Neuroscience and Biobehavioral Reviews 104 (2019) 209–221

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

Review article

P-curving the fusiform face area: Meta-analyses support the expertise hypothesis

Edwin J. Burns[a,*], Taylor Arnold[b], Cindy M. Bukach[a]

[a] Department of Psychology, University of Richmond, Richmond, USA
[b] Department of Mathematics, University of Richmond, Richmond, USA

ARTICLE INFO

Keywords:
P-curve
Meta-analysis
Expertise
Domain general
Domain specific
FFA

ABSTRACT

Psychologists have debated whether the right fusiform face area's (FFA) responses are domain specific to faces, or domain general for certain object categories that we have visual expertise with. This latter domain general expertise account has been criticised for basing its assumptions upon studies that suffer from small participant numbers, small effects, and statistically significant *p*-values that are close to .05. An additional criticism is that these findings are difficult to replicate. A modern reader familiar with the replication crisis may therefore question whether the FFA's expertise effect is real. The *p*-curve is a relatively new form of meta-analysis that enables researchers to identify whether there is evidential value for any given effect in the literature. We put the literature to the test by running *p*-curve analyses on all published expertise studies. Contrary to aforementioned criticisms, our meta-analyses confirm the right FFA's expertise effect is based upon evidential value. We therefore review the broader literature to address additional criticisms of the expertise account and propose ways to improve replicability.

1. Introduction to the modular and expertise hypotheses of the fusiform face area

Psychologists have been debating for decades whether the brain processes faces in a way that is unique from other objects. The modular hypothesis posits that faces are special stimuli as they are processed by domain specific neural networks, i.e., brain regions that only process faces (Kanwisher, 2000, 2017; McKone et al., 2007; Yovel and Kanwisher, 2004). The brain region typically cited as the most domain-specific for faces is the right fusiform face area (FFA; Kanwisher, 2017). The FFA was first explicitly identified in 1997 through an fMRI localizer task as a small area of the right fusiform gyrus that appeared most responsive when participants viewed faces in contrast to other forms of stimuli (Kanwisher et al., 1997). There are numerous studies, too many to mention here, that have replicated such enhanced selectivity in the FFA for faces versus scrambled faces, houses, animals, cars, hands, scenes and artificial objects (for a review, see Yovel and Kanwisher, 2004). Further support for the modular account of the right FFA comes from prosopagnosia cases that exhibit differentially greater deficits in face, versus non-face, recognition after lesions to their fusiform gyrus (e.g., Susilo et al., 2015; Susilo et al., 2013). When considered together, these data points provide a strong case that the right FFA is important for face perception.

In contrast to the modular account, the expertise hypothesis views the right FFA as a process-specific, rather than domain-specific, area and posits that this region becomes face selective because of our experience individuating faces (Gauthier et al., 1999; McGugin et al., 2012). According to this account, FFA responsiveness is determined by a combination of factors, including object properties (i.e., object categories that are visually homogenous), task demands (individuation), and experience (Bukach et al., 2006; Gauthier and Bukach, 2007). Support for the expertise perspective has come from a vast body of fMRI data showing that the right FFA activity in experts is modulated for many visual categories, including cars (Gauthier et al., 2000; McGugin et al., 2012), birds (Gauthier et al., 2000; Xu, 2005), radiographs (Bilalić et al., 2014), chessboards (Bilalić et al., 2011) and Greebles: artificial 'aliens' that are designed to require similar perceptual individuation as faces (Gauthier et al., 1999). Moreover, right FFA activity is directly correlated with participants' levels of behavioural expertise (i.e., how well they can identify said objects; Gauthier et al., 2000; McGugin et al., 2012; Ross et al., 2018; Wong and Gauthier, 2010b). There is, therefore, sufficient neuroimaging evidence in the literature to indicate that the expertise hypothesis is at least partly correct in reporting that the right FFA can develop experience related haemodynamic effects.

Despite this evidence, the expertise perspective has had many

* Corresponding author at: Department of Psychology, 209-B Richmond Hall, University of Richmond, 23173, USA.
E-mail address: edwinjamesburns@gmail.com (E.J. Burns).

https://doi.org/10.1016/j.neubiorev.2019.07.003
Received 17 October 2018; Received in revised form 30 June 2019; Accepted 1 July 2019
Available online 02 July 2019
0149-7634/ © 2019 Elsevier Ltd. All rights reserved.

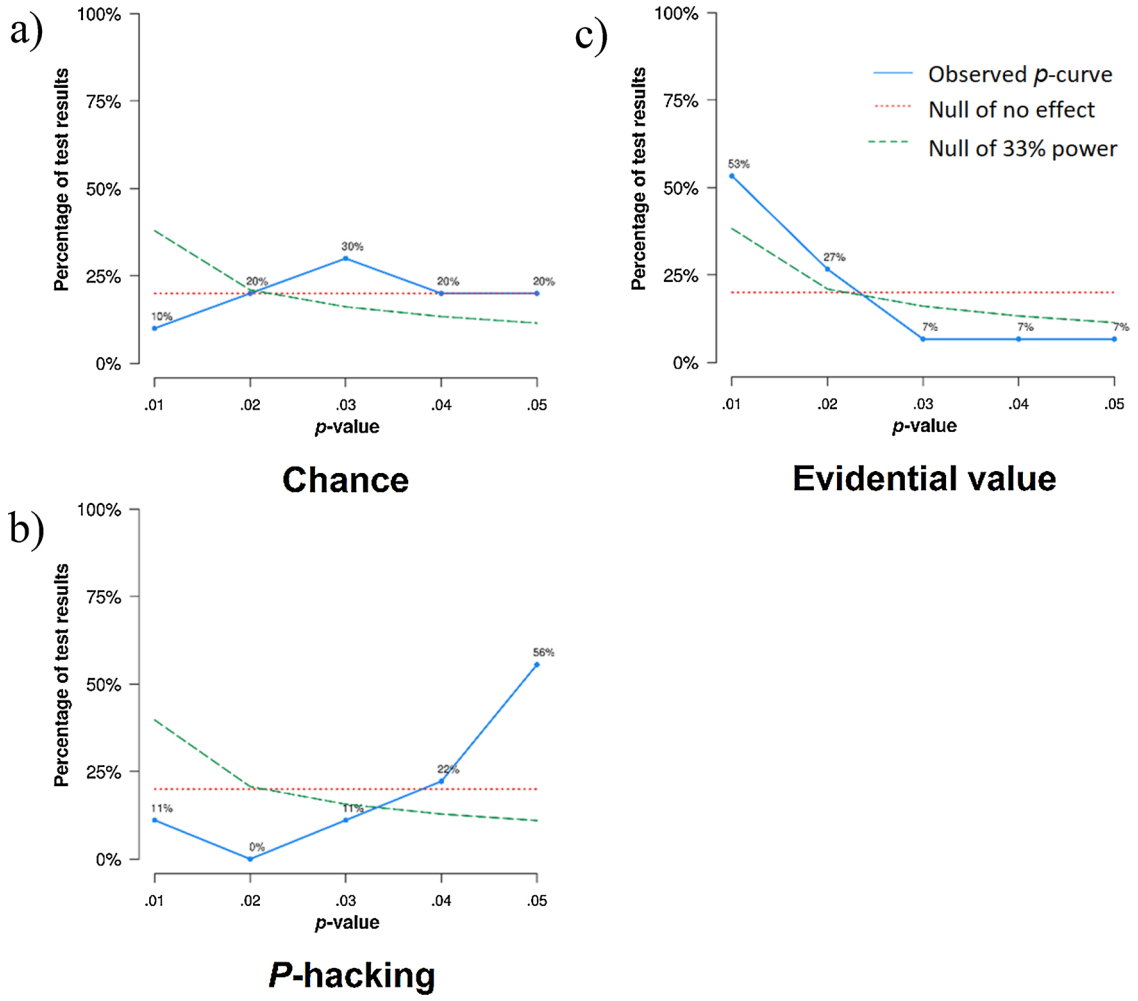

Fig. 1. Illustrations of the potential *p*-curves from the studies reporting expertise effects in the right FFA. The panel in the top-left (a) shows the flat distribution of *p*-values that we would expect if expertise effects were simply occurring through chance. The bottom-left panel (b) shows the left-skewed *p*-curve we would anticipate if the null hypothesis is true and some form of *p*-hacking had occurred in the expertise literature. The panel on the right (c) illustrates the *p*-curve we would expect if the literature provides evidential value for the presence of expertise effects in the right FFA. We used dummy data to create each of these *p*-curves.

criticisms levelled against it. These have included arguments that it relies upon small sample sizes, small effects, and statistically significant *p*-values that are close to .05 (Kanwisher, 2017; Kanwisher, 2006; McKone and Kanwisher, 2005; McKone et al., 2007; McKone and Robbins, 2011; Robbins and McKone, 2007). When these criticisms were first aired over 10 years ago, many readers would not have been aware that such factors would turn out to be responsible for the current replication crisis facing the field of psychology. This crisis is characterised by the fact that psychology researchers have been unable to replicate previously published effects in the literature (Collaboration, 2015; Pashler and Wagenmakers, 2012; Simons, 2014).

In the case of small sample sizes, modelling work has shown circumstances under which they are at greater risk of identifying false-positives than larger samples (Ioannidis, 2005; Simmons et al., 2011). If this is true, then data derived from small samples that support the expertise hypothesis may have wrongly identified effects as being present when the null hypothesis was true. Even if the expertise (i.e., alternative) hypothesis is true, underpowered studies will tend to produce significant results when they have overestimated effect sizes in cases of publication bias (Button et al., 2013). These data would then undermine future replication efforts that based their sample sizes on such underpowered work. This is because the replication attempts themselves would also be underpowered, and thus diminish the likelihood of finding the same result; i.e., making a Type II error where the alternative hypothesis is falsely rejected in favour of the null.

It has also been recently argued that when a body of literature relies upon *p*-values close to .05, then such an effect may have arisen due to *p*-hacking, which can increase the probability of finding a false positive when the null hypothesis is true (Simonsohn et al., 2014b). Examples of *p*-hacking include when an experimenter stops collecting data when their hypothesis has been supported by a *p*-value under .05, or when they have changed their hypothesis and/or purportedly pre-planned analyses post-hoc after extensive data dredging through repeated exploratory analyses (Simonsohn et al., 2014b). Looking back to claims that the expertise hypothesis relies upon 'a few close-to-significant' findings (McKone and Kanwisher, 2005), a modern reader may begin to wonder if this perspective is supported by studies that incorrectly rejected the null hypothesis due to *p*-hacking. These criticisms of the expertise account become exceedingly more pertinent to the modern reader when they appear further supported by studies showing that expertise effects in the right FFA do not replicate (Op de Beeck et al., 2006; Grill-Spector et al., 2004; Yue et al., 2006). It is therefore important to re-examine these criticisms of the expertise account in light of what we have learned from the replication crisis. This reassessment is possible through the use of modern statistical tools that have been developed to identify when the evidence for any given effect has genuine value, or is instead weak or problematic.

The recently developed *p*-curve is a form of meta-analysis that allows researchers to test whether there is any evidential value to support, or debunk, an observed effect in the literature (Simonsohn et al.,

2014a,b; Simonsohn et al., 2015; for similar alternatives, see: *p*-uniform, Van Assen et al., 2015; *z*-curve, Brunner and Schimmack, 2016; Schimmack and Brunner, 2017). Despite its young history, it has been extensively used in a wide variety of areas of psychological science (Chan et al., 2017; Cracco et al., 2018; Gildersleeve et al., 2014; Lakens, 2017; Ritchie and Tucker-Drob, 2018; Simmons and Simonsohn, 2017; Sala and Gobet, 2017; Steffens et al., 2017; Weingarten et al., 2016). One key attribute that makes the *p*-curve superior to traditional meta-analysis techniques is that it is not susceptible to publication bias; i.e., when authors do not publish null results or, to a lesser extent, failed replications. This is because the *p*-curve simply discards all non-significant results, and with them the inherent bias of unreported null findings that arise from journals' preference for publishing significant results (Ferguson and Heene, 2012). Instead, the *p*-curve focuses on assessing the evidential value of the data that *are* significant, independently of failed replications and unpublished works.

The *p*-curve is based upon the assumption that the distribution of *p*-values should be right-skewed when a true effect can be determined from the literature, thereby showing 'evidential value' for a particular hypothesis (Simonsohn et al., 2014a, 2014b; Simonsohn et al., 2015), e.g., does the right FFA respond to items of expertise? By contrast, under the null hypothesis the distribution of *p*-values supporting this purported effect (i.e., a Type I error) should be evenly distributed. Finally, if intense *p*-hacking has occurred when the null hypothesis is true, then we might expect to find a left-skewed distribution of *p*-values congregating just under $p = .05$ (see Fig. 1 for an illustration of these potential *p*-curves; Simonsohn et al., 2014a)[1] ; this is the outcome that we would expect to find if the criticism of the expertise account relying upon 'close-to-significant' *p*-values (McKone and Kanwisher, 2005) was a valid concern. Such a result would imply that expertise effects in the right FFA have arisen due to intense *p*-hacking when the modular account is actually true (Simonsohn et al., 2014b), and thus provide researchers with valid reasons for doubting the right FFA's role in visual expertise.

In order to assess the evidential value of the expertise hypothesis, we collated every fMRI/MRI study that appeared to link the right FFA to expert level object processing. After reviewing each paper against a set of inclusion/exclusion criteria, we submitted the suitable papers to two *p*-curve analyses: one based upon the data confirming the original authors' hypotheses, and another testing the linear relationship between behavioural expertise (i.e., participants' performance when recognising objects for which they have expert knowledge of) and right FFA activation. If there is evidential value for claims that the right FFA is responsive to items of expertise, then we should find a right-ward skew of *p*-values. However, if the distribution of *p*-values turned out to be significantly flat, or displayed a left-ward skew, then we would have to accept that there is no sufficient evidential value present in the literature to support the existence of expertise effects in the right FFA. Such a result would indicate why expertise effects in the FFA have not been replicated: it is because they simply do not exist.

2. Methods of our meta-analyses

2.1. Identifying relevant papers

As per the guidelines provided by the creators of the *p*-curve (Simonsohn et al., 2014b), we make our *p*-curve disclosure tables freely available to anyone that is interested at the Open Science Framework (Original Hypotheses Table: https://osf.io/x8vmw/; Correlation Table: https://osf.io/y49pe/; with a summarised disclosure presented in Table 1). We performed searches across April and May in 2018, on both Web of Science and Google Scholar, using such terms as "expertise", "fusiform face area", "FFA" and "right" in order to identify relevant studies. We further countered the possibility that we missed any studies by cross-referencing those that we had collected against papers cited in reviews on the expertise effect. For a flow diagram charting our study selection process, see Fig. 2.

We decided *a priori* to exclude any studies that examined expertise effects in the N170 event-related potential or M170 (e.g., Busey and Vanderkolk, 2005; Gauthier et al., 2003; Xu et al., 2005), due to the lack of certainty that they are entirely driven by the right FFA (Deffke et al., 2007; Eimer, 2011). These papers were easily identified, as were duplicates of already acquired papers, so we did not download or take a record of them. During the literature search stage, we only downloaded papers that explicitly employed fMRI/MRI methods and appeared to be testing expertise effects in the right FFA. We included structural MRI papers because if the underlying cortical thickness of a region is related to how well an individual can recognise any given object, then it would seem intuitive to assume that this region has a functional purpose in recognising that object (Gauthier, 2017). For one paper, we requested and received a copy from an author directly as we were unable to source it online (Righi et al., 2013). After performing our searches and downloads, we had 40 papers employing MRI methods that had potential for being used in our meta-analyses.

We employed a number of inclusion and exclusion criteria for determining whether each paper could be included in our meta-analyses. Our inclusion criteria allowed for papers that a) found significant right FFA effects[2] related to the visual processing of any object categories (e.g., cars, birds, Greebles), b) explicitly tested participants with perceptual expertise of said visual domains, and c) used a localizer task to identify the FFA. We excluded any papers that a) only tested expertise related to faces, b) tested neuropsychological populations, c) reused data that was not statistically independent from previously published research, d) did not provide relevant statistical results that could be used in the *p*-curve code, and e) examined word recognition; the expertise hypothesis has repeatedly stated that the right FFA is highly unlikely to be recruited for word recognition due to the fact that the processes involved in word and face perception are likely to be highly distinct[3] (Bukach et al., 2006; Gauthier et al., 2006; Wong & Gauthier, 2007; Wong et al., 2009). When we found papers that reused data from previously published work, we only included the original paper.

Using our exclusion criteria, we excluded one paper on the basis that they found expertise effects in the right FFA of a patient with agnosia (Behrmann et al., 2005), and another two that partially reused data from prior studies (Gauthier and Tarr, 2002; McGugin et al., 2016). We excluded a further five papers that found FFA non-face effects that were not directly testing expertise (Adamson and Troiani,

[1] Although *p*-hacking strategies involving multiple parallel tests can mimic a true effect (Ulrich & Miller, 2015), it has been demonstrated that when the alternative hypothesis is true, it is not possible to obtain a left-skewed distribution when effect sizes are medium to large (Hartgerink et al., 2016). By contrast, when effect sizes are zero to low, then a left-skew distribution could be interpreted as evidence that *p*-hacking has occurred and that a true effect size is not medium to large.

[2] We extracted both two-tailed and one-tailed results from the texts. In the *p*-curves in our results section, our analyses conservatively corrected the three papers reporting one-tailed results to two-tailed (i.e., doubled the *p*-values; Gauthier et al., 2005; McGugin et al., 2014a; Ross et al., 2018). All results were virtually unchanged when one-tailed values were used (see Supplementary Information for the results of 40000 *p*-curves with one-tailed corrections applied and the readme file on the Open Science Framework for the data and R code: https://osf.io/dkbxj/).

[3] While a number of papers seem to bear this prediction out (Burns et al., 2017a; Hill et al., 2015; Rubino et al., 2016; Susilo et al., 2015; Starrfelt et al., 2018), it should be noted that recent work has demonstrated general links between face perception, bilingualism (Burns et al., 2018; Fecher and Johnson, 2019; Fort et al., 2018; Kandel et al., 2016; Mercure et al., 2019; Singh et al., 2019) and word recognition (Behrmann and Plaut, 2012; Roberts et al., 2015; Sigurdardottir et al., 2015, 2018, 2019).

Table 1
The 18 studies that identified expertise effects in the right FFA. All studies were included in our first *p*-curve, asterisked papers used for our correlational *p*-curve directly linking perceptual expertise with the right FFA; the full *p*-curve disclosure tables can be found at the OSF (Original Hypotheses: https://osf.io/x8vmw/; Correlations: https://osf.io/y49pe/). The right column presents each test statistic entered into the *p*-curve and their resulting two-tailed *p*-values generated by the *p*-curve.

Author/Year	Expertise	Statistic entered into the original hypothesis *p*-curve and their *p*-curve generated *p*-value
Bartlett et al. (2013)	Chess	$F(1,19) = 6.49$, $p = .0197$
Bilalić et al. (2011),	Chess	Exp 1: $F(1,13) = 5.4$, $p = .037$; Exp 2: $F(1,12) = 7.9$, $p = .0157$; Exp 3: $F(1,11) = 6.6$, $p = .0261$
Bilalić (2016)	Chess	$t(33) = 2.06$, $p = .0474$
Bilalić et al. (2016)	Radiographs	$t(29) = 2.8$, $p = .009$
Gauthier et al. (1999)	Greebles	$F(4) = 88.9$, $p = .0007$
Gauthier et al. (2000)*	Cars and Birds	Car: $r(4) = .75$, $p = .0859$; Bird: $r(4) = .82$, $p = .0457$
Gauthier et al. (2005)*	Cars	Exp 1: $r(5) = .7$, $p = .0799$; Exp 2: $r(4) = .96$, $p = .0024$; Exp 3: $F(1,9) = 6.32$, $p = .0331$
Harel et al. (2010)*	Cars	Exp 1: $F(2,42) = 2.52$, $p = .0926$; Exp 2: $F(1,21) = 5.5$, $p = .0289$
Harley et al. (2009)	Radiographs	$r(18) = .55$, $p = .012$
McGugin et al. (2012)*	Cars	Exp 1: $r(16) = .57$, $p = .016$
McGugin et al. (2014a)*	Cars	$r(17) = .54$, $p = .017$
McGugin et al. (2014b)*	Cars	$r(25) = -.39$, $p = .0443$
Moore et al. (2006)	Blocks	$t(8) = 2.23$, $p = .0563$
Rhodes et al. (2004)*	Butterflies	$r(6) = .8$, $p = .0096$
Ross et al. (2018)*	Cars	$r(19) = .57$, $p = .007$
Wong et al. (2009)*	Ziggerins	$r(16) = .7$, $p = .001$
Wong et al. (2010)*	Musical Notation	Experts: $r(7) = -.94$, $p < .001$; Novices: $r(7) = .9$, $p < .001$
Xu (2005)*	Cars and Birds	Bird expertise across all participants: $r(8) = .74$, $p = .0144$

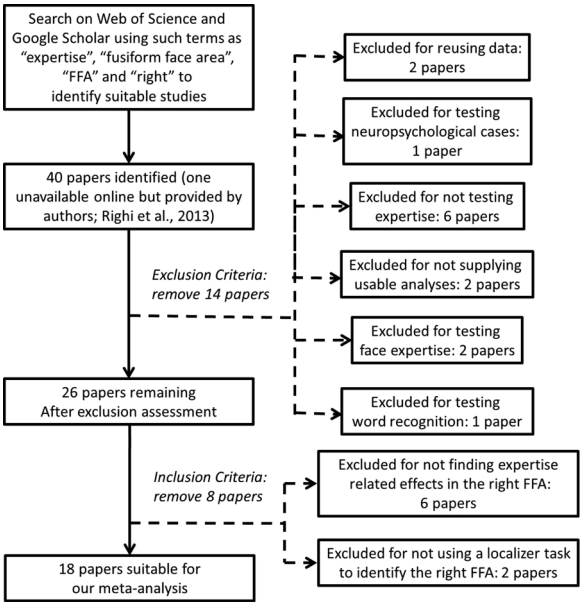

Fig. 2. The flow diagram charting our steps taken to identify and exclude studies for our meta-analyses.

2018; Çukur et al., 2013; Grill-Spector et al., 2006; Hanson and Schmidt, 2011; Slotnick et al., 2013), and another paper as it found right FFA Greeble effects that were associated with their similarity to faces, rather than expertise (Brants et al., 2011). From the remaining papers, two were excluded as they did not report the exact statistical values required (Gauthier et al., 2000a; Martens et al., 2018), and another two studies that appeared to test expertise for unusual faces (James and James, 2013; McGugin et al., 2017). A final paper was excluded as it examined expertise for words (Wong et al., 2009). From our original 40 papers, our exclusion criteria removed 14 papers altogether.

After our exclusion process, we examined the remaining 26 papers using our inclusion criteria to ensure each paper was suitably acceptable for our meta-analyses. From this, two papers failed to meet inclusion as they did not explicitly identify the right FFA with a localizer task (Liu et al., 2009; Righi et al., 2013). A further six papers were not included as they did not find expertise effects in the right FFA (Gilaie-Dotan et al., 2012; Grill-Spector et al., 2004; Krawczyk et al., 2011; Op de Beeck et al., 2006; Wong et al., 2009; Yue et al., 2006). This left us with a final collection of 18 papers that found expertise related effects in the right FFA (see Table 1).

2.2. P-curve: the method and the selection of p-values

Once we narrowed down the literature, we performed two *a priori* determined *p*-curve analyses. As per the general *p*-curve rules our first analysis was, to the best of our ability, based upon the original authors' specific hypotheses regarding their testing of expertise effects in the right FFA. Sometimes authors were not terribly specific about their predictions, and at times hypotheses were blurred by the fact that many papers reported their findings in their introductions. In such instances, we attempted to identify the statistical test that best fitted the overall story of their manuscript. The second *p*-curve analysis was based upon our own desire to test whether expertise effects in the right FFA index behavioural expertise (i.e., experts' performance in recognising their objects of interest). In this second analysis, we therefore included only the significant relationships between right FFA activity and behavioural performance. This relationship has been highlighted in a prior review supporting the expertise hypothesis (Gauthier and Bukach, 2007), so we felt it could be a more objective way of testing the linear relationship between the right FFA and behavioural expertise independent of researchers' original hypotheses; although it should be noted that in many papers, these *p*-values are the same (11 papers, see asterisked studies in Table 1).

The *p*-curve developers recommend some instances where it is preferable to use interaction effects in the *p*-curve over simple effects (e.g., in a 2 × 2 interaction where you only expect one simple effect to be significant, you use the interaction *p*-value; Simonsohn et al., 2014 p.543 and Fig. 5). However, for attenuated cross-over interactions (i.e., car > bird FFA activity in car experts, bird > car FFA activity in bird experts), the *p*-curve developers recommend using simple effects instead. We therefore initially planned to include these simple effects as per the *p*-curve developers' advice, but were largely unable to since most authors did not report these subsidiary analyses on their interactions. Instead, we included the interactions where available. However, if there were significant correlations between participants' behavioural expertise levels and their FFA activations, then these always took precedence to be included in our Disclosure Table and *p*-curve analyses. This is based upon the fundamental premise that behavioural levels of

object expertise should linearly predict activation levels in the FFA in experts, and because the *p*-curve developers recommend using linear trends in the *p*-curve (Simonsohn et al., 2014b).

Sometimes, there were multiple correlation *p*-values present in papers that were not statistically independent from one another; i.e., they used the same behavioural expertise measure to correlate with their haemodynamic responses across separate scanner runs or multiple FFA regions. In these instances, we followed the *p*-curve developers' guidelines and picked a single *p*-value in order to maintain statistical independence between our studies. Moreover, when there were multiple of these *p*-values to choose from, we picked the first *p*-value that was found in the text as per the developers' advice. After we had performed these main *p*-curve analyses, we replaced the first of these multiple *p*-values with the last available in the text, and reran our analyses (these additional *p*-curves are known as *Robustness Tests*); if the data supporting the expertise hypothesis is robust, then changing the selected *p*-values in this manner should have little effect on the overall *p*-curve results. The test statistics used can be found in a file on the Open Science Foundation (https://osf.io/q8x96/). Sections 3.1 and 3.2 of our Results can be replicated by copy and pasting these values into the *p*-curve online app (we used version 4: http://www.p-curve.com/app4/). By contrast, our other analyses used R code (see readme file here: https://osf.io/dkbxj/, data and code: https://osf.io/s2g5v/).

2.2.1. Full p-curve and half-p-curve

A regular *p*-curve analysis provides two results: a full-curve and a half-curve. The full-curve assesses the evidential value of all *p*-values identified by *p*-curvers; i.e., the authors running the *p*-curve analysis. The half-curve only includes *p*-values between 0 and .025. The reason for performing the half-curve is to counter what the *p*-curve developers call 'ambitious' *p*-hacking (Simonsohn et al., 2015). This is based upon the assumption that some *p*-hackers may lower their *p*-values as much as they can below the .05 cut-off of statistical significance in order to make readers less suspicious that they were *p*-hacking in the first place: e.g., $p = .033$ could be less questionable to a reader than $p = .049$. The *p*-curve developers suggest that attaining *p*-values below .035 through *p*-hacking is difficult, so using a threshold of .025 for the half-curve should exclude such results. While the half-curve is less powerful than the full-curve, it is reassuring when both curves show evidential value for a given effect. For the first original hypotheses, our *p*-curve analysis excluded four of our *p*-values from the full curve analysis as not significant, and a further seven from our half curve. The reason for the non-significant exclusions in the full-curve is because the *p*-curve does not rely upon the actual *p*-values reported in papers, but instead recalculates *p*-values from other details extracted from the statistical result, e.g., $t(34) = 2.5$, $F(2, 168) = 5.9$, or $r(29) = .57$. This means that authors' reported *p*-values can be slightly different from those calculated by the *p*-curve.

2.2.2. P-curve tests for flatness

In addition to the full- and half-curve tests for evidential value, the *p*-curve also performs tests for flatness. This test analyses whether the distribution of *p*-values is significantly flatter than what we would expect from a *p*-curve comprised of papers with an average of 33% power. A *p*-value of less than .05 for the test of flatness for the full-curve would indicate that evidential value is absent or weak.

2.2.3. P-curve robustness tests

When there were multiple correlations, and thus multiple *p*-values, we picked the first *p*-value that was found in the original texts to be used in the analyses in our results section (Figs. 3 and 4), as per the *p*-curve developers' advice. After we had performed these primary *p*-curve analyses, we replaced the first of these multiple *p*-values with the last available in the text, and reran our analyses; if the data supporting the expertise hypothesis is robust, then changing the selected *p*-values in this manner should have little effect on the overall *p*-curve results. We found nine instances where we needed to replace the correlations from our main *p*-curves (see files Original Hypotheses: https://osf.io/x8vmw/; Correlations: https://osf.io/y49pe/).

2.2.4. Performing 10,000 p-curves based upon all p-values supporting domain general effects in the right FFA

In addition to traditional *p*-curves, a more comprehensive way of examining the literature could be through the analysis of all relevant *p*-values that support the expertise hypothesis. This would counter any potential bias that may exist when authors present their most convincing findings (i.e., lowest *p*-values) at the beginning and end of their results sections, and their remaining significant *p*-values (i.e., values closer to .05) in between. Similarly, authors may have changed their hypotheses and predictions *post-hoc* to match the *p*-values that were the lowest from their analyses. To counter these potential biases, we identified every *p*-value that could be interpreted as linking the right FFA to non-face processing. We then ran 10,000 *p*-curves where each *p*-curve randomly selected one *p*-value from each independent experiment (see readme file for a guide on the files: https://osf.io/dkbxj/, R code and data files for these analyses can be found here: https://osf.io/s2g5v/). If the bulk of these *p*-curves are significant, then it could be taken as good evidence that the expertise literature is robust in providing evidential value.

It should be stressed that the distribution of these analyses should not be interpreted as meaningful in the same way that a single *p*-curve analysis is. This is because the 10,000 *p*-curves are not independent of one another, as required for single *p*-curve analyses. Rather, each generated *p*-curve in this robustness analysis, when significant, simply indicates a right skewed distribution and therefore evidential value from the randomly sampled *p*-values from all papers. This means that if these 10,000 *p*-curve results display a flat distribution, but 100% are significant, the important finding is that the literature always produces evidential value when potential biases are accounted for. It is therefore inappropriate to interpret a skew from our 10,000 *p*-curve results as a presence or absence of evidential value.

3. Results: do expertise effects exist in the right FFA? *P*-curving the literature

3.1. P-curves based upon the authors' original hypotheses

Our first *p*-curve analysis tested the literature based upon the *p*-values associated with the authors' original hypotheses. From our 18 studies, we identified 25 *p*-values to be entered into this meta-analysis (see Table 1). This *p*-curve produced a significant rightward skew with respect to the *p*-values' distribution suggesting evidential value was present for expertise effects in the right FFA (Fig. 2a: Full curve $Z = -2.31$, $p = .011$; Half *p*-curve, $Z = -2.07$, $p = .019$). The tests for flatness did not indicate an absence of evidential value (Full curve $Z = .21$, $p = .58$; Half *p*-curve, $Z = 4.45$, $p > .99$). Finally, the studies yielded an estimated power of 38% [*90% CI* = [9, 69%]], which suggests that this collection of experiments was underpowered relative to the generally recommended levels of power of 80% with alpha levels set at .05 (Button et al., 2013). Planned robustness tests, replacing the first nine reported *p*-values in the text with the last still yielded significant full- and half-curves (Fig. 2b: both $ps < .02$; Power = 37% [*90% CI* = [8, 69%]]). In summary, the results of our *p*-curves indicate that the distributions of *p*-values are generally what we would expect if expertise effects do exist in the right FFA.

3.2. P-curves based upon the correlations between behavioural expertise and the right FFA

While our first *p*-curve analyses suggested that expertise effects were present in the right FFA, we wanted to perform a second *p*-curve to assess whether the right FFA's activity is linearly related to behavioural

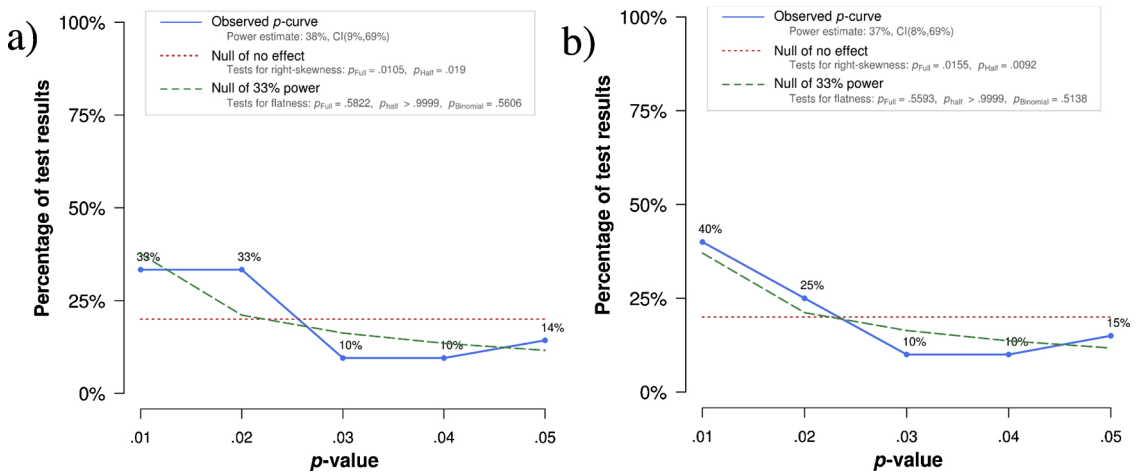

Fig. 3. *P*-curve results from the original researchers' hypotheses. Panel a) Illustrates our first *p*-curve based upon the authors' original hypotheses, Panel b) reflects the same analysis but we replaced the first correlations found in the text with the last: both the full- (both *ps* < .02) and half- (both *ps* < .02) curves were significant. Overall these meta-analyses indicate evidential value for the expertise hypothesis. You can replicate these results with data on the Open Science Framework (https://osf.io/q8x96/) and the online *p*-curve app (http://www.p-curve.com/app4/).

performance for recognising objects as reported in a previous review paper (Gauthier and Bukach, 2007). As correlational analyses can account for individual differences in expertise that may be obscured when analysing across categorical groups (Bukach et al., 2012), we anticipated that such data may have higher levels of power.

We identified each instance of an association between behavioural expertise performance and the right FFA, yielding 11 papers and 14 correlations: two were excluded from the full *p*-curve as they were deemed non-significant, with an additional two excluded from the half-curve. The results of this correlation *p*-curve analysis showed that the current literature does indeed contain evidential value for a linear relationship between FFA activity and behavioural expertise (see the rightward skew in Fig. 3a: Full curve $Z = -2.78$, $p = .003$; Half *p*-curve, $Z = -2.09$, $p = .018$). Second, the tests for flatness did not indicate an absence of evidential value (Full curve $Z = 1.16$, $p = .87$; Half *p*-curve, $Z = 3.91$, $p > .99$). Finally, this collection of studies yielded an estimated power of 63% [*90% CI* = [21, 88%]], which is just below the level of power modern researchers require to adequately detect an effect (Button et al., 2013; Cohen, 1992). Moreover, the mean correlation co-efficient linking the right FFA's activation and behavioural expertise was $r = .71$ (95% *CI* = [.61, .81]). When we replaced the nine 'first' correlations found in each manuscript with the last, we replicated these results (Fig. 3b: both *ps* < .008; Power = 64% [*90% CI* = [19, 90%]]). Our meta-analyses of the correlational data supports the right FFA's haemodynamic response as being directly linked to perceptual expertise.

3.3. Performing 10,000 p-curves when sampling from all relevant p-values

To avoid any possibility that we had somehow selected the most biased (i.e., lowest) *p*-values for our previous *p*-curves, we identified every single *p*-value that could be interpreted as supporting the idea that the right FFA is linked to non-face processing. We then ran 10,000 *p*-curves by randomly selecting one *p*-value from each independent experiment. Roughly > 99% of the 10,000 *p*-curves sampled from all potential *p*-values supporting domain general effects in the right FFA were significant (Table 2). Moreover, when we ran 10,000 *p*-curves randomly sampling only the correlations between behavioural expertise and FFA activity, roughly > 99% of these *p*-curves were also significant. Finally, it has been suggested that we should have more confidence in the evidential value of *p*-curves when they are still significant after lowest *p*-values are dropped; i.e., those that may be unrealistically

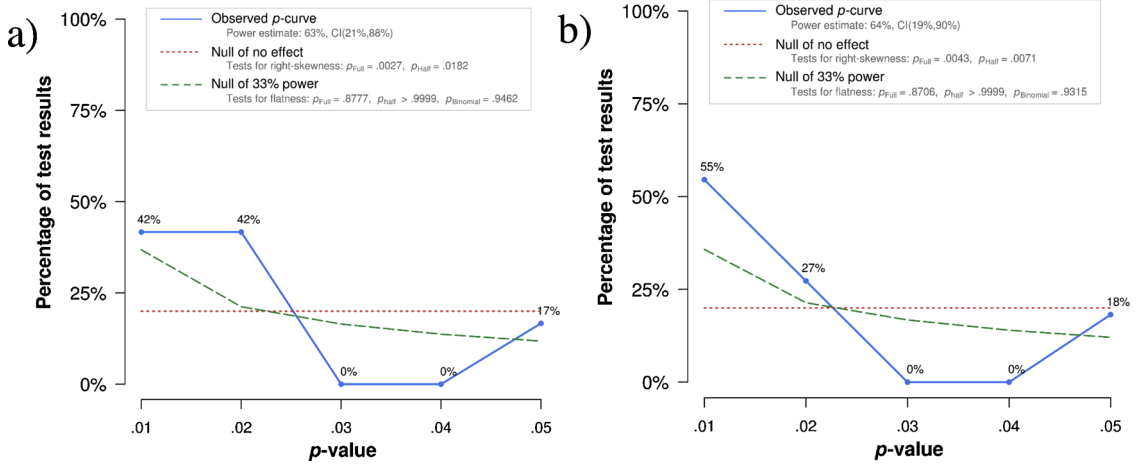

Fig. 4. *P*-curve results from our hypothesis that the right FFA can be directly correlated with behavioural expertise. Panel a) was based on the first correlations found in the expertise texts, Panel b) was created by replacing the first *p*-values with the last. The right skew in both *p*-curves indicates that there is evidential value for expertise effects' existence in the right FFA (full-curves *ps* < .005; half-curves *ps* < .008). You can replicate these results with data on the Open Science Framework (https://osf.io/q8x96/) and the online *p*-curve app (http://www.p-curve.com/app4/).

Table 2
Each row reports the percentage of 10,000 *p*-curve results that were significant and non-significant when sampling relevant *p*-values from the literature at random (see readme guide on the Open Science Framework: https://osf.io/dkbxj/; R code and data files for these analyses can be found here: https://osf.io/s2g5v/). The first row indicates the results from *p*-values sampled from each independent experiment using all available *p*-values, the third row sampled only the correlations between expertise and FFA activity. The second and fourth rows are the same results but with the lowest *p*-value dropped on each run. The fact that roughly > 99% of *p*-curves were significant indicates that the evidential value found in the literature is robust. Please note that these *p*-curve results will change on each sampling run as *p*-values are randomly selected from the literature. These outcomes were largely replicated using one-tailed values where appropriate (see Supplementary Information).

P-curve Results	$p < .01$	$p = .01$–.02	$p = .02$–03	$p = 03$–.04	$p = .04$–.05	$p > 05$
All available *p*-values	47%	21%	12%	7%	13%	≈0%
All available *p*-values (lowest *p*-value dropped)	44%	22%	12%	8%	13%	≈1%
All available correlations	32%	27%	30%	1%	9%	≈1%
All available correlations (lowest *p*-value dropped)	22%	31%	35%	2%	10%	≈0%

extreme (Simonsohn et al., 2017). While we do not believe that any of the reported *p*-values were impossibly low, we conservatively ran our 10,000 sampling runs again but dropped the lowest *p*-value on each *p*-curve; the results were largely similar to our runs where we did not drop any *p*-values (i.e., roughly > 99% of the 10,000 *p*-curves were significant; Table 2). In summary, the neuroimaging expertise literature has robust evidential value when all relevant *p*-values are taken into account.

4. Discussion

4.1. Meta-analyses support the existence of expertise effects in the right FFA

Our primary and robustness *p*-curves provide support for the right FFA's responsiveness to items of expertise, and that these responses can be correlated with behavioural performance. Moreover, > 99% of 80,000 *p*-curves taken from random sampling of all relevant *p*-values in the literature (40,000 in Table 2, 40,000 in Supplementary Information) were significant too. These findings suggest that based upon the neuroimaging literature, the expertise hypothesis is a theory that has evidential value.

The expertise hypothesis has been criticised for relying upon 'close-to-significant' findings (McKone and Kanwisher, 2005). We now know that when a collection of studies have *p*-values around .05, they are likely to have been *p*-hacked (Simonsohn et al., 2014b). To the modern reader, this is a serious criticism as it could potentially imply that statistical significance may have occurred due to the experimenter's degrees of freedom (Simmons et al., 2011), rather than from a true effect. Our analyses reject this criticism of the expertise hypothesis: expertise *p*-values were not likely to have occurred by chance (i.e., a flat distribution), or through questionable research practices (i.e., a left skew). The right skews we observed in our *p*-curves show evidential value for linking behavioural expertise to the right FFA. In summary, suggestions that we can reject the expertise hypothesis due to low sample sizes and *p*-values close to .05 do not appear to be valid. However, this may beg the question: why do failed replications exist in the literature?

4.2. Failures to replicate do not negate support for the expertise hypothesis

It has been suggested that both direct replications (e.g., identifying FFA car expertise effects across two separate studies with negligible differences in design) and conceptual replications (e.g., finding expertise FFA effects in different studies where major changes in design have occurred, such as across different stimulus domains or task demands) are essential for fighting the current replication crisis facing the field of psychology (Crandall and Sherman, 2016). In this respect, the expertise hypothesis has been exceptionally successful. For example, direct replications have been found for multiple domains of visual expertise, such as birds (Gauthier et al., 2000; Xu, 2005), cars (Gauthier et al., 2000; McGugin et al., 2012), chess related stimuli (Bilalić, 2016; Bilalić et al., 2011; Bartlett et al., 2013) and radiographs (Bilalić et al., 2014; Harley et al., 2009). Moreover, the heterogeneity of these stimulus domains shows the expertise hypothesis as being supported by numerous conceptual replications. In addition to replicating expertise effects in the FFA across multiple domains, there have been further conceptual replications when task designs have been extensively varied too. For example, while FFA expertise effects can be found during individuation (e.g., Gauthier et al., 1999; Harley et al., 2009), they also appear present during clutter (e.g., irrelevant stimuli crowding the item of expertise, McGugin et al., 2014b) and attentional manipulations (e.g., when attention is diverted away from the item of expertise or divided, McGugin et al., 2014a).

Considering the results of our meta-analyses, coupled with the breadth of stimulus domains and study designs under which right FFA expertise effects can be observed, one may ask why failures to replicate exist. While we do not want to dissect each of these failures in detail, it is worth remembering that the current expertise literature is generally underpowered (see Results 3.1.). This means that the early expertise studies with exceptionally small sample sizes may have overestimated effects in order to achieve significant results (Button et al., 2013). If the authors of the failed replications had based their participant sample sizes upon these studies, then it would have increased the likelihood that they would then fail to find significant results themselves.

To illustrate this point, many early expertise studies with small sample sizes (i.e., $n < 10$; Gauthier et al., 2005, 2000; Rhodes et al., 2004; Wong et al., 2009) found strong correlations between behavioural object expertise and FFA activation (mean $r = .83$, 95% *CI* [.72, .94]). A power analysis with power set at 80% and alpha levels of .05, would suggest a sample size of only eight participants would be required to replicate these results. However, if we were to base the strength of this relationship upon studies from 2014 onwards, that had much larger sample sizes with some performing power analyses from the wider literature (see Table 1; McGugin et al., 2014a,b; Ross et al., 2018), then the mean correlation co-efficient drops to ($r = .5$, 95% *CI* [.26, .74])[4] . The same power analysis on this correlation coefficient would recommend a sample size of 29 participants (23 one-tailed). When we go back and look at the failed replication attempts, we find that none of these six studies met this criterion, with the largest sample size comprising of only 12 experts ($Mean^{6\ studies} = 8$ experts; Gilaie-Dotan et al., 2012; Grill-Spector et al., 2004; Krawczyk et al., 2011; Op de Beeck et al., 2006; Wong et al., 2009; Yue et al., 2006). In these authors' defence, we are examining their data with the benefit of hindsight given to us by the replication crisis. They would not have realised these issues at the time when designing their experiments as they were likely basing their sample sizes on overestimated effects

[4] As significant results can overestimate effect sizes (Button et al., 2013), we calculated the mean correlation strength by including all significant and non-significant correlations between expertise and FFA activation in these papers. This yielded a lower mean correlation coefficient ($r = .33$, 95% *CI* [.21, .45]) which would require larger sample sizes than previously considered in the literature: two-tailed 70 participants, one-tailed 56 participants.

found in early work. It should therefore be stressed that these failures to replicate are not inconsistent with the expertise hypothesis, nor do they negate the evidence base that supports it; they are simply an inevitable consequence of underpowered studies (Cohen, 1990).

4.3. Arguments based on 'attention' do not allow us to reject the expertise hypothesis

While we have shown that fMRI expertise studies are likely to be replicable, our analyses have not addressed criticisms that the right FFA's responsiveness to expertise is simply due to enhanced attention (Duchaine & Yovel, 2015; Harel et al., 2010; Kanwisher, 2017). This attention hypothesis posits that those individuals who devote the most attention towards their items of expertise will inevitably have the greatest levels of object recognition as a result of this increased interest. Moreover, the right FFA's response to these items of interest is thus an artefact of enhanced attention that occurs across the whole brain when the viewer attends to any visual stimulus of interest (Wojciulik et al., 1998; Murray and Wojciulik, 2004). The reader is therefore asked to reject the expertise account of the non-face effects in the right FFA because this region is not performing any functional process involved in recognising such items.

These claims, however, do not allow us to reject the fact that the right FFA responds to non-face stimuli, nor do they allow us to reject the expertise hypothesis. First, a recent commentary outlines many of the key problems with this attention argument (Gauthier, 2017)[5] . These issues include the fact that modulating attention during a task still produces robust expertise effects in the right FFA (McGugin el al., 2014b). Moreover, faces and items of expertise appear to elicit right FFA effects that are more similar to one another in experts versus non-experts (McGugin el al., 2014b), suggesting that the same process or aggregate processes are being performed in this region for both sets of stimuli. Finally, Gauthier (2017) points out that cortical thickness in the right FFA predicts performance when recognising both faces and vehicles (McGugin et al., 2016). If the right FFA activity to items of expertise was simply an artefact of attention, then we should not find this region's morphological properties linked to behavioural performance. When considered together, these points pose serious problems for attention based criticisms of the expertise hypothesis. This is especially true when we find that acquired prosopagnosia cases with lesions to this region display object recognition deficits relative to their levels of expertise knowledge (Barton and Corrow, 2016; Barton et al., 2019; Barton et al., 2009); thus suggesting that the FFA is contributing towards object expertise recognition. Similarly, the attention hypothesis is further undermined by studies demonstrating non-face FFA effects in novices, as presumably these individuals do not have any high levels of attention towards the objects eliciting these effects (Çukur et al., 2013, Zachariou et al., 2018).

Remarkably, attention-based criticism of the expertise hypothesis can be turned on its head to suggest that the FFA's *face selectivity* is largely due to attention itself. For example, reducing attention can diminish the right FFA's haemodynamic responsiveness to faces (Williams et al., 2005). This shows that the right FFA's apparent face selectivity can also be explained in part as a result of attention being allocated during the individuation of faces. Similarly, prosopagnosia cases exhibit abnormalities in the way that they attend to faces (Bobak et al., 2017; Van Belle et al., 2010; Van Belle et al., 2011) again suggesting that the cortical face network is involved in the attentional processes required for expert level face individuation. The other race effect is characterised by superior performance in recognising own race faces over those of other races (Bate et al., 2018; Burns et al., 2018; Estudillo et al., 2019; Meissner and Brigham, 2001), with the FFA being cited as one of the neural loci for this effect (Golby et al., 2001). Strikingly, a number of behavioural studies have suggested that the other race effect may be driven by suboptimal allocation of attention towards other race faces (Hills et al., 2013; Hills and Lewis, 2006, 2011). When these points are all considered together, they infer that the right FFA's apparent face selectivity could be related to how attention is allocated. If this is true, then associations between face/object recognition and the right FFA may actually be due to the fact that this region is performing an important attentional process that drives individuation ability. When this attentional process is compromised, as it is in prosopagnosia, we see similar deficits in face and object processing becoming apparent (Barton and Corrow, 2016; Barton et al., 2009). This conclusion therefore still supports the expertise hypothesis in claiming a process-specific, rather than modular, account of the right FFA, as this region is driving the ability to attend to the salient object properties enabling expert level item individuation.

4.4. Neuropsychological evidence does not reject the expertise hypothesis

The right FFA's role in object recognition has also been criticised due to neuropsychological evidence (Duchaine & Yovel, 2015; Kanwisher & Yovel, 2006; McKone et al., 2007). Patients with brain lesions have a long history of providing researchers with insights into the functional nature of distinct neural regions: if a patient suffers damage to a particular area, and exhibits concurrent behavioural problems, then we can typically infer that said brain region must have a role to play in producing that behaviour (Heilman and Valenstein, 2010; Passingham et al., 2002; Rorden and Karnath, 2004). Damage to the right FFA typically leads to an inability to recognise facial identity (Barton et al., 2002; Dalrymple et al., 2011), a condition known as acquired prosopagnosia. Prosopagnosia cases have therefore provided strong support to the notion that the right FFA is essential for face recognition. If these individuals were to also demonstrate comorbid deficits in object recognition, then this could be taken as good evidence that the FFA is providing critical operations required for the processing of non-face items.

Classically, studies showed that acquired prosopagnosia cases were generally spared in object recognition (Farah, 1991). Despite this, we now know that the bulk of this research did not test their patients with objects that were equivalently matched in complexity to faces, nor were their tests likely to have been sensitive enough to detect impairment (Campbell & Tanaka, 2018; Gauthier et al., 1999). Moreover, early work failed to take into account participants' level of semantic knowledge related to a category of objects. This is important, as semantic knowledge for a non-face category can predict perceptual expertise to that category (Van Guklick et al., 2016). When acquired prosopagnosia cases with right FFA lesions were assessed for their semantic knowledge for non-face objects, it became apparent that these individuals also suffered deficits visually recognising these objects when compared to neurotypical individuals with similar levels of knowledge (Barton and Corrow, 2016; Barton et al., 2019[6] ; Barton et al., 2009). Furthermore, in addition to object recognition deficits, right FFA lesions also result in impairments in processing the stylistic aspects of text, such as fonts and handwriting (Hill et al., 2015), further confirming that the FFA must be providing some concrete perceptual operations for non-face stimuli. Finally, acquired prosopagnosia cases have been shown to attain both face and object recognition through atypical strategies (Bukach et al., 2012), thus suggesting that face and object processes may rely upon

[5] We urge anyone with an interest in the expertise and modular hypotheses to read this paper. It gives a very illuminating account of how the first two successful FFA expertise replications actually originated from a previously unreported collaboration between authors who disagreed on the FFA's functionality.

[6] Although this paper simply reported their cases' lesions as 'occipito-temporal' or 'fusiform', an earlier publication from the same lab (Hills et al., 2015) confirms that they had FFA lesions.

shared neural networks.[7]

Support for the expertise hypothesis has also come from developmental prosopagnosia cases. These individuals suffer severe deficits in face recognition (Bate et al., 2014; Burns et al., 2014, Burns et al., 2017a; Burns et al., 2017b; Duchaine & Nakayama, 2006), with both genetic and developmental (Behrmann and Avidan, 2005; Grüter et al., 2008; Duchaine et al., 2007; Schmalzl Palermo & Coltheart, 2008; Susilo and Duchaine, 2013) hypotheses proposed for their origins. Moreover, these individuals have been shown to exhibit a broad range of abnormalities associated with their cortical face perception network, including their right FFAs (Garrido et al., 2009; Song et al., 2015; Zhang et al., 2015). If such cases also exhibit deficits in object recognition, then we could infer that their associative cortical atypicalities in face processing regions are contributing towards these issues. Indeed, a number of studies have shown that many of these cases also suffer non-face recognition problems too (Behrmann et al., 2005; Biotti et al., 2017; Duchaine et al., 2007). A recent review estimated that around 80% of developmental cases in the literature may evince object recognition deficits (Geskin & Behrmann, 2017; although see Garrido et al., 2018), with many suffering specific difficulties related to visual expertise (Barton et al., 2019). When considered together with the data from the acquired cases, it becomes apparent that the right FFA is to some extent utilized during object recognition.

4.5. The expertise hypothesis does not claim expertise effects are restricted to the FFA

We believe that our meta-analyses and review have demonstrated that the right FFA functionally contributes to expert recognition of objects. With this information, we hope that researchers will be in a better position to objectively assess claims that the FFA is not recruited for object recognition (Kanwisher, 2017; Duchaine & Yovel, 2015). However, we do not wish to leave the reader with the misconception that the expertise hypothesis is concerned only with the right FFA, as is sometimes suggested (Duchaine & Yovel, 2015). It is true that early papers from the expertise debate focused on the FFA in an effort to explain *why* it appears face selective (e.g., Gauthier et al., 1999, 2000a); i.e., is it face-selective because of our vast experience with faces? Similarly, the expertise account aimed to test whether face selectivity could also help explain how the brain becomes specialized for objects (Bukach and Peissig, 2010); i.e., is the FFA involved in object recognition? These early neuroimaging papers extended prior behavioural comparisons of face (Yin, 1969) versus object expertise (Diamond & Carey, 1986) recognition into the realm of neuroscience in order to test claims that the FFA was simply face-specific (Kanwisher et al., 1997). Similarly, in the present manuscript at least, we sought to clarify the expertise hypothesis with respect to how it may refer to the right FFA specifically. As mentioned in the introduction, this was borne out of our desire to assess the expertise account using modern meta-analysis techniques that could support, or question, the evidence present in the literature.

What about the claims that expertise researchers have focused solely on the FFA (Duchaine & Yovel, 2015)? When we reviewed the expertise papers collated from our meta-analysis search, we found that the expertise account has never claimed that the FFA is the sole locus of object expertise in the brain, nor is it the only region expertise researchers have examined. Indeed, an early review paper on the expertise hypothesis from over 10 years ago explicitly outlines these facts: "*However, it is not the case that all types of expertise rely on identical mechanisms. Expertise with stimuli that vary radically from the geometry and functional goals of homogeneous object individuation do not engage the FFA, but recruit other, functionally appropriate, regions.*" (Bukach et al., 2006). As testament to this statement, expertise studies have highlighted many other brain regions outside of the FFA that may support object recognition (e.g., Gauthier et al., 1999, 2000; McGugin et al., 2014b; Ross et al., 2018), including areas linked to face perception (e.g., Gauthier et al., 1999, 2000; McGugin et al., 2014a; Ross et al., 2018). The expertise account of the right FFA's function is therefore meant to illustrate a broader general principal of brain plasticity related to visual experience.

5. Recommendations for future work

While our *p*-curves and review of the literature support the expertise hypothesis, we recognize a number of steps that future researchers could take in order to improve replication efforts. Some of these areas of improvement are specific to replicating expertise studies, while others have been taken from the wider replication crisis literature (e.g., Munafò et al., 2017; van Aert et al., 2016; Shrout and Rodgers, 2018; Simonsohn et al., 2014a; Simmons et al., 2011; Wicherts et al., 2016); we urge every researcher reading the current paper to familiarize themselves with these original works. The advice given in these papers is relevant to virtually everyone conducting scientific research, irrespective of discipline. We summarize the advice for the expertise literature in Table 3.

Assuming researchers ensure their studies are well powered with *a priori* determined sample sizes, and that they avoid questionable research practices such as data peeking[8] and *p*-hacking, then the next step for improving the expertise literature's replicability is through the preregistration of experiments (Nosek et al., 2018; Wicherts et al., 2016). This includes as much *a priori* detail as possible regarding the behavioural measures of expertise, exclusion criteria for 'outlier' participants and an outline of how behavioural expertise and neuroimaging data will be specifically analysed (see researcher's checklist: Wicherts et al., 2016). None of the studies we identified to include in our meta-analysis included any comments regarding preregistration. This, however, is unsurprising as designing, implementing, analyzing, and publishing fMRI projects can take many, many years. We would therefore expect a considerable lag between when researchers started recommending preregistration, and such standards becoming the norm in the literature. With this being the case, we should start to find preregistered studies becoming standard within the next few years.

The next recommendation to improve the replicability of expertise research is to share data. Data sharing helps replicability as it ensures researchers are both careful and transparent in their analyses, while also allowing others in the field to check published work is accountable (Nichols et al., 2017; Poldrack and Gorgolewski, 2014; Poline et al., 2012). Moreover, it can lead to new discoveries, particularly when large amounts of data from multiple studies can be combined (Van Horn and Gazzaniga, 2013). As with the case of preregistration, none of the 40 studies we obtained for our meta-analysis made any comment regarding data sharing, even though a national repository for neuroimaging data opened in the US some 20 years ago (D'Esposito, National Science Foundation et al., 2000; Van Horn and Gazzaniga, 2013)[9] .

[7] It should be noted that one study of acquired prosopagnosia cases showed that they could exhibit intact learning and recognition effects with Greebles (Rezlescu et al., 2014). While this study is commonly cited as evidence against the expertise hypothesis, it is difficult to accept this interpretation without first showing that prosopagnosia cases are also abnormal on the equivalent tests using face stimuli.

[8] i.e., analyzing data after each participant and stopping data collection when significant results have been achieved.

[9] It should be noted that the arguments for and against data sharing in neuroscience are more complex than those considered for behavioural work typical in psychology. For example, due to the expense and time it takes to collect, analyse and report neuroimaging data, and that data can be reused across multiple papers over many years (e.g., univariate analyses in one paper, multivoxel pattern analyses in another, structural MRI findings in a separate paper again), it has been argued that neuroscientists should not be forced to share

Table 3
Recommendations for improving the replicability of studies testing expertise/modularity.

Recommendation	Why does it help the replication crisis?
Preregister participant/data rejection criteria (e.g., not identifying an FFA, ± 3 SDs from group mean, no difference between expertise and non-expertise item accuracy) and FFA activation methods.	Reduces false positives by preventing researchers from changing participant rejection thresholds or activation measures post-hoc in order to 'achieve' significant results.
Preregister all expertise and non-expertise tasks, plus the analyses linking them to neural data.	Reduces false positives (or false negatives if the desire is to retain the null) by preventing 'ghost' variables, such as when authors run a batch of tests assessing performance (e.g., expertise accuracy, expertise holistic perception, expertise accuracy divided by non-expertise accuracy, inside or outside scanner performance) but only report those that support the hypothesis. Ensures proper correction of alpha for multiple comparisons when non-significant tests are included.
Preregister sample sizes based upon *a priori* power analyses.	Avoids data peeking/stopping when researchers stop collecting data once significant results have been found.
Specify which analyses were determined *a priori* (i.e., preregistered) and which were performed post-hoc; these latter tests should be reported as exploratory.	Improves transparency and reduces false positives by constraining researchers to *a priori* analyses that were originally motivated by their hypotheses. Exploratory analyses should be interpreted with caution prior to later replication.
Share data, materials, code both with co-authors and external sites (e.g., Open Science Framework or NSF/Keck Foundation National FMRI Data Center).	Encourages transparency and replicability as it allows other researchers (including co-authors) to confirm your analyses and rerun your study
Adversarial collaborations: researchers with differing views work together on preregistered studies with a commitment to publish irrespective of results and write review papers together.	Prevents suppression and/or distortion of knowledge which impedes scientific progress. Reduces partisan ideological divides between researchers. Can lead to more a balanced and nuanced understanding of the evidence.
Ensure stimuli are relevant to participants' expertise (e.g., test their conceptual knowledge) and employ standardized tests sensitive enough for demonstrating differences between experts and novices.	Reduces the risks of false negative results when participants are tested on items that fall outside the limits of their expertise (e.g., testing Australian bird experts with European birds).
Ensure that task demands are equated when comparing categories (e.g., using homogenous stimuli that would require expert level individuation).	Reduces the risk of false negative results by ensuring tasks require specialist individuation associated with expertise processing.

Despite the lack of explicit comments in papers regarding data sharing principles, there are instances of researchers working in the expertise field sharing data (Kanwisher, 2000) and materials (Gauthier, 2017). Moreover, there has been a history of 'adversarial collaborations' (Kahneman, 2003; Mellers et al., 2001), where scientists with differing views on the FFA's functionality have designed and analysed studies together in order to test the expertise hypothesis's claims (Gauthier et al., 2000; Gauthier, 2017; Xu, 2005). One would therefore assume that adversarial researchers in that position would not be able to engage in questionable research practices due to the transparency required to satisfy everyone involved. Similarly, many researchers who are a part of the FFA specificity debate have admirably published data at a later time point even when it is at odds with their own earlier findings and claims regarding the FFA's function (de Beeck et al., 2006d; Susilo and Duchaine, 2013; Hill et al., 2016; Martens et al., 2018). Given such circumstances it seems unlikely, at least in our opinion, that the expertise hypothesis has been built upon questionable research practices. This belief, however, does not mean that the field would not benefit from the above suggestions and those in Table 3 going forward.

Moreover, it is worth noting that there is not a single review paper written in collaboration between authors with differing expertise and modular views. Researchers with adversarial views in other fields have written collaborative review papers that have identified common ground, areas needing more work, and arguably helped provide a more balanced and nuanced debate surrounding their topics (Ariely et al., 2000; Kahneman, 2003; Wixted & Wells, 2017). It is probably worth mentioning that the first author (EB) and the third author (CB) have published work respectively supporting the modular and expertise perspectives in the past, so to some extent this meta-analysis paper started as an adversarial collaboration, with both authors committing to publish the results irrespective of outcome. Our different backgrounds have, we hope, therefore benefited the objectivity of our review here, but we would like to see such reviews becoming more commonplace.

(*footnote continued*)
their data (although these are not their only concerns, see Editorial, 2000; Marshall, 2000; Toga, 2002). We understand these concerns and are not arguing the case for or against them, but instead highlighting that data sharing is something that in theory should benefit replication efforts.

6. Conclusions: object expertise effects in the FFA are replicable

The field of psychology has recently been criticized due to a failure to replicate many published findings. The expertise hypothesis of the right FFA was one such effect that some have claimed is not replicable. Our meta-analyses have however shown that the right FFA is in some way related to how the brain processes both faces and objects. Moreover, our analyses suggest such work is likely to be replicable when studies are well-powered. It is worth noting that when we consider the studies that we excluded from our *p*-curves, and one published since our meta-analyses (Zachariou et al., 2018), the total number of papers that show non-face FFA effects is five times that of the papers cited as failed replications (30 versus 6). This fact may be rather surprising to some as when one reads the literature, this lopsided weight of evidence in favour of the FFA's domain generality is never made so clearly apparent. As we have shown that the data in support of the expertise account is valid, further denying the existence of non-face effects in the right FFA seems unhelpful to scientific progress. Instead, efforts would be better served in trying to identify what functional operations the right FFA is actually performing when we view non-face stimuli. We anticipate recent work demonstrating purportedly face-specific effects for rewarding objects (Burns and Wilcockson, 2019; Adamson and Troiani, 2018) will help further enhance our multi-dimensional understanding of face and object processing in the future.

Acknowledgements

Authors would like to acknowledge the generous funding provided by the James S. McDonnell Foundation Scholars Award for Understanding Human Cognition and the MacEldin Trawick Endowed Professorship of Psychology. We would also like to thank Marcel van Assen for his advice that improved earlier versions of our manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neubiorev.2019.07.003.

References

Adamson, K., Troiani, V., 2018. Distinct and overlapping fusiform activation to faces and food. NeuroImage 174, 393–406.

Ariely, D., Kahneman, D., Loewenstein, G., 2000. Joint commentary on the importance of, duration in ratings of and choices between, sequences of outcomes. J. Exp. Psychol. Gen. 129, 524–529.

Bartlett, J., Boggan, A.L., Krawczyk, D.C., 2013. Expertise and processing distorted structure in chess. Front. Hum. Neurosci. 7, 825.

Barton, J.J., Albonico, A., Susilo, T., Duchaine, B., Corrow, S.L., 2019. Object recognition in acquired and developmental prosopagnosia. Cogn. Neuropsychol. 1–31.

Barton, J.J., Corrow, S.L., 2016. Selectivity in acquired prosopagnosia: the segregation of divergent and convergent operations. Neuropsychologia 83, 76–87.

Barton, J.J., Press, D.Z., Keenan, J.P., O'connor, M., 2002. Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. Neurology 58 (1), 71–78.

Barton, J.J., Hanif, H., Ashraf, S., 2009. Relating visual to verbal semantic knowledge: the evaluation of object recognition in prosopagnosia. Brain 132 (12), 3456–3466.

Bate, S., Cook, S.J., Duchaine, B., Tree, J.J., Burns, E.J., Hodgson, T.L., 2014. Intranasal inhalation of oxytocin improves face processing in developmental prosopagnosia. Cortex 50, 55–63.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., Dudfield, G., 2018. The limits of super recognition: an other-ethnicity effect in individuals with extraordinary face recognition skills. J. Exp. Psychol. Hum. Percept. Perform.

Behrmann, M., Plaut, D.C., 2012. Bilateral hemispheric processing of words and faces: evidence from word impairments in prosopagnosia and face impairments in pure alexia. Cereb. Cortex 24 (4), 1102–1118.

Behrmann, M., Avidan, G., 2005. Congenital prosopagnosia: face-blind from birth. Trends Cogn. Sci. 9 (4), 180–187.

Behrmann, M., Marotta, J., Gauthier, I., Tarr, M.J., McKeeff, T.J., 2005. Behavioral change and its neural correlates in visual agnosia after expertise training. J. Cogn. Neurosci. 17 (4), 554–568.

Bilalić, M., 2016. Revisiting the role of the fusiform face area in expertise. J. Cogn. Neurosci. 28 (9), 1345–1357.

Bilalić, M., Grottenthaler, T., Nägele, T., Lindig, T., 2014. The faces in radiological images: fusiform face area supports radiological expertise. Cereb. Cortex 26 (3), 1004–1014.

Bilalić, M., Langner, R., Ulrich, R., Grodd, W., 2011. Many faces of expertise: fusiform face area in chess experts and novices. J. Neurosci. 31 (28), 10206–10214.

Biotti, F., Gray, K.L., Cook, R., 2017. Impaired body perception in developmental prosopagnosia. Cortex 93, 41–49.

Bobak, A.K., Parris, B.A., Gregory, N.J., Bennetts, R.J., Bate, S., 2017. Eye-movement strategies in developmental prosopagnosia and "super" face recognition. Q. J. Exp. Psychol. 70 (2), 201–217.

Brants, M., Wagemans, J., Op de Beeck, H.P., 2011. Activation of fusiform face area by Greebles is related to face similarity but not expertise. J. Cogn. Neurosci. 23 (12), 3949–3958.

Brunner, J., Schimmack, U., 2016. How Replicable Is Psychology? A Comparison of Four Methods of Estimating Replicability on the Basis of Test Statistics in Original Studies.

Bukach, C.M., Gauthier, I., Tarr, M.J., Kadlec, H., Barth, S., Ryan, E., et al., 2012. Does acquisition of Greeble expertise in prosopagnosia rule out a domain-general deficit? Neuropsychologia 50 (2), 289–304.

Bukach, C.M., Peissig, J.J., 2010. How Faces Became Special. Perceptual Expertise: Bridging Brain and Behavior. pp. 11–39.

Bukach, C.M., Gauthier, I., Tarr, M.J., 2006. Beyond faces and modularity: the power of an expertise framework. Trends Cogn. Sci. 10 (4), 159–166.

Burns, E.J., Bennetts, R.J., Bate, S., Wright, V.C., Weidemann, C.T., Tree, J.J., 2017a. Intact word processing in developmental prosopagnosia. Sci. Rep. 7 (1), 1683.

Burns, E.J., Martin, J., Chan, A.H., Xu, H., 2017b. Impaired processing of facial happiness, with or without awareness, in developmental prosopagnosia. Neuropsychologia 102, 217–228.

Burns, E.J., Tree, J.J., Weidemann, C.T., 2014. Recognition memory in developmental prosopagnosia: electrophysiological evidence for abnormal routes to face recognition. Front. Hum. Neurosci. 8, 622.

Burns, E.J., Wilcockson, T.D., 2019. Alcohol usage predicts holistic perception: a novel paradigm to explore addiction. Addict. Behav.

Burns, E.J., Tree, J., Chan, A.H., Xu, H., 2018. Bilingualism shapes the other race effect. Vision Res.

Busey, T.A., Vanderkolk, J.R., 2005. Behavioral and electrophysiological evidence for configural processing in fingerprint experts. Vision Res. 45 (4), 431–448.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14 (5), 365.

Chan, M.P.S., Jones, C.R., Hall Jamieson, K., Albarracín, D., 2017. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol. Sci. 28 (11), 1531–1546.

Cohen, J., 1990. Things I have learned (so far). Am. Psychol. 45 (12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304.

Cohen, J., 1992. A power primer. Psychol. Bull. 112 (1), 155.

Collaboration, O. S, 2015. Estimating the reproducibility of psychological science. Science 349 (6251), aac4716.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., et al., 2018. Automatic imitation: a meta-analysis. Psychol. Bull. 144 (5), 453.

Crandall, C.S., Sherman, J.W., 2016. On the scientific superiority of conceptual replications for scientific progress. J. Exp. Soc. Psychol. 66, 93–99.

Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2013. Functional subdomains within human FFA. J. Neurosci. 33 (42), 16748–16766.

Op de Beeck, H.P., Baker, C.I., DiCarlo, J.J., Kanwisher, N.G., 2006. Discrimination training alters object representations in human extrastriate cortex. J. Neurosci. 26 (50), 13025–13036.

Duchaine, B., Germine, L., Nakayama, K., 2007. Family resemblance: ten family members with prosopagnosia and within-class object agnosia. Cogn. Neuropsychol. 24 (4), 419–430.

Eimer, M., 2011. The face-sensitive N170 component of the event-related brain potential. The Oxford Handbook of Face Perception 28. pp. 329–344.

Estudillo, A.J., Lee, J., Mennie, N., Burns, E.J., 2019. No Evidence of Other-Race Effect for Chinese Faces in Malaysian Non-Chinese Population. https://doi.org/10.31234/osf.io/8xg26.

Dalrymple, K.A., Oruc, I., Duchaine, B., Pancaroglu, R., Fox, C.J., Iaria, G., et al., 2011. The anatomic basis of the right face-selective N170 IN acquired prosopagnosia: a combined ERP/fMRI study. Neuropsychologia 49 (9), 2553–2563.

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., Lueschow, A., 2007. MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. Neuroimage 35 (4), 1495–1501.

D'Esposito M National Science Foundation and W.M. Keck Foundation, 2000. Special Issue Celebrating the Launching of the NSF/Keck Foundation National FMRI Data Center.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J. Exp. Psychol.: Gen. 115 (2), 107.

Duchaine, B., Nakayama, K., 2006. The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia 44 (4), 576–585.

Duchaine, B., Yovel, G., 2015. A revised neural framework for face processing. Ann. Rev. Vision Sci. 1, 393–416.

Editorial, 2000. A debate over fMRI data sharing. Nat. Neursci. 3, 845–846.

Ferguson, C.J., Heene, M., 2012. A vast graveyard of undead theories: publication bias and psychological science's aversion to the null. Perspect. Psychol. Sci. 7 (6), 555–561.

Farah, M.J., 1991. Cognitive neuropsychology: Patterns of co-occurrence among the associative agnosias: Implications for visual object representation. Cogn. Neuropsychol. 8 (1), 1–19.

Fecher, N., Johnson, E.K., 2019. Bilingual infants excel at foreign-language talker recognition. Developmental science 22 (4), e12778.

Fort, M., Ayneto-Gimeno, A., Escrichs, A., Sebastian-Galles, N., 2018. Impact of bilingualism on infants- ability to learn from talking and nontalking faces. Language Learning 68, 31–57.

Garrido, L., Furl, N., Draganski, B., Weiskopf, N., Stevens, J., Tan, G.C.Y., et al., 2009. Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. Brain 132 (12), 3443–3455.

Garrido, L., Duchaine, B., DeGutis, J., 2018. Association vs dissociation and setting appropriate criteria for object agnosia. Cogn. Neuropsychol. 35 (1-2), 55–58.

Gauthier, I., Bukach, C., 2007. Should we reject the expertise hypothesis? Cognition 103 (2), 322–330.

Gauthier, I., Curby, K.M., Skudlarski, P., Epstein, R.A., 2005. Individual differences in FFA activity suggest independent processing at different spatial scales. Cogn. Affect. Behav. Neurosci. 5 (2), 222–234.

Gauthier, I., Curran, T., Curby, K.M., Collins, D., 2003. Perceptual interference supports a non-modular account of face processing. Nat. Neurosci. 6 (4), 428.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3 (2), 191.

Gauthier, I., Tarr, M.J., 2002. Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. J. Exp. Psychol. Hum. Percept. Perform. 28 (2), 431.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. Nat. Neurosci. 2 (6), 568.

Gauthier, I., Tarr, M.J., Moylan, J., Anderson, A.W., Skudlarski, P., Gore, J.C., 2000a. Does visual subordinate-level categorisation engage the functionally defined fusiform face area? Cogn. Neuropsychol. 17 (1–3), 143–164.

Gauthier, I., 2017. The quest for the FFA led to the expertise account of its specialization. arXiv preprint arXiv:1702.07038.

Gauthier, I., Wong, A.C., Hayward, W.G., Cheung, O.S., 2006. Font tuning associated with expertise in letter perception. Perception 35 (4), 541–559.

Geskin, J., Behrmann, M., 2017. Congenital prosopagnosia without object agnosia? A literature review. Cogn. Neuropsychol. 1–51.

Gilaie-Dotan, S., Harel, A., Bentin, S., Kanai, R., Rees, G., 2012. Neuroanatomical correlates of visual car expertise. NeuroImage 62 (1), 147–153.

Gildersleeve, K., Haselton, M.G., Fales, M.R., 2014. Meta-Analyses and P-Curves Support Robust Cycle Shifts in Women's Mate Preferences: Reply to Wood and Carden (2014) and Harris, Pashler, and Mickes (2014).

Golby, A.J., Gabrieli, J.D., Chiao, J.Y., Eberhardt, J.L., 2001. Differential responses in the fusiform region to same-race and other-race faces. Nat. Neurosci. 4 (8), 845.

Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nat. Neurosci. 7 (5), 555.

Grill-Spector, K., Sayres, R., Ress, D., 2006. High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat. Neurosci. 9 (9), 1177.

Grüter, T., Grüter, M., Carbon, C.C., 2008. Neural and genetic foundations of face recognition and prosopagnosia. J. Neuropsychol. 2 (1), 79–97.

Hanson, S.J., Schmidt, A., 2011. High-resolution imaging of the fusiform face area (FFA) using multivariate non-linear classifiers shows diagnosticity for non-face categories. NeuroImage 54 (2), 1715–1734.

Harel, A., Gilaie-Dotan, S., Malach, R., Bentin, S., 2010. Top-down engagement modulates the neural expressions of visual expertise. Cereb. Cortex 20 (10), 2304–2318.
Harley, E.M., Pope, W.B., Villablanca, J.P., Mumford, J., Suh, R., Mazziotta, J.C., et al., 2009. Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. Cereb. Cortex 19 (11), 2746–2754.
Hartgerink, C.H., van Aert, R.C., Nuijten, M.B., Wicherts, J.M., Van Assen, M.A., 2016. Distributions of p-values smaller than. 05 in psychology: what is going on? PeerJ 4, e1935.
Heilman, M.K.M., Valenstein, E., 2010. Clinical Neuropsychology. Oxford University Press.
Hills, P.J., Cooper, R.E., Pake, J.M., 2013. Removing the own-race bias in face recognition by attentional shift using fixation crosses to diagnostic features: an eye-tracking study. Vis. Cogn. 21 (7), 876–898.
Hills, P.J., Lewis, M.B., 2006. Reducing the own-race bias in face recognition by shifting attention. Q. J. Exp. Psychol. 59 (6), 996–1002.
Hills, C.S., Pancaroglu, R., Duchaine, B., Barton, J.J., 2015. Word and text processing in acquired prosopagnosia. Ann. Neurol. 78 (2), 258–271.
Hills, P.J., Lewis, M.B., 2011. Reducing the own-race bias in face recognition by attentional shift using fixation crosses preceding the lower half of a face. Visual Cogn. 19 (3), 313–339.
Ioannidis, J.P., 2005. Why most published research findings are false. PLoS Med. 2 (8), e124.
James, T.W., James, K.H., 2013. Expert individuation of objects increases activation in the fusiform face area of children. NeuroImage 67, 182–192.
Kahneman, D., 2003. Experiences of collaborative research. Am. Psychol. 58 (9), 723.
Kanwisher, N., 2000. Domain specificity in face perception. Nat. Neurosci. 3 (8), 759.
Kanwisher, N., 2006. What's in a face? Science 311 (5761), 617–618.
Kanwisher, N., 2017. The quest for the FFA and where it led. J. Neurosci. 37 (5), 1056–1061.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.
Kanwisher, N., Yovel, G., 2006. The fusiform face area: a cortical region specialized for the perception of faces. Philos. Trans. Biol. Sci. 361 (1476), 2109–2128.
Krawczyk, D.C., Boggan, A.L., McClelland, M.M., Bartlett, J.C., 2011. The neural organization of perception in chess experts. Neurosci. Lett. 499 (2), 64–69.
Kandel, S., Burfin, S., Méary, D., Ruiz-Tada, E., Costa, A., Pascalis, O., 2016. The impact of early bilingualism on face recognition processes. Front. Psychol. 7, 1080.
Lakens, D., 2017. Professors are Not Elderly: Evaluating the Evidential Value of Two Social Priming Effects Through p-Curve Analyses.
Liu, J., Tian, J., Li, J., Gong, Q., Lee, K., 2009. Similarities in neural activations of face and Chinese character discrimination. Neuroreport 20 (3), 273–277.
Marshall, E., 2000. A ruckus over releasing images of the human brain. Science 289 (5484), 1458–1459.
Martens, F., Bulthé, J., van Vliet, C., de Beeck, H.O., 2018. Domain-general and domain-specific neural changes underlying visual expertise. NeuroImage 169, 80–93.
McGugin, R.W., Gatenby, J.C., Gore, J.C., Gauthier, I., 2012. High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. Proc. Natl. Acad. Sci. 109 (42), 17063–17068.
McGugin, R.W., Newton, A.T., Gore, J.C., Gauthier, I., 2014a. Robust expertise effects in right FFA. Neuropsychologia 63, 135–144.
McGugin, R.W., Ryan, K.F., Tamber-Rosenau, B.J., Gauthier, I., 2017. The role of experience in the face-selective response in right FFA. Cereb. Cortex 1–14.
McGugin, R.W., Van Gulick, A.E., Gauthier, I., 2016. Cortical thickness in fusiform face area predicts face and object recognition performance. J. Cogn. Neurosci. 28 (2), 282–294.
McGugin, R.W., Van Gulick, A.E., Tamber-Rosenau, B.J., Ross, D.A., Gauthier, I., 2014b. Expertise effects in face-selective areas are robust to clutter and diverted attention, but not to competition. Cereb. Cortex 25 (9), 2610–2622.
McKone, E., Kanwisher, N., 2005. 17 does the human brain process objects of expertise like faces? Paper Presented at the From Monkey Brain to Human Brain: A Fyssen Foundation Symposium.
McKone, E., Kanwisher, N., Duchaine, B.C., 2007. Can generic expertise explain special processing for faces? Trends Cogn. Sci. 11 (1), 8–15.
McKone, E., Robbins, R., 2011. Are faces special. Oxford Handbook of Face Perception. pp. 149–176.
Meissner, C.A., Brigham, J.C., 2001. Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol. Public Policy Law 7 (1), 3.
Mellers, B., Hertwig, R., Kahneman, D., 2001. Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. Psychol. Sci. 12 (4), 269–275.
Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M.H., MacSweeney, M., 2019. Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. Dev. Sci. 22 (1), e12701.
Moore, C.D., Cohen, M.X., Ranganath, C., 2006. Neural mechanisms of expert skills in visual working memory. J. Neurosci. 26 (43), 11187–11196.
Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., Du Sert, N.P., et al., 2017. A manifesto for reproducible science. Nat. Hum. Behav. 1 (1), 0021.
Murray, S.O., Wojciulik, E., 2004. Attention increases neural selectivity in the human lateral occipital complex. Nat. Neurosci. 7 (1), 70.
Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., et al., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nat. Neurosci. 20 (3), 299.
Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. 115 (11), 2600–2606.
de Beeck, H.P.O., Baker, C.I., DiCarlo, J.J., Kanwisher, N.G., 2006d. Discrimination training alters object representations in human extrastriate cortex. J. Neurosci. 26 (50), 13025–13036.
Pashler, H., Wagenmakers, E.J., 2012. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect. Psychol. Sci. 7 (6), 528–530.
Passingham, R.E., Stephan, K.E., Kötter, R., 2002. The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3 (8), 606.
Poldrack, R.A., Gorgolewski, K.J., 2014. Making big data open: data sharing in neuroimaging. Nat. Neurosci. 17 (11), 1510.
Poline, J.B., Breeze, J.L., Ghosh, S.S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., et al., 2012. Data sharing in neuroimaging research. Front. Neuroinform. 6, 9.
Rezlescu, C., Barton, J.J., Pitcher, D., Duchaine, B., 2014. Normal acquisition of expertise with greebles in two cases of acquired prosopagnosia. Proc. Natl. Acad. Sci. 111 (14), 5123–5128.
Rhodes, G., Byatt, G., Michie, P.T., Puce, A., 2004. Is the fusiform face area specialized for faces, individuation, or expert individuation? J. Cogn. Neurosci. 16 (2), 189–203.
Righi, G., Tarr, M.J., Kingon, A., 2013. Category-selective recruitment of the fusiform gyrus with chess expertise. In: Staszewski, J.J. (Ed.), Carnegie Mellon Symposia on Cognition. Expertise and Skill Acquisition: The Impact of William G. Chase. Psychology Press, New York, NY, US, pp. 261–280.
Ritchie, S.J., Tucker-Drob, E.M., 2018. How much does education improve intelligence? A meta-analysis. Psychol. Sci. 29 (8), 1358–1369.
Robbins, R., McKone, E., 2007. No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103 (1), 34–79.
Roberts, D.J., Ralph, M.A.L., Kim, E., Tainturier, M.J., Beeson, P.M., Rapcsak, S.Z., Woollams, A.M., 2015. Processing deficits for familiar and novel faces in patients with left posterior fusiform lesions. Cortex 72, 79–96.
Rorden, C., Karnath, H.O., 2004. Using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat. Rev. Neurosci. 5 (10), 812.
Ross, D.A., Tamber-Rosenau, B.J., Palmeri, T.J., Zhang, J., Xu, Y., Gauthier, I., 2018. High-resolution functional magnetic resonance imaging reveals configural processing of cars in right anterior fusiform face area of Car experts. J. Cogn. Neurosci. (Early Access) 1–12.
Rubino, C., Corrow, S.L., Corrow, J.C., Duchaine, B., Barton, J.J.S., 2016. Word and text processing in developmental prosopagnosia. Cogn. Neuropsychol. 33 (5–6), 315–328. https://doi.org/10.1080/02643294.2016.1204281.
Sala, G., Gobet, F., 2017. Working memory training in typically developing children: a meta-analysis of the available evidence. Dev. Psychol. 53 (4), 671.
Schimmack, U., Brunner, J., 2017. Z-Curve.
Schmalzl, L., Palermo, R., Coltheart, M., 2008. Cognitive heterogeneity in genetically based prosopagnosia: a family study. J. Neuropsychol. 2 (1), 99–117.
Shrout, P.E., Rodgers, J.L., 2018. Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. Annu. Rev. Psychol. 69, 487–510.
Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22 (11), 1359–1366.
Simons, D.J., 2014. The value of direct replication. Perspect. Psychol. Sci. 9 (1), 76–80.
Simmons, J.P., Simonsohn, U., 2017. Power posing: P-curving the evidence. Psychol. Sci. 28 (5), 687–693.
Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014a. P-curve and effect size: correcting for publication bias using only significant results. Perspect. Psychol. Sci. 9 (6), 666–681.
Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014b. P-curve: a key to the file-drawer. J. Exp. Psychol. Gen. 143 (2), 534.
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).
Singh, L., Quinn, P.C., Xiao, N.G., Lee, K., 2019. Monolingual but not bilingual infants demonstrate racial bias in social cue use. Dev. Sci. e12809.
Slotnick, S.D., White, R.C., 2013. The fusiform face area responds equivalently to faces and abstract shapes in the left and central visual fields. Neuroimage 83, 408–417.
Song, S., Garrido, L., Nagy, Z., Mohammadi, S., Steel, A., Driver, J., et al., 2015. Local but not long-range microstructural differences of the ventral temporal cortex in developmental prosopagnosia. Neuropsychologia 78, 195–206.
Steffens, N.K., Haslam, S.A., Schuh, S.C., Jetten, J., van Dick, R., 2017. A meta-analytic review of social identification and health in organizational contexts. Personal. Soc. Psychol. Rev. 21 (4), 303–335.
Sigurdardottir, H.M., Fridriksdottir, L.E., Gudjonsdottir, S., Kristjánsson, Á., 2018. Specific problems in visual cognition of dyslexic readers: face discrimination deficits predict dyslexia over and above discrimination of scrambled faces and novel objects. Cognition 175, 157–168.
Sigurdardottir, H.M., Hjartarson, K.H., Gudmundsson, G.L., Kristjánsson, Á., 2019. Own-race and other-race face recognition problems without visual expertise problems in dyslexic readers. Vision Res. 158, 146–156.
Sigurdardottir, H.M., Ívarsson, E., Kristinsdóttir, K., Kristjánsson, Á., 2015. Impaired recognition of faces and objects in dyslexia: Evidence for ventral stream dysfunction? Neuropsychology 29 (5), 739.
Starrfelt, R., Klargaard, S.K., Petersen, A., Gerlach, C., 2018. Reading in developmental prosopagnosia: evidence for a dissociation between word and face recognition. Neuropsychology 32 (2), 138.
Susilo, T., Duchaine, B., 2013. Dissociations between faces and words: comment on Behrmann and Plaut. Trends Cogn. Sci. 17 (11), 545.
Susilo, T., Wright, V., Tree, J.J., Duchaine, B., 2015. Acquired prosopagnosia without word recognition deficits. Cogn. Neuropsychol. 32 (6), 321–339.
Toga, A.W., 2002. Neuroimage databases: the good, the bad and the ugly. Nat. Rev. Neurosci. 3 (4), 302.
Ulrich, R., & Miller, J. (2015). p-hacking by post hoc selection with multiple

opportunities: Detectability by skewness test?: Comment on Simonsohn, Nelson, and Simmons (2014).
van Aert, R.C., Wicherts, J.M., van Assen, M.A., 2016. Conducting meta-analyses based on p values: reservations and recommendations for applying p-uniform and p-curve. Perspect. Psychol. Sci. 11 (5), 713–729.
Van Assen, M.A., van Aert, R., Wicherts, J.M., 2015. Meta-analysis using effect size distributions of only statistically significant studies. Psychol. Methods 20 (3), 293.
Van Belle, G., Busigny, T., Lefèvre, P., Joubert, S., Felician, O., Gentile, F., Rossion, B., 2011. Impairment of holistic face perception following right occipito-temporal damage in prosopagnosia: converging evidence from gaze-contingency. Neuropsychologia 49 (11), 3145–3150.
Van Belle, G., Lefèvre, P., Laguesse, R., Busigny, T., De Graef, P., Verfaillie, K., Rossion, B., 2010. Feature-based processing of personally familiar faces in prosopagnosia: evidence from eye-gaze contingency. Behav. Neurol. 23 (4), 255–257.
Van Horn, J.D., Gazzaniga, M.S., 2013. Why share data? Lessons learned from the fMRIDC. Neuroimage 82, 677–682.
Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., Albarracín, D., 2016. From primed concepts to action: a meta-analysis of the behavioral effects of incidentally presented words. Psychol. Bull. 142 (5), 472.
Wicherts, J.M., Veldkamp, C.L., Augusteijn, H.E., Bakker, M., Van Aert, R., Van Assen, M.A., 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. Front. Psychol. 7, 1832.
Williams, M.A., McGlone, F., Abbott, D.F., Mattingley, J.B., 2005. Differential amygdala responses to happy and fearful facial expressions depend on selective attention. Neuroimage 24 (2), 417–425.
Wixted, J.T., Wells, G.L., 2017. The relationship between eyewitness confidence and identification accuracy: A new synthesis. Psychol. Sci. Public Interest 18 (1), 10–65.
Wojciulik, E., Kanwisher, N., Driver, J., 1998. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. J. Neurophysiol. 79 (3), 1574–1578.
Wong, A.C.N., Gauthier, I., 2007. An analysis of letter expertise in a levels-of-categorization framework. Visual Cogn. 15 (7), 854–879.
Wong, A.C.-N., Jobard, G., James, K.H., James, T.W., Gauthier, I., 2009. Expertise with characters in alphabetic and nonalphabetic writing systems engage overlapping occipito-temporal areas. Cogn. Neuropsychol. 26 (1), 111–127.
Wong, Y.K., Gauthier, I., 2010a. Holistic processing of musical notation: dissociating failures of selective attention in experts and novices. Cogn. Affect. Behav. Neurosci. 10 (4), 541–551.
Wong, Y.K., Gauthier, I., 2010b. A multimodal neural network recruited by expertise with musical notation. J. Cogn. Neurosci. 22 (4), 695–713.
Xu, Y., 2005. Revisiting the role of the fusiform face area in visual expertise. Cereb. Cortex 15 (8), 1234–1242.
Xu, Y., Liu, J., Kanwisher, N., 2005. The M170 is selective for faces, not for expertise. Neuropsychologia 43 (4), 588–597.
Yin, R.K., 1969. Looking at upside-down faces. J. Exp. Psychol. 81 (1), 141.
Yovel, G., Kanwisher, N., 2004. Face perception: domain specific, not process specific. Neuron 44 (5), 889–898. https://doi.org/10.1016/j.neuron.2004.11.018.
Yue, X., Tjan, B.S., Biederman, I., 2006. What makes faces special? Vision Res. 46 (22), 3802–3811.
Zachariou, V., Safiullah, Z.N., Ungerleider, L.G., 2018. The fusiform and occipital face areas can process a nonface category equivalently to faces. J. Cogn. Neurosci. 30 (10), 1499–1516.
Zhang, J., Liu, J., Xu, Y., 2015. Neural decoding reveals impaired face configural processing in the right fusiform face area of individuals with developmental prosopagnosia. J. Neurosci. 35 (4), 1539–1548.