Individual differences in visual science: What can be learned and what is good experimental practice?

John D. Mollon[a], Jenny M. Bosten[b], David H. Peterzell[c], Michael A. Webster[d,*]

[a] Department of Psychology, University of Cambridge, United Kingdom
[b] School of Psychology, University of Sussex, United Kingdom
[c] College of Psychology, John F. Kennedy University, United States
[d] Department of Psychology and Graduate Program in Integrative Neuroscience, University of Nevada, Reno, United States

ARTICLE INFO

ABSTRACT

We all pass out our lives in private perceptual worlds. The differences in our sensory and perceptual experiences often go unnoticed until there emerges a variation (such as 'The Dress') that is large enough to generate different descriptions in the coarse coinage of our shared language. In this essay, we illustrate how individual differences contribute to a richer understanding of visual perception, but we also indicate some potential pitfalls that face the investigator who ventures into the field.

1. Introduction

Visual science continues to generate an enormous body of empirical data on the characteristics and mechanisms of visual processing. Most such studies are designed to test different observers under nominally the same conditions, to understand the effects of those conditions and their implications for underlying processes. Multiple observers are included to ensure that the results are general, for example, to confirm that the findings can be replicated with naïve observers who are unaware of the aims of the study. The use of multiple observers also ensures that the results are significant and reliable. The data from different observers provide the estimate of measurement error. In this regard, the differences between observers are treated as a nuisance factor to be ignored – as mere noise in the measurements. And in very many studies these differences have accordingly lain unexploited.

However, the patterns of variations between observers are often systematic, and often arise from real differences in the very optical, neural, and cognitive processes that mediate the perceptions that the researchers are interested in. In this regard, individual differences provide a largely unmined treasure trove of information about these processes (de-Wit & Wagemans, 2016; Peterzell, 2016; Wilmer, 2008). Today, visual scientists have available to them extensive collections of archival data from recent times and from decades past. These data often

include individual variability that is reported but left unexamined. Compared to many areas of experimental psychology, these psychophysical data can have very low measurement (intra-observer) error, and in many cases can be evaluated against precisely quantified properties or well-defined models of the visual system. They therefore promise powerful new insights. As new questions are pursued, there is also potential for experimental designs that yield richer information by exploiting inter-observer variation, and visual scientists are increasingly turning to studies focused on measuring and analysing the visual differences between observers.

Yet all datasets also include variations that are not of interest to the experimenter, that reflect random noise or introduce confounds unrelated to the tested hypotheses (see §5 below). These spurious variations may mask or impersonate the target of inquiry. As we discuss below, these confounds can be especially problematic in studies of individual differences. Thus investigators must mine individual variations cautiously, or they risk lining their pockets with fool's gold. We hope this review will both highlight the power of individual differences in vision research, and provide some prescriptions for best practices.

2. Differences between individuals

Individual differences can be defined and interpreted in a variety of

---

ways. One important distinction is between 'individual differences *in data*', and '*true* individual differences'. The former phrase refers to differences obtained in actual measurements, and these can arise from real differences between individuals, but also from measurement error (e.g. from random variability, systematic biases, instrumental variation, and more). 'True individual differences' refers to variability that remains after the effect of measurement error has been excluded. It is a hypothetical construct, and an aspirational goal of measurement. In other definitions, 'true individual differences' involve more than zero measurement error, because they include only variability that is intrinsic to individuals, reflecting differences in a trait and not merely a state. As we discuss below, these distinctions are important for deciding how to design and interpret experiments probing observer differences.

When individuals who differ from others in a consistent way are categorized as a group, individual differences lead to group differences. Very many studies have investigated such group differences. Much of clinical vision is concerned with understanding disease by comparing control individuals to different patient populations; studies of development or lifespan compare individuals grouped by age; and studies on demographic factors might classify participants by gender or ethnicity. In designing these studies, individuals are often classified by predefined criteria, and results are then analysed in terms of the discrete groups. This has of course been a very fruitful approach, but can miss opportunities for a richer understanding of the observer differences because the within-group differences are again treated as noise. Research on true individual differences treats each observer as an individual. This can and often does still allow for observers to be classified by different criteria (e.g. according to their age or visual acuity), but importantly, also allows measurements to be analysed when the relevant classification is unknown (e.g. in the case of a set of observers all defined as 'normal' on some assessment).

The distinction between groups and individuals also has important implications for how a visual process is characterized or modelled. Colour science, and especially applied fields like colorimetry, rely heavily on the concept of a standard observer, defined by the average of the measurements for a large number of individuals. Similar models have been developed for other visual attributes such as spatial sensitivity (Watson & Ahumada, 2005). The standard observer provides an important working assumption for studying or predicting visual performance, but also has important limits, since it may rarely describe the properties of an actual observer. As we note below, for some applications the standard observer is of little value because it does not allow sufficient specification of the impact of the stimulus. Moreover, the mean alone provides no information on the range of tolerances that might be acceptable, say, to a given proportion of the population for an application like colour rendering. New observer models are being developed that explicitly incorporate estimates of normal variation in colour vision to better predict how a given individual or group might experience colour (Asano, Fairchild, & Blondé, 2016).

3. Sources of individual differences in vision

Variations in visual processing arise from many sources and are likely to be a prevalent characteristic at all levels of visual coding and all stages of the visual pathways. Even in the very first steps of image formation there are large, stable, and consequential variations in the optical aberrations of the eye, which affect the quality and form of the individual's idiosyncratic retinal image (Castejon-Mochon, Lopez-Gil, Benito, & Artal, 2002; Porter, Guirao, Cox, & Williams, 2001).

Colour vision is a case where patterns of individual differences have been extensively characterized (Webster, 2015b). The eye's optics differ in spectral quality, owing to pigment in the crystalline lens that screens light of shorter wavelengths. The density of the lens pigment varies markedly across observers and also increases steadily with age (Pokorny, Smith, & Lutze, 1987; Weale, 1988; Werner, 1982). Similarly, observers vary widely in the density of the macular pigment screening

the central fovea (Bone & Sparrock, 1971; Werner, Donnelly, & Kliegl, 1987. These pre-receptoral filters strongly bias the spectrum of the light reaching the photoreceptors and are in fact the primary source of inter-observer variations in colour matching (Webster & MacLeod, 1988). Moreover, the spectral sensitivities of the cone photoreceptors vary reliably in the positions of their peaks (λ_{max}) (Winderickx et al., 1992) and in their bandwidths (e.g. because of variations in optical density) (Wyszecki & Stiles, 1980). As is well known from studies of colour deficiencies, there can also be large and diverse differences in both the number and nature of the cone types (Neitz & Neitz, 2011). Also, there are striking differences in the relative numbers of different cone types. For example, it is often noted that there are on average twice as many L cones as M cones in humans, yet in individuals with normal colour vision the ratio of L to M cones has been reported to vary from 1:1 to 16.5:1 (Hofer, Carroll, Neitz, Neitz, & Williams, 2005).

There are also large and reliable individual differences in subjective judgments of colour, i.e. in how colours are reported or categorized. The stimulus spectra that observers describe as unique hues (pure red, green, blue, or yellow), or that they experience as achromatic, vary widely from one observer to the next (Bosten, Beer, & MacLeod, 2015; Kuehni, 2004; Webster, Miyahara, Malkoc, & Raker, 2000b). Moreover, there are very large differences in the patterns of colour naming. Anthropological studies of colour naming have focused on cross-linguistic differences in order to understand the aetiology of colour categories and whether they are more strongly determined by universal (e.g. biological) or relative (e.g. cultural) processes (Kay & Regier, 2006). However, these analyses have tended to overlook the enormous variations in colour naming within a language. A re-examination of the World Colour Survey found that individuals varied widely in their patterns of colour naming and that these basic 'motifs' were often more similar across speakers from different languages than among members from the same linguistic group (Lindsey & Brown, 2009). Recent analyses have also pointed to the importance of characterizing individual differences for understanding the representation of colour in a culture. For example, some languages are characterized by few colour terms and high levels of uncertainty at the level of the individual, yet include a rich parcellation of colour at the level of the society (Lindsey, Brown, Brainard, & Apicella, 2015).

A further important source of variation in colour vision – and indeed all vision – is variation in the observer's environment. While natural visual environments have many characteristic properties that are thought to have shaped visual coding (Geisler & Ringach, 2009; Simoncelli & Olshausen, 2001), the world also varies across both space and time. For example, observers are exposed to very different colours in lush or arid environments, and colours in the same location can cycle with the seasons (Webster, Mizokami, & Webster, 2007; Webster & Mollon, 1997). Similarly, the diet of faces experienced by an individual varies widely depending on his or her social environment. Vision routinely adapts to the prevailing stimulus characteristics of the environment (Webster, 2015a). Potential examples of such contextual effects are seasonal changes in colour appearance (Welbourne, Morland, & Wade, 2015) or 'other-race' effects in the perception of faces (Meissner & Brigham, 2001).

As the forgoing examples suggest, the causes of individual differences in vision are many. Some can be highly stable and tied directly to genes. Others depend on lifestyle and experience. For example, age-related changes in lens pigment density are largely a consequence of exposure to light (Lindsey & Brown, 2002), while the density of macular pigment (consisting of the retinal carotenoids lutein and zeaxanthin) varies with the amount of carotenoids in the individual's diet (Hammond et al., 1997). The sources of differences can also be intricately intertwined. For instance, an indirect genetic effect on macular pigment density could arise if polymorphism of the taste receptors mediated differences in diet, leading to a knock-on effect on macular pigment and colour vision. Similarly, an individual's culture or profession will determine the distribution of colours or faces he or she is

exposed to, while an individual's height, determined genetically or environmentally, could have an indirect effect on visual perception by affecting visual input such as patterns of optic flow. These interactions complicate the interpretation of the individual differences measured in any task, but also suggest that such differences are a rich resource for exploring different processes and their interplay.

4. The value of research on individual differences

In this section we illustrate some of the ways that individual differences have been utilized in vision research.

4.1. Selection for particular tasks

Often an interest in individual differences has had its origin in practical considerations, in the need to select individuals who were either particularly gifted or particularly weak in some perceptual skill. A celebrated example was the proliferation of tests for colour deficiency following the Lagerlunda disaster and the subsequent introduction of screening for railway employees in both Europe and America (Holmgren, 1877; Mollon & Cavonius, 2012; Stilling, 1877). In modern times, formal tests have revealed the large variation in the ability to recognise faces (Duchaine & Nakayama, 2006; Russell, Duchaine, & Nakayama, 2009; Wilmer, 2017) and such tests recommend themselves for the selection of border control officials and those whose profession requires them to recognise many individuals (Robertson, Noyes, Dowsett, Jenkins, & Burton, 2016; White, Kemp, Jenkins, Matheson, & Burton, 2014). The development of visual tasks that are sensitive and selective enough to measure specific abilities is discussed in Section 4.7.

4.2. Comparison of populations

As noted above, the most commonly studied individual differences in visual traits are group differences. There have been many reports of visual differences related to sex (e.g. Held, 1989; McGuinness, 1976) and age (e.g. Owsley, 2011; Werner, Peterzell, & Scheetz, 1990). Famously, it has been claimed that individuals with autism spectrum disorder differ from controls on many visual measures (for a review see Simmons et al., 2009) such as finding embedded figures (e.g. Jolliffe & Baron-Cohen, 1997), and sensitivity to coherent motion (e.g. Bertone, Mottron, Jelenic, & Faubert, 2003), though many other suggested differences have not consistently replicated (e.g. Happe & Frith, 2006). Similar characteristic differences have been reported for other patient groups such as people with schizophrenia (Butler, Silverstein, & Dakin, 2008). Though the classification of patient and control groups is categorical, some 'group' differences simply represent a categorical boundary in what is likely to be a continuously distributed set of traits between individuals. This is true for many visual deficits, which vary in severity and progression. It may also be true for many psychiatric syndromes. Autism, for example, is recognized as a spectrum of disorders. Even when clear grouping is evident, there can be blurred lines in the category boundaries. For example, sex is dimorphic, but could in some circumstances be a red herring (for example, if differences in male and female brains are a more relevant correlate of variation due to sex in visual tasks than is the presence or absence of a Y chromosome).

A potentially valuable application of individual differences in vision for clinical questions is their use as endophenotypes (Gottesman & Gould, 2003; Ritsner & Gottesman, 2009), where a visual trait is associated with a clinical condition but is more simply related to the causal mechanisms of the condition than is the broader clinical phenotype. For example, it has been suggested that smooth pursuit is an endophenotype for schizophrenia (Calkins & Iacono, 2000; Allen, Griss, Folley, Hawkins, & Pearlson, 2009). Investigating the neurological and genetic basis of individual differences in a candidate endophenotype may provide insights into the biological basis of a clinical condition more easily than attempting to correlate the condition itself with

biological variation.

4.3. Tailoring stimuli for the individual

An experimental design incorporating individual differences can be important even when the differences themselves are not of interest, for example to control for unwanted confounds. It has become standard practice in colour science to correct for the luminance sensitivity of the individual observer (Kaiser, 1988). This is necessary in order to isolate mechanisms that respond only to the chromatic information in the stimulus, for sensitivity to luminance contrast can be very high and thus a residual luminance error could dominate the performance. Photopic luminance sensitivity depends on the summed responses of the L and M cones, and as we noted above both the spectral sensitivities and the cone ratios vary widely in the normal population, so that mismatches based on a standard observer cannot be ignored (Lennie, Pokorny, & Smith, 1993). Moreover, the luminance match can also depend on the task and conditions such as the observer's state of adaptation (Webster & Mollon, 1993). Accordingly, a number of techniques have been devised to equate the sensation luminance of different chromaticities (e.g. Anstis & Cavanagh, 1983). Related approaches have been applied – perhaps less often than they should be - to eliminate confounding factors in other properties of colour vision (e.g. to isolate different chromatic mechanisms; Webster, Miyahara, Malkoc, & Raker, 2000a; Smithson, Sumner, & Mollon, 2003) or to control for other visual variations. For example, the aging eye suffers both optical and neural changes. To isolate the latter, it is important to define stimuli that are equated at the level of the retinal image (Werner, Bieber, & Schefrin, 2000). Similarly, studies sometimes adjust stimuli in terms of multiples of an individual observer's thresholds in order to equate the visibility of the patterns (Delahunt, Hardy, & Werner, 2008). However, such corrections are not currently the norm, and we do not know what the consequences might be for not making them. For example, studies of face perception often aim to explore attributes such as identity or expression. Yet we do not currently have a good understanding of the stimulus variations that the visual system uses to encode these dimensions in a standard observer, let alone an individual. This could potentially complicate the interpretation of a study, for the stimuli may fail to isolate appropriately the specific mechanisms targeted for study.

4.4. Identifying visual mechanisms

As we have noted, one of the most important applications of individual differences is as a means to unearth fundamental mechanisms of vision (Peterzell, 2016; Wilmer, 2008). The finest example may be the derivation of the spectral sensitivities of the retinal cones from the colour matches of normal and dichromatic observers (König & Dieterici, 1886) (Fig. 1). Later, in the twentieth century, it was again the existence of variant forms of colour vision that allowed Nathans, Thomas, and Hogness (1986) to identify the genes encoding the L and M photopigments – work that led to our detailed modern understanding of the opsin array on the X-chromosome. It was soon found that a polymorphism of the 180th codon of the L opsin gene was correlated with the proportion of red light that normal observers need in a Rayleigh match (Winderickx et al., 1992): those with the version encoding serine, rather than alanine require less red light to match a monochromatic orange light of 590 nm. This remains a remarkable result: the smallest possible genetic difference, a difference of one nucleotide, means that different individuals unknowingly spend their lives in different phenomenological worlds.

4.5. Correlating measurements

Individual differences designs typically rely on correlational analyses. In several cases, the pattern of correlations between individuals on a battery of tests has suggested which aspects of the stimulus domain
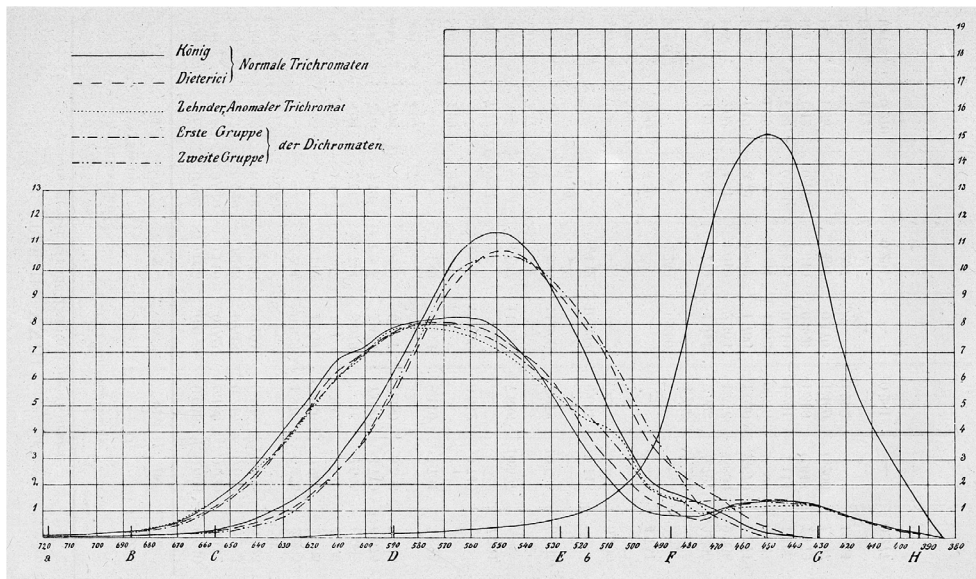
Fig. 1. König's estimates of cone sensitivities (König & Dieterici, 1886). The spectrum is plotted with wavelength decreasing to the right, and the capital letters indicate Fraunhofer lines. The solid lines represent König's own long-wave, middle-wave and short-wave sensitivities and the dashed lines those of his collaborator, Dieterici (whose middle-wave sensitivity differs from König's). Other curves are for anomalous and dichromatic observers.

are processed by independent mechanisms. Thus Peterzell and Teller (2000) showed that there were strong correlations between spatial contrast sensitivities for equiluminant red-green gratings of a range of spatial frequencies, but that these sensitivities were less systematically correlated with sensitivities for black-yellow gratings, suggesting independent channels for chromatic and achromatic gratings. Wilmer and Nakayama (2007) inferred from individual differences that two different components of ocular tracking were driven by different mechanisms of motion perception: pre-saccadic acceleration was correlated with the precision of low-level (motion-energy-based) estimates of speed, whereas the precision of post-saccadic pursuit was correlated with the precision of an observer's speed estimates for high-level (position tracking) motion. Similarly, the generalised concept of 'magno-cellular function' – a concept frequently invoked with respect to dyslexia, development disorders, and schizophrenia – was challenged by Goodbourn and colleagues (2012) who tested 1054 healthy participants and found little correlation between two of the most favoured measures of magnocellular function.

4.5.1. Correlational associations and dissociations and phenotypic variability

The correlational method may also help bridge 'explanatory gaps' between different levels of description for particular traits. A particular behaviour must be determined or influenced at many biological levels, including genetic, molecular, neurochemical, anatomical and systems levels. Correlating individual variation in a behavioural phenotype with variation at these diverse biological levels can elucidate the connection between mechanisms at different levels and the behaviour.

An example of this approach is the investigation of the neuroanatomical basis of traits in visual perception. For example, Duncan and Boynton (2003), found that the cortical magnification factor in primary visual cortex correlates with visual acuity. Recent work (see Kanai & Rees, 2011 for a review) has revealed correlations between brain structure and biological motion detection (Gilaie-dotan, Kanai, Bahrami, Rees, & Saygin, 2013), rate of perceptual alternation in binocular rivalry (Kanai, Bahrami, & Rees, 2010), motion assimilation and contrast (Takeuchi, Yoshimoto, Shimada, Kochiyama, & Kondo, 2017), susceptibility to geometrical illusions (Axelrod, Schwarzkopf, Gilaie-dotan, & Rees, 2017; Schwarzkopf, Song, & Rees, 2011), susceptibility to the tilt illusion (Song, Schwarzkopf, & Rees, 2013b; Song et al., 2013a), and orientation discrimination (Song et al., 2013b). The *lack* of correlations between some behavioural measurements has also drawn attention to explanatory gaps that might have gone unnoticed. An important example is the relationship between colour appearance and chromatic sensitivity. The large documented differences in spectral sensitivity often have surprisingly little effect on how people report their colour sensations. Thus young and old observers choose similar stimuli for white or for different hues despite viewing the stimuli through lenses that filter the spectrum in very different ways (Hardy, Frederick, Kay, & Werner, 2005; Wuerger, Xiao, Fu, & Karatzas, 2010; Werner & Schefrin, 1993). Moreover, the unique hues show little dependence on the large variations in cone ratios (Brainard et al., 2000; Jordan & Mollon, 1997; Miyahara, Pokorny, Smith, Baron, & Baron, 1998). Unique hue settings are also uncorrelated with one another (i.e. the choices for blue and yellow vary independently across observers), yet most factors affecting retinal spectral sensitivities have broad effects across the light spectrum and thus should produce covarying changes in unique hue settings (Webster et al., 2000b). The fact that chromatic sensitivity and colour appearance can vary independently has been important to the debate over the nature of the unique hues, and to whether the hues more closely reflect special properties of the observer or special properties of the environment (Mollon & Jordan, 1997).

Dissociations between sensitivity and appearance are evident for many other visual attributes, for example in the differences between subjective focus and visual acuity (Sawides, de Gracia, Dorronsoro, Webster, & Marcos, 2011), between threshold sensitivity and supra-threshold contrast (Georgeson & Sullivan, 1975), and in the relatively uniform appearance across the visual field despite the large changes of visual sensitivity with increasing eccentricity (Webster, Halen, Meyers, Winkler, & Werner, 2010). These dissociations reveal the presence of normalization processes that compensate or correct for many of the sensitivity limits of the observer, and thus across observers (Webster, 2015a). At the outset we noted that routine variations in the visual system place us each in different perceptual worlds, and this can be easy to demonstrate, for it is possible to choose stimuli that are discriminable for one person while indistinguishable to another. However, while less easy to verify, there is probably also a remarkable sense in which our percepts are actually much more similar than differences in the optical and neural mechanisms mediating those percepts might predict, because many aspects of perception are calibrated to discount the properties and limitations of the observer. A telling example is the #thedress. This image – which some insisted appeared blue-black while others white-gold - generated worldwide interest because it exposed how little we can infer from our own percepts about the visual experience of others (Brainard & Hurlbert, 2015). However, the public might also be (or at least should be!) as surprised at how similarly we

describe the colours in this image, given that we each view the world through such markedly different eyes. Again, the similarities (as much as the differences) in our percepts is an insight that owes much to the study of individual differences. That is, it is from individual differences that we have learnt much of what we know about how visual perception is normalized to compensate for the observer's unique visual apparatus.

Correlational approaches have also been used to predict associations and dissociations across multiple dimensions or domains. For example, Dobkins, Gunther, and Peterzell (2000) determined that measures of luminance contrast sensitivity, chromatic contrast sensitivity, and luminance spectral sensitivity were mostly uncorrelated across the three types of measures, but positively correlated within a type of measure. In another example, Wilmer and Backus (2008) found that variability in self-reported autostereogram skill predicts stereoacuity. Other studies examine whether or not variability in visual tasks predict non-visual abilities or performance. Orlansky et al. (2015) found that variability in astigmatism measured in preschool-aged children predicted various forms of reduced academic readiness in various developmental and educational domains, and Wilmer and Buchanan (2009) found that nearpoint phorias after near work predict ADHD symptoms in college students.

4.5.2. Behavioural genetics and heritability of visual traits

Correlations between genetic variability and individual differences in vision have been studied in two broad ways.

The first, which originated with Galton (1883), involves attempts to separate the cause of visual variation among individuals into genetic versus environmental components. The most common and well-established research methodologies are family studies, twin studies, and adoption studies: For example, concordances between identical twins compared to fraternal twins and other siblings provide measures of heritability (Fuller & Thompson, 1960; Knopik, Neiderhiser, DeFries, & Plomin, 2017). Recently, researchers have examined heritability of visual processing using modern behavioural-genetic methods (Wilmer, 2008). Examples of this work include studies reporting strong and significant heritability for strabismus, human face recognition, and refractive error, but weak heritability for phoria or aesthetic preferences for faces (Germine et al., 2015; Sanfilippo et al., 2012; Wilmer & Backus, 2009; Wilmer, Germine, Chabris, et al., 2010; Wilmer, Germine, Loken, et al., 2010; Yovel, Wilmer, & Duchaine, 2014).

A second type of investigation seeks genetic associations that link specific genes or other genomic variations to visual processing. Behavioural geneticists credit vision scientists with initiating the search for such associations (Greenspan, 2008), in early visual studies of phototaxis and octomotor responses in flies (Brown & Hall, 1936; Gavel, 1939; Kalmus, 1943; Hecht & Wald, 1934). As we noted above, advances in molecular genetics made it possible to identify and sequence the photopigment genes (e.g., Nathans, Piantanida, Eddy, Shows, & Hogness, 1986; Nathans et al., 1986). Genetic variation in DNA has been correlated with individual variation in visual perceptual and neural traits, using either a candidate-gene approach or genome-wide association (e.g. Goodbourn et al., 2014 on contrast sensitivity; Bosten et al., 2014 on phorias; Verhallen et al., 2014 on face perception; see also Mollon, 1986). New genetic methods such as pathway analysis (Torkamani, Topol, & Schork, 2008) can be similarly applied. Correlational studies have also been applied to investigate the genetic basis of the visual neural apparatus (e.g. Bakken, Roddey, Djurovic, Akshoomoff, & Amaral, 2012; for a review see Gu & Kanai, 2014).

4.6. Factor analytic approaches

The correlational approach to studying visual mechanisms can be especially powerful when the observed measurements are made for a large set of stimuli or tasks, and particularly for systematic sets of stimuli that differ quantitatively along a stimulus dimension, such as wavelength or spatial frequency. In such cases, individual differences for nearby stimulus levels are likely to show stronger correlations because they are more likely to depend on a common underlying source of variance. These underlying sources can be estimated from statistical techniques such as principal components analysis or factor analysis. Factor analysis is widely used in the biological and social sciences to characterize the bases for individual differences (e.g. to try to identify the underlying dimensions along which respondents vary based on their answers to a set of survey questions). However, the techniques have special advantages for psychophysical measurements, especially when the independent variables are metrical variations. Despite this, applications of factor analysis in the visual sciences are relatively uncommon (Peterzell, 2016; Thurstone, 1944; Webster & MacLeod, 1988).

To appreciate the benefits of factor analysis and correlational techniques more generally, note that almost all experiments involve an s (stimulus) X n (observers) matrix of data. Though widely discussed in other fields, in visual science it has been emphasized only infrequently that an s X n set of data contains two distinct types of information (Woodworth, 1938; Thouless, 1951; Cronbach, 1957; McCall, 1990; Peterzell, 1993; Peterzell, Werner, & Kaplan, 1993). Nearly all vision researchers extract and examine the first distinct type, *representative functions*, and they report significant differences between experimental conditions or groups (see §1 above). In contrast, most vision researchers rarely examine the second type of information in s X n data sets, which include how *individual differences in data* obtained for one variable relate to individual differences obtained for other variables. High correlations result when individual differences are stable between variables, pointing to a common underlying or latent variable. Low correlations result when individual differences are not stable between variables, pointing to the possibility that the two variables instead reflect the influence of separate factors. Importantly, *representative functions* and *individual differences in data* provide completely independent sources of information. One cannot infer the pattern of individual differences, such as the s X s correlations, from the representative functions. Nor can one infer the representative functions from the pattern of individual differences. When we write in this essay about the potential to uncover a trove of information in archival data, we are referring to latent information about visual processes that may lie dormant but discoverable in the neglected individual differences in data from old experiments.

Sometimes the underlying structure of individual differences can be easily visualized simply by plotting the individual data, allowing a non-statistical or *intuitive* factor analysis (Peterzell, 2016). For example, in a plot of spectral sensitivities, an individual who has weaker than average sensitivity to a 430 nm light (e.g. because of a higher than average density of lens pigment), is also likely to have lower sensitivity to a 420 or 440 nm light. Conversely, if lens pigment is the source of the short-wavelength differences, then a different ordering of the participants' sensitivities might occur at longer wavelengths, where the lens pigment has little effect. Such patterns can also be visualized by comparing the matrix of scatterplots of each pair of s X s stimuli. This can reveal different regions of the matrix (e.g. a cluster at shorter wavelengths and none at longer) that exhibit a similar pattern of correlations across the n observers (Peterzell, 2016).

Formal factor analyses involve a decomposition of the correlation or covariance matrix into a set of underlying factors or latent variables that are often far fewer in number than the s measurements, and thus provide a more parsimonious representation of the data. Again, this is because the observed variables are often redundant measures of the same underlying processes, and thus factor analysis can potentially reveal the number and nature of these processes. One form of this approach is to use *confirmatory* factor analysis, to test whether the factors underlying individual differences in the data are consistent with processes already known or assumed to be influencing the observed behaviour. An example is the study by Webster and MacLeod (1988), who applied factor analysis to the archival colour matching data of Stiles and Burch (1959). The resulting factors corresponded closely to the

pattern predicted by variations in pre-receptoral screening pigments and the cone sensitivities, confirming the posited sources of variation and also quantifying the extent to which they varied and how this affected individual differences in colour matching (MacLeod & Webster, 1988). Analogous analyses have been applied to understanding variations in spatial contrast sensitivity (Peterzell, Werner, & Kaplan, 1991, 1993, 1995; Peterzell, Dougherty, & Mayer, 1997; Peterzell & Kelly, 1996; 1997; Dobkins et al., 2000; Peterzell, Chang, & Teller, 2000; Peterzell, Schefrin, Tragear, & Werner, 2000; Peterzell & Teller, 1996, 2000; see also Owsley, Sekuler, & Siemsen, 1983; Sekuler, Wilson, & Owsley, 1984).

However, factor analysis is also important as an *exploratory* tool, in which one uses individual differences to investigate and perhaps discover previously unknown processes or sources of variability (Costello & Osborne, 2005). For example, in the analysis of Webster and MacLeod (1988) there were factors representing systematic variability in the data and thus candidate sources of sensitivity variations, but which were not predicted and could not be readily identified (though they accounted for only a small proportion of the variance). As other examples of exploratory factor analysis, we can include any study in which factor analysis is performed without having strong a priori knowledge, theory or even an educated guess about what the factors might be. As described in the next section, Thurstone (1944) conducted an exploratory analysis of individual responses to 60 widely differing tests. Prior to collecting the data, he had no idea what the underlying factors might be. This is often the case with factor analyses in our emerging literature, and one must be especially cautious not to prematurely interpret factors as real or meaningful. Instead, the resulting factors can be used to generate hypotheses about the number, nature and tuning of mechanisms, with the caveat that evidence for underlying processes must be obtained through replications and studies using other methods. Without replication of factors and validation using alternative methods, the resulting factors should be viewed, at best, as preliminary evidence of underlying processes. With that strong caveat in mind, it remains to be seen what new processes and insights might be revealed by performing factor analysis on data lying on the shelves of visual science.

4.7. Identifying "specific abilities"

As noted in Section 4.1, since the advent of individual differences research, including in vision and visual perception, investigators have been interested in measuring and understanding human "abilities," and thus measuring individual differences in performance or potential. Early psychometric researchers investigated human visual abilities (e.g. Thurstone, 1944, 1950), and this type of inquiry has been rekindled by Wilmer and colleagues (Cho et al., 2015; Wilmer, Germine, & Nakayama, 2014; Yovel et al., 2014). The various goals of such investigations are to identify processes that mediate visual abilities, as in the previous section, and to develop psychometrically sound assessments of these abilities for use in basic research and applied settings.

Intrinsic to these investigations have again been questions about correlations and underlying factors. In the domain of intelligence, researchers have attempted to explain cognitive abilities in terms of a single broad ability or statistical factor. The classic and controversial "*g*" factor refers to a single factor obtained from factor analysis, which purportedly reflects significant shared correlations among an entire set of variables. There is a considerable literature that attempts, with at best limited success, to associate genetic, neural, clinical, academic, professional, and personal variability with *g* (Wilmer, Germine, & Nakayama, 2014; Wilmer et al., 2012). Attempts to find a comparable "*v*" for visual and perceptual abilities have been elusive (Thurstone, 1944, 1950; Cappé, Clarke, Mohr, & Herzog, 2014). Thurstone showed that general factors such as *g* are often small and misleading, and thus created the first factor-analytic methodologies to identify primary or "specific" abilities, or separate factors (Thurstone, 1938).

In the first study designed to identify specific *visual* abilities,

Thurstone (1944) measured 194 individuals' performance on 60 visual tests. This ultimately led him to postulate seven sources of individual variation that reflected specific or primary abilities, in that they (1) could be replicated over different studies, (2) were clearly differentiated from general intelligence, and (3) were separate from each other and other known specific cognitive abilities (Thurstone, 1950). Two specific visual abilities he discovered were factors related to perceptual speed and visual memory. Three more reflected individuals' specific abilities to perceive orientation in space (to recognise an object from different angles; imagine the movement or parts within a configuration; and think about spatial relations relative to the observer's own orientation). The remaining two factors related to closure, including abilities to impose or maintain a perceptual organization of the image.

The discovery and identification of specific visual abilities, while remaining of interest to psychometricians, seems to have been neglected by much of vision science until Wilmer et al. (2012) and Wilmer, Germine, and Nakayama (2014). Wilmer et al. (2014) proposed three criteria for identifying specific visual abilities: the measures of a construct should be consistent with existing theory, should exhibit reliability, and should show convergent and discriminant validity. Two additional criteria involved establishing extensive norms for a measure of a specific ability, and establishing that measures provide precise, error-free values for particular individuals. Wilmer et al. proceeded to demonstrate that face recognition, as measured by the Cambridge Face Memory Test, meets these criteria. As part of this effort, they have examined and documented the robust psychometric properties of the CFMT (Cho et al., 2015; Duchaine & Nakayama, 2006).

5. What is required to demonstrate stable individual differences in perception?

Despite the promise of individual differences as an approach, a second aim of our review is to discuss what is optimal practice in this field. Rather often, experimenters claim that a particular perceptual measure exhibits individual differences when in fact the only evidence is drawn from the variation in the scores obtained from participants in a single session. This is seldom a secure argument, as was already recognized in the mid-twentieth century by Robert Thouless (1951). To demonstrate true individual differences, we argue that best practice is either (i) to show test-retest reliabilities for scores obtained on well-separated occasions or (ii) to show that the variations in individual scores are correlated with some independently measured trait, such as a genetic polymorphism or clinical diagnosis. And even in the latter case, it is desirable also to know the test-retest reliabilities, in order properly to interpret the size of any correlation that is found between perceptual scores and the independent trait – or in order properly to interpret an absence of correlation (*v* §5.2.3).

5.1. Test-retest reliability

When each participant is tested on a single occasion, at least four sources of variance can contribute to variation in the scores obtained from different participants:

(a) Constitutional, or other long-term, between-individual, variation. It is this variation that often interests the experimenter, but only with care can it be separated from other sources of variance.

(b) Within-individual variation. A given participant will vary over time in his or her scores or settings on a perceptual task. Some time-dependent factors have been formally studied, such as history of light exposure (e.g. Belmore & Shevell, 2008), diurnal variation (e.g. Barnard, Hattar, Hankins, & Lucas, 2006), and phase of the menstrual cycle (e.g. Farage, Osborn, & MacLean, 2008). But there are always less systematic factors of physical health, fatigue, mood, personal events, and random distractions that will change a participant's attentiveness and performance from one session to another. Most modern experimenters will have been faced with participants who surreptitiously

consult their telephones during testing. Lund and MacKay (1983) reported that a measure of perceptual plasticity – the initial strength of the McCollough effect – was weaker the fewer the hours of sleep that the participant had enjoyed the previous night. Owing to such factors, the consistency of a participant's responses during one short session cannot be taken as a measure of the within-individual variance over a longer period. Moreover, even if the participant is alert and undistracted, the consistency might arise only because in the short term he adopts a particular strategy or criterion – a response bias that may or may not change on a subsequent occasion. As Thouless (1951) wrote with regard to his own study of individual differences in phenomenal regression for shape: '…it might be the case that this self-consistency was only the result of each participant maintaining during his experimental session a level of response that was the one he had happened to adopt at the beginning of the experiment.' Of course, the participant may adopt the same criterion on second test, or may be in a similar state of sleep deprivation or of light exposure or of motivation, but it is surely likely that repeating measurements on different occasions will reduce the contaminating effects of within-individual variation.

(c) Instrumental variation: The instruments that are used in visual science are not absolutely steady in their outputs. Not only is there always some intrinsic variation that is due to photon noise, but calibrations may also vary systematically with factors such as total operating hours, fluctuations in mains supply, and fluctuations in room temperature; and these variations may be confounded with which participant is being tested. The luminous and chromatic outputs of CRT displays vary demonstrably both in the short-term and the long-term (Mollon & Baker, 1995). Similarly, the chromaticities of LEDs vary with ambient temperature (Raypah, Devarajan, & Sulaiman, 2016). So sensitive is the Nagel anomaloscope to ambient temperature that an experienced observer could use it to measure room temperature to within one or two degrees Celsius (Jordan & Mollon, 1993): an instrumental variation of this generic kind would be a serious confound in any pedigree study in which all members of a given family were examined on the same occasion, while different families were tested on different occasions. Included in the category of instrumental variation is variation deliberately introduced by the experimenter, for example, in the random ordering of trials or conditions.

(d) Variation in the experimenter. Although the laboratory computer allows us to present (approximately) the same stimuli to different participants and protects us from some of the experimenter effects that exercised our predecessors (Rosenthal, 1963), variations in the experimenter's behaviour may still affect the results. If many participants are being tested, the experimenter may over time become subtly more skilled in explaining the task and the procedures, or conversely may become more perfunctory in processing each participant. Using forced-choice procedures in visual performance tasks will distinguish variations in sensitivity from variations in criterion, but it cannot protect against variations in the participant's raw motivation, and the latter is likely to be influenced by social interaction with an experimenter at the beginning of the session. When the participants are participating for course credit or purely to secure token payments, their cooperation and motivation may be particularly frail, and particularly sensitive to slight variations in the experimenter's behaviour. In large population studies, such as modern genome-wide association studies, where thousands of participants are tested, there will necessarily be more than one experimenter, and so this is very likely to be a source of variance. Online testing attenuates the effects of experimenter influence, but is more vulnerable to sources of variance (b) and (c).

All these sources of variance contribute to the differences in measured performance between participants, and sources (b), (c) and (d) mask the size of any true individual differences (a). In an optimal experiment designed to examine stable, trait-like individual differences, the experimenter might test each participant multiple times, counterbalancing completely for time of day, season, experimenter etc. This has probably never been achieved. But in best practice, a minimal requirement is to obtain test-retest reliabilities, testing participants on the same task on at least two, well-separated, occasions.

It is not necessarily enough merely to correlate the scores on one test with those obtained on a second test administered on the same occasion, or indeed, to rely on split-half reliabilities from a single test, correlating alternate, or randomly divided, trials from a single testing session. For positive correlations between measures obtained on the same occasion – a 'positive manifold', to take a term from the classical psychometric literature (Spearman, 1904) – could arise from the several sources of time-dependent variation discussed above.

Of course, on a second test, the participant will no longer be in a state of experimental naïvety. If the task, say, requires recognition of images or aesthetic judgements, it will be necessary to have a sufficient stock of equivalent stimuli. But this is a problem that would in any case arise if reliability were estimated from a split-half procedure in a single session. This problem is reduced in many psychophysical experiments, where an adaptive procedure is used and where the task is, say, discriminating verniers or detecting gabors. Certainly perceptual learning will occur between sessions, but learning will also occur within a single extended session. One secondary advantage of repeated testing sessions is that it reduces the variance due to participants who simply didn't grasp the task during the initial trials. On perceptual learning, see also §5.2.7 below. However, as we noted above, some within-individual variation in 'state' factors is likely to persist across the test-retest interval, meaning that test-retest reliability offers only a partial estimate of time-varying sources of variance. The extent to which test-retest reliability differs from within-session (e.g. split half) reliability is an empirical question, which has not often been addressed. It would be helpful for researchers who have the available data to publish both reliabilities in order to generate a more precise picture of how and under what circumstances they differ.

5.2. Correlation with an independent measure

This second way to establish true individual differences requires that the correlated measures are independent.

5.2.1. Inter-test reliability

Goodbourn and colleagues (2012) have introduced the useful concept of *inter-test reliability*. In their study of 'magnocellular function', there were small residual, positive, correlations between all pairs of tasks when measurements taken on the same occasion were considered, and these correlations were all significant, owing the large cohort being tested; but when performance on Task A in experimental session 1 was correlated with performance on Task B in session 2 (i.e. when *inter-test reliability* was measured), then the correlations between different magnocellular tasks – with the exception of two tasks using low-frequency gratings – were close to zero and were insignificant, even though the test-retest reliabilities for any given task ranged from 0.52 to 0.77 (Spearman's rank-order coefficients). The use of inter-test reliability eliminates at least part of the variance that is specific to a particular occasion of testing. The study of Goodbourn and colleagues is more generally instructive for the present essay, since it reveals how a spurious 'positive manifold' can be apparent if only a single test session is considered.

5.2.2. Correlation with an independent trait

If inter-test reliabilities are not available, then – if individual differences are to be demonstrated – the second variable must be firmly independent of the primary measure. It might be sex or a genetic polymorphism or a clinical condition or a phenotypic trait that has been measured independently. Even with variables such as sex or genetics the usual issues of sampling must be considered. Few studies of sex differences, for example, sample randomly from the parent male and female population: to draw only from a university population, and worse, to rely on self-volunteered participants, is to introduce unknown
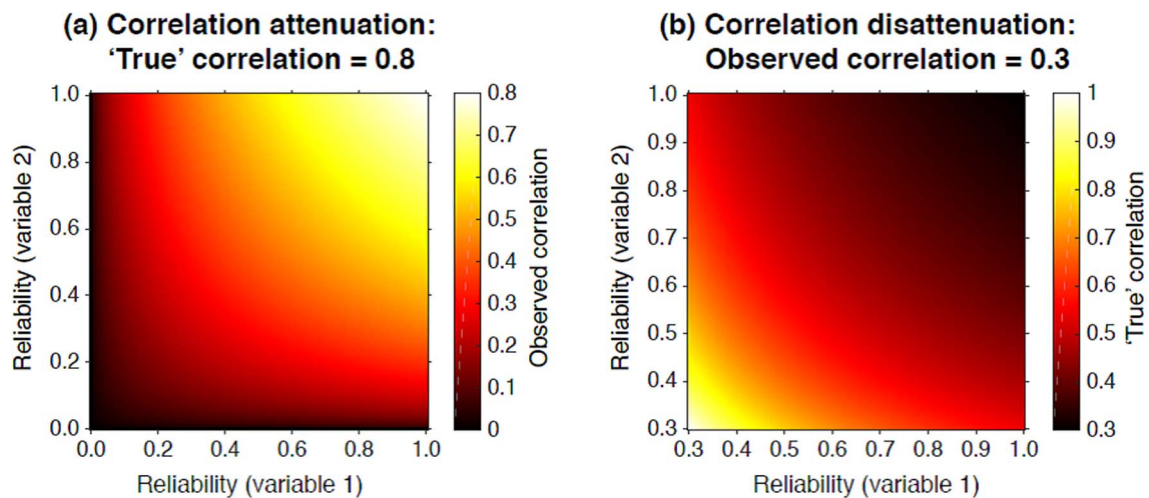
Fig. 2. Panel (a) shows the attenuation of a 'true' correlation (the correlation between hypothetical 'universe scores') of 0.8 between two variables, as a function of the test-retest reliabilities of each of the two variables. Panel (b) shows the converse: the disattenuated 'true' correlation between two variables given an observed correlation of 0.3, as a function of the test-retest reliabilities of each of the two variables.

biases that may be the true cause of the apparent 'sex difference'. In genetic work, an ever-present problem is that of 'stratification' (Price et al., 2006): the polymorphism nominally studied may only be an indicator of two latent but genetically distinct populations within the sample, and it may be other differences between the two sub-populations – differences elsewhere in the genome or environmental differences – that are the true source of the perceptual variation that is of interest.

5.2.3. Estimating effect size

A knowledge of the test-retest reliability of each individual trait is critical for interpreting the effect size of any correlation found between two traits. Without reliabilities, the absence of a correlation between two variables is meaningless, since the two correlated measures may themselves contain only random variation. On the other hand, a small significant correlation between two variables may represent a large underestimation of the true relationship. Test-retest reliabilities provide an estimate of the noise in the correlated variables, which can be used to estimate the true effect size (Spearman, 1904; Schmidt & Hunter, 1996). Fig. 2 shows how the observed correlation between two measures depends on the underlying test-retest reliabilities, and gives an estimate of the true effect size for a given observed correlation as test-retest reliability varies.

A particularly interesting case arises when the 'effect size' under debate is in fact heritability, i.e. the proportion of the variance in a phenotypic trait that is of genetic origin. Methods have recently been developed to estimate the heritability of a trait directly, by taking into account all the single nucleotide polymorphisms in the genome (Davies et al., 2011; Yang, Lee, Goddard, & Visscher, 2011). This is a lower-bound estimate, of course, since it does not include other forms of genomic variation; but if only a single measure is available of the phenotypic trait – as is often the case in genome-wide studies of large cohorts – then the reliability of the phenotypic measure is unknown and the heritability will be a lower-bound estimate for this second reason.

5.2.4. Parametric or non-parametric statistics in correlational studies?

The uncritical application of correlational statistics can lead to both false negatives and false positives. These occur when outliers exert leverage to misleadingly inflate or deflate the correlation coefficient (Rousseau & Leroy, 2005). Leverage is a particular problem for smaller samples, but also affects large samples if specific data points are outliers for both correlated variables – if, for example, a handful of participants were badly sleep-deprived or were affected by one of several other types of time-varying factors considered in §5.1. Though not unaffected

by leverage, Spearman's rho offers substantial protection in comparison to Pearson's r. Therefore, as an alternative, or in addition to parametric statistics (noting the marginal power advantage for Pearson's r), it seems wise also to calculate Spearman's rho, or to empirically quantify the null distribution for the particular data by permutation testing (Churchill & Doerge, 1994). Additionally, when using parametric (or non-parametric) statistics, it is essential to examine the data for outliers (Judd, McClelland, & Ryan, 2017), by generating and inspecting frequency histograms, scatterplots, and scatterplot matrices. Particular vigilance is advisable for pairs of data that are outliers for both correlated variables. If such pairs are identified, statistical results should be interpreted with caution.

5.2.5. The need to interleave the testing of case and control participants

One large class of studies on individual differences are those where a clinically defined group are compared with controls (see §4.2). In the present context, the independent second variable is the presence or absence of a formal clinical diagnosis. In reports of such studies, the method of matching of the two groups in demographic variables is usually discussed and justified. But what is often neglected is the temporal and spatial equation of the two groups with regard to the time and place of testing. It is not unknown for experimenters interested in a developmental disorder to first test the clinical cases on a visual task and then to set out to find suitable controls. By then, of course, the experimenters have become more skilled in explaining the task to children. Conversely, controls may be tested first, as part of the development of the test or because controls are easy to find and the testing of patients has to be opportunistic, upon referral. All the many dangers mentioned in §5.1 are relevant here. It is essential that the testing of controls and cases should be fully interleaved in time and space. We fear that this is seldom achieved, since it is seldom discussed.

5.2.6. The danger of interleaving conditions

An experimenter might judge that it is always desirable randomly to interleave trials of different conditions, for example in an experiment that tests different hemifields or different wavelengths or different spatial frequencies. Yet this ostensibly correct design will generate spurious individual differences in any cases where different strategies are optimal for different conditions. For example, in an experiment that compares hemifields, individual participants may selectively and arbitrarily attend to one hemifield or the other. A subsequent factor analysis would then reveal a factor corresponding to hemifield and the experimenter might compose a paper about 'left-hemisphere' and 'right-hemisphere' participants. But the individual differences might dissolve

if hemifields were tested in separate blocks (fixation being monitored with an eye-movement recorder). This class of pitfall is especially dangerous in experiments where the targets in different conditions differ in their liminal appearance. Human observers may not be able to distribute their attention uniformly over the full domain of potential targets, i.e. they may not be able to search for multiple types of target concurrently. An observer typically has a specific *Suchbild*, a template for the target that she seeks. Thus in an experiment on spatial frequency, where different frequencies are interleaved, some participants may selectively attend for low-frequency targets and some for high.

5.2.7. Should the order of tests or of conditions be fixed or should it be randomised?

Here the design that recommends itself in research on individual differences may differ from the design that is favoured in the more common type of experiment, where conditions are being compared in a within-subjects design. In research on individual differences, it may be better to give the tests in fixed order to all participants, rather than in a randomised or counterbalanced order. Otherwise, there is the danger that participants tested in one order will learn skills, or will adopt strategies or criteria, that carry forward and are different from those learnt or adopted by participants tested in a different order. And then spurious individual differences will emerge. Even if the transfer is positive between tests or conditions (i.e. a skill or strategy or criterion acquired in one task is also optimal in the second), the transfer may be asymmetric, as when a skill acquired on an easy task transfers positively to a more difficult task, but little or no transfer occurs in the opposite direction. It was the possibility of such asymmetric effects that led Poulton (1966) completely to eschew within-group designs in experimental psychology.

However, the experimenter must make a nice judgement here. If the tests or the conditions are presented to all participants in the same fixed order, then time-varying sources of noise (of the type discussed in §5.1) may produce artificial correlations between pairs of tests or conditions that are close to one another in the sequence of testing. Certainly, if the conditions being tested are ones that lie along a continuum, such as wavelength or frequency, it would be inappropriate to use a fixed order that mapped systematically on to the continuum. For then, time-varying noise would produce apparent correlations between adjacent wavelengths or frequencies.

One way in which participants undoubtedly differ is in speed of learning. If tests or conditions are presented in fixed order, then differences in rate of acquiring a generic skill may manifest themselves as apparent differences in perceptual ability: Fast and slow learners might appear matched on an initial test but appear to differ later in the sequence of testing. Counterbalancing or randomising the order of tests across participants may allow these learning effects to cancel – provided the transfer of skill is always positive and is of similar average magnitude in the different directions of transfer. If a fixed order of tasks were chosen, then individual differences in rates of learning would lead to apparent individual differences for tasks that occur later in the test battery. It is worth remarking here that very few studies of individual differences in vision have thus far explicitly distinguished between individual differences in rate of perceptual learning and individual differences in asymptotic performance. Owing to the time taken to test large numbers of participants, and the known perceptual learning that occurs over hundreds or thousands of trials, we suspect asymptotic performance has seldom been measured. However, a test-retest design will at least bring participants closer to asymptote.

In sum, we have set out in the preceding sections some of the considerations that an experimenter may wish to take into account in deciding whether to test different participants in the same fixed order. In an ideal experiment, the order of conditions would be fixed but each participant would be tested on each condition for enough trials to ensure that he or she reaches asymptotic performance on each condition. And the conditions would be repeated in counterbalanced order after an interval of at least one day, allowing the process of reminiscence (Mollon & Danilova, 1996) to occur and inter-test reliabilities to be derived. This, of course, is an ideal.

5.3. Correction for multiple testing

Owing to the time and difficulty of achieving the large samples necessary for individual differences research, researchers understandably often measure multiple variables in their participants, some of which may be tangential to the primary scientific hypothesis. This is particularly the case in genome-wide association studies or brain imaging studies, where the costs of genotyping or imaging may be large relative to the costs of additional behavioural testing. The authors of a study may then publish separately their correlational findings for particular perceptual measures. There may be justification for this, in that a paper containing many different results, with the several accompanying discussions, would be too large and too clumsy for most journals; and indeed, different perceptual traits may be of interest to different audiences.

However, there has been little discussion of whether a statistical correction for multiple comparisons (e.g. Bonferroni correction) should then take into account all the different behavioural measures that have been included in the study – in addition to the corrections needed to allow for the large number of single-nucleotide polymorphisms correlated with each behavioural measure in GWAS studies, or the large number of individual (for example voxel-based) analyses applied in brain imaging studies. In principle, the absence of the former correction could lead to spurious results. This problem is compounded by publication bias (Open Science Collaboration, 2015). If only 'significant' correlations are published, readers are unable to assess whether multiple comparisons have been applied correctly. If researchers conduct many correlations, in the same or separate studies on individual differences and publish only those that show a p-value of < .05, the reader will be unable to estimate the true probability of a false positive in their research, which is a function of the number of unpublished negative findings (Ioannidis, 2005; Rosenthal, 1979).

5.4. On the need to consider response bias as a possible confounding factor

Since the advent of signal detection theory (SDT) in the 1950s, it is widely recognized that decision processes as well as sensory processes determine measures of detection, discrimination, and response time (Green & Swets, 1966; Harvey, 1992; Link, 1992). Such systematic response biases can therefore also influence individual variability in data, e.g. between observers who are more cautious or liberal in their decisions. If individuals hold to such a bias across stimulus conditions, then spuriously positive correlations and broad general statistical factors will result (Dobkins et al., 2000; Peterzell et al., 1993). Criterion effects can obscure individual differences in other ways. For example, when observers are asked to categorize stimuli that vary along a dimension (e.g. from blues to greens), they may tend to locate the category boundary near the middle of the stimulus set, and such "range effects" may mask the individual differences in the perceived boundaries (Wright, 2011).

It is often assumed that two types of method decouple decision criteria from sensory and discrimination measures, thereby controlling or minimizing response bias (Harvey, 1992). One, an SDT method, requires that there are at least two types of decision trials, including one with the stimulus and at least one without (i.e., "yes/no" paradigms). The second is the forced-choice paradigm, which is widely assumed to require all observers to use the same response criterion. However, such measures are not bias free. For example, different test participants might be biased to different degrees towards picking either the first or second interval in a two-interval forced-choice (2IFC) experiment, or biased towards responding left- or right-oblique in a single interval orientation discrimination task. Such biases result in individual variations in proportion-correct performance (Klein, 2001; Witt, Taylor,

Sugovic, & Wixted, 2015), which must be dissociated with care from differences in the sensory, perceptual or cognitive measure of interest. This might be achieved using signal detection theory. However, although SDT discriminates between sensitivity and bias, it cannot determine whether the underlying source of the bias reflects perceptual biases, response biases, or some combination of the two (Witt et al., 2015). While there is currently no solution to disentangle these two, it is important to realize that true perceptual effects may appear not in measures like d', but rather in measures of criterion. With such limitations, the importance of testing models across studies, converging operations, and laboratories becomes even more paramount.

6. Conclusions

Individual differences provide a large yet seldom opened window into the mechanisms and processes underlying how we see. Appropriate measurement and analysis of these differences can yield insights that may be difficult to achieve through conventional approaches focused on defining a standard observer or representative function. However, a major challenge for individual differences research is to separate the wheat of real observer differences from the chaff of random noise and systematic confounds. Few studies meet all of the requirements of the best practices advocated, though many have yielded plausible and compelling accounts of the variations in vision and the processes responsible. Each study – both those from the past and those that follow in the future – should be weighed carefully to assess the strength of the design, and the conclusions that can be drawn from the results.

References

Allen, A. J., Griss, M. E., Folley, B. S., Hawkins, K. A., & Pearlson, G. D. (2009). Endophenotypes in schizophrenia: A selective review. *Schizophrenia Research, 109*(1), 24–37.

Anstis, S., & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In J. D. Mollon, & L. T. Sharpe (Eds.). *Colour vision: Psychophysics and physiology* (pp. 155–166). London: Academic Press.

Asano, Y., Fairchild, M. D., & Blondé, L. (2016). Individual colorimetric observer model. *PLoS One, 11*(2), e0145671.

Axelrod, V., Schwarzkopf, D. S., Gilaie-Dotan, S., & Rees, G. (2017). Perceptual similarity and the neural correlates of geometrical illusions in human brain structure. *Scientific Reports, 7*, 39968.

Bakken, T. E., Roddey, J. C., Djurovic, S., Akshoomoff, N., & Amaral, D. G. (2012). Association of common genetic variants in GPCPD1 with scaling of visual cortical surface area in humans. *Proceedings of the National Academy of Sciences of the United States of America, 109*(10), 3985–3990.

Barnard, A. R., Hattar, S., Hankins, M. W., & Lucas, R. J. (2006). Melanopsin regulates visual processing in the mouse retina. *Current Biology, 16*(4), 389–395.

Belmore, S. C., & Shevell, S. K. (2008). Very-long-term chromatic adaptation: Test of gain theory and a new method. *Journal of the Optical Society of America A, 25*, 411–414.

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2003). Motion perception in autism: A '"complex"' issue. *Journal of Cognitive Neuroscience, Neurosurgery, and Psychiatry, 15*, 218–225.

Bone, R. A., & Sparrock, J. M. (1971). Comparison of macular pigment densities in human eyes. *Vision Research, 11*(10), 1057–1064.

Bosten, J. M., Beer, R. D., & MacLeod, D. I. (2015). What is white? *Journal of Vision, 15*(16), 5.

Bosten, J. M., Hogg, R. E., Bargary, G., Goodbourn, P. T., Lawrance-Owen, A. J., & Mollon, J. D. (2014). Suggestive Association WithOcular Phoria at Chromosome 6p22GWAS of Phorias. *Investigative Ophthalmology & Visual Science, 55*(1), 345–352.

Butler, P. D., Silverstein, S. M., & Dakin, S. C. (2008). Visual perception and its impairment in schizophrenia. *Biological Psychiatry, 64*(1), 40–47.

Brainard, D. H., & Hurlbert, A. C. (2015). Colour vision: understanding# thedress. *Current Biology, 25*(13), R551–R554.

Brainard, D. H., Roorda, A., Yamauchi, Y., Calderone, J. B., Metha, A., Neitz, M., ... Jacobs, G. H. (2000). Functional consequences of the relative numbers of L and M cones. *Journal of the Optical Society of America. A: Optics, Image Science, and Vision, 17*(3), 607–614.

Brown, A., & Hall, B. V. (1936). The directive influence of light upon *Drosophila melanogaster* Meig and some of its eye mutants. *Journal of Experimental Zoology, 74*, 205–220.

Calkins, M. E., & Iacono, W. G. (2000). Eye movement dysfunction in schizophrenia: A

heritable characteristic for enhancing phenotype definition. *American Journal of Medical Genetics, 97*, 72–76.

Cappé, C., Clarke, A., Mohr, C., & Herzog, M. H. (2014). Is there a common factor for vision? *Journal of Vision, 14*(8).

Castejon-Mochon, J. F., Lopez-Gil, N., Benito, A., & Artal, P. (2002). Ocular wave-front aberration statistics in a normal young population. *Vision Research, 42*(13), 1611–1617.

Cho, S. J., Wilmer, J. B., Herzmann, G., McGugin, R. W., Fiset, D., Van Gulick, A. E., ... Gauthier, I. (2015). Item response theory analyses of the Cambridge Face Memory Test (CFMT). *Psychological Assessment, 27*(2), 552.

Churchill, G. A., & Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics, 138*, 965–971.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1–10.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*(11), 671–684.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry, 16*(10), 996–1005.

Delahunt, P. B., Hardy, J. L., & Werner, J. S. (2008). The effect of senescence on orientation discrimination and mechanism tuning. *Journal of Vision, 8*(3) 5-1–9.

de-Wit, L., & Wagemans, J. (2016). Individual differences in local and global perceptual organization. In J. Wagemans (Ed.). *Oxford Handbook of Perceptual Organization*. Oxford: Oxford University Press.

Dobkins, K. R., Gunther, K., & Peterzell, D. H. (2000). What mechanisms underlie red/green isoluminance, luminance contrast sensitivity and chromatic contrast sensitivity at various spatial and temporal frequencies? *Vision Research, 40*, 613–628.

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.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron, 38*(4), 659–671.

Farage, M. A., Osborn, T. W., & MacLean, A. N. (2008). Cognitive, sensory, and emotional changes associated with the menstrual cycle: A review. *Archives of Gynecology and Obstetrics, 278*, 299.

Fuller, J. L., & Thompson, W. R. (1960). *Behavior genetics*. New York: John Wiley and Sons.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan & Co.

Gavel, L. V. (1939). Die 'Kritische Streifenbreite' als Mass der Sehscharfe bei *Drosophila melanogaster*. *Zeitschrift für vergleichende Physiologie, 27*, 80–135.

Geisler, W. S., & Ringach, D. (2009). Natural systems analysis. Introduction. *Visual Neuroscience, 26*(1), 1–3.

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology, 252*(3), 627–656.

Germine, L., Russell, R., Bronstad, P. M., Blockland, G. A. M., Smoller, J. W., Kwok, H., ... Wilmer, J. B. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology, 25*, 2684–2689.

Gilaie-Dotan, S., Kanai, R., Bahrami, B., Rees, G., & Saygin, A. P. (2013). Neuroanatomical correlates of biological motion detection. *Neuropsychologia, 51*, 457–463.

Goodbourn, P. T., Bosten, J. M., Hogg, R. E., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. D. (2012). Do different 'magnocellular tasks' probe the same neural substrate? *Proceedings of the Royal Society B-Biological Sciences, 279*(1745), 4263–4271.

Goodbourn, P. T., Bosten, J. M., Bargary, G., Hogg, R. E., Lawrance-Owen, A. J., & Mollon, J. D. (2014). Variants in the 1q21 risk region are associated with a visual endophenotype of autism and schizophrenia. *Genes, Brain and Behavior, 13*(2), 144–151.

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636–645.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Press.

Greenspan, R. J. (2008). The origins of behavioral genetics. *Current Biology, 18*, R192–R198.

Gu, J., & Kanai, R. (2014). What contributes to individual differences in brain structure? *Frontiers in Human Neuroscience, 8*, 262.

Hammond, B. R., Jr., Johnson, E. J., Russell, R. M., Krinsky, N. I., Yeum, K. J., Edwards, R. B., & Snodderly, D. M. (1997). Dietary modification of human macular pigment density. *Investigative Ophthalmology & Visual Science, 38*(9), 1795–1801.

Happe, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36*(1), 5–25.

Hardy, J. L., Frederick, C. M., Kay, P., & Werner, J. S. (2005). Color naming, lens aging, and grue: What the optics of the aging eye can teach us about color language. *Psychological Science, 16*(4), 321–327.

Harvey, L. O., Jr. (1992). The critical operating characteristic and the evaluation of expert judgment. *Organizational Behavior & Human Decision Processes, 53*(2), 229–251.

Hecht, S., & Wald, G. (1934). The visual acuity and intensity discrimination of *Drosophila*. *Journal of General Physiology, 17*, 517–547.

Held, R. (1989). Perception and its neuronal mechanisms. *Cognition, 33*, 139–154.

Hofer, H., Carroll, J., Neitz, J., Neitz, M., & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *Journal of Neuroscience, 25*(42), 9669–9679.

Holmgren, F. (1877). *Om färgblindheten i dess förhållande till jernvägstrafiken och sjöväsendet*. Upsala: Berlings Boktryckeri.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8).

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology and Psychiatry, 38*, 527–534.

Jordan, G., & Mollon, J. D. (1993). The Nagel anomaloscope and seasonal variation of colour vision. *Nature, 363*(6429), 546–549.

Jordan, G., & Mollon, J. D. (1997). Unique hues in heterozygotes for protan and deutan deficiencies. In *Colour vision deficiencies VIII, documenta ophthalmologica proceedings series*, Vol. 59, pp. 67–76.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model approach to regression, ANOVA, and beyond* (3rd ed.). New York: Routledge.

Kaiser, P. K. (1988). Sensation luminance: A new name to distinguish CIE luminance from luminance dependent on an individual's spectral sensitivity. *Vision Research, 28*(3), 455–456.

Kalmus, K. (1943). The optomotor responses of some eye mutants of *Drosophila. Journal of Genetics, 45*, 206–213.

Kay, P., & Regier, T. (2006). Language, thought and color: Recent developments. *Trends in Cognitive Sciences, 10*(2), 51–54.

Kanai, R., Bahrami, B., & Rees, G. (2010). Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current Biology : CB, 20*(18), 1626–1630.

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience, 12*(April), 231–242.

König, A., & Dieterici, C. (1886). The modern development of Thomas Young's theory of colour-vision. *Report of the British Association for the Advancement of Science, 56*, 431–439.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*, 1421–1455.

Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics* (7th ed.). New York: Worth Publishers.

Kuehni, R. G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research and Application, 29*, 158–162.

Lennie, P., Pokorny, J., & Smith, V. C. (1993). Luminance. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 10*(6), 1283–1293.

Lindsey, D. T., & Brown, A. M. (2002). Color naming and the phototoxic effects of sunlight on the eye. *Psychological Science, 13*(6), 506–512.

Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences of the United States of America, 106*(47), 19785–19790.

Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-Gatherer color naming provides new insight into the evolution of color terms. *Current Biology, 25*(18), 2441–2446.

Link, S. W. (1992). *Wave theory of difference and similarity*. Psychology Press.

Lund, N. J., & MacKay, D. M. (1983). Sleep and the McCollough effect. *Vision Research, 23*(9), 903–906.

MacLeod, D. I., & Webster, M. A. (1988). Direct psychophysical estimates of the cone-pigment absorption spectra. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 5*(10), 1736–1743.

McCall, R. B. (1990). Infancy research: Individual differences. *Merrill-Palmer Quarterly, 36*, 141–157.

McGuinness, D. (1976). Away from a unisex psychology: individual differences in visual sensory and perceptual processes. *Perception, 5*, 279–374.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3–35.

Miyahara, E., Pokorny, J., Smith, V. C., Baron, R., & Baron, E. (1998). Color vision in two observers with highly biased LWS/MWS cone ratios. *Vision Research, 38*(4), 601–612.

Mollon, J. D. (1986). Molecular genetics: Understanding colour vision. *Nature, 321*, 12–13 (News & Views).

Mollon, J. D., & Baker, M. R. (1995). The use of CRT displays in research on colour vision. In B. Drum (Ed.). *Colour vision deficiencies XII* (pp. 423–444). Dordrecht: Kluwer Academic.

Mollon, J. D., & Cavonius, L. R. (2012). The Lagerlunda collision and the introduction of color vision testing. *Survey of Ophthalmology, 57*(2), 178–194.

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision, 10*, 51–58.

Mollon, J. D., & Jordan, G. (1997). On the nature of unique hues. In C. Dickenson, I. Maurray, & D. Carden (Eds.). *John Dalton's colour vision legacy*. London: Taylor and Francis.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B., & Hogness, D. S. (1986). Molecular genetics of inherited variation in human color vision. *Science, 232*, 203.

Nathans, J., Thomas, D., & Hogness, D. S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science, 232*, 193–202.

Neitz, J., & Neitz, M. (2011). The genetics of normal and defective color vision. *Vision Research, 51*(7), 633–651.

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

Orlansky, G., Wilmer, J. B., Taub, M. B., Rutner, D., Ciner, E., & Gryczynski, J. (2015). Astigmatism and early academic readiness in preschool children. *Optometry & Vision Science, 92*(3), 279–285.

Owsley, C. (2011). Aging and vision. *Vision Research, 51*, 1610–1622.

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision research, 23*(7), 689–699.

Peterzell, D. H. (1993). Individual differences in the visual attention of human infants: Further evidence for separate sensitization and habituation processes. *Developmental Psychobiology, 16*, 207–218.

Peterzell, D. H. (2016). Discovering sensory processes using individual differences: A review and factor analytic manifesto. *Electronic Imaging, 16*, 1–11.

Peterzell, D. H., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for

red-green and luminance-modulated gratings: Psychophysical data from human adults. *Vision Research, 40*(4), 417–430.

Peterzell, D. H., Chang, S. K., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red-green and luminance-modulated gratings: Psychophysical data from human infants. *Vision Research, 40*, 431–444.

Peterzell, D. H., Dougherty, R. F. & Mayer, M. J. (1997). Temporal tuning of flicker-sensitive channels derived from individual differences in de Lange functions. *Vision science & its applications: Technical digest v.1.* (pp. 218–221). Washington, D.C.: Optical Society of America.

Peterzell, D. H., & Kelly, J. P. (1996). Spatial frequency channels revealed by individual differences in contrast sensitivity functions: Visual evoked potentials from adults and infants. *Vision Science & Its Applications: Technical Digest v.1* (pp. 10–13). Washington, D.C.: Optical Society of America.

Peterzell, D. H., & Kelly, J. P. (1997). Development of spatial frequency tuned "covariance" channels: individual differences in the electrophysiological (VEP) contrast sensitivity function. *Optometry & Vision Science, 74*, 800–807.

Peterzell, D. H., Schefrin, B. E., Tragear, S. J., & Werner, J. S. (2000). Spatial frequency tuned covariance channels underlying scotopic contrast sensitivity. In D. C. Washington (Ed.). *Vision science & its applications: OSA technical digest* (pp. 39–42). Optical Society of America.

Peterzell, D. H., & Teller, D. Y. (1996). Individual differences in contrast sensitivity functions: The coarsest spatial channels. *Vision Research, 36*, 3077–3085.

Peterzell, D. H., & Teller, D. Y. (2000). Spatial frequency tuned covariance channels for red-green and luminance-modulated gratings: Psychophysical data from human adults. *Vision Research, 40*, 417–430.

Peterzell, D. H., Werner, J. S., & Kaplan, P. S. (1991). Individual differences in the contrast sensitivity functions of human adults and infants: A brief review. In P. Bagnoli, & W. Hodos (Eds.). *The changing visual system: maturation and aging in the central nervous system* (pp. 391–396). NY: Plenum.

Peterzell, D. H., Werner, J. S., & Kaplan, P. S. (1993). Individual differences in contrast sensitivity functions: The first four months of life in humans. *Vision Research, 33*, 381–396.

Peterzell, D. H., Werner, J. S., & Kaplan, P. S. (1995). Individual differences in contrast sensitivity functions: Longitudinal study of 4-, 6- and 8-month-old human infants. *Vision Research, 35*, 961–979.

Pokorny, J., Smith, V. C., & Lutze, M. (1987). Aging of the human lens. *Applied Optics, 26*, 1437.

Porter, J., Guirao, A., Cox, I. G., & Williams, D. R. (2001). Monochromatic aberrations of the human eye in a large population. *Journal of the Optical Society of America. A: Optics, Image Science, and Vision, 18*(8), 1793–1803.

Poulton, E. (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin, 66*, 1.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*, 904–909.

Raypah, M. E., Devarajan, M., & Sulaiman, F. (2016). Influence of injection current and ambient temperature on intensity and wavelength of low-power SMD LED. *2016 Ieee 37th international electronics manufacturing technology (Iemt) & 18th electronics materials and packaging (Emap) conference.*

Ritsner, M. S., & Gottesman, I. I. (2009). Where do we stand in the quest for neuropsychiatric biomarkers and endophenotypes and what next? In M. S. Ritsner (Vol. Ed.), *The handbook of neuropsychiatric biomarkers, endophenotypes, and genes: Vol. 1,* (pp. 3–22). New York: Springer.

Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., & Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PLoS One, 11*(2).

Rosenthal, R. (1963). On social psychology of psychological experiment – experimenters hypothesis as unintended determinant of experimental results. *American Scientist, 51*(2), 268–283.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rousseau, P. J., & Leroy, A. M. (2005). *Robust regression and outlier detection*. Wiley Interscience.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review, 16*(2), 252–257.

Sanfilippo, P. G., Hammond, C. J., Staffieri, S. E., Kearns, L. S., Liew, M. S. H., Barbour, J. M., ... Mackey, D. A. (2012). Heritability of strabismus: Genetic influence is specific to eso-deviation and independent of refractive error. *Twin Research and Human Genetics, 15*, 624–630.

Sawides, L., de Gracia, P., Dorronsoro, C., Webster, M. A., & Marcos, S. (2011). Vision is adapted to the natural level of blur present in the retinal image. *PLoS One, 6*(11), e27031.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2).

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience, 14*(1), 28–30.

Sekuler, R., Wilson, H. R., & Owsley, C. (1984). Structural modeling of spatial vision. *Vision research, 24*(7), 689–700.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1216.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research, 49*(22), 2705–2739.

Smithson, H. E., Sumner, P., & Mollon, J. D. (2003). How to find a tritan line. In J. D. Mollon, J. Pokorny, & K. Knoblauch (Eds.). *Normal and Defective Colour Vision* (pp. 279–287). Oxford: Oxford University Press.

Song, C., Schwarzkopf, D. S., Lutti, A., Li, B., Kanai, R., & Rees, G. (2013a). Effective connectivity within human primary visual cortex predicts interindividual diversity in

illusory perception. *The Journal of Neuroscience, 33*(48), 18781–18791.

Song, C., Schwarzkopf, D. S., & Rees, G. (2013b). Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nature Communications, 4,* 2201.

Spearman, C. (1904). 'General Intelligence', Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201–292.

Stiles, W. S., & Burch, J. M. (1959). N.P.L. colour matching investigation: final report (1958). *Optica Acta, 6,* 1–26.

Stilling, J. (1877). *Die Prüfung des Farbensinnes beim Eisenbahn- und Marine-personal.* Cassel: Theodor Fischer.

Takeuchi, T., Yoshimoto, S., Shimada, Y., Kochiyama, T., & Kondo, H. M. (2017). Individual differences in visual motion perception and neurotransmitter concentrations in the human brain. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 372,* 20160111.

Thouless, R. (1951). Individual differences in perception and their significance in psychology. In G. Ekman, T. Husén, G. Johansson, & C. I. Sandström (Eds.). *Essays in psychology dedicated to David Katz* (pp. 240–247). Uppsala: Almqvist & Wiksells.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Thurstone, L. L. (1944). A factorial study of perception, *Psychometric monographs* no. 4, Chicago, IL: University of Chicago Press and Psychometric Society.

Thurstone, L. L. (1950). *Some primary abilities in visual thinking* (Rep. No. 59)Chicago IL: University of Chicago, Psychometric Laboratory.

Torkamani, A., Topol, E. J., & Schork, N. J. (2008). Genomics pathway analysis of seven common diseases assessed by genome-wide association. *Genomics, 92,* 265–272.

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Bargary, G., Lawrance-Owen, A. J., & Mollon, J. D. (2014). An online version of the Mooney Face Test: Phenotypic and genetic associations. *Neuropsychologia, 63,* 19–25.

Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision, 5*(9), 717–740.

Weale, R. A. (1988). Age and the transmittance of the human crystalline lens. *Journal of Physiology, 395,* 577–587.

Webster, M. A. (2015a). Visual adaptation. *Annual Review of Vision Science, 1,* 547–567.

Webster, M. A. (2015b). Individual differences in color vision. In A. Elliott, M. Fairchild, & A. Franklin (Eds.). *Handbook of color psychology* (pp. 197–215). Cambridge University Press.

Webster, M. A., Halen, K., Meyers, A. J., Winkler, P., & Werner, J. S. (2010). Colour appearance and compensation in the near periphery. *Proceedings of the Royal Society B-Biological Sciences, 277*(1689), 1817–1825.

Webster, M. A., & MacLeod, D. I. (1988). Factors underlying individual differences in the color matches of normal observers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 5*(10), 1722–1735.

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000a). Variations in normal color vision. I. Cone-opponent axes. *Journal of the Optical Society of America. A: Optics, Image Science, and Vision, 17*(9), 1535–1544.

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000b). Variations in normal color vision. II. Unique hues. *Journal of the Optical Society of America. A: Optics, Image Science, and Vision, 17*(9), 1545–1555.

Webster, M. A., Mizokami, Y., & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network, 18*(3), 213–233.

Webster, M. A., & Mollon, J. D. (1993). Contrast adaptation dissociates different measures of luminous efficiency. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 10*(6), 1332–1340.

Webster, M. A., & Mollon, J. D. (1997). Adaptation and the color statistics of natural images. *Vision Research, 37*(23), 3283–3298.

Welbourne, L. E., Morland, A. B., & Wade, A. R. (2015). Human colour perception changes between seasons. *Current Biology, 25*(15), R646–R647.

Werner, J. S., Bieber, M. L., & Schefrin, B. E. (2000). Senescence of foveal and parafoveal cone sensitivities and their relations to macular pigment density. *Journal of the Optical Society of America. A: Optics, Image Science, and Vision, 17*(11), 1918–1932.

Werner, J. S. (1982). Development of scotopic sensitivity and the absorption spectrum of the human ocular media. *Journal of the Optical Society of America, 72,* 57–58.

Werner, J. S., Donnelly, S. K., & Kliegl, R. (1987). Aging and the human macular pigment density; appended with translations from the work of Max Schultze and Ewald Hering. *Vision Research, 27,* 257–258.

Werner, J. S., Peterzell, D. H., & Scheetz, A. J. (1990). Light, vision and aging. *Optometry and Vision Science, 67,* 214–229.

Werner, J. S., & Schefrin, B. E. (1993). Loci of achromatic points throughout the life span. *Journal of the Optical Society of America A, 10*(7), 1509–1516.

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *PLoS One, 9*(8).

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision, 21*(6), 561–579.

Wilmer, J. B. (2017). Individual differences in face recognition: a decade of discovery. *Current Directions in Psychological Science, 26,* 225–230.

Wilmer, J. B., & Backus, B. T. (2008). Self-reported magic Eye™ stereogram skill predicts stereoacuity. *Perception, 37,* 1297–1300.

Wilmer, J. B., & Backus, B. T. (2009). Genetic and environmental contributions of strabismus and phoria: evidence from twins. *Vision Research, 49,* 2485–2493.

Wilmer, J. B., & Buchanan, G. M. (2009). Nearpoint phorias after nearwork predict ADHD symptoms in college students. *Optometry and Vision Science, 86,* 971–978.

Wilmer, J. B., & Nakayama, K. (2007). Two distinct visual motion mechanisms for smooth pursuit: evidence from individual differences. *Neuron, 54*(6), 987–1000.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Gerbasi, M., & Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: The example of face recognition. *Cognitive Neuropsychology, 29*(5–6), 360–392.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., & Nakayama, K. (2010). Human face recognition ability is highly heritable. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 5238–5241.

Wilmer, J. B., Germine, L., Loken, E., Guo, X. M., Chatterjee, G., Nakayama, K., ... Duchaine, B. (2010). Response to Thomas: Is human face recognition entirely genetic? *Proceedings of the National Academy of Sciences of the United States of America.*

Wilmer, J. B., Germine, L. T., & Nakayama, K. (2014). Face recognition: a model specific ability. *Frontiers in Human Neuroscience, 8.*

Winderickx, J., Lindsey, D. T., Sanocki, E., Teller, D. Y., Motulsky, A. G., & Deeb, S. S. (1992). Polymorphism in red photopigment underlies variation in colour matching. *Nature, 356*(6368), 431–433.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception, 44,* 289–300.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Henry Holt & Co. Inc.

Wright, O. (2011). Effects of stimulus range on color categorization. In Biggam, Hough, Kay, & Simmons (Eds.). *Chapter in new directions in color studies.* Amsterdam/Philadelphia: John Benjamin Publishing Company.

Wuerger, S., Xiao, K., Fu, C., & Karatzas, D. (2010). Colour-opponent mechanisms are not affected by age-related chromatic sensitivity changes. *Ophthalmic and Physiological Optics, 30*(5), 653–659.

Wyszecki, G., & Stiles, W. S. (1980). High-level trichromatic color matching and the pigment-bleaching hypothesis. *Vision Research, 20*(1), 23–37.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics, 88*(1), 76–82.

Yovel, G., Wilmer, J. B., & Duchaine, B. (2014). What can individual differences reveal about face processing? *Frontiers in Human Neuroscience, 8.*