Evolution and Human Behavior 40 (2019) 269–280

Contents lists available at ScienceDirect

Evolution and Human Behavior

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

Do 'watching eyes' influence antisocial behavior? A systematic review & meta-analysis

Keith Dear*, Kevin Dutton, Elaine Fox

Department of Experimental Psychology, University of Oxford, United Kingdom

ARTICLE INFO

Keywords:
Watching eyes
Eye cues
Eye images
Surveillance cues
Eye primes
Implicit social cues

ABSTRACT

Eye cues have been shown to stimulate rapid, reflexive, unconscious processing and in many experimental settings to cue increased prosocial and decreased antisocial behaviour. Eye cues are being widely applied in public policy to reduce crime and antisocial behaviour. Recently, failed replication attempts and two meta-analyses examining the eye cue effect on generosity have raised doubts regarding earlier findings. Much of the wider evidence on eye cues has still not been systematically reviewed, notably that which is most relevant to its practical application: the effect of eye cues on antisocial behaviour. Given the evidence of humans' heightened sensitivity to threat and negative information, we hypothesized that the watching eyes effect would be more consistent in studies examining antisocial behaviour. In our meta-analysis of 15 experiments from 13 research papers we report a reduction in the risk of antisocial behaviour of 35% when eye cues are present. By contrast, systematic reviews have suggested CCTV cameras reduce crime by only 16%. We conclude that there is sufficient evidence of a watching eyes effect on antisocial behaviour to justify their use in the very low-cost and potentially high-impact real-world interventions that are proliferating in public policy, particularly in the UK.

Public significance statement: Our meta-analysis of 15 experiments involving 2035 participants shows that photographs and/or stylized images of eyes reduced antisocial behaviour by 35%. Our findings support public policy initiatives employing pictures of 'watching eyes' to reduce crime. Furthermore, in an age when we are watched more than at any time in modern history – both online and on the street – our findings highlight an urgent need to fully understand the effect that perceived surveillance, feeling watched, has on our decisions and actions.

1. Introduction

The 'watching eyes' effect refers to a strange phenomenon whereby the mere presence of pictures of eyes or stylized eye images (hereafter: eye cues) is enough to cause people to adjust their behavior (e.g. Bateson, Nettle, & Roberts, 2006; Burnham & Hare, 2007; Haley & Fessler, 2005). The 'watching eyes effect' suggests that just feeling watched may be enough to make us modify our actions independent of deliberative, explicit, conscious, evaluation of the costs and benefits of an action.

Much of the research on the influence of pictures of eyes on behaviour has focused on prosocial behaviour. For instance, a picture of human eyes added to charity donation buckets in supermarkets increased donations by 48% in comparison with a control condition (Powell, Roberts & Nettle 2012).

In addition to field experiments, laboratory studies have also addressed the question of whether eye cues increase generosity, utilizing the Dictator Game (Nettle et al., 2013) and similar economic games. Two of the earliest studies in the field serve as examples. In 2005, Haley & Fessler found subtle watching eye cues increased generosity in the Dictator game by 31%, and when the eye cues were more clearly visible by 55%. A similar eye cues & economic game design run by Burnham & Hare in 2007 suggested that the presence of eyes increased altruistic contributions in an economic simulation by 29% in comparison with a control.

More recent research has challenged earlier findings on the effect of eyes cues on prosocial behaviour. A series of replication attempts and new experiments failed to find any consistent evidence for an eye cue effect (Beyfus et al., 2016; Brudermann, Bartel, Fenzl, & Seebauer, 2015; Bush, Erlich, Prather, & Zeira, 2016; Cai, Huang, Wu, & Kou, 2015; Carbon & Hesslinger, 2011; Fehr & Schneider, 2010; Fujii, Takagishi, Koizumi, & Okada, 2015; Gaiani, Rose, & Roberts, 2014; Golja, 2013; Huang, Liu, Zheng, Tan, & Zhao, 2015; Jackson, 2015; Jolij & de Haan, 2014; Kuliga, Tanja-Dijkstra, & Verhoeven, 2011;

* Corresponding author at: Department of Experimental Psychology, University of Oxford, University Department of Experimental Psychology, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX1 3UD, United Kingdom.
E-mail address: keith.dear@psy.ox.ac.uk (K. Dear).

https://doi.org/10.1016/j.evolhumbehav.2019.01.006
Received 13 October 2017; Received in revised form 29 January 2019; Accepted 29 January 2019

Lamba & Mace, 2010; Matland & Murray, 2016; Matsugasaki, Tsukamoto, & Ohtsubo, 2015; Northover, Pedersen, Cohen, & Andrews, 2017b; Palomäki, Modic, & Yan, 2015; Raihani & Bshary, 2012; Rose, Gaiani, & Roberts, 2014; Sparks, 2010; Sparks & Barclay, 2015; Stella et al., 2013; Tane & Takezawa, 2011; Vogt, Efferson, Berger, & Fehr, 2015; Waktare & Roberts, 2014; White, 2015).

There has been a broad consensus since the first eye cues experiments (Burnham, 2003; Burnham & Hare, 2007) that if eye cues work it is by making us feel watched. Burnham and Hare called this 'the evolutionary legacy hypothesis', the idea that we evolved a pronounced responsiveness to feeling watched and adapt our behaviour initially to protect us from predation, and later in social groups, to protect our reputation. There is now a significant body of neuropsychological research which suggests this is indeed the case. In their unified theory aiming to explain the mechanism through which eye cues influence behaviour, Conty, George, and Hietanen (2016) suggest that watching eyes prompt a subconscious reputation management mechanism to change our behaviour. First, eye cues' highly salient nature captures attention which, second, triggers heightened self-referential processing. In triggering self-referential processing, eye cues are unconsciously prompting heightened concern over how we are socially evaluated. These subconscious reputational concerns can lead us to moderate our behaviour (Conty et al., 2016). In short, this false cue can alter behaviour by changing the way we process information, leading us to act as if are watched and our reputation is at stake even when in fact it is not. If this is so, we should consider the reputational consequences of behaviours when predicting whether eye cues are likely to elicit a strong effect on decision making and also in predicting the direction of any effect.

We argue that the inconsistent results in watching eyes effects studies and in recent meta-analyses may be due to the varying effects of prosocial behaviour on reputation. In contrast we both propose, and test, the hypothesis that eye cues should more reliably reduce antisocial behaviour, since antisocial behaviour is more consistently reputationally damaging.

In exploring our hypothesis, we first review the evidence for the 'watching eyes' effect on prosocial behaviour highlighting its inconsistency. Second, we provide evidence for our argument that antisocial behaviour may be more consistent in damaging reputation than prosocial behaviour is in enhancing it. We then conduct a meta-analysis to test our hypothesis before reviewing our findings.

The first reviews in the field suggested there was a reliable influence of eye cues on prosocial behaviour. (Nettle et al., 2013; Sparks & Barclay, 2013). However, Nettle et al.'s (2013) conclusion that in dictator games the presence of eye cues increased not the mean donation but the probability of donating something rather than nothing was recently challenged by Northover, Pedersen, Cohen, and Andrews (2017a) whose larger meta-analysis and review found no effect of eye cues on 'the proportion who gave' in generosity tasks across 27 experiments. Furthermore, in a separate analysis Northover and colleagues also found that the effect size was extremely small when comparing the difference between mean donations in the eyes and no eyes conditions across 26 dictator game and charitable giving experiments (Northover et al., 2017a). Northover and colleagues also investigated the effect of eye cues of moral judgement. They analysed six studies and found no statistically significant eye cue effect, an effect size close to zero and a confidence interval crossing zero (Northover et al., 2017b). These inconsistent findings might also be explained by the variable effects of generosity on reputation.

2. Eye cues and antisocial behaviour

Evidence for the inconsistent effect of generosity on reputation can be found in a series of 11 experiments conducted by Klein and Epley (2014). These show that generosity is not always good for one's reputation. In fact, while some generosity enhances reputation, greater generosity delivers diminishing reputational returns and extreme generosity can even be damaging.

Klein and Epley ran scenarios in which participants rated the reputation of a fictional character "Tom" after reading scenarios that described how much he had donated for a charity concert ticket: below the suggested amount, at the suggested amount or above it. No significant difference was found in ratings of Tom's competence across the 3 levels of donation. This is as expected given the description was identical barring the figure for how much he donated. In contrast, the amount he donated did matter to his reputation for warmth. Behaving fairly by donating as suggested rather than unfairly, donating below the recommended amount, improved ratings of Tom's 'warmth'. But greater generosity did not bring higher 'warmth' ratings. Warmth is related to judgements of friendliness, trustworthiness, and morality and is more dominant in the formation of reputation than competence (Fiske, Cuddy, & Glick, 2007; Wojciszke, Bazinska, & Jaworski, 1998).

Klein and Epley conducted additional experiments to further test the relationship between prosociality and reputation via the online crowdsourcing site Amazon Mechanical Turk, and in laboratory-based settings with similar results. Competence ratings varied little, but warmth ratings did, with extreme generosity decreasing rather than increasing reputational ratings. As the authors put it, 'it pays to be nice, but pays no more to be really nice'. An extension of this research across 7 countries and cultures has found similar results (Klein, Grossmann, Uskul, Kraus, & Epley, 2015). There is no consistent effect of generosity on reputation, so we should not expect eye cues to cause a consistent increase in generous behaviour.

Surveillance cues, if they provoke reputational concerns, should have more consistent effects in reducing antisocial and selfish behaviour than in promoting prosocial acts because of our acute sensitivity to negative information (Rozin & Royzman, 2001). In an extensive review, instructively titled *Bad is Stronger than Good*, Baumeister, Bratslavsky, Finkenauer, and Vohs (2001) describe a number of experimental findings that suggest why this might be so. They show: how negative stereotypes are more prevalent, more enduring and more influential than positive; how our self-esteem is more sensitive to criticism than it is to praise; how we respond more to critical feedback than to praise, learning faster from negative experiences than positive; how we prioritise the processing of negative over positive information, allocating greater attention and cognitive resource to it; how we tend to remember bad events longer so antisocial acts echo longer in the memory of the observer than prosocial behaviours, and how negative experiences and negative language in relationships exert more influence on ratings of relationship happiness than positive experiences and positive language – thus we might be keener to avoid providing negative signals about ourselves for others to gossip about than we are to seek to provide positive signals and be talked about positively. They also suggest that negative affect (emotion) and emotional distress have a greater and more enduring influence on us than positive affect and pleasant emotions, so we would be more motivated to avoid the censure that follows signalling bad character than to signal good character to experience warm emotions. Thus, the evidence from the psychological research summarised by Baumeister et al. (2001) suggests negative information should have consistently damaging – and lasting – effects on our reputation.

This contention that the reputational threat from displaying antisocial behaviour should be greater and more consistent than that from failing to show generous behaviour is supported by evidence from evolutionary psychology too. Punishment in social groups is necessary for group success. It helps prevent free-riding and social infractions, enforce hierarchy to enable group functioning and it sustains group cohesion (Fehr & Gachter, 2002; Gürerk, Irlenbusch, & Rockenbach, 2006). Indeed, an acute fear of being watched may have been essential for social groups to succeed in order to enforce norms without constant combat and punishment. The difficulty of extending surveillance beyond the small groups of our early evolutionary history may have been

responsible for the emergence of belief in all-seeing, punishing supernatural agents – gods and spirits – who could ensure we all felt watched even in the absence of others (Johnson, 2016). We should therefore expect stronger, more consistent responses to reputational threat (being seen to behave antisocially) than ambiguous reputational opportunity (being seen to be generous).

In the evolutionary context it is interesting to note that participants in Klein and Epley's experiments expected to earn a better reputation by being more generous. This common-sense assumption underpins experiments examining the effect of watching eyes on generosity. While the assumption that greater generosity generates consistent and linear reputational benefits is unsound it may still be adaptive. Expecting generosity of others creates an explicit social norm that would be of benefit to the group – we should all give, think highly of those who are generous and thus both individual and the group gain if we give when watched. However, an unconscious mechanism promoting prosociality even when not watched would encourage maladaptive behaviours for the individual – leading them to give away resources without prospect of reward.

Given this logic it is unsurprising that cues to surveillance such as watching eyes produce varied responses on people's generosity. In contrast, having an unconscious mechanism to avoid unfairness and antisociality when watched would protect the individual's reputation and thus avoid damaging reprisals. Thus, there are sound theoretical reasons to focus our meta-analysis on the effect of watching eyes on anti-social behaviour.

There are sound practical reasons for this focus too. The eye cue effect on antisocial behaviour is already being put to practical use across Britain. In 2006 West Midlands Police used 100 buses to promote the message 'We're Keeping an Eye on Crime' along with a picture of eyes to try to reduce bicycle and other theft (West Midlands Police, 2006). British Police in the Nottinghamshire town of Hucknall claim posters of eyes reduced crime by 40% (Flanagan, 2013). Eye cues have been added to trees to reduce littering in the Forest of Dean, and are being used by Britain's HM Revenue and Customs Service to discourage tax evasion (BBC News, 2015; Knapton, 2016; Nelson, 2013). Press reports indicate that the posters have been used across Britain's rail network to deter crime (Basildon, Canvey, & Southend Echo, 2013). At the time of writing eye cues are appearing at motorway service stations across Britain as Keep Britain Tidy expands an initiative to reduce littering by drivers, with preliminary results suggesting that they have caused a 23% reduction in littering (Extra Services, 2015; H.M. Government (UK), 2017).

In fact, so well established is the idea in policy circles that the UK Government's, 2017 National Anti-Littering Strategy recommends the use of watching eyes interventions to reduce littering (H.M. Government (UK), 2017). Thus, given our hypothesis and its practical implications, we focus our meta-analysis on the watching eyes' effect on antisocial behaviour.

2.1. Methods

PRISMA 2009 flow diagram.

2.2. Types of intervention

We reviewed 113 full-text articles comparing between groups exposed to eye cues and those in a control condition. We examined both laboratory and field experiments. We considered for meta-analysis 19 articles identified as examining the watching eyes effect on antisocial behaviour, excluding a further 6 due to issues with the experimental design or the unavailability of data as described below.

2.3. Types of outcome measures

The outcome measure for the meta-analysis was antisocial behaviour.

2.4. Search strategy

Initially, we conducted a full search without limitation by outcome measure to document all eye-cue relevant research in one place for the first time. In order to take as full a view of the watching eyes field as possible we searched Web of Science by topic using the search terms "watching eyes", "eye cues", "eye cue", "eye spot", "eyespot", "eye-like", "social cues", "eye-images", "watching you", "surveillance cues", "images of eyes", "observation cues", "perception of human face", "face cues", "cues of observation", "gaze cues", "implicit gaze", "implicit observability", "eye primes", and by title using the terms "implicit social", "watching you", "social eyes OR social-eyes" (See Additional Materials for full details of our search strings). We excluded duplicates counting only studies not previously identified using earlier search terms. We identified studies in the bibliography of the most recent and thorough review (Northover et al., 2017a).

Noting that Northover et al. had just completed an extensive search for unpublished research at the same time as our review was written we did not duplicate this effort by announcing our own formal call for unpublished papers. However, we did conduct our own private search for unpublished material - searching by author name when alerted to their interest in the watching eyes effect in other papers - and in doing so identified 4 additional studies to those picked up by keyword searches,. Two studies (Palomäki et al., 2015; Traver & Cordell, 2014) were identified via the abstract booklet of the *Decepticon* conference attended by a colleague. Palomäki et al.'s unpublished study was received via correspondence with the author and is included in the meta-analysis.

We searched studies using Google Scholar's 'cited by' function to identify further research in different domains that we might not otherwise have found (generosity in field and lab, voting behaviour, minimal cues), identifying a further 15 studies. Seeking real-world applications of the research we conducted a general web search using Google, finding a further large-scale field experiment (and a number currently underway).

An extensive search was made of Newcastle University's online repository (http://research.ncl.ac.uk) which yielded a number of unpublished Masters' research studies.

We repeated our search prior to submitting for publication in March 2017 adding a further 13 studies.

All searches are documented in the Excel tabulation in 'Additional Materials' and provide a comprehensive overview of the 48 different outcome measures identified ranging from the watching eyes effect on generosity and hand-washing to their influence on choice justification and poker risk-taking.

This is, to the best our knowledge, the most comprehensive database of watching eyes studies yet compiled.

2.5. Meta-analysis study selection

All studies containing the search terms were reviewed by title and abstract. Where their relevance remained unclear the full text was accessed. Studies were eliminated from, or included in the review in this way on the first search. Further refinement for the meta-analysis was undertaken following the recording and categorisation of the studies [See Excel spreadsheet in additional materials].

With all watching eyes studies documented in Excel format we filtered the table to identify watching eyes studies examining antisocial behaviour, selecting from the column 'outcome measure' the antisocial behaviour categories: 'Advice Giving', 'Amount Taken', 'Antisocial Behaviour', 'Cheating', 'Corruption', 'Cycle Theft', 'Dishonesty', 'Dog fouling', 'Engine Idling', 'Honesty', 'Littering', and 'Lying'. This selection process returned 19 papers. All others were excluded as they did not examine antisocial behaviour [this process can be replicated via the Excel database in additional materials]. The 19 papers selected all

studied behaviours that met the Oxford English Dictionary definition of antisocial, that is, 'contrary to the laws and customs of society, in a way that causes annoyance and disapproval in others'.

From these 19 papers we excluded six. Bateson et al.', 2006 study of the effect of images of watching eyes on donations to a coffee bar honesty box was excluded because its design did not allow us to know whether the experiment examined conformity to norms, antisocial behaviour or generosity. It did not, for example, count or track how much individuals donated, or how many did or did not donate. We therefore cannot know whether eye cues increased the generosity of those who would have donated anyway or whether eye cues caused people who would have free-loaded and failed to donate, to contribute. We are also unable to ascertain whether this was a test of the eye cue effect on antisocial behaviour or whether their result was a combination of both the eye cue effect increasing generosity in some and reducing antisocial free-riding in others The same criteria resulted in the exclusion of Brudermann et al.'s (2015) field experiment examining contributions to newspaper honesty boxes.

We excluded Oda, Kato, and Hiraishi's (2015) study examining the eyes effect on lying principally because their design construed lying as pro-, as opposed to antisocial behaviour.

Traver and Cordell (2014) was excluded as we could not create a dichotomous variable suitable for our odds ratio-based meta-analysis. These authors examined the eye cue effect in a coin toss and a dice roll experiment (where flipping a head or rolling a higher number brought a greater reward) reporting the number of outcomes against the expected frequencies to identify how frequently people lied to gain personal advantage. We could not calculate the frequency of lying vs not lying from the proportions as all frequencies lay within one standard deviation of the expected frequencies, and thus could have been down to chance.

One study (Chowdhury, Jeon, & Saha, 2014), was excluded as the authors were unwilling to share the data on which their study – cited here - was based because they were seeking re-publication of their paper in another journal. There was insufficient information reported in the available paper to assess dichotomously who did and did not behave antisocially [data requested on 12 Oct 2016 and refused in writing on 13 October 2016 & again on 23 August 2017]. Data was requested from the authors for Li et al. (2018) but no reply was received at the time of submission for publication.

The 13 studies remaining reported results for a total of 15 eligible experiments. We transformed all findings into dichotomous variables to test the simple effect of eye cues on antisocial behaviour via the log odds ratio. This measure is intuitively meaningful because it indicates the relative change in frequency of antisocial behaviour in the experimental condition or area compared with the control condition or area. For example, in the study of littering in a café by Ernest-Jones, Nettle, and Bateson (2011) we transformed the data into dichotomous 'littered or did not litter' variables from the original data. This comes at the cost of hiding significant interactions e.g. in the Ernest-Jones' et al. example they found a significant effect of eye cues but only when > 6 people were present. Such an effect is hidden when we transform the data to the dichotomous did or did not litter by eyes and control.

Within some of the papers selected some experiments were eligible for inclusion in the meta-analysis while others were not. For example, in Zengerink (2013) Experiment 1 involved the completion of a series of surveys with eye cues present but did not examine antisocial behaviour and was therefore excluded. Experiment 2 was a field experiment examining the effect of eye cues on antisocial behaviour and therefore was included. The same approach was taken with two other experiments examining the eye cue effect on litter (Bateson et al., 2015; Zengerink, 2013).

We reinterpreted the data from Cai et al. (2015) examining whether eye cues affected participants' tendency to lie for personal advantage to show the frequency of lying or not lying in the eyes and control condition. Palomäki et al. (2015) also examined the effect of eye cues on lying, in their case in an online simulated insurance claims process. We re-analysed their data to provide a dichotomous outcome: lied or did not lie. We reinterpreted Hoffman et al.'s data from their study of the effect of robot eyes on lying (2015) to provide dichotomous proportions of those that cheated or did not cheat and reinterpreted the proportions as frequencies.

Huang et al. (2015) examined whether participants would bribe or not bribe in scenario-based economic games studying corruption. Their Experiments 2 and 3 include an 'eye cue or no eyes' condition enabling this data to be translated for meta-analysis.

In Baillon, Selim, and van Dolder' (2013) game participants could pay a small cost to destroy part of a payoff to another participant gaining nothing in the process except the possible satisfaction of damaging the other's interest. We transformed their data to report a dichotomous outcome "destroyed" or "did not destroy" in the eyes and control conditions.

Finally, in an extended analysis, we included Nettle, Nott, and Bateson's (2012) examination of the effect of eyes on bicycle crime and Keep Britain Tidy's (2014) eye cue intervention designed to reduce dog fouling. In Nettle et al. (2012), Bateson, Callow, Holmes, Roche, and Nettle (2013), and some of Keep Britain Tidy's (2014) crime reduction interventions the eye images were used along with text explicitly stating that participants were being watched. For example, in Nettle et al.'s study the eye cues sign included the text 'Cycle Thieves We're Watching You' and a police logo. Bateson et al. (2013) used the same signs as Nettle et al. (2012) in their experiment. Similarly, Keep Britain Tidy's intervention posters all included text and various organisational logos. Consequently, it might be argued that either the logos or the text alone, or the two in combination, caused the reduction in the bicycle theft, and not the watching eyes. Indeed just erecting signs alone can reduce crime (McNees, Egli, Marshall, Schnell and Risley, 1976) and so we may not be able to attribute any crime reduction to the presence of eye cues in these two studies. Additionally, it is plausible that people seeing the signs could have thought they suggested CCTV was present, which perhaps might also reduce crime although the evidence for the crime-reducing effects of CCTV is inconsistent (Welsh & Farrington, 2009).

One could make a case for excluding Nettle et al. (2012), and some of Keep Britain Tidy's (2014) studies due to the presence of these potential confounds. However, given that Nettle et al. (2012) is perhaps the most influential paper in this area on public policy in the UK, and also represents one of the baseline studies in the research field, many would be surprised to see it excluded. Indeed, reviewers have offered conflicting advice on this point. Consequently, we calculate log odds ratios with and without the inclusion of Nettle et al. (2012), Bateson et al. (2013) and Keep Britain Tidy (2014) to satisfy both perspectives.

Keep Britain Tidy's study had 4 types of poster, all with eyes, but with differing text to examine whether this increased or decreased the effectiveness of the intervention. As well as having pictures of eyes, these four poster designs also displayed the text 'Thoughtless Dog Owners, We're Watching You'. While the first type had no further text, the three others added different messages to try to select the most effective wording to deter dog owners from allowing their dogs to defecate in the area. Because in this meta-analysis we are interested only in the main effect of eyes, we collapsed the data for the four eyes posters into one experimental condition. Furthermore, we reinterpreted Keep Britain Tidy's 'displacement areas' as controls as they were functionally equivalent to Nettle et al.'s (2012) control areas.

2.6. Analytic methodology & statistical approach

2.6.1. Summary measures

For the first of our two analyses we identified studies comparing the number of participants behaving antisocially or not in an eyes and a control condition in order to enable the calculation of the log odds ratio as a measure of effect size. The odds ratio is used in one of the landmark

'watching eyes and antisocial behaviour' experiments by Nettle et al. (2012) and was selected as the measure of effect size as recommended by Welsh and Farrington (2009) for meta-analysis of location-based crime interventions, and by Lipsey and Wilson (2001) for meta-analysis of inherently dichotomous variables. Our statistical methods used in the fixed effects and random effects analyses mirror those of Northover et al. in their two meta-analyses of the watching eyes effect on generosity (Northover et al., 2017a) and on moral judgement (Northover et al., 2017b), thus enabling comparison.

Odds ratios are centred around 1 rather than zero (where 1 indicates no relationship, values 0–1 a negative relationship and values > 1 a positive relationship). The log odds ratio is distributed approximately normally with a mean of zero and a standard deviation of 1.83, thereby allowing us to represent a negative effect (a reduction in antisocial behaviour when eyes are present) with a negative value and a positive effect with a positive value, thus simplifying interpretation and calculation.

	Antisocial behaviour (ASB)_YES	*ASB_NO*
Eyes condition (Group 1)	(a)	(b)
Control (Group 2)	(c)	(d)

The log odds ratio effect size is calculated after Lipsey and Wilson (2001):

$$OR = \frac{(a * d)}{(b * c)} \quad (1)$$

$$OR_{log} = \log_e(OR) \quad (2)$$

The logs odds ratio standard error and inverse variance weight are calculated as:

$$SE_{LOR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad (3)$$

$$W_{LOR} = \frac{1}{SE^2_{LOR}} \quad (4)$$

We applied Cochran's *Q* test to estimate whether the individual effect sizes were representative of the population values:

$$Q = (\sum W_i ES_i^2) - \frac{(\sum W_i ES_i)^2}{\sum W_i} \quad (5)$$

Finding a statistically significant *Q* indicates a heterogeneous distribution greater than that expected from subject level sampling error alone. A significant *Q* value therefore suggests it may be appropriate to apply a random effects model to account for variance introduced by study differences such varying experimental designs or outcome measures. We first calculated the random effects variance (V_θ).

$$V_\theta = \frac{Q - (k - 1)}{\sum W_i - (\sum W_i^2 / \sum W_i)} \quad (6)$$

where *k* is the number of effect sizes, and $W_i$is the inverse variance weight for each effect size. The random variance was then added to the sampling error and the inverse variants weights recalculated as

$$\frac{1}{(V_\theta + SE^2_{LOR})} \quad (7)$$

The log odds ratio was then recalculated using the new weights.

Next, we calculated the lower and upper bounds of the 95% confidence interval of the log odds ratio:

$$95\%CI\ Lower = OR_{log} - (1.96 * SE_{LOR}) \quad (8)$$

$$95\%CI\ Upper = OR_{log} + (1.96 * SE_{LOR}) \quad (9)$$

We calculated the *z*-statistic and *p*-values for our log odds ratios:

$$Z = \frac{OR_{log}}{SE_{LOR}} \quad (10)$$

P-values were than calculated after the method recommended by Altman and Bland (2011):

$$P = \exp(-0.717 * Z - 0.416 * Z^2) \quad (11)$$

However, Nettle et al.'s (2012) field study examining the eyes effect on bicycle theft could not report how frequently people decided *not* to behave antisocially (i.e. how many were 'exposed' to the experimental and control conditions and chose *not* to steal bikes). Instead, they report only the frequency of antisocial behaviour across experimental and control locations. On account of this we selected the incidence ratio (sometimes called the rate ratio) as the appropriate measure of effect size. Such a rate ratio is calculated by controlling for any change in behaviour in both experimental and control locations before and after the intervention (CTSPedia, 2009).

Similarly, the experiment run by Keep Britain Tidy (2014) examined only before and after counts of dog faeces found at each of the 240 sites where signs were put up, comparing these the total number of dog faeces found over the same before and after period at the 240 adjacent control areas (or displacement areas, as they termed them).

	Before (ASB = Bicycles stolen or dog faeces in area)	After (ASB = Bicycles stolen or dog faeces in area)
Experimental Locations (sheds or areas where eyes signs were introduced for the second 3 weeks of the six week experiment)	(a)	(b)
Control Locations (sheds or areas where no signs were introduced during the six weeks of the experiment)	(c)	(d)

$$\text{Incidence Rate Ratio} = \frac{a/b}{c/d} \quad (13)$$

Analyses were performed in R 3.1 using the METAFOR package's random effects model function. R Script is provided in Additional Material to aid with replication of our analysis (R Development Core Team, 2016). Calculation of 95% confidence intervals, *Z*-scores and *P*-values was completed in Microsoft Excel. *P*-curve analysis was conducted after Simonsohn, Nelson, and Simmons (2014a, 2014b) using their online app provided at http://www.p-curve.com. PET & PEESE analysis was undertaken in R using code provided by Hilgard (2016).

3. Results

Table 1 summarises out metanalytic output, providing log odds ratios and associated statistics for the dichotomous data from all studies included in the analyses that follow.

We conducted a homogeneity analysis test to examine the assumption that all of our effect sizes were estimating the same population mean. The fixed effects model on the data without Nettle et al. (2012), Bateson et al. (2013) and Keep Britain Tidy (2014) returned a result of $OR_{log} = -.42$, (SE = 0.08, 95% CI [−0.59 to −0.26]); exponentiating the OR_{log} we get an OR = 0.65, suggesting a 35% reduction in antisocial behaviour by condition. However, since our *Q*-test for homogeneity of variance was significant ($Q(11) = 33.99$, $p < .001$), suggesting significant heterogeneity across our studies it proved necessary to fit a random effects model to the data.

The random effects model produced a meta-analytic effect size of $OR_{log} = -.45$, (SE = 0.18, 95% CI [−0.81, −0.09]) $p = .01$. Transforming our effect size to aid with interpretation we found an odds ratio (via $OR = e^{OR_{log}}$) of 0.64, suggesting a 36% reduction in the

Table 1
Studies in the antisocial behaviour (ASB) meta-analysis.

	Citation & experiment number	Outcome variable	Eye cues present			Control			OR_{log}	Var_{LOR}	SE_{LOR}	95% CI lower	95% CI upper	Z	P
			N	ASB yes	ASB no	N	ASB yes	ASB no							
1	Ernest-Jones et al. (2011)	Litter	273	56	233	289	98	175	−0.85	0.04	0.20	−1.23	−0.46	−4.34	**0.01**
2	Bateson et al. (2013)	Litter	305	80	225	297	83	214	−0.09	0.03	0.18	−0.45	0.27	−0.47	1.28
3	Baillon et al. (2013), Experiment 1	Destruction of others money	51	9	42	49	19	30	−1.08	0.22	0.47	−2.00	−0.16	−2.31	0.57
4	Zengerink (2013), Experiment 2	Litter	630	297	333	314	170	144	−0.28	0.02	0.14	−0.55	−0.01	−2.02	0.78
5	Bateson et al. (2015), Experiment 1	Litter	147	8	139	137	23	114	−1.25	0.18	0.43	−2.10	−0.41	−2.92	0.23
6	Bateson et al. (2015), Experiment 2	Litter	216	31	185	97	25	72	−0.73	0.09	0.30	−1.32	−0.14	−2.41	0.50
7	Cai et al. (2015), Experiment 1	Dishonesty	66	30	36	65	31	34	−0.09	0.12	0.35	−0.78	0.60	−0.26	1.17
8	Cai et al. (2015), Experiment 3	Dishonesty	64	18	46	66	22	44	−0.25	0.15	0.38	−0.99	0.50	−0.64	1.34
9	Hoffman et al. (2015)	Dishonesty	20	7	13	20	9	11	−0.42	0.42	0.65	−1.69	0.85	−0.64	1.34
10	Palomäki et al. (2015)	Dishonesty	105	13	92	86	9	77	0.19	0.21	0.46	−0.71	1.09	0.41	0.69
11	Huang et al. (2015), Experiment 2	Corruption	24	10	14	24	9	15	0.17	0.35	0.59	−0.98	1.33	0.30	0.78
12	Huang et al. (2015), Experiment 3	Corruption	30	11	19	31	24	7	−1.78	0.33	0.57	−2.90	−0.66	−3.11	0.17
13	Meleady et al. (2017), Experiment 1	Engine idling	112	83	21	104	82	39	0.63	0.10	0.31	0.02	1.24	2.02	**0.04**

		Experimental locations				N	Control locations		OR_{log}	Var_{LOR}	SE_{LOR}	95% CI lower	95% CI upper	Z	P
		Outcome variable	N	Frequency of ASB before	Frequency of ASB after		Frequency of ASB before	Frequency of ASB after							
14	Nettle et al. (2012)	Bicycle theft	N/A	31	51	N/A	39	15	−1.45	0.14	0.38	−2.20	−0.71	−3.83	**0.04**
15	Keep Britain Tidy (2014)	Dog fouling	N/A	861	434	N/A	2159	1208	0.10	0.00	0.07	−0.03	0.24	1.51	0.13

Bold text refers to those studies where $p < .01$.

risk of antisocial behaviour in the eyes condition.

Extending our analysis to include Nettle et al. (2012), Bateson et al. (2013) and Keep Britain Tidy's (2014) dog fouling intervention, we expected, and found, greater heterogeneity rendering the fixed effects model an inappropriate measure. However, there was little change in the effect size since Keep Britain Tidy's large-scale experiment found a small increase in antisocial behaviour in the eyes areas in comparison with the control which was sufficient to offset the large effect size in the opposite direction found in Nettle, Nott and Bateson's bicycle crime experiment.

Our fixed effects model including these studies returned a result of $OR_{log} = -.13$, (SE = 0.05, 95% CI [−0.23, −0.03]); exponentiating the OR_{log} we get an $OR = 0.88$, suggesting a c. 12% reduction in ASB in the eyes when compared with the control condition. The *Q*-test for homogeneity of variance was significant ($Q(14) = 68.54$, $p < .01$) again requiring us to fit a random effects model to the data.

The random effects model suggested that there was a 35% reduction in the risk of antisocial behaviour between the eyes and control condition ($OR_{log} = -.43$, ($SE = 0.17$, 95% *CI* [−0.76, −0.11]) $p < .01$; $OR = 0.65$).

Neither a forest or funnel plot show any evidence of over-dispersion (See Figs. 1 and 2 below). In addition, a *trim and fill* test for missing values (after Duval & Tweedie, 2000) finds no evidence to suggest either missing reports or publication bias.

The *Trim and Fill* method (Duval & Tweedie, 2000) relies on the assumption that selective reporting of studies, that is publication bias, is driven by effect size. This is vulnerable to criticism since the principle barrier to publication is widely acknowledged to be achieving statistical significance below $p = .05$. *P*-curve analysis explores the consistency and distribution of statistically significant results and outperforms other methods in identifying publication bias (Simonsohn, Nelson, and Simmons 2014a & 2014b).

To enable a *p*-curve analysis we calculated *z*-scores from our log odds ratios (see Table 1). Since our analysis relied on data transformed to provide dichotomous results the *p*-values in original published studies were not relevant. We ran this test inputting to the *p*-curve application the *z*-scores for the fifteen studies in Table 1. The application recalculates statistical significance based on this input. Fig. 3 shows the 8 studies of 15 that were statistically significant after this re-analysis. The application excludes non-significant results. The full output from the *p-curve analysis* can be obtained via the appendix Additional Materials.

P-Curve analysis allows us to draw inferences about evidential value. If the half *p*-curve test – which is much more robust to ambitious *p*-hacking than the simple full *p*-curve test (Simonsohn et al., 2014a, 2014b) - is right-skewed with $p < .05$, or both the half and full test are

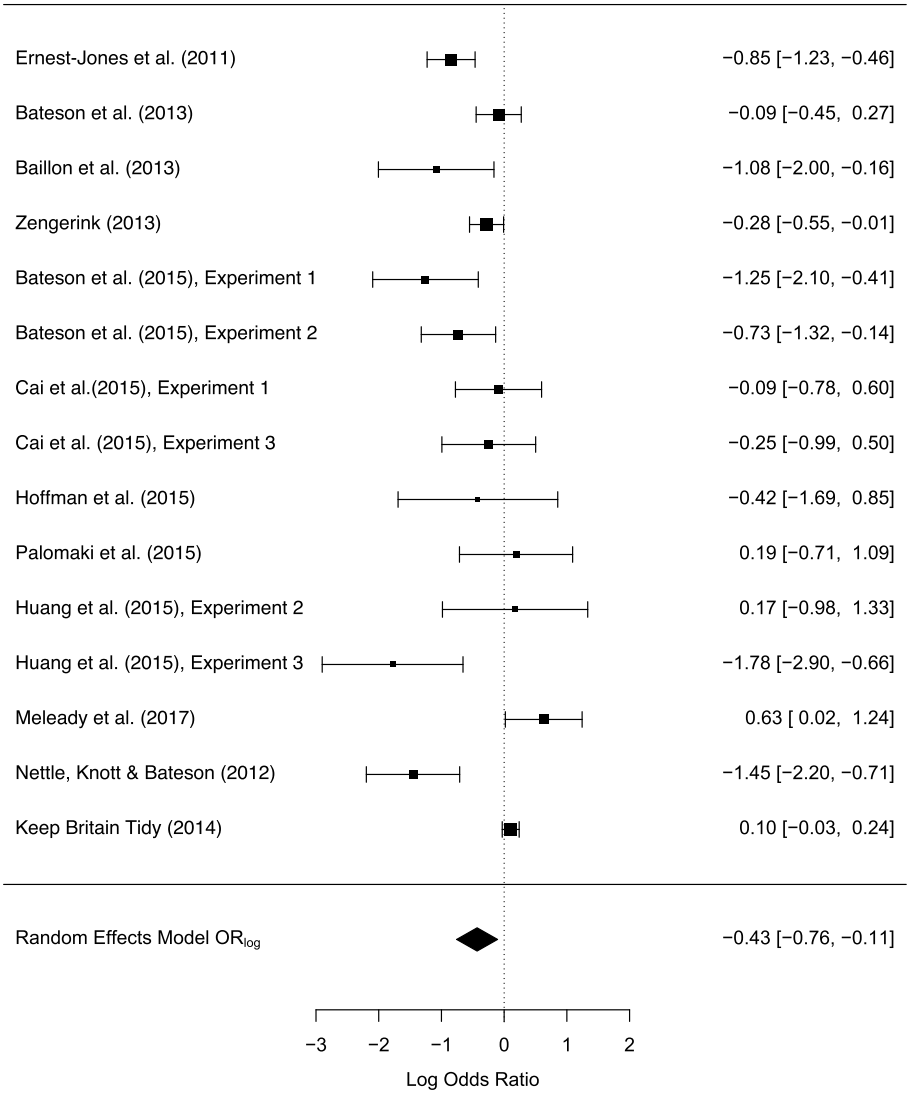

Fig. 1. Forest plot of experiments included in both the limited and extended meta-analyses. Plot is centered on zero (null effect) and displays effect size in ORlog with corresponding 95% CI.

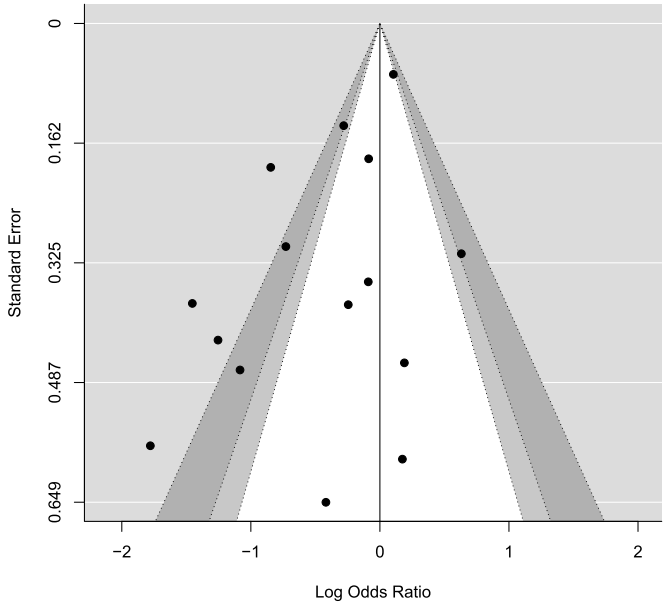

Fig. 2. Contour-enhanced funnel plot of experiments included in the limited and extended meta-analyses. Plot is centered on a log odds ratio of zero (null effect) and the dots plot the dispersion of effect sizes on the x-axis against the standard error on the y-axis. The shading indicates p-values. Those to the in the unshaded (white) region correspond to p-values > .1, the light-grey shaded region indicates p-values between 0.1 and 0.05, the dark grey region indicates p-values between 0.05 and 0.01. The region outside the funnel corresponds to p-values below 0.01.

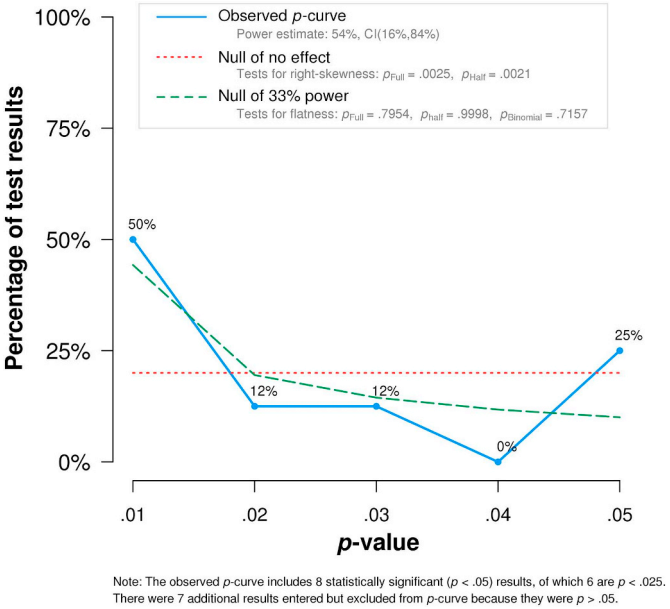

Fig. 3. P -curve results examining z -scores of our 15 studies for publication bias.

right-skewed with $p < .1$, then p -curve analysis indicates the presence of evidential value (Simonsohn, Simmons, & Nelson, 2015). Here both conditions are met indicating evidential value, as they were when entering only our statistically significant results.

Similarly, p-curve analysis indicates that evidential value is inadequate or absent if the 33% power test is $p < .05$ for the full p-curve, or both the half p-curve and binomial 33% power test are $p < .1$. Here neither condition is met so the p-curve does not indicate that evidential value is inadequate nor absent. The same also proved true when entering only our statistically significant studies. This P-curve analysis provides some reassurance that our findings are not the product of publication or other bias.

As a further confirmation check on our results we conducted a Precision Effects Test (PET; Stanley, 2005) which estimates effect size via a weighted least squares regression model and which provides '*…a valid basis for determining whether there is a genuine empirical effect beyond publication selection bias.*' (Stanley & Doucouliagos, 2013). Because PET can over-correct for publication bias, underestimating the true effect where one exists (i.e. it is most accurate when the effect size is zero) (Stanley, 2007) it is best treated as a pass or fail test rather than an estimator of effect size (Stanley & Doucouliagos, 2012, pp. 63, 78–79). We therefore added also a Precision Effect Estimate with Standard Errors (PEESE) which uses variance (standard error squared) as a predictor instead of the standard error (Stanley & Doucouliagos, 2007). PEESE is more accurate when there is a true effect (i.e. the true effect is non-zero) and should be selected over the PET output as the more accurate estimate of effect size.

Since random effects models skew PET & PEESE analysis in the direction of any publication bias, we chose a Fixed Effects Model following the advice of Stanley and Doucouliagos (2012, p. 83). Because both PET & PEESE require an N value as input we excluded Nettle et al. (2012) and Keep Britain Tidy (2014) from this analysis. We also exclude Bateson et al. (2013) in line with our previous analysis. The R-code and our full PET-PEESE analysis is available at the appendix Additional Materials.

Using the Standard Error of the Log Odds ratio and Z-Scores (Table 1) in a PET analysis with a Fixed Effects Model and R Code from Hilgard (2016), our results suggested that the presence of watching eyes significantly decreased the frequency of antisocial behaviour, $b_0 = -3.14$, $SE = 0.18$, 95% CI[−3.49,-2.78], $p < .0001$, $r = -0.99$.

Our fixed-effects PEESE analysis (after Hilgard, 2016) again found a significant effect, $b_0 = -2.47$, $SE = 0.12$, 95% CI[−2.70, −2.24], $p < .0001$, $r = -0.99$ where $r = 0.50$ is considered a large effect against conventional heuristics (Cohen, 1992) again suggesting that the presence of watching eyes significantly decreases antisocial behaviour. See Fig. 4 for a funnel plot of the Fixed Effects PET-PEESE model.

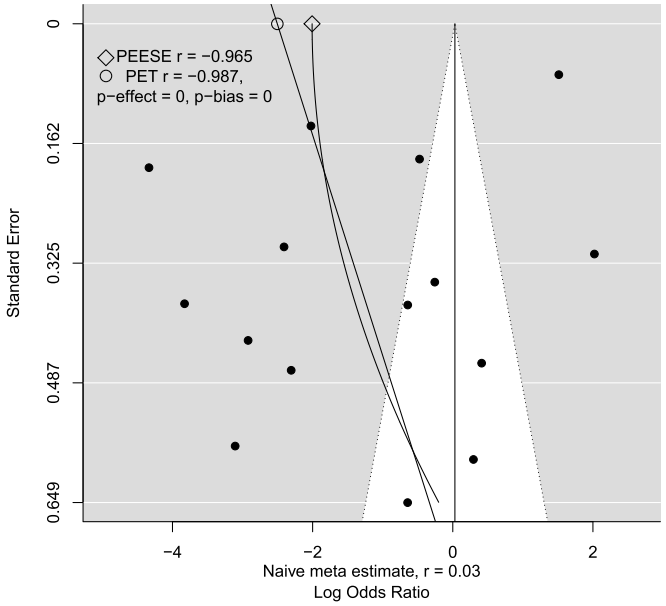

Fig. 4. Funnel plot the Fixed Effects PET-PEESE analysis of experiments 1–13. Plot is centered on the naive meta estimate and displays effect size on the x-axis in ORlog.

4. Discussion

In this section we first seek to interpret our findings, second reconcile our meta-analytic output with that of Northover et al. (2017a) and third, extend our eye cues and antisocial behaviour analysis descriptively.

Though our meta-analytic effect sizes are quite widely dispersed they consistently show that the presence of eyes cues correlate with a reduction in antisocial behaviour. Our first fixed effects analysis suggests there is a 35% reduction in the risk of antisocial behaviour in the presence of watching eyes. The Random Effects Model puts the effect at 36%; adding in Nettle et al. (2012) and Keep Britain Tidy (2014) leads the Fixed Effects Model output to suggest the presence of watching eyes may reduce antisocial behaviour by 12%; and the Random Effects Model including the additional studies puts the reduction at 35%. Given the difference in experimental design between Nettle et al. (2012) and Keep Britain Tidy (2014) and the other studies included in our analysis we report our overall effect size as a 35% reduction in the risk of antisocial behaviour in the presence of eye cues based on our first, more carefully selective analysis and using the more restrictive fixed effects model to report the more conservative result.

Our *p*-curve and *trim and fill* analyses find no evidence for publication bias. Our PET-PEESE analysis finds a significant effect in the same direction as our previous analysis – the presence of eye cues correlates with a reduction in antisocial behaviour.

However, there are reasons to be cautious of relying on PET-PEESE. Stanley (2017), the originator of the method, highlights the limitations of the analysis and related methods for small sample studies, areas with few studies, and areas with high heterogeneity. Simulations suggest PET-PEESE misses real effects around 80% of the time, and when heterogenous effects are present but real (i.e. there is not a single underlying effect that never differs regardless of context, and heterogeneity is not just noise and measurement error) or when publication bias is present, PET-PEESE under-estimates the true effect size 60–80% of the time (Gervais, 2016). The analysis also performs poorly when sample size is not reasonably consistent (Hilgard, 2015; Simonsohn, 2017).

The watching eyes literature is characterised by wide variations in sample size, includes many small samples, and finds heterogenous effects that may well be due to context and individual differences not just measurement error and noise. Thus PET-PEESE may not be the best method by which to examine our data. That said, all meta-analytic techniques have limitations and should be used only to understand broad trends in a given field and to guide the further development of the primary literature, not to establish the incontrovertible truth of a particular hypothesis (for a discussion of the limitations of meta-analytic methods in psychology see: Carter, Schönbrodt, Gervais, & Hilgard, 2017). Best practice to try to triangulate on results using multiple meta-analytic methods as we have sought to do (Carter et al., 2017).

We were surprised to find no evidence of publication bias in our statistical analyses. However, perhaps we should not have been. Northover et al. (2017a) had conducted a search for missing papers concurrent with the period that we completed our meta-analysis finding just 2 unpublished studies. Our own extensive search for unpublished studies found only four additional papers. It may be that there are many more watching eyes studies in the file drawer, but neither extensive web-searching, specific calls for unpublished papers nor statistical analysis has yet been able to find any evidence for this.

Northover et al.'s (2017a) analysis is the benchmark for meta-analysis in the watching eyes field. It finds null effects where our effects are significant. Consequently, we shall orientate our discussion around it. However, it is important to note that, valuable as Northover et al.'s analysis is there are two weaknesses within it. The first is their decision to include two studies from Tane and Takezawa (2011) which were conducted in darkness. These two experiments were implicitly testing the boundary conditions of the watching eyes effect rather than its existence– if watching eyes cue people to feel watched presumably darkness makes people feel less conspicuous. There is evidence to support this common-sense assumption. Zhong, Bohns, and Gino (2010) report that darkness can induce illusory anonymity leading to more cheating in economic games. Therefore, the inclusion of Tane & Takewaza's studies skews the analysis. The second is the inclusion of Mifune, Hashimoto, & Yamagishi (2010) experiment in Northover's analysis. Mifune et al.'s study was explicitly conducted to show the boundary conditions where eye cues would *not* work. The outcome of the experiment was as predicted by Mifune and colleagues: eye cues increased generosity to in-groups but not to out-groups. However, this nuance is lost by simply including their study as 'null results'.

Finally, it should also be noted that no-one, including Northover and colleagues, suggests that their meta-analysis provides definitive evidence that eye cues have no effect on generosity.

Our meta-analysis finds a large effect size, indicating a reduction in antisocial behaviour in the presence of eye cues. In marked contrast, in both of the meta-analyses by (Northover et al., 2017a, 2017b) a vanishingly small effect of eye cues on generosity was found. This is true both when comparing the standardised mean difference and when examining the probability of donating something rather than nothing.

How, then, do we reconcile our findings that eye cues have a large effect in reducing antisocial behaviour with those of Northover et al. that show no effect of eye cues on generosity?

In seeking to reconcile our findings with Northover et al.'s, we first consider recent experimental evidence that upholds the intuitive idea that eye cues cause people to feel watched. If, as the evolutionary legacy hypothesis suggests, eye cues make us feel watched, and thereby make us more aware of our reputation, then the difference between Northover's meta-analytic findings and our own may be best explained by the variable effect of generosity on reputation and the more consistent effect of antisocial behaviour on reputation, as outlined earlier.

Pfattheicher and Keller (2015) have provided the best available evidence that eye cues make us feel watched. Across two experiments their results also suggest that those most sensitive to being watched are the most responsive to eye cues. In the first experiment, participants read a scenario in which they were asked to imagine wearing an embarrassing T-shirt while walking in a busy corridor with 30 people present or on a train with 50 people in the compartment. The participants were then asked to report how many people they thought would notice them. Results showed that those entering their estimation on a page with eye cues in the header reported significantly more people would notice them than those in a 'no eyes' control. Eye cues seemed to make people feel more watched.

In the second experiment the results suggested that participants' sensitivity to being watched predicted their responsiveness to eye cues. Measuring participants' sensitivity to being watched on the public self-awareness scale (Fenigstein, Scheier, & Buss, 1975; completed online via Amazon Mechanical Turk) their results indicated that those with very high or chronic self-awareness scores were more likely to donate some of their participant payment to a charity in the eye cues condition than in the control, whereas those with low public self-awareness showed no difference between conditions.

To summarize, then, Keller and Pfattheicher's findings suggest that it is by making us feel watched that eye cues affect behaviour. First, because eye cues make people feel a higher state of public self-awareness, that is more aware of how they are presenting themselves to others. Second, because people's differing trait levels of public self-awareness predict their responsiveness to eye cues, with those more sensitive to how they present themselves to others more strongly affected by eye cues than those who are less aware of the impression they create on others. If this is true, then we can expect eye cues to have similar effects on our behaviour to the presence of other people.

There is a large body of evidence showing that the 'mere presence' of others can affect behaviour (Markus, 1978; Zajonc, Heingart, & Herman, 1969), that gaining cooperative benefits depends often on

sustaining a good reputation (Barclay, 2013; Izuma, 2012; Van Vugt, Roberts, & Hardy, 2007), and that reducing anonymity can make people more cooperative as it emphasizes reputational risk (Yoeli, Hoffman, Rand, & Nowak, 2013). Thus if eye cues make people feel watched and thereby affect decisions and change behaviour, it is likely that they do so by prompting people to give greater consideration to their reputation.

Consequently, we suggest that Northover et al. (2017a) may have found no effect of eye cues on generosity because whether one should give more or less when watched is not clear cut. This finds some support from Nettle et al.'s (2013) dictator game: when compared with those in a 'no eye cues' control, participants making their decisions in the presence of eye cues were more likely to give something than nothing, but less likely to give all or a lot. The eye cue effect seemed to be to reduce variation in donations, not to promote prosociality per se (Nettle et al., 2013), perhaps suggestion that the very generous were cued to avoid the damaging reputational effects of being seen to be over-generous.

Similar effects have been found in a taking game (an inverse dictator game). Participants who took a lot in the control condition took less when eye cues were present, whereas participants who took very little in a control condition took more in the presence of eye cues (Chowdhury et al., 2014). Thus, although Northover et al. (2017a) showed there is no consistent shift in the proportion who gave when eye cues were present, it does seem that eye cues can both increase and reduce generosity dependent on the situation, suggesting that perhaps we should not expect consistent eye cue effects on generosity.

Such a deduction accords with the logic that being too conspicuously generous is not always reputation-enhancing (Klein et al., 2015; Klein & Epley, 2014) and, we contend, perhaps explains why the effects across the different experimental designs analysed in Northover et al. (2017a) reflect conflicting patterns and a neutral overall effect size.

In conclusion our meta-analysis suggests watching eyes produce a more consistent and robust effect on antisocial behaviour – the result, we suggest, of there being few situations, regardless of individual differences, in which being observed engaging in antisocial behaviour might be seen to be advantageous.

4.1. Summary of evidence

Our meta-analysis has shown that the presence of watching eyes correlates with a 35% reduction in the risk of antisocial behaviour. We note that this contrasts with a recent robust meta-analysis on the effect of eye cues on generosity, which showed an effect size close to zero (Northover et al., 2017a). We suggest this may be a product of individual differences and the complexity of the concept of generosity in the context of observation.

4.2. Limitations

Our meta-analytic model incorporates varied experimental designs which may weaken the overall findings. Stronger evidence for the eye cue effect on antisocial behaviour would have been forthcoming had there been more attempts to replicate existing eye-cue experiments and less design diversity. Our meta-analysis also rests on narrow foundations and it would be premature to place too much weight on our findings without further research.

Additionally, we might be seeing a typical spate of strong effects in early research in the antisocial behaviour field (de Winter & Happee, 2013; Ioannidis & Trikalinos, 2005; Monsarrat & Vergnes, 2018; Moonesinghe, Khoury, & Janssens, 2007) and thus an exaggerated effect size.

Finally, in looking only at the main effect of eye cues we have coded data as having shown no effect in Keep Britain Tidy's, 2014 experiment when the authors did indeed report an effect in more nuanced analyses accounting for mediating and moderating variables. As more studies are published examining the watching eyes effect on antisocial behaviour, more nuanced meta-analysis will be possible providing greater insight than we are able to at present.

4.3. Further research

Further research to demonstrate that eye cues make people feel more watched is needed in order to provide a more robust theoretical underpinning for the more elaborate theories that rest on this assumption, and the concomitant idea that eye cues prompt unconscious reputation management or heighten fear of punishment. Pfattheicher and Keller's (2015) finding that public self-awareness may moderate any effect of eye cues is a useful step in this direction. A replication of this finding and an examination as to whether eye cues prompt greater self-regulation would be helpful.

An additional point of note here is that one possible limitation of laboratory-based studies is the difficulty in avoiding participants feeling watched even when they aren't under surveillance. It is difficult to see how participants cannot be aware that their actions, even if indirectly and/or anonymously, will not be the subject of scrutiny by the researcher. Consequently, exploring the effect of eye cues in more ecologically valid and/or through carefully designed field studies might yield greater insight into eye cues' effect on behaviour.

4.4. Conclusions

Our meta-analysis of the eye cue effect on antisocial behaviour suggests eye cues may reduce the risk of antisocial behaviour by 35%. What matters here is not whether an effect size is large or small according to statistical guidelines but rather whether the effect size is meaningful (Aberson, 2011; Coe, 2002). In criminological analysis, for example, even smaller effect sizes may suggest significant reduction in crime (Weisburd, Telep, Hinkle, & Eck, 2010).

Our effect size, a 35% risk reduction for antisocial behaviour when eye cues are present, represents a meaningful effect in real-world terms. In its last *Costs of Crime* report the UK Home Office estimated that crime costs the UK economy between £35bn and £60bn a year (Brand & Price, 2000). Taking the lower end of the estimate, a 1% reduction in crime might be said to equate to a saving of £350,000,000. While we do not claim the watching eyes effect is anything like a panacea, the evidence so far suggests it could be a highly cost-effective criminal deterrent in some circumstances.

Funding

This research was funded by the Royal Air Force (Grant Number DDS/44) and United States Air Force (Grant Number: FA9550-16-1-0090) as part of a wider project examining the effect of surveillance on behavior

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2019.01.006.

References

Aberson, C. L. (2011). *Applied power analysis for the behavioural sciences.* Routledge.
Altman, D. G., & Bland, J. M. (2011). How to obtain the P value from a confidence interval. *BMJ,* 343. https://doi.org/10.1136/bmj.d2304.
Baillon, A., Selim, A. I., & van Dolder, D. (2013). On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behaviour, 34*(2), 146–154.
Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behaviour, 34*(3), 164–175.
Basildon, Canvey, & Southend Echo (2013). *Poster trial on c2c line to target bike thieves.* Basildon, Canvey, Southend Echo. Retrieved from http://www.echo-news.co.uk/news/local_news/10373770.Poster_trial_on_c2c_line_to_target_bike_thieves/?ref=nt.

Bateson, M., Callow, L., Holmes, J. R., Roche, M. L. R., & Nettle, D. (2013). Do images of 'watching eyes' induce behaviour that is more pro-social or more normative? A field experiment on littering. *PLoS One, 8*(12), e82055.
Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412–414.
Bateson, M., Robinson, R., Abayomi-Cole, T., Greenlees, J., O'Connor, A., & Nettle, D. (2015). Watching eyes on potential litter can reduce littering: Evidence from two field experiments. *PeerJ, 3*, e1443.
Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology, 5*(4), 323–370.
BBC News (2015). Tax crackdown by HMRC unit nets 60% more money, report says. Retrieved from BBC News website http://www.bbc.co.uk/news/business-30973191.
Beyfus, T. A., Dawson, N. L., Danner, C. H., Rawal, B., Gruber, P. E., & Petrou, S. P. (2016). The use of passive visual stimuli to enhance compliance with handwashing in a perioperative setting. *American Journal of Infection Control, 44*(5), 496–499.
Brand, S., & Price, R. (2000). *The economic and social costs of crime.* 217. Retrieved from http://webarchive.nationalarchives.gov.uk/20110218135832/rds.homeoffice.gov.uk/rds/pdfs/hors217.pdf.
Brudermann, T., Bartel, G., Fenzl, T., & Seebauer, S. (2015). Eyes on social norms: A field study on an honor system for newspaper sale. *Theory and Decision, 79*(2), 285–306.
Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behaviour & Organization, 50*(1), 133–144.
Burnham, T. C., & Hare, B. (2007). Engineering human cooperation: Does involuntary neural activation increase public goods contributions? *Human Nature, 18*(2), 88–108.
Bush, S. S., Erlich, A., Prather, L., & Zeira, Y. (2016). The effects of authoritarian iconography: An experimental test. *Comparative Political Studies, 49*(13), 1704–1738.
Cai, W., Huang, X., Wu, S., & Kou, Y. (2015). Dishonest behaviour is not affected by an image of watching eyes. *Evolution and Human Behaviour, 36*(2), 110–116.
Carbon, C. C., & Hesslinger, V. M. (2011). Bateson et al.'s (2006) cues-of-being-watched paradigm revisited. *Swiss Journal of Psychology, 70*(4), 203–210.
Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2017). *Correcting for bias in psychology: A comparison of meta-analytic methods.* https://doi.org/10.31234/osf.io/9h3nu.
Chowdhury, S., Jeon, J. Y., & Saha, B. (2014). Eye-image in experiments: Social cue or experimenter demand effect? Retrieved from http://econpapers.repec.org/RePEc:uea:aepppr:2012_67.
Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *Paper presented at the annual conference of the British Educational Research Association*UK: University of Exeterhttp://www.leeds.ac.uk/educol/documents/00002182.htm.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.
Conty, L., George, N., & Hietanen, J. K. (2016). Watching eyes effects: When others meet the self. *Consciousness and Cognition, 45*, 184–197. https://doi.org/10.1016/j.concog.2016.08.016.
CTSPedia (2009, 4 May). Measure of association - Rate ratio. Retrieved from https://www.ctspedia.org/do/view/CTSpedia/RateRatio.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel- plot– based method of testing and adjusting for publication bias in meta- analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.
Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behaviour: A field experiment. *Evolution and Human Behaviour, 32*(3), 172–178.
Extra Services, K. B. T (2015). Litterers: We're watching you. [Press release]. Retrieved from http://www.extraservices.co.uk/article/extra-msa-keep-britain-tidy/.
Fehr, E., & Gachter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140.
Fehr, E., & Schneider, F. (2010). Eyes are on us, but nobody cares: Are eye cues relevant for strong reciprocity? *Proceedings of the Royal Society B, 277*(1686), 1315–1323.
Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness - assessment and theory. *Journal of Consulting and Clinical Psychology, 43*(4), 522–527.
Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11*(2), 77–83.
Flanagan, P. (2013, 2 June). *Crime... The answer's staring you in the face: The striking new police poster that has slashed thefts by up to 40 per cent.* Daily Mail. Retrieved from http://www.dailymail.co.uk/news/article-2334577/Crime–answers-staring-face-The-striking-new-police-poster-slashed-thefts-40-cent.html.
Fujii, T., Takagishi, H., Koizumi, M., & Okada, H. (2015). The effect of direct and indirect monitoring on generosity among pre-schoolers. *Scientific Reports,* 5.
Gaiani, A., Rose, E., & Roberts, G. (2014). How and why do watching eyes enhance prosocial behaviour. http://research.ncl.ac.uk/expeditionresearchscholarships/postergalleries/2014posters/Annika.Gaiani.pdf.
Gervais, W. M. (2016). *Putting PET-PEESE to the test.* Will Gervais Blog. Retrieved from http://willgervais.com/blog/2015/6/25/putting-pet-peese-to-the-test-1.
Golja, A. (2013). How cues of being watched affect moral judgements. *Paper presented at the MEi: CogSci conference 2013, Budapest.*
Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*(5770), 108–111.
H.M. Government (UK) (2017). Litter strategy for England. Retrieved from UK: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630999/litter-strategy-for-england-2017-v2.pdf.
Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26*(3), 245–256.
Hilgard, J. (2015, 29 June). Putting PET-PEESE to the test, part 1A. Crystal prison zone. Retrieved from http://crystalprisonzone.blogspot.com/2015/06/putting-pet-peese-to-test-part-1a.html.
Hilgard, J. (2016). *Functions for R to make PET-PEESE modeling easier.* GitHub. Retrieved from https://github.com/Joe-Hilgard/PETPEESE.
Hoffman, G., Forlizzi, J., Ayal, S., Steinfeld, A., Antanitis, J., Hochman, G., ... Finkenaur, J. (2015). Robot presence and human honesty: Experimental evidence. *Paper presented at the proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction.*
Huang, Z.-w., Liu, L., Zheng, W.-w., Tan, X.-y., & Zhao, X. (2015). Walking the straight and narrow: The moderating effect of evaluation apprehension on the relationship between collectivism and corruption. *PLoS One,* (3), 10.
Ioannidis, J. P. A., & Trikalinos, T. A. (2005). Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *Journal of Clinical Epidemiology, 58*(6), 543–549. https://doi.org/10.1016/j.jclinepi.2004.10.019.
Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience Research, 72*(4), 283–288.
Jackson, D. (2015). *Eye donated to charity: The effect of visual monitoring on charitable donations. Poster.* Hanover College.
Johnson, D. (2016). *God is watching you : How the fear of God makes us human.* New York: Oxford University Press.
Jolij, J., & de Haan, T. (2014). Being watched doesn't make you nicer: No effect of visible and invisible eye primes on prosocial behaviour in a masked priming study. Working Paper, Department of Experimental Psychology. University of Groningen. Retrieved from https://pure.rug.nl/ws/files/16082044/JolijDeHaan2014.docx
Keep Britain Tidy (2014). Keeping an eye on it. Retrieved from http://innovate.keepbritaintidy.org/keeping-an-eye-on-it/2463/2/14/1944/152.
Klein, N., & Epley, N. (2014). The topography of generosity: Asymmetric evaluations of prosocial actions. *Journal of Experimental Psychology. General, 143*(6), 2366.
Klein, N., Grossmann, I., Uskul, A. K., Kraus, A. A., & Epley, N. (2015). It pays to be nice, but not really nice: Asymmetric reputations from prosociality across 7 countries. *Judgment and Decision making, 10*(4), 355.
Knapton, S. (2016). *Trees given faces to prevent littering in Forest of Dean.* The Telegraph. Retrieved from http://www.telegraph.co.uk/science/2016/05/11/trees-given-faces-to-prevent-littering-in-forest-of-dean/.
Kuliga, S., Tanja-Dijkstra, K., & Verhoeven, F. (2011). Wa (tc) sh out! The effects of cues of being watched on implicit activation of norms and hand disinfection behaviour. *Paper presented at the proceedings of the 9th Biennial Conference on environmental psychology, Eindhoven.*
Lamba, S., & Mace, R. (2010). People recognise when they are really anonymous in an economic game. *Evolution and Human Behaviour, 31*(4), 271–278.
Li, J., Zhan, Y. L., Fan, W., Liu, L., Li, M., Sun, Y., & Zhong, Y. P. (2018). Sociality mental modes modulate the processing of advice-giving: An event-related potentials study. *Frontiers in Psychology, 9*, 11. https://doi.org/10.3389/fpsyg.2018.00042.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, Calif. London: Sage Publications.
Markus, H. (1978). Effect of mere presence on social facilitation - unobtrusive test. *Journal of Experimental Social Psychology, 14*(4), 389–397.
Matland, R. E., & Murray, G. R. (2016). I only have eyes for you: Does implicit social pressure increase voter turnout? *Political Psychology, 37*(4), 533–550.
Matsugasaki, K., Tsukamoto, W., & Ohtsubo, Y. (2015). Two failed replications of the watching eyes effect. *Letters on Evolutionary Behavioural Science, 6*(2), 17–20.
Meleady, R., Abrams, D., Van de Vyver, J., Hopthrow, T., Mahmood, L., Player, A., ... Leite, A. C. (2017). Surveillance or self-surveillance? Behavioral cues can increase the rate of drivers' pro-environmental behavior at a long wait stop. *Environment and Behavior, 49*(10), 1156–1172.
McNees, M. P., Egli, D. S., Marshall, R. S., Schnelle, J. F., & Risley, T. R. (1976). Shoplifting Preventions: Providing InformationThrough Signs. *Journal of Applied Behavior Analysis, 9*(4), 399–405. https://doi.org/10.1901/jaba.1976.9-399.
Mifune, N., Hashimoto, H., & Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior, 31*(2), 109–117. https://doi.org/10.1016/j.evolhumbehav.2009.09.004.
Monsarrat, P., & Vergnes, J.-N. (2018). The intriguing evolution of effect sizes in biomedical research over time: Smaller but more often statistically significant. *GigaScience, 7*(1), gix121. https://doi.org/10.1093/gigascience/gix121.
Moonesinghe, R., Khoury, M. J., & Janssens, A. C. J. W. (2007). Most published research findings are false—But a little replication goes a long way. *PLoS Medicine, 4*(2), e28. https://doi.org/10.1371/journal.pmed.0040028.
Nelson, F. (2013). *Big Brother (and HMRC) is watching you.* The Spectator: Coffee House Blog. Retrieved from http://blogs.spectator.co.uk/2013/01/big-brother-and-hmrc-is-watching-you/.
Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the dictator game: it's not how much you give, it's being seen to give something. *Evolution and Human Behaviour, 34*, 35–40. https://doi.org/10.1016/j.evolhumbehav.2012.08.004.
Nettle, D., Nott, K., & Bateson, M. (2012). Cycle thieves, we are watching you': Impact of a simple signage intervention against bicycle theft (bicycle theft). *PLoS One, 7*(12), e51738.
Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017a). Artificial surveillance cues do not increase generosity: Two meta-analyses. *Evolution and Human Behavior, 38*(1), 144–153. https://doi.org/10.1016/j.evolhumbehav.2016.07.001.
Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017b). Effect of artificial surveillance cues on reported moral judgment: Experimental failures to replicate and two meta-analyses. *Evolution and Human Behavior, 38*(5), 561–571. https://doi.org/10.1016/j.evolhumbehav.2016.12.003.
Oda, R., Kato, Y., & Hiraishi, K. (2015). The watching-eye effect on prosocial lying. *Evolutionary Psychology,* (3), 13.
Palomäki, J., Modic, D., & Yan, J. (2015, August). Deterring deception in online insurance

claims. *Paper presented at the Decepticon: International conference on deceptive behaviour, Cambridge*.
Pfattheicher, S., & Keller, J. (2015). The watching eyes phenomenon: The role of a sense of being seen and public self-awareness. *European Journal of Social Psychology, 45*(5), 560–566. https://doi.org/10.1002/ejsp.2122.
Powell, K. L., Roberts, G., & Nettle, D. (2012). Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology, 118*(11), 1096–1101.
R Development Core Team (2016). *R: A language and environment for statistical computing (Version 3.3.1)*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.
Raihani, N. J., & Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proceedings of the Royal Society B, 279*(1742), 3556–3564.
Rose, E., Gaiani, A., & Roberts, G. (2014). Watching eyes and cooperation: Does attractiveness of gender of watching eyes increase donations? http://research.ncl.ac.uk/expeditionresearchscholarships/postergalleries/2014posters/Eve.Rose.pdf.
Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality & Social Psychology Review, 5*(4), 296–320.
Simonsohn, U. (2017). *PET-PEESE is not like homeopathy. Data colada: Thinking about evidence and vice versa. Vol. 59*. Retrieved from http://datacolada.org/59.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*(6), 666–681. https://doi.org/10.1177/1745691614553988.
Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). *P-curve: A key to the file-drawer*. US: American Psychological Association143. https://doi.org/10.1037/a0033242.
Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-curves: Making P-curve analysis more robust to errors, fraud, and ambitious P-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology-General, 144*(6), 1146–1152. https://doi.org/10.1037/xge0000104.
Sparks, A. (2010). *Subtle cues and economic games*. Hamilton, Ontario, Canada: McMaster University. Retrieved from http://hdl.handle.net/11375/9175.
Sparks, A., & Barclay, P. (2013). Eye images increase generosity, but not for long: The limited effect of a false cue. *Evolution and Human Behaviour, 34*(5), 317–322.
Sparks, A., & Barclay, P. (2015). No effect on condemnation of short or long exposure to eye images. *Letters on Evolutionary Behavioural Science, 6*(2), 13–16.
Stanley, T. (2005). Beyond publication bias. *Journal of Economic Surveys, 19*(3), 309–345. https://doi.org/10.1111/j.0950-0804.2005.00250.x.
Stanley, T. (2007). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics, 70*(1), 103–127. https://doi.org/10.1111/j.1468-0084.2007.00487.x.
Stanley, T. (2017). Limitations of PET-PEESE and other meta-analysis methods. *Social Psychological and Personality Science, 8*(5), 581–591. https://doi.org/10.1177/1948550617693062.
Stanley, T., & Doucouliagos, C. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *Economics Series, 11*, 2007.
Stanley, T., & Doucouliagos, C. (2012). *Meta-regression analysis in economics and business*. Routledge.
Stanley, T., & Doucouliagos, C. (2013). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*(1), 60–78. https://doi.org/10.1002/jrsm.1095.
Stella, S. A., Stace, R., Miller, A., Keniston, A., Burden, M., Price, C. S., & Albert, R. K. (2013). Eye Images Improve Hand Hygiene Compliance in an Emergency Department. *Paper presented at the society of general internal medicine 36th annual meeting, Denver, Colorado, USA*http://www.sgim.org/File.Library/SGIM/Meetings/Annual Meeting/Meetign Content/AM13 presentations/Abstract-Session-F3-Clinical-Epi-Healthcare-Effectiveness.pdf.
Tane, K., & Takezawa, M. (2011). Perception of human face does not induce cooperation in darkness. *Letters on Evolutionary Behavioural Science, 2*(2), 24–27.
Traver, C., & Cordell, T. (2014). Look me in the eyes: Examining the effects of priming with eyes on deception. Retrieved from http://christophertraver.com/assets/eye-primes.pdf.
Van Vugt, M., Roberts, G., & Hardy, C. (2007). Competitive altruism: Development of reputation-based cooperation in groups. In R. Dunbar, & L. Barrett (Eds.). *Handbook of evolutionary psychology* (pp. 531–540). Oxford Handbooks Online: Oxford University Press.
Vogt, S., Efferson, C., Berger, J., & Fehr, E. (2015). Eye spots do not increase altruism in children. *Evolution and Human Behaviour, 36*(3), 224–231.
Waktare, C., & Roberts, G. (2014). Can 'watching eyes' influence punishment behaviour? *Paper presented at the undergraduate poster presentation, Newcastle University*http://research.ncl.ac.uk/expeditionresearchscholarships/postergalleries/2014posters/Catriona.Waktare.pdf.
Weisburd, D., Telep, C. W., Hinkle, J. C., & Eck, J. E. (2010). Is problem-oriented policing effective in reducing crime and disorder? *Criminology & Public Policy, 9*(1), 139–172.
Welsh, B., & Farrington, D. (2009). Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Quarterly, 26*(4), 716–745.
West Midlands Police (2006). W Mids gain 'momentum' in fight against crime [press release]. Retrieved from http://www.policeoracle.com/news/W-Mids-Gain-Momentum-In-Fight-Against-Crime_11284.html.
White, C. (2015). *Unexpected events as a cue to social surveillance*. London, Ontario, Canada: (BSc), King's University College, Western University. Online. Retrieved from http://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1018&context=psychK_uht.
Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251–1263.
de Winter, J., & Happee, R. (2013). Why selective publication of statistically significant results can be effective. *PLoS One, 8*(6), e66463. https://doi.org/10.1371/journal.pone.0066463.
Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences, 110*(Supplement 2), 10424–10429.
Zajonc, R. B., Heingart, A., & Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. *Journal of Personality and Social Psychology, 13*(2), 83–92.
Zengerink, L. (2013). *The image of eyes: psychological mechanisms and effects on littering in a real-world setting*. (MSc), University of Twente. Online. Retrieved from http://purl.utwente.nl/essays/64617.
Zhong, C.-B., Bohns, V. K., & Gino, F. (2010). Good lamps are the best police darkness increases dishonesty and self-interested behavior. *Psychological Science, 21*(3), 311–314. https://doi.org/10.1177/0956797609360754.