology, 6*, 115–121.

Singh, N. & Mishra, R. K. (2012). Does language proficiency modulate oculomotor control? Evidence from Hindi-English bilinguals. *Bilingualism, 15*(4), 771-781.

Skrzypek, A., Szeliga, M., Jabloński, K. & Nowakowski, M. (2017). Effect of night shift-

related tiredness on eye saccades. *Neurophysiology, 49*(12) 1-4.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Tan, G., Rintala, D. H., Jensen, M. P., Fukui, T., Smith, D., & Williams, W. (2015). A randomized controlled trial of hypnosis compared with biofeedback for adults with chronic low back pain. *European Journal of Pain*, *19*(2), 271–280. http://doi.org/10.1002/ejp.545.

Terhune, D. B., Cleeremans, A., Raz, A. & Lynn, S. J. (2017). Hypnosis and top-down regulation of consciousness. *Neuroscience and Biobehavioral Reviews*, **81**(A), pp. 59-74.

Thompson, Trevor; Terhune, Devin Blair; Oram, Charlotte; Sharangparni, Joseph; Rouf, Rommana; Solmi, Marco; Veronese, Nicola and Stubbs, Brendon. 2019. The effectiveness of hypnosis for pain relief: A systematic review and meta-analysis of 85 controlled experimental trials. *Neuroscience & Biobehavioral Reviews*, **99,** pp. 298-310

Wagstaff, G. F., Cole, J. C., & Brunas-Wagstaff, J. (2007). Effects of hypnotic induction and hypnotic depth on phonemic fluency: A test of the frontal inhibition account of hypnosis. *International Journal of Psychology and Psychological Therapy, 7,* 27-40.

Wenban-Smith M. G.,& Findlay J. M. (1991). Express saccades: Is there a separate population in humans?*Experimental Brain Research*, **87** (1), 218–222.

Weitzenhoffer, A. M. (1980). Hypnotic susceptibility revisited. *American Journal of Clinical Hypnosis, 22*, 130-146.

Weitzenhoffer, A. M, & Hilgard, E. R. (1962). *Stanford Hypnotic Susceptibility Scale: Form C*. Palo Alto, CA: Consulting Psychologists Press.

Woodhouse, J. M. (1975). The effect of pupil size on grating detection at various contrast levels. *Vision Research*, *15*(6), 645–648.

Table 1: Saccade latencies with standard deviations in brackets for each condition in

milliseconds.

	Suggestion Absent	Suggestion Present
Response set	516 (98)	470 (111)
Non-Response set	462 (79)	523 (93)
Neutral	444 (100)	469 (116)
Congruent	430 (78)	424 (74)

Table 2: Magnitudes of Stroop effects (in ms) for each condition.

Stroop effects	*Suggestion Absent*	*Suggestion Present*	*Difference*
Interference (Response set - Neutral)	72	-1	73*^
Response conflict (Response set – Non-response set)	54	-53	107*^
Semantic conflict (Non-response set - Neutral)	18	54	-36
Facilitation (Neutral - Congruent)	14	46	-32

*p<.05 for change following the suggestion.

^Bayes Factor >3 for change following the suggestion.

Figure captions

Figure 1. The trial sequence and pupil sampling periods.

Figure 2. Scatterplot (with density displayed above) of the observed Word Blindness Suggestion Effect showing the Stroop interference effect (Incongruent – Neutral trials) for saccade latency was smaller in the Suggestion Present condition compared to the Suggestion Absent condition for the majority of the participants.

Figure 3. Bar chart showing pupil size changes in both conditions and for each trial type in the intratrial period. Only the main effects (of Suggestion condition (Absent, Present) and Word Type (response set, non-response set, neutral, congruent)) were significant.

Figure 4. Scatterplot (with density displayed above) showing increased overall intratrial pupil sizes in the Suggestion Present condition which is taken as evidence for the recruitment of effortful control to achieve the word blindness suggestion effect.