Self-Report Mental Imagery and Social Desirable Responding

John Allbutt

School of Health Sciences, University of Salford (UK)

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, May 2019.

Contents page

List of tables and figures

Table 1.

Abbreviations used in text

Abbreviation	Term
AM	Agency management
BIDR	Balanced inventory of desirable responding
CFA	Confirmatory factor analysis
CIDR	Comprehensive inventory of desirable responding
CM	Communion management
EEG	Electroencephalogram
EMS	Egoistic and moralistic self-enhancement scale
E-SDR	Egoistic social desirable responding scale
ERG	Electroretinogram
ERPs	Event-related potentials
FMM	Fixed factor modelling
HEXACO-PI-R	HEXACO personality inventory – revised
IM	Impression management
IRM	Item response modelling
MC	Marlowe-Crowne social desirability scale
MIRM	Multidimensional item response modeling
MMPI	Minnesota multiphasic personality inventory
MTMM	Multitrait multimethod
M-SDR	Moralistic social desirable responding scale
NEO-PI-R	NEO-personality inventory revised
OCQ	Overclaiming questionnaire
OSIQ	Object-spatial imagery questionnaire
OSIVQ	Object-spatial imagery and verbal questionnaire
Psi-Q	Plymouth sensory imagery questionnaire
QMI	Questionnaire upon mental imagery
rCBF	Regional cerebral blood flow
SD	Self-deception'
SDD	Self-deceptive denial
SDE	self-deceptive enhancement
TVIC	Test of visual imagery control
UIRM	Unidimensional item response modeling
VOSI	Vividness of object and spatial imagery questionnaire
VMIQ	Vividness of movement imagery questionnaire
VMIQ-2	Vividness of movement imagery questionnaire version 2
VVIQ-1	Vividness of visual imagery questionnaire version 1
VVIQ-2	Vividness of visual imagery questionnaire version 2
VVQ	Verbalizer-visualizer questionnaire

Word count

	Word count
Abstract	297
Introduction	1019
Literature review	27 315
Introduction & literature review total	28 334

Critical review
Micro-level review

Paper 1: Allbutt, Shafiullah and Ling (2006)	2 995
Paper 2: Allbutt, Ling and Shafiullah (2006)	1 349
Paper 3: Allbutt, Ling, Heffenan and Shafiullah (2008)	1 322
Paper 4: Allbutt, Ling, Rowley and Shafiullah (2011)	2 132
Paper 5: Allbutt and Ling (2009)	2 135
Micro-level review sub total	9 959
Macro level review	5 917
Critical review total [micro + macro]	15 876
Thesis total	44 210 (excluding words in the 5 papers)

Acknowledgements

I would like to thank the following people:

My friends, family, supervisor and co-authors for your kindness, love and support, I could not have made it this far without your help. Also all the individuals who gave up their time to answer questionnaires for me.

<u>Thank you</u> all.

Candidate & co-authors contributions to the papers

The tables below show the contribution of authors to the five papers.

Paper 1: Allbutt, Shafiullah and Ling (2006)

Contribution of authors to paper 1

Author	Affiliation	Approx. % contribution	Roles
John Allbutt	University of Salford	90%	• Read relevant literature • Originator of paper's ideas • Sought ethics approval • Selected & prepared questionnaires • Collected data • Inputted data into SPSS and conducted analysis • Wrote draft of paper • Edited paper • Corresponding author with journal
Mohammed Shafiullah	Leeds Metropolitan University	2.5%	• Proofread paper
Jonathan Ling	Keele University	7.5%	• Acted as mentor & 'sounding board' for ideas • Gave comments on drafts of paper & revisions

Paper 2: Allbutt, Ling and Shafiullah (2006)

Table 3.
Percentage contribution of authors to paper 2

Author	Affiliation	Approx. % contribution	Roles
John Allbutt	University of Salford	90%	• Read relevant literature • Originator of paper's ideas • Sought ethics approval • Selected & prepared questionnaires • Collected data • Inputted data into SPSS and conducted analysis • Wrote draft of paper • Edited paper • Corresponding author with journal
Jonathan Ling	Keele University	7.5%	• Acted as mentor & 'sounding board' for ideas • Gave comments on drafts of paper & revisions
Mohammed Shafiullah	Leeds Metropolitan University	2.5%	• Proofread paper

Paper 3: Allbutt, Ling, Heffenan and Shafiullah (2008)

Table 4.
Percentage contribution of authors to paper 3

Author	Affiliation	Approx. % contribution	Roles
John Allbutt	University of Salford	85%	• Read relevant literature • Originator of paper's ideas • Sought ethics approval • Selected & prepared questionnaires • Collected data for 2 studies • Inputted data into SPSS and conducted analysis • Wrote draft of paper • Edited paper • Corresponding author with journal
Jonathan Ling	Keele University	7.5%	• Acted as mentor & 'sounding board' for ideas • Gave comments on drafts of paper & revisions
Thomas M. Heffernan	Northumbria University	5%	• Collected data for 1 study • Proofread paper
Mohammed Shafiullah	De Montfort University	2.5%	• Proofread paper

Paper 4: Allbutt, Ling, Rowley and Shafiullah (2011)

Table 5.
Percentage contribution of authors to paper 4

Author	Affiliation	Approx. % contribution	Roles
John Allbutt	University of Salford	85%	• Read relevant literature • Originator of paper's ideas • Sought ethics approval • Selected & prepared questionnaires • Inputted data into SPSS and conducted analysis • Wrote draft of paper • Edited paper • Corresponding author with journal
Jonathan Ling	University of Sunderland	7.5%	• Acted as mentor & 'sounding board' for ideas • Gave comments on drafts of paper & revisions
Mohammed Shafiullah	De Montfort University	2.5%	• Proofread paper
Martin Rowley	Keele University	5%	• Collected data for study 1 • Proofread paper

Paper 5: Allbutt and Ling (2009)

Table 6.
Percentage contribution of authors to paper 5

Author	Affiliation	Approx. % contribution	Roles
John Allbutt	University of Salford	95%	• Read relevant literature • Originator of paper's ideas • Sought ethics approval • Selected & prepared questionnaires • Collected data • Inputted data into SPSS and conducted analysis • Wrote paper • Corresponding author with conference • Presented paper
Jonathan Ling	University of Sunderland	5%	• Acted as mentor & 'sounding board' for ideas

Abstract

The relationship between scores on self-report imagery questionnaires and sub-types of socially desirable responding was investigated across five papers. Paulhus (2002) has proposed a distinction between two forms of social desirable responding: the egoistic bias which is the tendency to claim positive social and intellectual qualities, and the moralistic bias which is the claiming of positive moral qualities. The research made use of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002) which has separate scales to measure both biases.

The papers found the following: 1) 'greater' imagery ability was always associated with higher social desirable responding; 2) correlations with self-report imagery scales and egoism were always higher than those with moralism; 3) the magnitude of correlations between imagery scales and egoism was generally in the r= .20 to .35 range and so sometimes exceeded a r=.25 criterion for acceptability set by McKelvie (1994); 4) significant correlations between imagery properties and egoism were seen for scales measuring vividness of visual imagery and for some other visual properties of images such as ease of image generation; 5) the size of imagery scale–egoism correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which the imagery property was desired; 6) the imagery scale–egoism correlations were of a similar size for imaging real world items and spatial ability test type items; and 7) the imagery scale–egoism correlations were largely independent of correlations with trait anxiety.

The findings suggest that the self-report imagery questionnaire–egoistic bias correlations mostly reflect distortion of scores, and that the size of the relationship is stronger than previously thought. However, the size of this relationship does not appear to reach to the extent where it impacts notably on imagery questionnaire performance.

Introduction

The research reported here is carried out in the area of individual differences in cognition, specifically individual differences in mental imagery. Mental images are quasi-perceptual mental representations of the physical world that can occur in all sensory modalities (Eysenck, 2012; Pearson, Deeprose, Wallace-Hadrill, Burnett Hayes & Holmes, 2013; Thomas, 2003). Mental images are an important aspect of human thought and they provide a useful resource for cognitive functions such as perception, memory, planning, navigation, learning, decision making and problem solving (Eysenck, 2012; Pearson, Naselaris, Holmes & Kosslyn, 2015). Visual imagery has received the most attention from psychologists and refers to the experience of mentally visualising the appearance of something usually without it being present. Individual differences in imagery experience, such as its vividness, are often reported by individuals. Psychologists have commonly assessed these differences using subjective self-report imagery scales that request participants to form mental images and then rate some aspect of their imagery experience. Self-report imagery questionnaires have appeal because they offer the possibility of a simple, easy and flexible way to measure differences in imagery ability. People can fill them in anywhere there is quiet and no timing is necessary. Further, for some aspects of imagery experience, such as image vividness, they appear to be the only measurement option available.

Self-report imagery questionnaires have been used in pure research (McKelvie, 1995a & 1995b) and applied settings such as clinical psychology (Pearson et al., 2013), sports psychology (Munzert & Lorey, 2013) and medical surgeon training (Sevdalis, Moran & Arora, 2013). However, the usefulness of self-report imagery questionnaires has been debated (e.g. Lacey & Lawson, 2013) and correlations with objective spatial tests are usually low (Dean & Morris, 2003).

The thesis describes a research project carried out across five papers, four that are published journal articles in peer reviewed journals, and one peer-reviewed conference paper. The research investigated the relationship between scores on self-report imagery questionnaires and sub-types of socially desirable responding. Paulhus (2002, p.50) defines social desirable responding as "…the tendency to give overly positive self-descriptions". Reviewers have typically argued that self-report imagery questionnaires are not heavily influenced by social

desirable responding. For example, in a meta-analytical review, McKelvie (1995b) concluded that the most widely used imagery scale, the VVIQ-1 (Vividness of Visual Imagery Questionnaire, version 1; Marks, 1973), was not seriously affected by social desirable responding. However, this thesis argues that past research on self-report imagery questionnaires and social desirable responding is limited.

The five papers making up the thesis systematically explored the relationship between scores on self-report imagery questionnaires and measures of social desirable responding in a series of samples of psychology students tested anonymously in classroom settings. Their results provide empirical data to support the position that:

1) Social desirable responding should be conceptualised as a multifactorial phenomenon with the need for distinction between egoistic and moralistic biases being crucial (Paulhus, 2002; Paulhus & Trapnell, 2008). The egoistic bias relates to the tendency to claim positive social and intellectual qualities, while the moralistic bias relates to the claiming of positive moral qualities;

2) Respondents answering self-report imagery questionnaires most likely view imagery as a valued ability being tested rather than as a 'value neutral' aspect of their cognition, and therefore will value responses they perceive to indicate 'better' imagery ability. As such egoistic forms of social desirable responding are more of a 'threat' to the validity of self-report imagery questionnaires;

3) Past studies have underestimated the size of the relationship between self-report imagery questionnaires and social desirable responding because they have used inappropriate measures of social desirable responding that primarily measure moralistic forms of bias. The research conducted in the thesis papers made use of the Balanced Inventory of Desirable Responding (BIDR; Paulhus 1994, 1998, 2002) which has separate scales to measure the two biases and so does not have this limitation. Also, most imagery researchers have conceptualised social desirable responding as distortion and overlooked the possibility that it might reflect substantive relationships with imagery;

4) The series of five research papers reported in this thesis have produced evidence that responses to self-report imagery questionnaires show small but consistent positive correlations with the egoistic form of social desirable responding in the approximate range of $r=.20$ to $.35$, with 'greater' imagery ability associated with higher levels of the

egoistic bias;

5) While there is evidence e.g. Connelly and Chang (2015) that questionnaire measures of social desirable responding are not pure measures and tap a combination of substance and distortion, the findings of the research reported here suggest that the self-report imagery questionnaire–egoistic bias correlations mostly reflect distortion. Specifically: 1) the size of imagery scale–egoistic bias correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which an imagery property was desired; and 2) the finding that the correlation between the Vividness of Visual Imagery Questionnaire (Marks, 1974) and the egoistic bias was found to be largely independent of its correlation with trait anxiety;

6) However, while the thesis data suggests that the relationship between self-report imagery questionnaire responses of valued imagery properties and measures of social desirable responding is larger than previously thought, the size of this relationship does not appear to reach to the extent where it impacts notably on imagery questionnaire performance. Despite the relationship with social desirable responding being larger than previously thought, self-report imagery questionnaires do still correlate with other related measures. Furthermore, in the thesis, the results of a partial correlational analysis reported in paper 1 suggests that removing the variance associated with social desirable responding from the relationship between imagery questionnaires largely leaves the effect sizes unaffected. However, this conclusion needs to be confirmed using more advanced statistical techniques such as Multiple Item Response Modelling (MIRM) which have the capacity to weight items and take account of cross-loadings across the items of different questionnaires.

It is argued that the five papers making up the thesis represent original peer reviewed research that has increased our understanding of the relationship between self-report imagery questionnaires and social desirable responding and extended the forefront of the mental imagery questionnaire literature.

The thesis is organised into the following sections:

- General introduction;
- Literature review;
- Introduction to the five papers;
- Presentation of five research papers, four journal articles published in peer reviewed journals and one conference poster. Following each paper is a micro level evaluation of each paper describing the paper's strengths and limitations;
- Macro critical evaluation.

The literature review which follows, comprises of an introduction to the area of mental imagery followed by updated literature reviews of the self-report visual imagery questionnaire literature and social desirable responding literature. The literature on self-report visual imagery questionnaires and social desirable responding was located from a search of the PsycINFO database. Prior to taking on this thesis the author had conduced database searches for the five articles. The current search was conducted on 25th July 2017 and included the database content from this dating back to 2009. Search terms used for imagery included 'imagery questionnaire', 'vividness and imagery', 'VVIQ'. Search terms used for social desirable responding included 'social desirable responding', 'social desirability' and 'balanced inventory of desirable responding', 'Paulhus deception scales', 'Comprehensive Inventory of Desirable Responding', 'egoistic bias', 'agentic bias' and 'self-deceptive enhancement'. Reference lists of articles produced by the search were studied as were key author websites. Also the content lists of journals dedicated to mental imagery were looked at for relevant articles back to 2009. These journals were the Journal of Mental Imagery and Imagination, Cognition and Personality.

Literature review

1) Mental imagery

This section briefly introduces the topic of mental imagery. It describes the history of research into the topic and the key research question of 'What are mental images?' and the extent to which they can be considered to be 'picture-like' in nature. The section finishes by discussing applications of mental imagery.

Mental images are quasi-perceptual mental representations of the physical world that can occur in all sensory modalities (Eysenck, 2012; Pearson, Deeprose, Wallace-Hadrill, Burnett Hayes & Holmes, 2013; Thomas, 2003). Visual imagery has received the most attention from psychologists. A visual image can be defined as a mental representation that is experienced as if it were a picture in one's head. The representation can be of events experienced in the past or the construction of imagined events in the future. Typically a visual image is experienced as being adjustable as if it really were a picture, and as being less vivid than its corresponding visual percept. Common examples of visual imagery include visualising a scene while reading a book and dreaming or daydreaming about a sporting victory. People report differences in the vividness of their mental imagery, ranging from those that experience imagery as if they were perceiving to those that report experiencing no imagery at all.

1.1)   Mental imagery research and applications

Historically mental imagery and its role in human thought processes has been discussed and 'researched' for over 2000 years. Aristotle regarded imagery as the medium of thought (Leahy, 2013). He called images 'phantasma' and thought of them as being like paintings or wax impressions. Orators in ancient Greece used imagery-based mnemonics to help memorise speeches.

Francis Galton (1883), a leading scientific figure in British 19th century society and half-cousin of Charles Darwin, conducted one of the earliest studies into psychological individual differences. He gave a questionnaire he had designed to one hundred successful male

acquaintances which asked them to imagine the appearance of their breakfast table from that morning or some other familiar object or scene. He reported that that many of his participants experienced very vague imagery or no imagery at all and so concluded that over-readiness to form clear mental images hampered a person's intellect. Galton's study was successful in showing the potential of a questionnaire approach for studying mental imagery.

From this early Victorian era work, the study of visual imagery had a controversial status through parts of the twentieth century (Leahy, 2013). The rise of behaviourism lead to the rejection of the study of mental imagery because the private nature of experiencing images meant they could not be seen, counted or controlled by others. But in the 1960's, the popularity of behaviourism waned and cognitive psychology emerged. Psychologists accepted that it was necessary to study how information is acquired, processed and stored in order to fully understand our minds, so psychologists began to study imagery again.

The key research question that psychologists since the 1970's have tried to answer is: 'What are mental images?' (Eysenck, 2012; Kosslyn, 1994). When we experience visual images they seem to be like pictures in our mind that we are actually perceiving in the same way as we might see external things in our environment, but are they really like pictures and if so in what way? Answers to these kinds of questions fall into two types:

1)   Yes, images do have picture-like qualities, they operate in their own special medium and the way they work is quite different from verbal representations;

2)   No, they do not have picture-like qualities. Underlying both verbal representations and imagery representations are propositional representations (a single universal, abstract, amodal code or 'mentalese' which is neither verbal nor visual).

Research into this issue has come to be known as 'the imagery debate'. Finke (1989) suggested 5 principles of visual imagery that are meant to describe the fundamental nature and properties of visual imagery, and offered evidence from cognitive psychological studies in support of them:

Principle 1: 'Imagery is fundamentally equivalent to perception to the extent that similar mechanisms in the visual system are activated when objects or events are imaged as when the same objects or events are actually perceived' (Finke, 1989, p41). This principle relates to whether parts of the brain which are active during visual perception are also active during visual imagery;

Principle 2: 'The structure of mental images corresponds to that of the actual perceived objects, in the sense that the structure is coherent, well organised, and can be reorganised and reinterpreted' (Finke, 1989, p. 120). This principle has to do with the parts of an image that make its structure, for example, the ways that the different part of an image are put together to form an image's structure and whether the parts of an image can be moved around to change its nature and whether we can recognise what it has changed into. For example, drawings of objects are put together piece by piece and objects that have more parts and are larger take longer to draw. There also seem to be parallels to these phenomena in imagery because images appear to be constructed piece by piece and take longer to generate when they are more complex or larger;

Principle 3: 'Mental imagery is instrumental in retrieving information about the physical properties of objects, or about physical relationships among objects, that was not explicitly encoded at any previous time' (Finke, 1989, p. 7). When we perceive things visually and store memories of what we see, this information may include more than what we were actually attending to or intending to remember. This 'extra' information can be accessed in our images. So, for example, in the same way that a person can look at a visual scene or a picture again and see something new, so they can mentally scan a mental image and 'see' new things;

Principle 4: 'The spatial arrangement of the elements of a mental image corresponds to the way objects or their parts are arranged on actual physical surfaces or in an actual physical space' (Finke, 1989, p. 61). This principle has to do with how spatial information such as location, distance, and size is represented. For example, visually scanning between different parts of a visual image appears to mirror the process of visually scanning between different parts of an actual visual scene in that the time taken to scan reflects the distance between the elements; and

Principle 5: 'Imagined transformations and physical transformations exhibit corresponding dynamic characteristics and are governed by the same laws of motion' (Finke, 1989, p. 93). This principle relates to the movement and transformation of images. For example, the mental rotation of visual images shares some of the features of the physical rotation of objects such as greater rotations take longer periods of time.

Equally, though, there are limits to the picture metaphor. For example, Pylyshyn (1973) has argued that pictures are different to images in several ways:

1) It is possible to look at a picture without knowing what it is i.e. you can struggle to recognise or 'make out' what it is you are looking at, but you cannot do this with an image;

2) Pictures and images are disruptable and disrupted in different ways. You can cut a picture in any two arbitrary halves, but images are organised more meaningfully so when parts are forgotten, only meaningful parts disappear. For example if you image a dog smoking a cigar, later on you may only remember the dog, not the dog smoking half a cigar; and

3) Images seem more easily distorted by the viewer's interpretations. For example, Carmichael, Hogan and Walter (1932) found that viewer's visual images can be distorted by the verbal labels they are given. They presented 'ambiguous' shapes to participants and varied the verbal label that went with them e.g. 'crescent moon' versus 'the letter C'. When participants later recalled the shapes by drawing them they altered the shapes so that they became more like the verbal label they had been presented with.

Psychological research in the 1970's and 80's attempted to investigate the issues raised by the imagery debate using introspective and behavioural techniques. However, researchers such as Anderson (1978) argued that these techniques were unable to distinguish conclusively whether images genuinely have picture-like qualities or are propositional in nature. For example, it is possible that participants tacit knowledge of mental imagery might influence the nature of their responses on experimental tasks such that their responses show picture-like qualities. But, from the 1990's onwards psychologists' ability to tackle

these issues benefited greatly from the use of brain scanning and brain stimulation techniques, and studies investigating the functioning of neurological patients. Ganis (2013) describes the findings of this research and how it provides evidence that visual perception and visual imagery share some brain structures and neural processes, and that visual imagery operates in its own special medium and is not simply propositional in nature. He summaries how neuroscientific research provides evidence that 'mental images of shapes sometimes activate topographically organized areas of the early visual cortex' (p16) such as V1 that are responsible for the processing of low-level visual details including spatial frequency and edges, but that such activity is weaker than for visual perception. Later visual areas in the ventral system also show overlapping involvement in both visual perception and visual imagery, for example in the processing of faces and animals. In these later visual areas activity from visual imagery involves a sub-set of the areas that are used in perception, and for early areas, activity is weaker than that seen for perception. It appears visual images, like percepts, reply on representations that are produced by all levels of the visual processing hierarchy, and that strength of activity becomes more similar higher up the visual processing hierarchy. Ganis reports that studies of the degree of overlap in neural systems estimate that visual perception and visual imagery share over 90% of their brain regions. So visual imagery can be thought of as a weak, or noisy, form of top-down perception.

Mental imagery has applications in a range of applied settings, particularly within sport and clinical settings. Munzert and Lorey (2013) describe how mental imagery can play a role in two main areas of sport: motor learning and preparation for sporting events. There is strong evidence (experimental and correlational) that mental imagery can be used to improve motor performance and acquisition of a range of motor skills and appears able to partially replace physical practice. This is especially true where the skills include cognitive elements and when mental practice is alternated with physical practice, although physical practice on its own is more effective than mental practice on its own. The ability of imagery to have an effect has been found to depend on an athlete's ability to form vivid mental images (e.g. Munroe, Giacobbi, Hall & Weinberg, 2000). Effect sizes for mental practice are small or moderate, and so are likely to be of more importance to elite athletes where even small gains are significant. Notable physical strength gains have also been observed from mental practice which remain stable over short periods of time such as a week. The gains are thought to be due to neural effects, for example it has been suggested that mental practice

may activate neural pathways from the brain to the muscles to prime movements or facilitate coordination, or it may act to facilitate memory for key elements of a task. There is also evidence from correlational studies that mental imagery can also help athletes' preparation prior to sporting events. Imagery can be used to enhance motivation and confidence, and reduce anxiety so that arousal levels are at their optimum for performance, and aid athletes in rehabilitation after injuries. For example, using visual imagery to imagine oneself being successful in a competition. However, experimental work is needed in the future to allow causal interpretations of this data.

Some other applied settings where mental imagery has been discussed include the training of medical techniques which involve psycho-motor skills to trainee doctors and nurses (Sevdalis, Moran & Arora, 2013; Wright, Hogard, Ellis, Smith & Kelly, 2008). Here mental simulation methods have been used in the learning and practice of surgical techniques by doctors such as suturing and knot tying, and they have been argued to offer a cost-effective alternative to the traditional surgeon apprenticeship approach i.e. repeated supervised practice. Also pre-registration student nurses mastery of blood pressure measurement technique has been shown to be enhanced by imagery based practice.

Pearson, Deeprose, Wallace-Hadrill, Burnett Hays and Holmes (2013) also describe how mental imagery plays a role in a range of mental disorders. In post-traumatic stress disorder (PTSD) flashbacks (intrusive mental images) are thought to occur because of a survivor's impaired information processing during a traumatic event. People experiencing a social phobia often report viewing recurrent distressing imagery of past social events from an observer perspective which cause anxiety and act to maintain the disorder. In schizophrenia hallucinatory imagery can overtake the perception of reality when it occurs in its most extreme form. Depression has been associated with intrusive negative imagery for past events and the lack of positive imagery for the future, while in bipolar disorder it has been suggested that imagery plays a role in moderating mood swings between depression and mania. Laing, Morland and Fornells-Ambrojo (2016) observed that the perceived likelihood of generating positive, future oriented imagery and its vividness was negatively correlated with levels of depression and social anxiety in people who had early psychosis.

Pearson et al (2015) describe how intrusive negative imagery experiences are more than just simple symptoms of a disorder because they can act as cognitive mechanisms that help maintain it. Important aspects of such imagery include: 1) images can trigger stronger emotional reactions than verbal thoughts; 2) images, such as hallucinations, have a quality of appearing to be 'real' even when known not to be; and 3) images can affect our perceptions and memories. For example, repeatedly imaging a feared event can lead to the perception that it is more likely to occur in the future.

Mental imagery has also played a role in the treatment of mental disorders (Comer, 2015). Guided imagery is a technique where a person images thoughts suggested by a therapist (either present or through an audio recording), which can help aid relaxation and positive thinking. The therapist instructs the person to image themselves in a place where they have experienced feelings of relaxation and comfort e.g. on holiday, in the garden. The person is told to breath in and out deeply, relax their body and notice the sights and sounds around them. In the case of imagery desensitisation techniques, patients with anxiety disorders experience exposure to their fears by visualising them. In the area of health psychology, Kwekkeboom (2000) reported that cancer patients who had received guided imagery and reported it to be helpful to them in coping with pain to score higher on self-report measures of imagery vividness than patients who reported not experiencing such benefits. In imagery re-scripting therapy a person's thoughts around a clinically relevant issue are reimaged in a more adaptive way. For example, therapy with a person with a social phobia might encourage them to replace images of poor social performance with successful performance (Hirsch, 2003). In eye movement desensitization therapy (EMDR) individuals re-image past emotional experiences while making some form of bilateral sensory input such as side-to-side eye movements. The sensory input is thought to lessen the emotional impact of the images, making them easier to process (Van den Hout, 2012). Imagery has also been used in conjunction with hypnosis to attempt to retrieve lost memories. The therapist hypnotises a person and asks them to visualise themselves back in the past, then they ask them questions to try to elicit lost memories. However, this is a controversial technique as it may lead to the 'recovery' of false memories (Davis & Loftus, 2014).

1.2)   Summary of key points

- Mental images are quasi-perceptual mental representations of the physical world that can occur in all sensory modalities. Most mental imagery research has focused on visual imagery.

- Historically mental imagery and its role in human thought processes has been discussed and 'researched' for over 2000 years, although researchers' attitudes towards the phenomenon, its importance and psychologists' ability to study it have varied.

- The key research question that psychologists since the 1970's have tried to answer is: 'What are mental images?' For visual imagery, the positions psychologists have taken on this issue are: Yes, images do have picture-like qualities, they operate in their own special medium and the way they work is quite different from verbal representations; No, they do not. Underlying both verbal representations and imagery representations are propositional representations. Research into this issue has come to be known as 'the imagery debate'.

- Some researchers have articulated principles that describe the fundamental nature of visual imagery and its unique properties, while others have drawn attention to the differences between visual images and pictures.

- Research using modern psychophysiological and neurological approaches provide evidence that visual perception and visual imagery share brain structures and neural processes, and that visual imagery operates in its own special medium. So visual imagery can be thought of as a weak, or noisy, form of top-down perception.

- Mental imagery has applications in a range of applied settings, particularly within sport and clinical settings. There is evidence that the vividness of imagery influences some real-world imagery effects.

2) Self-report imagery scales

The purpose of the self-report imagery questionnaire section is to provide an up-to-date description of work in the area, introduce the reader to key imagery models, questionnaires and issues around their use which can be later drawn on in the critical review. Some of the imagery questionnaires described here were used in the papers making up the thesis. The

section starts with a detailed description of the Vividness of Visual Imagery Questionnaire version 1 (VVIQ-1, Marks 1973) and its research. This questionnaire has been the most extensively used self-report test and it was used in three of the five papers that make up the thesis. As such aspects of the test's reliability and validity are described in some depth. The end of the section introduces other important self-report imagery tests including some, such as the Shapes Questionnaire (Dean & Morris, 1991 & 2003), which have been developed directly out of imagery theory. Several of issues highlighted in this section will be picked up again in the in the final section of the literature review where self-report imagery questionnaires and social desirable responding are brought together, and in the critical review when discussing the five papers.

2.1) Types of imagery tests

Two types of test have been used to assess individual differences in imagery ability. These have been classified as either: 1) subjective self-report imagery tests, or 2) objective behavioural tests (Katz, 1983).

In subjective self-report tests participants are asked to introspect on their imagery experience and comment on selected aspects such as the vividness of their imagery, the ability to control their images or the extent to which they use imagery in certain specified situations such as when reading a book, solving a problem, thinking about a person's face or appearance. Numerous self-report tests have been developed with most focusing on visual aspects of mental imagery. Examples of these tests include the Vividness of Visual Imagery Questionnaire version 1 (VVIQ-1, Marks 1973) and the Verbalizer-Visualizer Questionnaire (VVQ, Richardson, 1977a). Self-report imagery questionnaires have the advantage of being quick and easy to use and for some aspects of imagery experience, such as image vividness, appear to be the only measurement option available.

In objective behavioural tests participants are typically given pictures of abstract geometric shapes and asked to perform some mental spatial manipulation upon them. When they have done this they select a response from among a set of possible alternative depictions of the shape. Usually, only one of the possible responses shows the shape as it would appear after the manipulation had been correctly performed. Examples of these tests include the Space

Relations Test (Bennett, Seashore & Wesman, 1974) and the Vandenberg Mental Rotation Test (Vandenberg & Kuse, 1978). Despite the less obvious role of imagery in these tests from a phenomenological perspective, it is typically assumed that they do involve imagery.

Observations that the two kinds of test at best show only weak inter-correlations e.g. (Poltrock and Brown, 1984) and load on orthogonal factors in factor analyses (Richardson, 1977b) has been interpreted to suggest that they are measuring different imagery processes. Subjective self-report imagery tests have the advantage of having a more direct link to the subjective experience of imagery, behavioural tests on the other hand have a more objective scoring method.

2.2) The Vividness of Visual Imagery Questionnaire

The most widely used self-report imagery scale is the Vividness of Visual Imagery Questionnaire version 1 (VVIQ-1, Marks 1973). The VVIQ aims to assess vividness of visual imagery using 16 items. Participants form a series of images of a friend or relative's face, the rising sun, a shop, and a country scene, with four images generated for each item. Ratings of image vividness are made on a five-point scale ranging from 1, (Perfectly clear and vivid as normal vision), to 5, (No image at all, you only 'know' you are thinking of the object). Traditionally, participants make 16 ratings with their eyes open, then 16 ratings with their eyes closed and the two sets of scores are added together.

Marks (1998) has argued that the VVIQ-1 is designed primary to be a measure of trait differences in visual imagery vividness, however, it can also be used to measure states of consciousness or mental contents. Marks introduced the VVIQ in a 1973 paper where he investigated whether the vividness of visual imagery played a functional role in visual memory.

McKelvie (1995a) produced an extensive critical quantitative review of the work that had been amassed on the VVIQ-1 in an article in the Journal of Mental Imagery which drew on over 250 pieces of empirical work. McKelvie used meta-analytical techniques to present estimates of key aspects of the test such as its reliability and validity. This was followed by a series of open peer commentaries from researchers active in the field. McKelvie (1995b)

then responded to the comments and made revisions to his estimates. The following review of the VVIQ-1 is based around McKelvie's work.

2.2.1) The VVIQ-1 items

The VVIQ is based on the view that visual imagery is a quasi-perceptual conscious experience. McKelvie (1995b) describes the most common themes listed by researchers in their assessment of what the VVIQ measures as being: similarity to perception, clarity, detail, liveliness and brightness. Thus, the VVIQ appears to equate vividness of visual imagery with closeness to the actual experience of seeing.

A unique aspect of the questionnaire is that participants are asked to make their responses twice, once with their eyes open and then once with their eyes closed. McKelvie (1995b) reported an effect size r of .04, 95% CI [-.09, .17] for eyes open versus closed suggesting no reliable difference existed between the two rating methods. Because of this McKelvie recommended the use of only one set of ratings with the choice left up to the participant.

The VVIQ response scale does not have equal numbers of vivid and non-vivid response options. The first three options on the response scale refer to different degrees of vivid imagery, the fourth to vague imagery and the fifth response to the inability to form an image at all.

2.2.2) Factors affecting the vividness of visual imagery

While not directly involving the VVIQ, Hishitani, Miyazaki and Motojama's (2011) have proposed a model of the mechanisms responsible for the vividness of visual imagery which is based on findings from cognitive psychology and neuroscience (see Figure 1). Hishitani et al define vividness of visual imagery as the amount of perceptual information contained in an image. Images that have more colour, detail, elements, elaborate structure and that are larger in size will be experienced as more vivid. Vivid images thus have a stronger resemblance to the reality they represent than non-vivid images and they are better able to function as a substitute for a percept in cognitive tasks such as tests of visual memory or problem solving. Studies support Hishitani et al's view, for example Cornoldi, De Beni,

Giusberti, Marucci and Mazzoni (1991) who explored the relationship between different features of an image, such the presence of colour, and the image's vividness.

In Hishitani et al's model the visual image generation process starts with the intention to generate an image of an object. Information in long-term memory (LTM) about the object is stored in terms of interconnected perceptual, affective, motor information and meaning. This organization is presented in terms of a pyramidal structure in the model which captures the fact that visual imagery is a composite experience, such that along with visual information, a person may also experience emotion and movement associated with the percept. Perceptual information about the visual appearance of the object to be imaged is transferred from LTM to working memory along an information channel. Flow of perceptual information along the channel is controlled by a mechanism called the 'suppressor'. Emotional information in LTM may also be retrieved and used in a process which computes the emotional value of the content to be imaged at a particular point in time / situation. If this emotional value is negative, the suppressor mechanism restricts the flow of information along the information channel which leads to a less vivid image. This element of the model captures the fact that images with negative content are reported to be less vivid than images with positive content. The suppressor mechanism acts to protect individuals from the emotional impact of negative imagery. Motor information and meaning information associated with the imaged object in LTM can also influence image vividness by acting on the image construction process. For example, the image formed of a favourite wilderness area might be perceived as more vivid than that of a prototypical neutral scene because of its personal relevance. A structure called the 'closer' acts to inhibit visual perception while imagery is taking place. The final stage of the model is the inspection of the visual image. The model thus captures how visual perception, affect, motor and meaning information and have all been reported to affect the vividness of visual imagery.

Psychophysiologcal functional Magnetic Resonance Imaging (fMRI) research by Japanese researchers has suggested brain locations for the suppressor mechanism and the closer. Motoyama, Matsumura and Hishitani (2010, cited in Hishitani et al, 2011) found part of the left posterior cingultate gyrus to be more active during negative imagery and for its activity to be correlated with imagery vividness (r=.82). As such, Hishitani et al suggested that it might be the site of the suppressor mechanism. The cingulate gyrus is a semi-circular fold of

brain tissue that covers the corpus callosum and forms part of the limbic system. Hishitani (2008, cited in Hishitani et al, 2011) found that a part of the cuneus, an occipital lobe brain area involved in primary visual processing, was less active during an imagery task than during perception, and activation in the area was negatively correlated with imagery vividness. Thus, this may be the site of the closer.
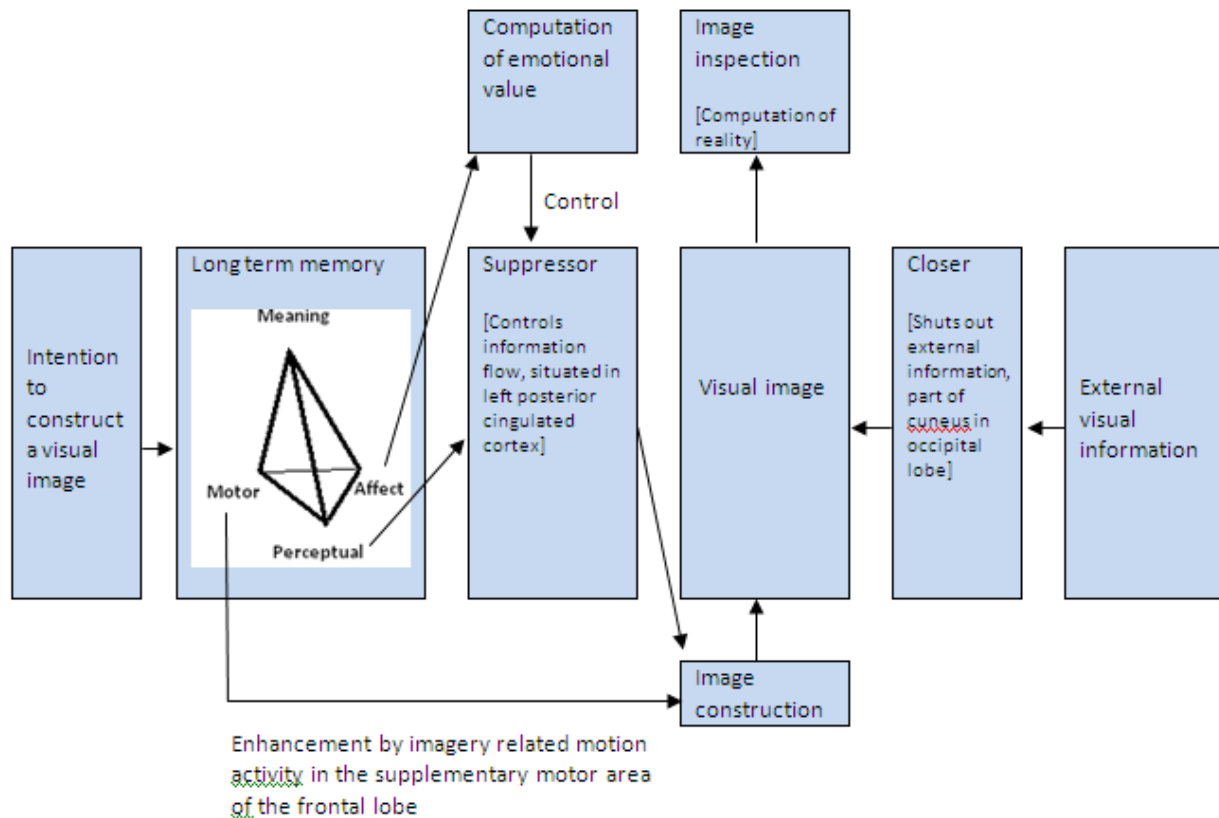
Figure 1. Hishitani, Miyazaki and Motojama's (2011) model of the mechanisms responsible for intrapersonal differences in the vividness of visual imagery (modeled from Hishitani et al, 2011)

Hishitani et al proposed two possible reasons for interpersonal differences in image vividness: 1) vivid imagers have greater capacity in the information channel responsible for transporting visual information from LTM to working memory; 2) vivid imagers possess greater visual information in their LTM. Hishitani et al argue that the first possibility is more likely because interpersonal differences in vividness still exist when images are formed of very familiar objects and scenes such as those used in the VVIQ. Intrapersonal differences in image vividness are attributed primarily to the action of the suppressor mechanism.

2.2.3) Norms

McKelvie (1995b) reported an estimate of the mean response on the 5-point VVIQ-1 response scale based primarily on student respondents to be 2.31, 95% CI [2.09, 2.52]. Killstrom, Glisky, Peterson, Harvey and Rose (1991) collected responses from 730 American psychology students and reported their distribution of scores had a positive skew with most participants reporting that their images had been at least 'moderately clear and vivid' (a mean rating of 3). Less than 3% of their sample reported that their images were 'vague and dim' (a mean rating of 4) or could form no images (a mean rating of 5). Thus the response distribution of the VVIQ appears to differ from a typical psychometric test which has a mean response falling near to the midpoint of the response scale and has a distribution that is normally distributed. McKelvie (1995b) quotes an effect size *r* of .16 for sex, 95% CI [.02 to .31], suggesting a reliable effect for females to report more vivid imagery than males.

2.2.4) Reliability

McKelvie (1995b) reported a VVIQ-1 split-half effect size *r* of .89 for internal consistency 95% CI [.85 to .91] which suggests good internal consistency, and an effect size *r* of .76 for test-retest reliability 95% CI [.70 to .82] suggesting acceptable test-retest reliability.

2.2.5) Validity

Factor structure

Studies of the factorial structure of the VVIQ have not produced consistent results. Early studies e.g. Dowling (1973, cited in White, Sheehan & Ashton, 1977) showed a simple unitary factor structure, but later analyses have suggested the existence of a more complex picture. Both Dean and Morris (1990, cited in Dean & Morris, 1991) and Killstrom et al. have reported that the VVIQ has four underlying factors, and that these factors correspond to the four groupings of questions. To date, no Confirmatory Factor Analysis (CFA) has been reported for the VVIQ scale.

Relationships with other self-report imagery questionnaires

There is evidence of convergent validity between the VVIQ and some of the other self-report imagery questionnaires. The largest correlations are seen with tests such as the Sheehan's QMI (Questionnaire of Mental Imagery, Sheehan, 1967) which also measures vividness, but does so across other modalities in addition to vision. McKelvie (1995b) reported an effect size *r* of .65 for convergent validity with Sheehan's QMI with 95% CI [.60 to .76]. The lowest correlations are seen with tests of habitual imagery use such as the Verbilizer-Visualizer Questionnaire (VVQ; Richardson, 1977a) which assesses where respondents lie on a hypothesised single verbal – visual dimension of thinking style. There McKelvie (1995b) reported an effect size *r* of .13 with 95% CI [.06 to .20]. Thus the strength of the relationship appears to depend on the similarity of the nature of the test with the VVIQ.

Relationships with perception and memory

McKelvie (1995b) reports that reliable relationships exist between the VVIQ-1 and performance on perceptual tasks and memory tasks. In terms of perception, vivid imagers are reported to scan pictures more quickly, make larger saccades, have more consistent scan paths, have a higher rate of information pickup per fixation, make less eye movements when recalling visual images, experience stronger imagery induced visual after effects, show more accurate cross-modal matching of randomly shaped polygons, and show stronger imagery induced modifications to colour-flash electroretinograms. For memory, vivid imagers are reported to have more accurate memory for the visual details of pictures, recognize pictures more accurately, detect differences between pictures more quickly, and recall both concrete words and paired-associates more accurately. For perceptual tasks McKelvie's estimate of effect size *r* was .44 and 95% CI [.31 to .56], and for memory tasks his estimate of effect size *r* was 0.18 and 95% CI [.06 to .18].

Psychophysiological studies

Research using psychophysiological techniques has also found overlapping relationships between VVIQ-1 scores and brain activity measured using a range of different techniques. For example, Farah, Peronnet, Weisberg and Monheit (1988) asked twelve participants to

either read a briefly presented visual word or read it and generate an image of its referent and recorded event-related potential brain activity (ERPs). Farah et al observed an ERP positivity relative to the baseline conditions which was maximal around 700 msec at occipital and occipito-temporal sites in the imagery condition and was greater over the left hemisphere sites involved. Because the activity occurred in the time period which Kosslyn (1980) has reported for visual image generation times (500-2000 msec), the authors suggested that that the activity observed resulted from image generation. Farah, Steinhauer, Lewicki, Zubin and Peronnet (1988) replicated and extended this study by also measuring VVIQ scores. Twenty participants were divided up according to their scores on the VVIQ into the top ten VVIQ scorers and bottom ten. The ERP positivity was again observed and was greater for the high vividness group than the low vividness group.

Overall the psychophysiological studies provide support for the validity of the VVIQ-1 using a range of different recording techniques. They show that activity in the visual parts of the brain varies between vivid and non-vivid imagers with greater vividness being associated with greater brain activation. Further, that this activation may reach as far down the visual processing hierarchy as areas V1 and V2 (Cui, Jester, Yang, Montague & Eagleman, 2007). Studies of this type support the view that imagery involves activation of some of the same processes involved in perception, and that vividness is a meaningful property of visual imagery.

Between participant comparisons versus within participant comparisons

Kaufman (1981) and Richardson (1980, 1988) have argued that self-report imagery questionnaires are better suited to making within-participant comparisons (comparing across rating items) than between-participant comparisons (comparing ratings across participants). This is because, when these tests are used to assess individual differences, participants have to make absolute judgments about the vividness of their imagery. Participants will struggle to do this though, because they have no absolute criteria upon which to base their judgments, and no way of comparing the vividness scale that they use to those adopted by other participants because images are internal subjective experiences. Because of this state of affairs, there is no way of knowing whether different participants are rating their imagery experiences in the same way. This is not a problem when within-

participant comparisons are made since no absolute judgments are required. Here participants can compare separate images in terms of their own rating criteria and so have a basis for making comparative judgments. In line with this argument, Walczyk (1995) found within-participant statistical analyses produced larger estimates of the relationship between measures of imagery vividness and performance on tasks tapping the accuracy of visual information in the images formed than between-participant analyses.

Chara and Verplank (1986) attempted to investigate inter-participant consistency in assigning vividness ratings to visual percepts that varied in clarity. Their experiment provides empirical verification of the problems facing participants when they make ratings on the VVIQ-1. Chara and Verplank required participants to select five projector slides, from a set of 15, which corresponded to five points spaced equidistantly along a scale ranging from the clearest slide to least clear slide. The slides were of a mountain cove and varied in clarity from 'very blurry' through to 'very sharp' and in illumination from 'very dark' to 'too bright'. Participants viewed the set of 15 slides three times through and then selected the five slides they felt corresponded to the five-point scale. All participants agreed on the slides that made up the end points of the scale, but showed considerable disagreement over which slides should represent the other three points.

2.3) The Vividness of Visual Imagery Questionnaire version 2

In response to debate around the VVIQ-1, Marks (1995) produced a new version of the VVIQ called the VVIQ version 2 (VVIQ-2; Marks, 1995). The VVIQ-2 was used in paper 3 of the thesis. It has 16 additional items so that the scale samples a wider range of imagery experiences than the VVIQ-1. The additional items involve respondents generating images of a being driven in a car, being on a beach, a railway station, and in a garden with lawns. The response scale has been reversed so that the greatest vividness of imagery was scored '5' rather than '1' and all ratings were made only once with eyes closed. However, the VVIQ-2 scale has not been widely used and most authors continue to use the VVIQ-1 in their studies.

2.4) Ahsen's adapted VVIQ

Ahsen has developed an adapted version of Mark's VVIQ-1 which he called Ahsen's Adapted Vividness of Visual Imagery Questionnaire (AA-VVIQ; Ahsen, 1985). The AA-VVIQ asks respondents to image the VVIQ items while keeping in mind people who are significant to them such as their mother or father. Ahsen (1985) reports that when using these parental 'filters' respondents report dimer imagery when keeping their father in mind than when keeping their mother in mind. Ahsen (1985, 1986, 1987, 1990) has argued that there is a lack of consistency in VVIQ imagery vividness, within a single image, across the images created by a single person and when using different parental filters that is inconsistent with a trait conceptualisation of imagery vividness. As such, he has criticized Mark's predominantly trait approach to the vividness of visual imagery, and argued that research should focus on the processes that lead to the creation of an image and its experienced vividness. Ahsen also has argued that dim, or unvivid imagery, is as significant as vivid imagery and labels this phenomenon the 'unvividness paradox'. Ahsen views unvivid imagery to be a separate process to vivid imagery and one that is driven by nonconscious self-inhibitory forces which act to reduce the vividness of images. He supports this contention using a range of evidence, for example, clinical observations that patients sometimes report an image to be perceptually clear and vivid (perceptual element) but at the same time emotionally dull and flat (affective element) and so experienced as unvivid.

2.6) Modern theory driven approaches to measuring subjective aspects of mental imagery

Some researchers e.g. Lacey and Lawson (2013), have argued that vividness may not be the best measure of subjective imagery experience and that questionnaires based on theoretical models of the underlying nature of mental imagery may be more appropriate. Lacey and Lawson argue that vividness is only weakly related to such models, and using only vividness as the measure of imagery ignores the fact that it may be a multifaceted phenomenon. One recent theory driven self-report imagery questionnaire is the Shapes Questionnaire which was used in was used in paper 2 of the thesis.

Dean and Morris (1991, 2003) developed the Shapes Questionnaire to assess a wide range of self-reported visual imagery properties including vividness, but also 17 other aspects of the

imagery experience. The ratings are of two broad types: 1) ratings of the subjective ease of hypothesized processes, such as the ease of generation, maintenance or rotation of an image, and 2) ratings of the pictorial aspects of an image, such as the vividness of an image or the degree of detail present in an image. The items used in the questionnaire are of the kind used on tests of spatial ability such as the 'block shapes' used in the Vandenberg test of mental rotation (Vandenberg, 1978) and may be either two- or three-dimensional shapes. For each item the participants are required to generate an image, maintain the image, and then rotate the image through the plane of the paper. Participants are allowed to see pictures of the items to be imaged, but must close their eyes when imaging, and ratings are made on appropriate nine-point rating scales. Dean and Morris have used two versions of their questionnaire. In Dean and Morris's 1991 paper, the Shapes Questionnaire was made up of ratings of five imaged items. In the 2003 paper only two of these items were retained, one a two-dimensional shape, and the other a three-dimensional shape, with analyses performed separately for each shape.

The rationale behind the development of the Shapes Questionnaire draws on the work of Kosslyn (1980, 1994) in suggesting that mental imagery may be best viewed as a collection of relatively independent sub-processes that should be assessed separately. Kosslyn's model emphasizes four key imagery processes: image generation, image maintenance, image inspection and image transformation / manipulation. Kosslyn at al (2004) provided neuroimaging evidence that imagery subprocesses show evidence of localization in different brain areas. Thus the Shapes Questionnaire represents an attempt to go beyond ratings of what has been the most popular measure of the imagery experience i.e. vividness. Further, the authors have argued that the inclusion of measures such as the ease of generation and ease of maintenance may prove to have more functional significance than pictoral aspects of images.

In their work Dean and Morris compared the size of the correlations seen between the Shapes Questionnaire and objective spatial tests compared to other self-report imagery questionnaires such as the VVIQ-1, in an attempt to understand why such self-report imagery questionnaires rarely correlate with spatial tests. In both papers they found scores on their questionnaire to show a stronger relationship with scores on spatial tests than other self-report imagery questionnaires and reported correlations as high as r=.51 for ease of

image maintenance. Dean and Morris suggested that this may be because of the better match in item type and / or type of imagery between the Shapes Questionnaire and spatial tests, and the limits of only measuring imagery in terms of vividness which they argued may only be an overall measure of imagery processes.

2.7) Other self-report imagery tests

Several other self-report imagery questionnaires are of relevance because they are either used in the papers which make up this thesis or are mentioned later on in the thesis. Brief background details of these questionnaires will now be given.

The Vividness of Movement Imagery Questionnaire (VMIQ; Isaac, Marks, & Russell, 1986) assesses vividness of movement imagery using 48 items. The VMIQ was used in paper 1 of the thesis. Respondents rate the vividness with which they can image movement in someone else carrying out a series of 24 activities involving basic body movements such as running or jumping, and movements controlling an object such as throwing a stone into water. Then they rate the vividness with which they can image kinesthetic sensations within themselves across the same 24 activities. Ratings of image vividness are made on the same five-point ratings scale that is used in the VVIQ and all of the images are made once with the eyes closed. A limitation in the VMIQ is that its two scales confound visual and kinesthetic imagery with the person perspective (self versus other). Roberts, Callow, Hardy, Markland and Bringer (2008) developed the VMIQ version 2 (VMIQ-2) to address this issue. The VMIQ-2 has three sub-scales: first-person visual imagery (image the activity as if seeing it through your own eyes), third-person visual imagery (image watching yourself performing the activity as a spectator) and first-person kinesthetic imagery (image feeling the sensation in your body of performing the activity), with 12 items per scale. Studies with the VMIQ-2 have tested a range of issues relevant to a sporting context such as: the relationship between athletes' imagery ability and preferences for using the three types of imagery (Callow & Roberts, 2010) , information differences between first-person and third-person imagery and their relevance for imagery interventions for different types of sport (Callow, Roberts, Hardy, Jiang & Edwards, 2013) and the neurological basis of the three imagery types (Jiang, Edwards, Mullins, & Callow, 2015).

The Auditory Imagery Scale (AIS; Guissurason, 1992) assesses the vividness of auditory imagery using seven items. The AIS was used in paper 3 of the thesis. Participants form a series of images of the sounds made by e.g. music, conversation at a party, footsteps coming up a flight of stairs. Ratings of image vividness are made on a four-point scale.

Also a number of tests of habitual visual thinking style have been developed. For example, the Individual Differences Questionnaire (IDQ-HIS; Cohen & Saslona, 1990) which assesses the extent to which people prefer a visual style of thinking using only 13 items. Responses are made on a five-point response scale. The IDQ-HIS was used in paper 3 of the thesis.

2.8) Summary of key points

- Two types of test have been used to assess individual differences in imagery ability. These have been classified as either: 1) subjective self-report imagery tests, or 2) objective behavioural tests. The two kinds of test at best show only weak inter-correlations.

- The most widely used self-report imagery scale is the Vividness of Visual Imagery Questionnaire version 1. The questionnaire has been used extensively. McKelvie (1995b) reports that reliable relationships exist between the VVIQ-1 and performance on perceptual tasks and memory tasks. Research using a range of psychophysiological techniques has also found meaningful relationships with brain activity in visual areas. Critics of the test have drawn attention to issues such as the difficulties facing participants making ratings on the VVIQ response scale and argued that it is more suited to making within-participant comparisons than between.

- Hishitani, Miyazaki and Motojama's (2011) have proposed a model of the mechanisms responsible for the vividness of visual imagery which is based on findings from cognitive psychology and neuroscience. In the model information in long-term memory (LTM) about the object is stored in terms of interconnected perceptual, affective, motor information and meaning. Perceptual information about the visual appearance of the object to be imaged is transferred from LTM to working memory along an information channel. Flow of perceptual information along the channel is controlled by a mechanism called the 'suppressor'. Emotional information in LTM may also be retrieved and used in a process which computes the emotional value of the content to be imaged at a

particular point in time / situation. If this emotional value is negative, the suppressor mechanism restricts the flow of information along the information channel which leads to a less vivid image. Motor and meaning information in LTM can also influence image vividness by acting on the image construction process. A structure called the 'closer' acts to inhibit visual perception while imagery is taking place. The final stage of the model is the inspection of the visual image.

- Ahsen has developed an adapted version of Mark's VVIQ-1 which he has used to explore clinical applications of imagery vividness. He has argued that dim, or unvivid imagery, is as significant as vivid imagery and he views unvivid imagery to be a separate process to vivid imagery.

- Some researchers have developed tests underpinned by modern imagery theory. For example, Dean and Morris (1991, 2003) have developed a self-report questionnaire called the 'Shapes Questionnaire' which assesses a wide range of self-reported visual imagery properties in addition to vividness. The authors have argued that subjective self-report imagery tests may better predict performance on objective spatial tests when the items to be imaged are better matched, and that measures of imagery processes such as the ease of maintenance show stronger relationships with spatial test performance than pictoral aspects such as vividness.

- Self-report imagery tests have also been developed to assess other aspects of mental imagery such as movement imagery, auditory imagery, and habitual visual thinking style.

3) Social desirable responding

The purpose of the section is to provide an up-to-date description of work in the social desirable responding area, introduce the reader to key models, questionnaires and issues around their use which can be later drawn on in the critical review. The section starts by introducing the topic of social desirable responding, situating it in the broad context of questionnaire response biases. Then the influential work of Canadian psychologist Delroy Paulhus is described who has argued that social desirable responding can occur in different forms. Paulhus's models of social desirable responding are introduced finishing with his most recent model, his two-tier model. This is followed by the questionnaire he developed to

measure the phenomenon, the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002), along with the most widely used measure of social desirable responding, the Marlowe-Crowne questionnaire (MC; Crowne & Marlowe, 1960). Paulhus's two-tier model of social desirable responding (Paulhus, 2002) theorises four sub-types of social desirable responding and forms the theoretical base for the conceptualization of social desirable responding in all five of the thesis papers. Equally, the BIDR was used in all of the papers, while the MC was used in paper 4.

Determining whether the BIDR scales measure departures from reality, 'substance' or a mixture of the two is a crucial issue for the use of the questionnaire and the thesis. So next, the three different conceptualisations of social desirable responding as substance are described, followed by a discussion of the two main kinds of evidence researchers have used to support the substantive position and the handful of studies that have attempted to test the extent to which the BIDR scales measure substance versus style.

This is followed by discussion of recent tests of Paulhus's questionnaire by a group of researchers based in Croatia, and evidence suggesting that social desirable responding occurs as a three-way interaction between the characteristics of the focal scale items, the respondent and the situation. Research carried out on the BIDR by the Croatian research group constitutes the most thorough test to date of the full BIDR questionnaire and Paulhus's two-tier model. The idea that social desirable responding occurs as a three-way interaction is discussed in paper 4 and linked there to the BIDR scales. It is also used in the last section of the literature review as a framework for organizing the factors that might potentially influence the occurrence of social desirable responding on self-report imagery questionnaires.

The social desirable responding section finishes by describing two modern componential indices of over-claiming followed by the Over-Claiming Questionnaire (OCQ; Paulhus, Harms, Bruce & Lysys; 2003) which was designed by Paulhus to be an objective measure of general cultural knowledge over-claiming. These measures offer possibilities for future directions the body of research described in the thesis might take, and so are described here and then picked up again on at the end of the critical review.

3.1) Social desirable responding and questionnaire response biases

Peoples' self-reports on their experiences are often considered to be valuable sources of information about their psychological functioning (Paulhus & Vazire, 2007). However, even when participants are seeking to provide accurate information, self-reports are open to a number of potential biases. Paulhus (1991, p17) defines a response bias as 'a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content' i.e. what the items were designed to measure. Response biases are traditionally classed as either 'sets' or 'styles'. Response sets are short-term biases affecting a person's responses on a questionnaire, caused by some temporary pressure in a situation such as having limited time to complete a questionnaire or the expectation that questionnaire responses will be viewed by others. In contrast, response styles are long-term biases that consistently affect a person's responses across time and different questionnaires. Examples of response biases include: acquiescence bias (the tendency to agree with questionnaire items regardless of their content), extremity bias (the tendency for responses to questionnaire items to be made from the extreme ends of available rating scale options) and social desirable responding.

Several definitions of social desirable responding have been put forward. Paulhus (2002, p.50) defines social desirable responding as "…the tendency to give overly positive self-descriptions". This definition does not tie the phenomenon to a particular methodological context such as responses on specific questionnaires, and it covers both responses styles and sets. Further it includes the concept of social desirable responding as being, not just an endorsement of desirable features and rejection of undesirable features, but being a departure from reality.

3.2) Social desirable responding as a multifaceted construct

Up until the late 1950's social desirable responding was generally viewed as a unitary phenomenon. However, studies which tested the relationships between social desirable questionnaires found low intercorrelations between different measures and factor analyses produced multiple factors (Cattell & Scheier, 1961; Edwards, Diers & Walker, 1962; Jackson & Messick, 1962; Messick, 1962; Wiggins, 1959). Thus it appeared that socially desirable

responding might occur in more than one form, and over time, this view has gained increasing acceptance. Work by the Canadian psychologist Delroy Paulhus and his colleagues has been particularly influential (for example: John & Paulhus 2000; Paulhus, 2002, 2006; Paulhus & Notareschi, 1993). Their work lead to the development of a series of increasingly more sophisticated models of the forms of social desirable responding, and a questionnaire to measure them.

What now follows is a description of Paulhus et al's work which attempts to capture the key developments in their work but which is also tailored to be in line with the concepts and terminology used in the five papers that make up this thesis. Table 7 summaries the key stages in the development of Paulhus et al's models.

Table 7.

Summary of the main developments of Paulhus's modelling of social desirable responding

Year	Name of the model	Number of elements in the model	Summary of key developments
1984	Two-factor model	Two	Social desirable responding conceptualized as having two forms, unconscious and conscious
1994	Revised two-factor model	Three	Self-deceptive denial (SDD) form of social desirability added
2002	Two-tier model	Four	Social desirable responding reconceptualised in a two-tier format with a primary content distinction (egoism versus moralism) and a secondary process distinction (unconscious versus conscious)

3.2.1) Two-factor model (Paulhus, 1984)

Early work by Paulhus involved the use of exploratory factor analysis to investigate the relationship between different measures of socially desirable responding and the

development of an initial version of his questionnaire. Paulhus (1984) factor analysed the scores from six questionnaires commonly used to measure social desirable responding and found that two relatively independent factors emerged. The first factor was labelled 'self-deception' (SD). Paulhus interpreted this as reflecting an honest, but overly positive self-presentation. The second factor was labeled 'impression management' (IM). Paulhus interpreted this as reflecting self-presentation tailored to an audience. Thus the two factors were viewed as differing primarily in terms of conscious awareness and the intentionality of the social desirability tendency. As part of the same study Paulhus developed the first version of a questionnaire called the Balanced Inventory of Desirable Responding (BIDR) to assess the two forms of social desirable responding. The BIDR went through several revisions until a manual was published for the sixth version in 1994 (BIDR-6; Paulhus, 1994).

3.2.2) Revised two-factor model (Paulhus, 1994)

The two-factor model and BIDR were revised in 1994 as a new form of SD was introduced. The measures of SD were self-deceptive enhancement (SDE) measuring the promotion of positive qualities, and self-deceptive denial (SDD) measuring the refutation of negative qualities. The measure of IM was simply called by the same name.

3.2.3) Two-tier model (Paulhus, 2002)

Later work by Paulhus (John & Paulhus, 2000; Paulhus & John, 1998 and Paulhus & Notareschi, 1993) saw a shift to new methodologies and a deeper reflection on the specific content of the BIDR scales. This lead to a reinterpretation of the number and meaning of the forms of socially desirable responding that were believed to exist and as a consequence Paulhus (2002) proposed a new two-tier model of social desirable responding (see Figure 2). This is the model of social desirable responding that is used in the five papers that make up the thesis.
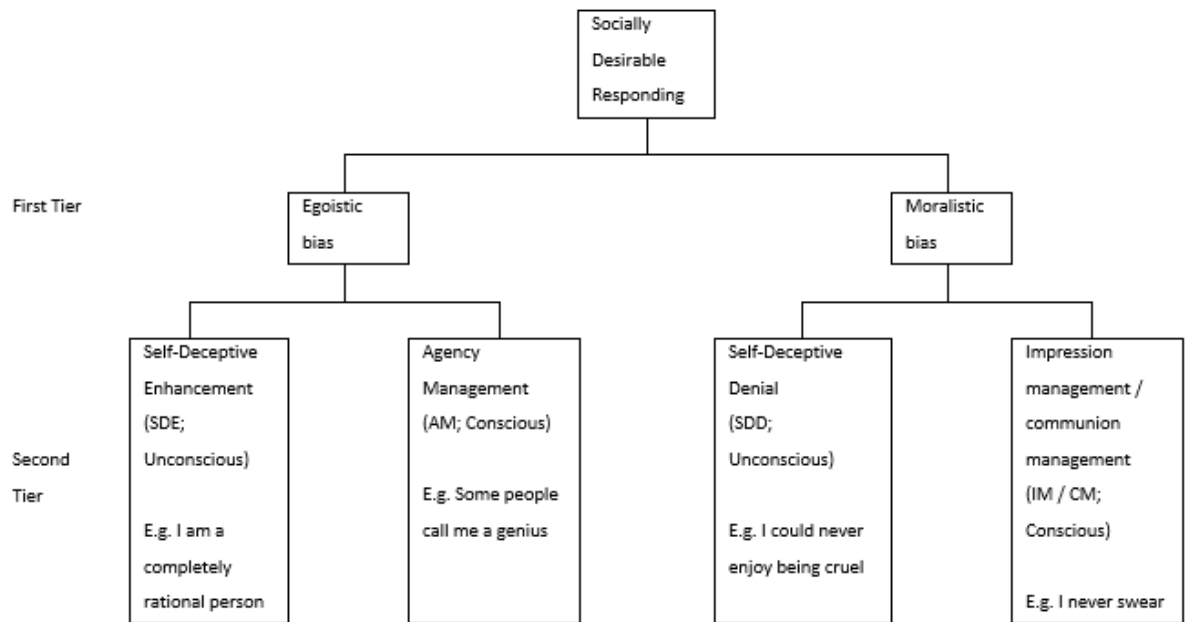
Figure 2. Paulhus's two-tier model (modified from Paulhus, 2002)

The 2002 two-tier model has a primary distinction in terms of 'personality content' (Paulhus, 2002, p. 63). In the first tier there is a division of social desirable responding into egoistic and moralistic biases. The egoistic bias refers to the tendency to claim positive social and intellectual qualities such as dominance, fearlessness and intellect. Egoism is therefore driven by a need for power. The moralistic bias refers to the claiming of positive moral qualities such as agreeableness, dutifulness and restraint. Moralism is driven by the valuing of intimacy, relationships benefiting others and a concern with following appropriate social norms and rules. The egoism – moralism distinction can be summarized as the difference between claiming to be a 'superhero' or a 'saint'. The original two-factor model distinction between conscious and unconscious forms of social desirable responding is retained in the second tier of the new model where both the egoistic and moralistic biases are split into conscious and unconscious forms. This process distinction is described as linking roughly to whether an audience for the self-presentation is present or not. The egoistic bias becomes SDE and agency management (AM), while the moralistic bias becomes SDD and communion management (CM). Note, to avoid confusion from this point on in the thesis the CM form of social desirable responding will be referred to solely as IM, because this is the term used within the five papers of the thesis and is the term other authors typically use. SDE and SDD have previously been explained, although note that in this new model they are now viewed

as measures of different forms of socially desirable responding (egoistic bias and moralistic bias).

The SDE and SDD scales of the BIDR-6 map straight forwardly on to this new model. Paulhus refers to them collectively as the self-deceptive scales and views them as best suited to measuring response styles i.e. unconscious, stable trait-like biases. Paulhus refers to the existing IM scale, and a new AM scale designed to measure egoism, as the impression management scales. He suggests that IM and AM scales are best suited to measuring response sets which are likely to be more extreme, conscious and context dependent. Collectively the four scales constitute the BIDR-8.

An 'uneven' aspect of the model is that the nature of the differentiation of conscious from unconscious levels differs between egoistic and moralistic biases. For the egoistic bias, SDE and AM scale items differ in the degree of exaggeration made about the ability claims such that AM claims are very extreme while SDE are less so. In contrast for the moralistic bias, SDD scale items cover more sensitive and emotionally charged themes than IM items such as the admission of negative emotions like anger, jealousy or sexual feelings, and have a defensive tone similar to that of psychoanalytic denial. In contrast IM items cover compliance with norms relating to socially acceptable behaviour such as not telling lies. Equally, because IM refers to overt behaviours a person should be aware of e.g. 'I never swear', low IM item non-endorsement rates are assumed to reflect distortion that a person has conscious awareness of.

Paulhus (2002) developed his two-tier model following the results from two different methodologies. The first method involved varying the kind of 'fake good' instructions that were given to participants (Paulhus & Notareschi, 1993). Instructions were either 'standard' fake-good instructions which participants reported interpreting to mean present yourself as a 'nice person' or 'good citizen' i.e. in a positive moralistic light, or instructions to answer in a way that would impress an experimenter how strong and competent they were which were more egoistic in nature. 'Moralistic' fake-good instructions lead to increases in IM scores, while 'egoistic' fake-good instructions lead to increases in SDE scores. The second method (Paulhus & John, 1998, and John & Paulhus, 2000) involved the factor analysis of residual scores derived for personality (differences between self and knowledgeable others'

estimates on personality scales) and intelligence (differences between self-estimates of IQ and the results of IQ tests) together with measures of socially desirable responding. The rationale behind this approach was that the estimates of knowledgeable others, such as peers and family, and IQ tests would act as objective standards to which self-estimates could be compared, so that residual scores would reflect departure from reality. Factoring of the residual scores for personality and intelligence produced solutions comprised of two factors which corresponded to the egoistic and moralistic biases.

The two-tier model is able to make sense of findings the older two-factor model struggled to explain, for example why SDD and IM scales have been observed to correlate highly together under most conditions. However, the empirical data Paulhus et al (2002) presents supports only the primary distinction between egoistic and moralistic biases, and only limited background data is available for the SDD and AM scales.

Acceptance of Pauhus's (2002) model necessitates a reinterpretation of previous findings. Paulhus's (1984) labels of SD and IM for the two forms of social desirable responding he identified through factor analysis would be reinterpreted as most likely reflecting egoistic bias and moralistic biases. Further, given the strong body of evidence suggesting that socially desirable responding is multidimensional in nature, it is important that this should be taken into account when researching the relationship between socially desirable responding and responses to imagery self-report questionnaires. It is possible that different kinds of socially desirable responding have independent and possibly even additive effects. Estimates of effect sizes should be made separately for different social desirable responding scales or at least different kinds of socially desirable responding.

In later work Paulhus and Trapnell (2008) provided an account of the two-tier model that positioned it in the broad context of self-presentation in everyday life. Paulhus and Trapnell conceptualised self-presentation as the motivated and effortful presentation of an inaccurate self-portrayal and so subsumes social desirable responding on questionnaires. They positioned the egoistic and moralistic biases as two overarching forms of self-presentation that occur in everyday life. They hypothesized that the two biases originate from an interactive developmental process whereby personality, in the form of the big 5

personality traits, provides a genetic basis for behavior on top of which society's socialization processes layer agency and communion meta-values.

Furthermore, it is possible to situate the egoism-moralism biases within the agency-communion content conceptual framework first proposed by Bakan (1966). Agency refers to a person's ability to strive to achieve their life goals ('getting ahead'), while communion refers to a person's ability to connect to other people and form social relationships ('getting along'). Abele and Wojciszke (2018) describe how the 'big 2' framework helps explain phenomena relating to the motives underlying human behaviour, personality, self-perception (self-concept and self-esteem), interpersonal perception, inter-group perception and stereotypes. For example, in interpersonal perception the detection of agency and communion in other people can act to help individuals achieve social status and identify candidates for long-term affiliative relationships. The perception of individuals who possess agency indicates people who may help us achieve our goals and so are potential resources, while the perception of individuals without communion signals people who may pose a threat to the social stability of the groups we belong to and so constitute potential threats. Each of the big 2 have been argued to have two facets with agency having competence and assertiveness, while communion has warmth and morality.

3.3) The two most popular social desirable responding questionnaires

A number of questionnaires have been developed to measure social desirable responding. The logic behind the tests is that they contain items that make claims which for most people would be unlikely to be true, such as 'my first impressions of people are always right' or 'I never swear'. So when people endorse many of them, it would seem likely that their answers are not correct and that they are exhibiting socially desirable responding. In modern research two tests have been particularly popular, the Marlowe-Crowne social desirability scale (MC; Crowne & Marlowe, 1960), and Paulhus's Balanced Inventory of Desirable Responding.

3.3.1) The Marlowe-Crowne social desirability scale

The MC questionnaire was developed by Marlowe and Crowne in the 1950's as a test of social desirable responding in self-reports. Beretvas, Meyers and Leite (2002) reported a literature search that found 1069 research papers and dissertations that had used the MC scale or one of its short forms. Marlowe and Crowne developed their questionnaire items so they were not associated with poor adjustment and psychopathology as some previous scales had been. This was important because when participants give 'no' responses to items which are confounded such as 'my sleep is fitful and disturbed', a questionnaire does not differentiate between a response indicating a denial of symptoms which is a socially desirable response, and response indicating a genuine absence of symptoms which is not. The test has 33 statements that relate to culturally approved or disapproved behaviours with response options of either 'true' or 'false'. Eighteen of the statements are positively phrased and 15 negatively phrased, and they are randomly ordered across the questionnaire. Examples of items from the scale include: item 2 'I never hesitate to go out of my way to help someone in trouble' and item 10 'On a few occasions, I have given up doing something because I thought too little of my ability'. Positively phrased item responses of 'true' score one point, and negatively phrased item responses of 'false' score one point. Paulhus (1991) reports a number of studies that support both the reliability and the convergent and discriminant validity of the test.

Over time Marlowe and Crowne's conceptualisation of what their test measures has evolved. When first published in 1960, the MC scale was intended to be a measure of social desirable responding in self-reports. The authors' then shifted their position to suggest that the scale taps a more general motivation in people to need the approval of others (Crowne & Marlowe, 1964). Later Crowne (1979) refined this idea to a motivation to avoid others' disapproval. Paulhus (1991, p. 29) states that the scale sustains 'a dual existence as a social desirable responding scale and a measure of the approval-dependent personality'.

Since its construction, a number of weaknesses have been identified in the MC test. For example, Strober (2001) argued that some of the items on the test may have become out of date and reflect the social standards of the 1950's. Helmes (2002) suggests that the test confounds the type of item (attribution of positive attributes and denial of negative

attributes) with direction of item scoring since most positively phrased items imply positive attributes, while most negatively phrased items imply negative attributes. Also, Leite & Beretvas (2005) found that confirmatory factor analysis supported neither a single nor a two-factor model for the scale.

In terms of Paulhus's two-tier model, data from Paulhus (1984) suggests that the MC test is primarily a measure of moralistic bias. Figures from the study (table 2 on page 601) show a loading of 0.68 on a factor labelled 'impression management' (moralism) and 0.40 on a factor labelled 'Self-deception' (egoism).

3.3.2) The Balanced Inventory of Desirable Responding

As well as developing a model of social desirable responding, Paulhus also developed a questionnaire, the Balanced Inventory of Desirable Responding (BIDR), to assess its different forms that was not confounded with psychopathology. The questionnaire was developed in parallel with Paulhus's models starting with the SDE and IM scales, then SDD being added later, and finally AM. The latest version of the BIDR is version 8. The BIDR questionnaire (either selected pairs of scales or all four scales) was used in all five of the papers making up the thesis and so is described in some detail in this section.

Each of the four subscales have 20 items, and example items can be seen in Figure 2. All scales are made up of alternating positively and negatively phrased propositions, so that there are 10 items of each type. Responses are usually made on a seven-point scale ranging from 1 'Not true' to 7 'Very true'. There are two ways to score the subscales: continuous scoring and dichotomous scoring. The continuous method is to recode responses to negative items and then sum scores on the seven-point response scale across items. The dichotomous method is to count only extreme responses in a person's score. In dichotomous scoring, negative items are recoded and then responses of either six or seven on an item are counted as one point, while responses of one to five do not score. Paulhus (1994) argues that the second method has the advantage of only scoring clearly exaggerated or managed responses.

Of the four BIDR scales, comprehensive data on reliability and validity is only currently available for the two most well established ones, SDE and IM. As such the information in this section, and those that follow, will focus primarily on these scales. Later on in the literature review the findings of recent studies using a Croatian version of the BIDR involving all four scales will be described.

Norms

Paulhus (2006) reports mean dichotomous scores for all four BIDR scales from a sample of 183 university introductory psychology students. Males scored more highly on both egoistic scales while females scored more highly on both moralistic scales. Male mean scores and standard deviations were: SDE 6.0 (3.2), AM 5.4 (3.0), SDD 4.9 (3.1) and IM 4.7 (3.1). Female mean scores and standard deviations were: SDE 5.5 (3.1), AM 5.0 (3.6), SDD 5.9 (3.2) and IM 6.9 (3.2).

Reliability

Li and Bagger (2007) conducted a meta-analysis for internal consistency on the SDE and IM scales when scored dichotomously. They reported a mean reliability coefficient for SDE from 90 studies as .68 (*SD*=.09), and for IM from 107 studies of .74 (*SD*=.09). In our research, Allbutt, Ling, Rowley and Shafiullah (2011), we found Cronbachs alpha under honest, anonymous, group testing conditions with a student sample for the SDE, AM, SDD and IM subscales scored dichotomously to be: .70, .69, .64, and .60; and scored continuously to be: .75, .82, .75, and .72. Paulhus (1994) reported SDE test-retest reliability over a five-week interval scored dichotomously was *r*=.69 while IM test-retest reliability over the same interval scored dichotomously was *r*=.77. As such the reliabilities of the BIDR scales lie towards the bottom end of the acceptable range for research purposes.

Validity

Paulhus (2002, 1998, 1994, 1991) reports a number of studies that support the convergent and discriminant validity of the SDE and IM scales. Exploratory factor analysis applied to multiple social desirable responding questionnaires has consistently produced a two-factor

structure with SDE and IM scales loading on separate factors (Paulhus & Reid, 1991; Borkenau & Ostendorf, 1989, and Nichols & Greene, 1980). However, CFA applied to the SDE and IM scale items has not always produced the egoism and moralism factors. Paulhus (1998) reported CFA support for a two-factor model with a Canadian sample of students, and Gignac (2013) factored scores on the SDE, SDD and IM scales with an American student sample and found support for a model composed of a general factor and two nestled factors corresponding to SDE and IM. Whereas Li and Li (2008) found a four-factor structure with a Chinese sample of students.

Paulhus argues that 'SDE appears to tap a form of confidence that is not contingent on accurate knowledge' (Paulhus, 1994, p. 15) i.e. overconfidence, rather than simply confidence. In support of this view, for the egoistic bias Paulhus presents four lines of evidence that individuals achieving high SDE scores demonstrate verifiable distortions from reality which they are unaware of. First, studies of memory for previously seen and unseen words shows that high SDE scorers show higher confidence in their memory judgments despite being no more accurate than low SDE scorers. Second, high SDE scorers show more hindsight bias than low SDE scorers when estimating their likely success at answering challenging multiple-choice trivia questions that are typically answered at chance level. Third, high SDE scorers are more likely than low SDE scorers to rate themselves as being familiar with non-existent items from a range of different categories such as historical names and events, the arts, books and poems. Fourth, comparison of self and other ratings of performance in discussion groups also show a relationship with SDE scores. In this setting, high SDE scorers tend to overestimate their own contribution to the group discussions relative to the mean judgments of other group members.

Paulhus (1994, 1998, 2002) provides other sources of support for the validity of the SDE and IM scales. In terms of divergent validity: the SDE scale shows greater discrepancy between 'self' and 'other' ratings than the IM scale does; the two scales show different patterns of responses to situational pressures, the SDE scale is less responsive than the IM scales to situational changes such as fake-good versus fake bad instructions; and the scales show logical patterns of relationships with other psychological variables which vary between the two scales. The SDE and IM scales have also been found to correlate meaningfully with other

egoistic and moralistic bias measures such as the Egoistic and Moralistic Self-Enhancement Scale (EMS; Vecchione, Alessandri & Barbaranelli, 2013).

Paulhus (1994) reports that SDE correlates positively with self-esteem, and negatively with trait anxiety, social anxiety, depression and personal distress. The highest of these was with trait anxiety (r=-.52). For relationships with the big 5 personality traits, estimates from Li and Bagger's (2006) meta-analysis of 30 articles are the most comprehensive. They found SDE to correlate with neuroticism (r=-.54), conscientiousness (r=.42) and extraversion (r=.31), while IM correlated with agreeableness (r=.42), conscientiousness (r=.42) and neuroticism (r=-.35) aspects of personality. However, some sources e.g. Paulhus (1994) cite openness as correlating with SDE, often in place of conscientiousness, while neuroticism is often cited as only correlating with SDE.

Holden and Passey (2010) report SDE and IM correlations with big 5 personality trait questionnaire facets as measured by the NEO-PI-R in a sample of 301 university students. SDE – big 5 facet correlations that reached a medium effect size or larger included: C1 competence (r=.32), N6 vulnerability (r=-.31) and E3 assertiveness (r=.31). All these three facets share a common thread of confidence in abilities. IM – big 5 facet correlations that reached a medium effect size or larger included: A2 straightforwardness (r=.46), C3 dutifulness (r=.40), A3 altruism (r=.37), A4 compliance (r=.36), N2 anger-hostility (r=-.35) and N5 impulsiveness (r=-.31). These facets share a common thread of positive moralistic qualities such as being constant, sincere, meticulous, gentile, generous, tolerant and controlled. The pattern of conscientiousness facet correlations with SDE and IM suggest that SDE correlates with facets reflecting the will to achieve, while IM correlates with facets reflecting the will to be reliable and dependable. Holden and Passeys' results also show how different neuroticism facets load on SDE and IM i.e. SDE is negatively associated with vulnerability, while IM is negatively associated with anger-hostility and impulsiveness. Holden and Passey observed a low, significant overall correlation between SDE and openness, r=.16, which was smaller in size than the SDE overall correlations with conscientiousness (r=.27), neuroticism (r=-.26) and extraversion (r=.23). This result is consistent with the NEO-PI-R's broad operationalization of openness which deemphasizes egoistic intellectual and creative aspects.

Because most of the research that has been carried out on social desirable responding has been within the area of individual differences, only a few studies have investigated the exact cognitive mechanisms that underlie such a thinking style. Two studies, Djikic, Peterson and Zelazo (2005) and Shane and Peterson (2004), have tested whether SDE is associated with effects at the encoding stages of new information via attention, and / or at the retrieval stage for already processed and stored information in memory. Neither study found SDE effects at the attentional level, but both found links between egoism and inferior retrieval of negative information about the self from memory and egoism and the superior retrieval of positive self information. These findings were interpreted as reflecting the operation of a cognitive filtering mechanism that serves a defensive, ego protective function. Further, Barrios et al (2008) investigated the neuronal basis of the egoistic bias using Transcranial Magnetic Stimulation (TMS) in a 'virtual lesion' design study. They presented egoistic trait adjectives to ten participants while TMS was applied to several brain locations. They found that only TMS applied to the medial prefrontal cortex (located at the very front of the frontal lobe) reduced levels of egoistic self-enhancement.

Debate on scoring method

Researchers have also compared the psychometric properties of the BIDR scales for the dichotomous and continuous scoring methods. Several researchers have found support for the categorical scoring system. Stöber, Dette and Musch (2002) compared the effectiveness of the two methods using a German version of the BIDR. They found the continuous method to be preferable, for example in that it showed: 1) larger Cronbach's alphas; 2) larger convergent correlations with other social desirable responding questionnaires; and 3) SDE continuous scores showed larger correlations with those traits of the Five-Factor model of personality for which substantial correlations were expected i.e. neuroticism, extraversion and conscientiousness.

Cervellione, Lee and Bonanno (2009) used Rasch modelling, to assess the SDE scale only in a sample of undergraduate students in low social desirability situations for use of response category, sample appropriateness and internal consistency. Rasch modelling, like the similar technique of Item Response Modelling (IRM), is a statistical approach for modelling the level of a latent trait possessed by persons taking the test and the test items, and is used for the

for the design, analysis and scoring of psychological tests and questionnaires (Embretson & Reise, 2000). Cervellione et al's analysis suggested that the continuous scoring system showed the best psychometric properties and was the best fit for their data and the sample. This was because, when dichotomous scoring was used, respondents tended not to score highly enough on the items for them to be useful in the measurement process.

Stöber *et al*. and Cervellione *et al*. have both criticized the dichotomous scoring system for treating social desirable responding as an 'all or nothing' phenomena rather than one that occurs in degrees, and that individuals may have a predisposition towards social desirable responding but not select the most extreme response options. They have argued that the dichotomization of the measurement scale leads to a loss of information and this may add to the error in measurement.

However, not all findings have supported the continuous scoring method. Leite and Beretvas (2005) and Gignac (2013) tested the two scoring methods using CFA and found that the dichotomous scoring system was a better fit for their data. Gignac also found that internal consistency estimates for BIDR scales made with the ordinal Cronbach's alpha statistic to exceeded those for continuous scoring. Gignac suggested that continuous scoring may only correlate more highly with the big 5 because it taps aspects of personality more so than dichotomous scoring does.

In summary it can be said that the BIDR is an important questionnaire in the area of social desirable responding and has close links to Paulhus's two-tier model. Of the four BIDR scales, comprehensive data on reliability and validity is only currently available for the two most well established scales, SDE and IM. Paulhus (1994, 1998, 2002) provides a range of sources of support for the validity of the SDE and IM scales. For example, evidence that individuals achieving high SDE scores demonstrate verifiable distortions from reality which they are unaware of. In addition there is debate over several aspects of the scales such as their scoring method and item discrimination ability.

3.4) Social desirable responding as distortion or error versus a valid personality trait

Traditionally social desirable responding has been viewed as distortion or error that should be eliminated or minimized (Domino & Domino, 2006). In terms of response style, individuals scoring high on social desirable questionnaires are argued to show a consistent tendency to overestimate their standing on desirable qualities. When applied to self-report questionnaires measuring a desirable characteristic, this tendency results in a consistent pattern of score distortion. This does not necessarily mean such a person is consciously lying, they may simply have a distorted, overly positive view of themselves or a high need to be valued and accepted by others. But, whatever the cause, the end result is the same, a departure from reality. As a consequence, self-report questionnaires that correlate highly with measures of social desirable responding are considered invalid, or at very least, suspect.

However, the view of social desirable responding as simply reflecting distortion has been questioned. Increasing numbers of psychologists have argued that social desirable responding in some sense reflects substantive or valid responding, particularly in low demand situations. Three main perspectives on social desirable responding as substance have been proposed, and these will now be described.

3.4.1) Social desirable responding as a as departure from reality, but a meaningful personality trait in its own right

Some psychologists have argued that social desirable responding is better viewed as a meaningful personality trait in its own right, or a set of traits such as egoism and moralism, that correlate with a variety of psychological constructs and behaviours. Paulhus views the egoistic and moralistic biases as reflecting 'two different personality constellations' (Blasberg, Rogers & Paulhus, 2013, p. 523). When viewed as an important substantive trait in its own right, the egoistic bias has been viewed as a self-enhancing bias and strong belief in one abilities that is related to traits such as the need to achieve, extraversion, openness and positive traits relating to self-reported psychological adjustment such as high self-esteem and confidence, and low neuroticism, depression, and social anxiety. The moralistic bias has been viewed as a social approval motive related to traits such as conscientiousness and agreeableness or interpersonal sensitivity. Some evidence supports Paulhus's position, for

example, the internal consistency and test-retest scores of the SDE and IM scales suggest that they are 'trait-like' in nature.

3.4.2) Social desirable responding as a departure from reality, but an integral part of many different aspects of personality

A second way substantive responding has been discussed is to argue that, although high scores on social desirable responding questionnaires represent a departure from reality, on a conceptual level they form an integral and inseparable part of many different constructs of personality. For example, it could be argued that an important part of the traits emotional stability, extraversion or conscientiousness is having a strong belief in one's abilities. Or from a moralistic perspective, having a strong desire for others approval, or possessing high levels of self-control are valid parts of conscientiousness or agreeableness. From this position Holden (2010) argues that 'if social desirability is a feature of many different constructs of personality e.g. extraversion, agreeableness… then these individual difference constructs are not truly and conceptually independent and should not be theorized, measured or reported as such' (p. 913).

Several researchers have offered view points on the process of determining whether social desirable responding is conceptually distinct from a construct a researcher is trying to measure. Zerbe and Paulhus (1987) argue that social desirable responding can only be considered to be 'contamination' when it is unrelated to the construct of interest. While Helmes (2000) argues the emphasis placed on the issue of whether social desirable responding is best viewed as distortion versus substantive responding is misplaced. This is because, 'whether variance associated with social desirability is due to a response style or to a meaningful construct, the net result is the same. The variance that is associated with social desirability is almost always different at a conceptual level from the variable(s) in question' (p. 35-36). However, Holden and Passey (2009) point out that 'discerning whether social desirability is part of the construct being measured or an interfering response style is not an easy task and is not well established either theoretically or empirically' (p. 443).

3.4.3) Social desirable responding as honest responding

A third way substantive responding has been discussed is to argue that, despite their excessive nature, high scores on social desirable responding questionnaires actually reflect genuinely accurate responses and reflect real virtues, at least when made under low demand circumstances. The substance – distortion debate has most often been discussed within the context of the measurement of the big 5 personality traits. There the proponents of the substantive position argue that egoistic and moralistic scales act more as proxy, or 'disguised', measures of the big 5 personality traits than as measures of genuine distortion. Thus correlations of SDE typically reflect realistic levels of neuroticism, extraversion or conscientious traits, while correlations of IM reflect genuine levels of agreeableness, conscientiousness or neuroticism traits.

In relation to imagery self-report questionnaires, the second and third perspectives on social desirable responding as substance are most relevant. The issue of whether social desirable responding forms a meaningful personality trait in its own right, or a set of traits, is not of major relevance to imagery work. Instead, key issues are whether it can be demonstrated that social desirable responding measures actually do index departures from reality, and that if this is possible, that such self-deceptive processes do not form an integral and inseparable part of the normal functioning imagery system.

Researchers that have argued for the substantive position have drawn on two main kinds of evidence to support their views:

1) data from other raters who are familiar with the target person; and
2) data relating to attempts to statistically control for the effects of social desirable responding.

In response, researchers backing the validity of social desirable responding measures have offered counter arguments. The debate on the substantive – bias issue has often been polarized with authors arguing that social desirable responding questionnaires tap into either substance or style. However, the consensus that has emerged is that social desirable responding questionnaires measure a combination of both.

The two main types of substantive evidence will now be discussed. This will serve to introduce the kinds of data that have been drawn on by researchers supporting a substantive position, discuss their strengths and weaknesses, and highlight potentially fruitful avenues for future exploration in imagery research that will be picked up again in the critical analysis. This will then be followed by a description of the studies that have attempted to address the more refined issue of: which is more dominant within social desirable responding measures, substance or distortion?

Data type one – data from other raters who are familiar with the target person

The first kind of data involves studies that have explored the relationship between self-ratings and ratings of a target person by others who have known the person for some time such as spouses and coworkers. Ratings by other people who are familiar with the target person are thought to offer an important objective criterion against which to judge the validity of self-ratings. Such data has shown that self-ratings made on social desirable responding questionnaires correlate with other-ratings made on the same questionnaires and also with other-ratings made on measures of the big 5 personality traits. For example, in a study of 1106 Dutch undergraduate psychology students, de Vries, Zettler and Hilbig (2014) observed a self-other agreement of $r=.34$ on the BIDR SDE scale and a self-other agreement of $r=.45$ for the IM scale. Because other-ratings are viewed as an objective measure, the correspondence between self- and other-ratings has been argued to provide support for a substantive interpretation of social desirable responding scores.

In response, other psychologists have pointed out that there are limits to the use of other-ratings as an objective measure of an individual's personality. First, correlations between self- and other-ratings are typically only moderate in size, which means that while there is notable agreement between self- and other-ratings, there are also areas of disagreement (Pauls & Stemmler, 2003). Second, agreement between self and other does not necessarily guarantee the accuracy of the impression held. Connelly and Chang (2015) argue that when making self-ratings on questionnaires, participants are projecting self-images to others. Self-other correlations measure the extent to which these match i.e. the extent to which an individual and others agree on the individual's reputation, or their social reality. Third, just as self-ratings may be influenced by multiple biases, so other-ratings may also be affected by

biases although they may differ in nature (Oh, Wang & Mount, 2010). For example, 'friendship biases' could affect other-ratings whereby affiliative ties operate to cause observers to favourably accept past self-presentations made to them by the target person as truthful. Or such biases could operate so as to maximise others reporting of a targets' desirable characteristics and minimize the reporting of the target's undesirable characteristics, perhaps because familiar others are viewed as part of the extended self. But either way, the size of self-other correlations would be artificially inflated. Other methodological issues of note for this paradigm include the fact that the person being rated is typically left to choose who will rate them which could bias selection (Oh et al, 2010). Also, the participant's anticipation that someone else will be rating them and their results compared, could influence how they respond, for example, reducing the social desirable responding inflation of their scores (Lönnqvist et al, 2007).

Other empirical findings also support the view that other-ratings cannot be taken as simple, objective measures. For example, Konstabel, Aavik and Allik (2006) found that partialling social desirable responding, operationalized as the sum of a personality trait item endorsement weighted by the item's perceived desirability value, from self-peer and peer-peer personality trait correlations only improved the correlations when partialled from both self- *and* peer-report data, not when partialled from self-ratings alone. Also, Connelly and Chang (2015) point out that scores from social desirable responding scales correlate more highly with self-rated personality traits than other-rated traits. Further, that the relationship between social desirable responding scales and self-rated personality remains after partialling out other-rated personality, suggesting that at least some of the variance in social desirable responding scores is due to distortion.

Data type two - data relating to attempts to statistically control for the effects of social desirable responding

The second kind of data involves studies that have explored the statistical effects of social desirable responding. Lönnqvist et al (2007) describe how, if social desirable responding is truly distortion, then there should be predictable statistical effects when its influence is controlled using correlational techniques. Where social desirable responding contributes variance in a similar fashion to both variables in a correlation, it can be thought of as a

unobserved common response variable that acts to spuriously increase the size of the correlation between the imagery scales, and removal of its influence should reduce the size of correlation. This situation might occur where both variables are self-report measures of constructs with desirability implications. But where social desirable responding contributes variance to only one of the variables in a correlation, it should operate as a suppressor variable, and removal of its influence should increase the size of the correlation. This situation might occur where one variable is a self-report measure of a construct with desirability implications and the other an objective measure of a construct. Suppressors increase the correlation between a predictor and a criterion because they remove the part of a predictor that is unrelated to the criterion and in doing so, increase its ability to predict. Ganster, Hennessey & Luthans (1983) add a third possible statistical effect of social desirable responding, moderation. In moderation, there is an interaction effect between the predictor variable and social desirable responding. Thus at high levels of social desirable responding one type of relationship between the predictor and the criterion variable exists, while at low levels a different relationship exists. For example, high levels of social desirable responding might be associated with loss of accurate measurement information or greater error variance. In moderation social desirable responding may or may not be correlated with the predictor and / or the criterion.

In relation to mental imagery, spurious correlation might occur when two self-report imagery questionnaires measuring valued aspects of imagery are correlated together. Partialling social desirable responding from this relationship should remove shared 'distortion' of imagery self-reports and so reduce the size of the correlation between the imagery scales. Suppression might occur in a regression analysis where a self-report imagery questionnaire is regressed onto an objective measure of performance that is related to the type of imagery, such as perception or sensory related memory. Here, the inclusion of a measure of social desirable responding in the regression should act to 'unsuppress' the relationship between imagery questionnaire responses and the objective measure. Moderation might occur in a similar regression analysis, but where a centred imagery test x social desirable responding measure interaction term is included. If such a term reached statistical significance, it would suggest that moderation was occurring. These issues and how they relate to mental imagery will be picked up again in the critical analysis where a partial correlation analysis was run on the self-report imagery data in paper 1. For the

moment, data focusing primarily on the relationship between the big 5 personality traits and performance in job and academic settings will be discussed.

Early research (up to 2006) attempting to statistically control for the effect of social desirable responding on the relationship between big 5 personality traits and performance showed that despite considerable research across a range of contexts, few notable statistical effects could be found. It appeared that when partialling techniques were applied to the big 5 trait – personality relationship, at least as much substance was being removed as distortion. For example, two meta-analytic papers investigated the ability of social desirable responding questionnaires to act as suppressor variables on the relationship between the big 5 personality traits and performance in job and academic settings backed up this viewpoint. Ones, Visesvaran and Reis (1996) reviewed data from 233 papers, treating social desirable responding as a unitary construct, while Li and Bagger (2006) reviewed the data from 30 studies that had used the BIDR and conducted separate analyses for SDE and IM scales. Despite their differences, both papers found that social desirable responding questionnaires did not act as effective suppressor variables. For example, Li & Bagger tested the effect of removing the shared variance between social desirable responding and the big 5 personality traits using a semi-partial correlational analysis on the relationship between the traits and performance. They found that changes in coefficients after controlling for social desirable responding were typically only in the range 0 to .04. Since suppressor effects should be observed if social desirable responding is distortion, both sets of authors interpreted their results to be consistent with a substantive view of social desirable responding scores.

However, it should be noted that meta-analytical studies have both strengths and weaknesses as measures of statistical effects such as suppression. Their strength is their ability to draw together the effects measured by multiple studies that meet their selection criteria and so test very large samples. Their weakness is that they are global measures. If either egoistic or moralistic biases show reliable statistical effects, but only in certain employment domains, then these more subtle effects could be lost in the meta-analysis 'averaging' process.

Equally, some researchers have discussed how methodological limitations in the social desirable responding questionnaires used might have impacted on the limited reports of significant statistical effects. Galic and Jerneic (2013) have suggested that current questionnaires lack the precision of measurement required to act as an effective suppressor or moderator variable. Certainly questionnaires such as the MC and BIDR scales typically show levels of internal consistency that are towards the lower end of acceptable values for these psychometric scale properties, and if social desirable responding questionnaire scores reflect a substantial mixture of substance and distortion, then the precision of their measurement would be impaired. Griffith and Peterson (2008) have argued that participants may interpret the 'always' or 'never' statements on social desirability scales in general terms. So, for example, the item 'My first impressions of others are always right' could be interpreted to mean that their impressions of others are usually correct. Where this happens high scores on social desirable responding scales would reflect accurate self-descriptions rather than distortion and so attempts to partial out social desirable responding scores would remove substance rather than style. Richman, Kiesler, Weisband & Drasgow (1999) also suggested that the effects of social desirable responding might take on complex non-linear forms such that 'reporting too much of any good trait looks like immodesty, and reporting too much of any bad trait looks like malingering. Hence someone giving a strategically socially desirable response could actually report less of a good trait than someone less strategically but fancifully giving an aggrandizing self-description' (p. 771).

A further line of argument has considered how the use of correlational techniques could have impacted on the results observed. Some researchers have suggested the failure of partialling techniques to prove effective in potential suppressor situations may reflect the rarity of genuine suppressor variables in psychology. Howit and Cramer (2011) note that in psychology in general few examples of notable suppressor effects have been seen and that this may be because of the relatively strict conditions required to show them. This is that, in 'classical suppression', there is a high correlation between a predictor variable and a criterion variable, and the third variable (the suppressor), will only suppress if it has a high correlation with the predictor (or more specifically with its error variance), while not correlating with the criterion. When this situation occurs, it leaves a substantial part of the predictor variable's unique variance to give good prediction of the criterion. But Conger and Jackson (1972) say this specific situation rarely occurs. In Ones et al and Li and Bagger's

meta-analytical studies, attempts were made to regress the big 5 personality traits onto job performance while controlling for social desirable responding. However, in both studies the largest correlation between a big five trait and performance was $r=.27$ and most correlations were smaller in size.

Paunonen and LeBel (2012) used Monte Carlo simulation methods to model the effects of social desirable responding on the predictive validity of personality trait scores for performance. Their analysis suggested that notable suppressor and moderator effects can only be detected when social desirable responding contamination reaches high levels. Paunonen and LeBels' analysis of computer generated samples for three models of social desirable responding compared to no contamination baselines found that at the level of the individual the construct validity of a personality test was compromised by social desirable responding at moderate effect sizes such that it became a mixed measure of the trait and desirable responding. However the test's predictive validity, as measured by correlation coefficients at the group level, was only noticeably affected when social desirable responding reached large effect sizes. The authors suggested the dissociation between construct and predictive validity occurred because, although individuals' personality trait scores were changed by social desirable responding, their rank position within a group on the trait was only substantially altered when social desirable responding reached high levels.

Some authors (Konstabel, Aavik, & Allik, 2006, Leite & Cooper, 2010, Watson, Baranowski, Thopmson, Jago, Baranowski, & Klesges, 2006; and Ziegler & Buehner, 2009) have suggested that social desirable responding may be able show effects such as suppression or moderation, but that approaches based on more advanced statistical techniques are needed to show them. Berry, Page and Sackett (2007) investigated the relationship between the big 5 personality traits and the job performance of 261 mid-level managers at a large American electrical firm while social desirable responding effects were accounted for using the SDE and IM scales. Berry et al argued that the meta-analytical analyses carried out by Ones et al and Li and Bagger were limited in scope because they only examined the semi-partial correlational effects of social desirable responding on the relationship between the big 5 traits and performance. Such analyses only consider the impact of removing shared variance between a trait and social desirable responding on the relationship between the trait and performance. Berry et al conducted moderated regression analyses on their data using

hierarchical regression analysis (HMR). They argued that such analyses were able to consider additional ways that social desirable responding could impact on the trait – performance relationship. For example, semi-partial correlational analysis would not necessarily be sensitive to interactions between social desirable responding and a trait that was at least partially caused by shared variance between social desirable responding and performance. Berry et al tested their critique of the semi-partial correlation approach used by Ones et al and Li and Bagger by first analysing their data using this approach, and then using HMR to see if it detected effects the semi-partial approach had not. Berry et al found no significant suppressor effects using the semi-partial correlation approach for SDE or IM. However, the HRM regression analyses found two significant effects involving SDE and extraversion (a suppression effect) and SDE and neuroticism (a moderation effect). The two effects showed that taking account of SDE improved prediction above that of using the big 5 personality traits alone.

Like Berry et al, Leite and Cooper (2010) have also discussed the impact of the statistical tested used. They have argued that advanced statistical techniques such as Factor Mixture Modelling (FMM) and multidimensional item response modeling (MIRM), which treat social desirable responding as a latent factor, show greater promise than those based on simple correlational techniques. Such advanced techniques have the capacity to reduce error in the measurement of latent constructs by: 1) weighting each scale item for how well it measures the relevant construct rather than assuming that all items measure the construct equally well, and 2) by taking account of the interrelationships between focal scale items and social desirable responding scale items. As a consequence, they measure the effects of social desirable responding more precisely.

Leite and Cooper tested the utility of the FMM approach for modelling social desirable responses on a student end of course evaluation questionnaire answered by a sample of 668 health studies students. FMM is an extension of the factor analysis technique that is used to identify unobserved latent classes (subpopulations) of participants in a data set. The participants answered a short 10-item attitude scale designed to measure the student's attitude towards interprofessional service learning and a short version of the MC scale at the end of a community based interprofessional educational module. In their analysis Leite and Cooper first observed that when using traditional correlational techniques the attitude scale

did not show an overall correlation with the MC scale. However, their subsequent FMM analysis found that the best fit for their data was a two-class FMM model which separated the students into those who showed social desirable responding (8%) and those who did not (92%).

In a similar vein, Watson et al. (2006) tested the ability of IRM and MIRM to reduce social desirable responding distortion in the measurement of latent constructs. While both IRM and MIRM are latent trait model approaches for the analysis of psychological questionnaires and are able to weight scale items for how well they measure the relevant construct, only MIRM can take account of the interrelationships between focal scale items and social desirable responding scale items. The authors collected responses from 473 boys aged 10-14 years on multiple measures of health self-efficacy for exercise and healthy eating, health related behaviours and social desirable responding measured by the lie scale from Reynolds and Paget's Revised Children's Manifest Anxiety Scale (Reynolds & Paget, 1981). Analysis of the data set showed only weak relationships with the lie scale, so in order to test the IRM and MIRM approaches, Watson et al selected a subset of their data to simulate a data set which showed a larger relationship with social desirable responding. Scores on the three types of variables were estimated using IRM and MIRM and then subjected to partial correlational analysis to explore whether social desirable responding acted to artificially boost the size of relationship between self-efficacy and health behavior i.e. that it acted as an unobserved common response variable that contributed variance to the two self-report variables. In the MIRM analyses the self-efficacy scales were modelled as made up of two dimensions, self-efficacy as a first dimension and social desirable responding as a second dimension. When IRM estimates were applied to the partial analysis, correlation coefficients were largely unaffected. However, when MIRM estimates were used the correlations between self-efficacy and health behavior showed a notable reduction in size when the variance associated with social desirable responding was removed. For example, the correlation between self-efficacy for eating fruit and vegetables and self-reported fruit and vegetable intake was reduced from r=.43 to r_p=.29. Given the fact that only estimates made by the MIRM technique showed a social desirable responding effect, it would appear that the ability to include details of the interrelationships between focal scale items and social desirable responding scale items in models improves model fit and is important in detecting

and correcting for social desirable responding effects. This interpretation is backed up by the fact that FMM also considers these relationships in its analysis.

Taken overall, it seems that despite early failures to find reliable statistical effects such as suppression with social desirable responding scales, research after the influential meta-analyses by Ones et al and Li and Bagger suggests that the picture may be more optimistic. The detection of statistical effects appears to be more likely under certain circumstances, such as when there is a strong relationship between the predictor variable and the criterion, the amount of social desirable responding is high and advanced statistical techniques are used. Techniques such as FMM and MIRM show promise, however, they also have limitations. They require large sample sizes, specialist statistical software and researchers with training in the techniques to carry them out. As of yet, they have not been widely used, tested or commented on beyond the authors who have proposed them.

3.5) Attempts to assess the degree to which social desirable responding scales assess substance and distortion (style)

Connelly and Chang (2015) argue that discussion on the substance – distortion debate has often been polarized and framed in an all-or-nothing way. For example, findings of significant correlations between self- and other-ratings on social desirable responding scales have been interpreted to mean that social desirable responding scales measure only substance. However, consensus on the debate has now moved towards the view that social desirable responding scales tap both substance and distortion. Given this, the key question becomes: how much of the variance in social desirable responding scales reflects substance and how much reflects distortion? Three studies provide data that is of relevance to this question (Connelly & Chang, 2015; Lönnqvist et al, 2007; and Pauls & Stemmler, 2003). Two of these studies tested the BIDR under anonymous, low stakes conditions, while Lönnqvist et al tested the BIDR under low and high stakes settings. These three studies will now be described.

Lönnqvist et al (2007) conducted research on Finnish versions of the BIDR SDE and IM scales to explore the extent to which they were measures of distortion versus substantive responding. In their research, Lönnqvist et al. distinguished between three types of influence

on responses: set, style and substantive individual differences, and is the only study reviewed here that tested set. 'Set' was conceptualized as patterns of socially desirable responding that occurred only in situations that created a high demand to present oneself in a positive manner. 'Style' was conceptualised as patterns of responding that were consistent across situations but that exaggerated levels of abilities or adherence to moral standards and so were inauthentic. 'Substance' was conceptualised as consistent responses that were authentic.

Lönnqvist et al collected data from non-student samples (military personnel and the general public), across three studies which included job applicant settings. In study one, 57 male Army officer recruits answered the BIDR scales, first when candidates for training as part of their application, and then 3 years later when they were officers. Study two, replicated study one a year later using a further sample of 62 male recruits with the addition that participants also completed measures of the big 5 personality dimensions in the form of a series of personal adjectives. This time there was a two-year interval between testing and retesting. In study three, 35 Finish couples selected at random from the Helsinki phone book, completed the BIDR scales and the NEO-PI-R. For each couple, one person acted as the participant and answered the scales for themselves, while their spouse answered the scales in relation to their judgment of the participant. Study one and two focused primarily on set and style, while study three focused primarily on substantive individual differences. A range of effects were measured across the three studies that had the potential to suggest whether a scale was tapping set, style or substance. In study one and two evidence suggesting a measure was sensitive to set included changes in mean BIDR scale scores and changes in sizes of correlations with desirable aspects of the big 5 measures from the job applicant situation, where there would be high pressure to give desirable answers, to the employee situations, where there would be less pressure. Evidence suggesting a measure was assessing style included consistency of BIDR scale mean scores across applicant and incumbent situations, and consistency of BIDR correlations with socially desirable poles of relevant big 5 measures. Also, participant rank order stability of BIDR scores across the two situations was taken as evidence for style and / or substantive individual differences. In study three correlations between participant and spouse responses on the BIDR and NEO-PI-R scales were taken as evidence for substantive individual differences. This was based on the assumption that spouse ratings were accurate measures. Across the three studies, the

relative impact of set, style and substance on SDE and IM scores was judged by comparing the sizes of effects and their patterns across conditions.

Lönnqvist et al.'s results were complex, but in essence suggested that that neither BIDR scales were pure measures, rather they were influenced by multiple sources of variance. The SDE and IM scale mean scores and correlations with the big 5 measures were affected by changes in situational demands, but these effects were larger and more consistent for the IM scale. The SDE scale showed generally larger and more consistent correlations with relevant desirable aspects of the big 5 measures than IM. Both BIDR scales showed considerable participant rank order stability of scores across the testing situations (*rs* of .68 and .68 for IM, and .44 and .71 for SDE). In comparison, smaller correlations were seen between participant and spouse ratings on both scales (*rs* of .35 for IM and .33 for SDE) and between self SDE scores and spouse ratings of neuroticism (*r*=-.24) and self SDE scores and spouse ratings of extraversion (*r*=.33). No correlations were seen between participant IM scores and spouse ratings of relevant big 5 traits.

Lönnqvist et al. concluded that both BIDR scales measured set, style and substance but to differing amounts. IM was 'largely' influenced by set, 'somewhat' by style and 'minimally' by substance. SDE was 'mostly' influenced by substance, 'notably' by style and 'somewhat' by set. Lönnqvist et al. suggested IM style took the form of a stable pattern of responding that was only partially corroborated by spouse IM ratings, and largely independent of the big 5 measures of personality as judged by the lack of correlation with spouse big 5 ratings. Individuals believed that they adhere to moral standards to higher level than others and consistently strive to present this impression but, in reality it was only partially true. SDE style responding was argued to take the form of a consistent personal striving across situations to be a 'superhero' that was projected to others but that again was only partially true. SDE substance took the form of higher SDE scores tapping into higher levels of extraversion and lower levels of neuroticism that were actually reflecting the participant's real personality.

Pauls and Stemmler (2003) also used self- and other-ratings to explore whether the SDE and IM BIDR scales measure substance or distortion, but unlike Lönnqvist et al they focused on the differences between self- and other-ratings rather than their similarities. They also

concluded that IM and SDE were influenced by multiple sources of variance. Their results suggested that when tested under the 'normal' anonymous low demand conditions used in their study, that SDE and IM scores measure both substance and response style. The starting point for their research was to argue that researchers have tended to focus on the similarities between self and other measures while overlooking differences between them, and they noted a trend in past studies for a person's scores on measures of social desirable responding to correlate more highly with self-ratings of personality than other-ratings. Pauls and Stemmler tested these ideas empirically in a sample of 67 German psychology students. They collected self- and other-reports (a near friend and at least one acquaintance) on measures of the big 5 personality traits, and egoistic and moralistic biases using shortened German versions of the SDE and IM BIDR scales (the SDE scale had 8 items while the IM had 9 items). Pauls and Stemmler conducted new analyses not conducted by past studies. Specifically they calculated separate self-rating inflation scores for the big five traits in their data by regressing self-ratings on to other-ratings for each of the big 5 traits, and saving the residuals as their indices. These residual scores captured the difference in ratings between self and others, and were then correlated with each BIDR scale. Their analysis showed that both SDE and IM were correlated in meaningful patterns with relevant inflation indexes. SDE correlated with bias indexes for neuroticism ($r=-.32$), extraversion ($r=.23$) and openness ($r=.23$), while IM correlated with bias indexes for agreeableness ($r=.38$) and conscientiousness ($r=.25$). Further, a principle components analysis of the bias scores from the big 5 traits produced two factors which corresponded to the egoistic bias and the moralistic bias. In light of their results, Pauls and Stemmler suggested that even under the anonymous conditions they used that distortions in self-reports exist, and that SDE and IM scales assess both substantive personality traits and distortion. The BIDR scale - inflation index correlations reported above fall in the range $r=.23$ to $.32$ for SDE and $r=.25$ to $.38$ for IM. If Pauls and Stemmler had reported their self-other BIDR correlations it would have been possible to directly compare the sizes of the BIDR scale - inflation index correlations with them to give an indication of whether either substance or style had a larger impact on BIDR scale scores. Another study conducted with German students, de Vries, Zettler and Hilbig (2014), observed a self-other agreement of $r=.34$ on the original 20 item BIDR SDE scale and a self-other agreement of $r=.45$ for the original IM scale, though the scales were double the length of those used by Pauls and Stemmler. Also, in Pauls and Stemmler's study the relevant correlations seen between the big 5 trait self- and other-ratings were on average

.24 larger than the corresponding BIDR - inflation index correlations described previously. Taken together, these patterns of correlations suggest that substance appears to be predominant in both SDE and IM scales.

Connelly and Chang (2015) conducted a meta-analytic multi-trait multimethod (MTMM) study to compile a large data set from eight studies to explore the relative balance between substance and response style in the SDE and IM BIDR social desirable responding scales. In partial alignment with the previous two studies the authors measured substance as the correspondence between self- and other-ratings of the big 5 personality traits. However, the authors measured style in a unique way. They conceptualized stylistic distortion as the common method variance observed across self-reports on the big 5 personality traits. Common method variance is the 'systematic error variance shared among variables measured with, and introduced as a function of, the same method and/or source' and can stem from various causes (Richardson, Simmering, & Sturman, 2009, p. 763). They argued that when CFA is used to separate common method variance from substantive big 5 trait variance, the method factor that emerges reflects participants' general tendency to rate themselves positively regardless of the trait being measured. Thus, it appears reasonable to view common method variance as an effective operationalization of social desirable responding. Also, as with Pauls and Stemmlers' use of big 5 trait self-rating inflation indexes, using an independent criterion of style allowed them to assess the extent to which it shared variance with responses on the BIDR SDE and IM scales.

Connelly and Chang's study used five types of data: 1) self-ratings of the big 5 personality traits; 2) other-ratings of the big 5 personality traits; 3) measures of performance (job and academic); 4) BIDR social desirable responding scales (SDE and IM); and 5) common method variance of the self-report big 5 personality traits as a measure of response style. Connelly and Chang drew together estimates of the size of correlations between their five constructs from several previous meta-analytical studies. The authors then applied SEM latent variable modelling to meta-analytical inter-correlational matrices. The common method variance was modelled as a single factor within each rater that spanned across all of the big 5 traits. In each analysis it was first tested as a whole and then, based on previous studies analysis of big 5 trait scores e.g. Digman (1997), then when split into sub-divisions of egoistic bias (common method variance of extraversion and openness) and moralistic bias (common

method variance of agreeableness, conscientiousness and neuroticism). The SEM explored two issues of relevance: 1) the relative balance between substance and style in BIDR scale responses under anonymous low demand conditions. This was tested by regressing each of the five big 5 personality trait factors estimated from the common variance of self- and other-ratings, and the common method variance factor, onto the criterion variables of the BIDR scales; 2) the extent to which the BIDR SDE and IM scales mirror the impact of common method variance on the predictive validity of self-report big 5 personality traits. This was assessed using a mediation analysis. First a regression model was tested whereby the common method variance factor predicted the BIDR scales, which in turn predicted performance. This model acted as a baseline. Next, a second model was tested which added a direct path from the common method variance to performance. If the second model did not substantially outperform the first in its ability to predict, this would suggest that the BIDR scales capture much of the effects of common method variance i.e. style.

The analysis of the study's data produced the following findings: 1) overall, the BIDR scales were observed to tap 'considerable' substantive trait variance (shared variance of self- and other-rated big 5 scores) and 'some' stylistic distortion (common method variance) i.e. more substance than style. In terms of substance, SDE correlated with neuroticism and conscientiousness, while IM correlated with agreeableness and conscientiousness; 2) the second analysis tested the extent to which the BIDR scales captured the suppressive effect of the common method variance on performance. The mediation analysis showed that both SDE and IM scales captured a substantial proportion of the common method variance effect (57% and 59%). Overall the second model, which included a direct path from the common method variance to performance, did not substantially outperform the mediation model in its ability to predict performance. In conclusion, Connelly and Chang stated that their analyses provided mixed support for the BIDR SDE and IM scales as their scores appear to reflect a combination of substance and style.

Taken together the three studies reviewed here, despite using different samples and methodologies, show a large degree of consistency in their results. All show that even under anonymous, low stakes response conditions, the SDE and IM scales tap into response style i.e. a consistent personal tendency to distort responses in a positive direction. The strength of the style effect lies towards the lower end of the medium effect size range. However, the

studies suggest that under these conditions, the BIDR scales are influenced to an even greater extent by substance. Results relating to the exact nature that 'substance' takes varied across the studies, though both Lönnqvist et al and Connelly and Chang found SDE to relate to other-rated neuroticism levels. The substantive effect strength lies towards the upper end of the medium effect size range. However, the estimates of substance effect size derived from these studies probably reflect an upper bound estimate given that factors such as the influence of friendship biases on others-ratings might act to inflate estimates of substance. Further, when situational factors increase the pressure to respond desirably, the proportion of variance in social desirable responding scales loading on distortion (style and set combined) is likely to be higher than in low demand situations.

3.6) Social desirable responding as a three-way interaction

Leite and Cooper (2010) have argued that past studies (e.g. Holtgraves, 2004; Lönnqvist et al., 2007; and Richman, Kiesler, Weisband & Drasgow 1999) suggest that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the characteristics of: 1) the items making up the focal scale; 2) the respondent i.e. individual differences in levels of response style; and 3) the situation that the testing takes place in. Details of these three elements are show in Table 8.

Table 8.

Description of the three elements in Leite and Coopers' (2010) conception of social desirable responding as a three-way interaction

Element	Details
Focal scale	Desirability of scale items, for example, items measuring intelligence are likely to carry the highest level of desirability
Respondent characteristics	Individual differences in levels of egoistic and moralistic biases
Situation characteristics	Situational factors such as: the importance of the situation e.g. job applicant versus incumbent; anonymous versus identifiable responses; individual versus group testing; ability to back-track on responses if a computerized questionnaire administration is used

For example, Richman et al (1999) conducted a meta-analysis of social desirable responding effects in non-cognitive questionnaires such as personality, attitude, and social desirable responding scales, administered by computer, paper and interviews in 61 studies. Past research had found conflicting results with some researchers finding less social desirable responding with computer administration, others finding more, and others no difference. Richman assessed the impact of social desirable responding on the questionnaires in several ways including testing whether mean responses on social desirable responding questionnaires showed relationships with the nature of their administration, and more indirect ways such as testing whether mean responses on other questionnaires increased in hypothesized social desirable meaningful directions in specific conditions. Based on the results of a series of hierarchical regression analyses, Richman concluded that computer administration did have effects on levels of social desirable responding. However, these effects depended on focal scale characteristics, specifically whether the scale contained sensitive items asking about personal weaknesses such as feeling lonely or ever wanting to kill a person or not. Also situational characteristics such as whether participants were tested on their own or in groups, whether participants were anonymous or identifiable and whether they could back track on their answers or not. Further, there was an interaction between the focal scale characteristics and the situation characteristics such that participants were less likely to admit personal weaknesses if they were in groups, identifiable and denied the ability to back track.

It is possible to apply Leite and Cooper's proposal that social desirable responding depends on a three-way interaction to the ideas behind Paulhus's (2002) two-tier model and the properties of the four BIDR scales. When situational pressures to respond in a desirable manner are low, the occurrence of social desirable responding would reflect a two-way interaction between the desirability of the focal scale items and the level of egoism or moralism possessed by an individual. However, when situational pressures are high, the occurrence of social desirable responding would reflect a three-way interaction between the desirability of the focal scale items, the level of egoism or moralism possessed by an individual and the demands of the situation. According to Paulhus (2002), the SDE and SDD scales are best suited to detecting social desirable responding effects in low demand situations, while the AM and IM scales are best suited to detecting social desirable responding effects in high demand situations.

3.7) Recent research (2009-2015) testing Croatian versions of the full BIDR scales

Four recent papers from a research group in Croatia have tested the two dimensions of Paulhus's (2002) two-tier model, the egoistic – moralistic and unconscious – conscious dimensions, and the performance of a Croatian version of Paulhus's four BIDR scales. The papers all used designs incorporating fake-good instructions in simulated job or student application situations designed to illicit either egoistic or moralistic biases. For example, instructing participants to make their responses on BIDR scales and / or big 5 trait questionnaires as if they were applying to be a manager of a business (egoistic) or to be a teacher in a school (moralistic). Responses made in these experimental conditions were compared to responses made in an honest responding baseline condition. The earlier Croatian papers used the fake-good paradigm to test the model with all four of Paulhus's BIDR scales, while the later papers used more sophisticated designs to test additional aspects of the AM and IM scales and to develop two new scales.

Galic, Jerneic and Belavic (2009, cited in Dodaj, 2012, and Kovacic, Galic & Jermeic, 2014) and Dodaj (2012) used the basic fake-good paradigm to test Paulhus's (2002) two-tier model using all four BIDR scales. Their results produced support for the egoistic – moralistic distinction. For example, Dodaj (2012) found in their honest response condition that the BIDR egoistic scales clustered together ($r=.53$) as did the moralistic scales ($r=.63$). Also the manager applicant condition (egoistic) produced the highest mean scores on the SDE and AM scales, while a teacher applicant condition (moralistic) produced the highest mean scores on the SDD and IM scales. However, in contrast, the studies showed less clear-cut support for the unconscious – conscious distinction. The authors used two criteria to test this dimension. They suggested that if the dimension was valid the BIDR scales should show the following: 1) in terms of how mean BIDR scale values are affected by a switch from low to relevant high demand instructions, it was predicted that the AM and IM scales should be sensitive and show a large increase, while the SDE and SDD scales should be insensitive and not change; and 2) in terms of how BIDR correlations with desirable aspects of the big 5 traits are affected by a switch from low to relevant high demand instructions, it was predicted that AM and IM should only show relationships with personality variables under conditions of high demand that require positive self-presentation, while the SDE and SDD scales should be insensitive and not change. The exact rationale behind the second

assumption was not made explicit in their work, but it would appear to reflect the idea that the impression management scales should tap exclusively into the set form of social desirable responding and so should only correlate with desirable aspects of the big 5 when a relevant situational incentive to respond in a desirable manner is present. However, in their analyses Galic et al and Dodaj both observed all four BIDR scales to be sensitive to situational pressures, all four scales' mean scores were higher in the relevant applicant conditions and correlations between the BIDR scales and the big 5 traits were larger when required to fake-good. Further, Dodaj found in the honest response condition that the AM and IM scales correlated with some of the big 5 personality traits which she interpreted to indicate substantive relationships. Finally, although not commented on by Dodaj, it is worthy of note that in her study the impression management scales (AM and IM) showed the largest increases in mean values between the honest and applicant response conditions, and a trend for correlations with desirable traits to increase more in size compared to the self-deceptive scales, although only by a small amount.

A third Croatian paper, Galic and Jerneic (2013), used the same general fake-good experimental paradigm as the two previous Croatian studies to test the ability of the two BIDR impression management scales (AM and IM), answered under applicant instructions requiring positive self-presentation, to offer effective faking interventions. First by attempting to correct big 5 personality scores for distortion, and second by the less ambitious strategy of identifying participants who have faked big 5 personality scores. Semi-partial correlation suppression analyses were used to test the effectiveness of the former, while moderation regression analyses were used to test the latter. Galic and Jerneic found that while neither AM nor IM scales were able to effectively correct faked big 5 trait questionnaire scores in their semi-partial correlation analyses, the two scales did show the ability to act as moderators of the relationship between the honest and applicant personality trait scores in some of their big 5 trait analyses.

A fourth Croatian paper, Kovacic, Galic and Jermeic (2014), attempted to overcome the problems seen with the full four BIDR scales performance in relation to the unconscious – conscious distinction by developing two new scales which would be primarily sensitive to unconscious self-deceptive processes in honest response conditions and primarily sensitive to conscious impression management processes in applicant situations which created

demands for positive self-presentation. One of the new scales would be egoistic while the other would be moralistic. Kovacic et al conducted two studies. In study 1 two new scales were developed, the egoistic and moralistic social desirable responding scales (E-SDR and M-SDR). Samples of participants answered the four BIDR scales and a measure of the big 5 personality traits under the same three response conditions as past Croatian studies: honest responding, and simulated applicant conditions designed to induce the egoistic bias and the moralistic bias. Items were selected for the new Croatian scale based on two criteria: 1) loadings of .3 or greater on egoistic and moralistic factors produced from a series of factor analyses of the BIDR scales under the different response conditions, 2) items which showed the largest difference in scores between honest and either egoistic or moralistic applicant conditions. These procedures resulted in the creation of two new ten item scales. The E-SDR was made up of six items from the AM scale and four items from the SDE scale, while the M-SDR was made up of six items from the IM scale and four items from the SDD scale. The reliabilities of the two scales were modest but argued to be acceptable given their length (E-SDR $\alpha=.70$ in the honest condition, M-SDR $\alpha=.66$ in the honest condition) and they showed only a low inter-correlation in the honest condition ($r=.18$).

In study 2 the two new scales were tested to see if they were primarily sensitive to unconscious self-deceptive processes in honest response conditions and conscious impression management processes in applicant situations which created demands for positive self-presentation. Kovacic et al tested this using other-ratings on big 5 personality traits sampled from four peers and by manipulating the situational demands to self-present as they had done in study 1. Differences in self-reports of personality traits in the honest condition from the other-reports was taken as evidence of unconscious self-deceptive processes, while differences in self-reports of personality traits between honest and relevant applicant conditions were taken as evidence of conscious impression management processes. The results of study 2 supported the two new scales. Across all conditions the E-SDR was most sensitive to egoistic relevant measures and the M-SDR to moralistic relevant measures. In honest conditions the SDR scales most strongly correlated with self-peer personality trait discrepancy scores (medium effect sizes) and less so with relevant other-ratings of personality (small effect sizes). In applicant conditions the SDR scales most strongly correlated with self honest – applicant personality trait discrepancy scores (up to large effect sizes) and less so with self-peer personality trait discrepancy scores (small effect

sizes) and relevant other-ratings of personality (small effect sizes). Based on these findings, Kovacic et al concluded that the new E-SDR and M-SDR scales were reliable and valid indicators of social desirable responding. They also argued that their findings provided support for both dimensions of Paulhus's (2002) two-tier model. Thus Kovacic et al argued that the previous studies using the Croatian version of the BIDR appear to have failed to find support for the unconscious-conscious dimension because of limitations of the BIDR scales rather than of the two-tier model.

It is pertinent to note that while Paulhus (1994, 1998) has reported correlations between the SDE, IM scales and the big 5 personality traits, corresponding correlations have not yet been reported for the AM and SDD scales. In the Croatian BIDR studies reported above in honest responding conditions, the AM scale had less of an association with neuroticism than the SDE scale, and SDD showed associations with all of the big 5 traits.

Taken together, the recent Croatian studies provide support for Paulhus's (2002) distinction between egoistic and moralistic biases. Less support was found for the distinction between unconscious and conscious biases, however, this may have been due to limits in Paulhus's questionnaires rather than the two-tier model itself. Evidence was found that the BIDR self-deceptive scales (SDE and SDD) are affected by situational factors, but to a lesser extent than the BIDR impression scales (AM and IM). Neither AM nor IM appear able to effectively correct faked big 5 trait questionnaire scores using semi-partial correlations, however, the scales show a degree of ability to act as moderators. Finally, the new E-SDR and M-SDR scales showed good performance in the one study that has tested them.

3.8) Indices of departure from reality

Researchers have also attempted to develop objective indexes of departure from reality. Two such types of indices are Paulhus, Harms, Bruce and Lysy's (2003) over-claiming measure of cultural knowledge, and Kwan et al (Kwan, John, Kenny, Bond & Robins, 2004; and Kwan, John, Robins & Kuang, 2008) and Leising, Locke, Kurzius and Zimmermann's (2016) indices of the general tendency to rate oneself more favourably than one is rated by others, and more positively than one rates others.

3.8.1) Paulhus et al's (2003) over-claiming measure of general cultural knowledge

Paulhus et al (2003) have developed an objective measure of academic / artistic general cultural knowledge and knowledge over-claiming called the over-claiming questionnaire (OCQ). On the OCQ, when all 150 items are administered, respondents rate their familiarity with items from 10 academically oriented general knowledge domains e.g. historical names and events, and social science and law. Ratings are usually made on a five-point response scale: 'never heard of it' to 'very familiar'. Of the items, 80% relate to factual entities while 20% are non-existent 'foils'. Examples of factual items from the historical OCQ category are Napoleon and Robespierre, and examples of foils are Prince Lorenzo and Queen Shattuck. OCQ responses can be analysed with signal detection formulas to generate indices of knowledge accuracy (extent of general knowledge) and knowledge over-claiming. To calculate indices, participants' responses are typically reduced to dichotomous scores: responses of 1-4 score 1 point i.e. the respondent expresses familiarity with an item regardless of level, while a response of 0 scores 0 points. A number of accuracy and over-claiming indices exist. Accuracy indices are based on participants' ability to distinguish real items from fictitious items, while knowledge over-claiming indices are based on the tendency to claim familiarity with either just foils (number of false alarms) or all items ('bias score', calculated by adding the number of hits and the number of false alarms together). The bias score index is based on the idea that participants will over-claim on all items, not just foils i.e. responses to factual entities will capture overclaiming when participants claim knowledge of entities that they are not actually familiar with. Paulhus et al. (2003) argue that the over-claiming bias score has the advantage that it is based on all 150 items. In contrast false alarms are based on only 20% of the items and so may have constrained power, particularly as some participants claim no familiarity with the foils. In contrast, Ludeke and Makransky (2016) argue that the false alarm measure has the advantage of being an unambiguous indicator of overclaiming.

The OCQ accuracy index has received support as a measure of general knowledge and has been found to correlate with IQ. Paulhus et al. (2003) has presented the OCQ over-claiming indices as measures of the unconscious egoistic bias and this claim has received some support. For example, Paulhus et al. (2003) found the OCQ to correlate significantly with SDE, $r=.30$, but not with IM or SDD. Amati, Oh, Kwan, Jordan, and Keenan (2010) have

investigated the neuronal basis of over-claiming responses using Transcranial Magnetic Stimulation (TMS) and a virtual lesion design. Participants were presented with 50 items taken from Paulhus's OCQ, one at a time, that they were told formed a 'popular cultural literacy' IQ test. Amati et al found that only TMS applied to the medial prefrontal cortex reduced overclaiming and reduced response times.

However, Dunlop et al (2017) tested the relationship between OCQ knowledge over-claiming and the big 5 personality traits, plus honesty-humility, using both self- and other-ratings on the HEXACO Personality Inventory – Revised (HEXACO-PI-R; Lee & Ashton, 2004) and found evidence that OCQ overclaiming scores are also influenced by other psychological mechanisms aswell. First, an indirect effect of openness whereby it acts as a correlate of accumulated cultural knowledge. It appeared that possessing a large amount of cultural knowledge leads to a memory bias whereby plausible cultural knowledge foils are more likely to be perceived as familiar. Second, a direct effect of openness independent of knowledge whose exact underlying mechanism was less clear. The authors speculated that this effect may stem from openness related individual differences in apophenia, the general tendency to see connections and patterns where none actually exists which might generalise to novel academic and artistic stimuli. Dunlop et al's data suggests that the strength of the openness effects were noticeably lower with the false alarm scores than the over-claiming bias index.

3.8.2) Kwan et al (2004, 2008) and Leising et al's (2016) indices

Kwan et al (2004, 2008) and Leising et al (2016) have adopted an interpersonal, componential approach to develop indices of general self-deceptive enhancement. The first version of the index was developed and tested by Kwan et al (2004 & 2008) in America. They used a sophisticated componential technique to isolate the general tendency to rate oneself more favourably than one is rated by others, and more positively than one rates others. The measure of self-deceptive enhancement was derived from a series of long-term small student study group interactions. For example, in Kwan et al (2004) the study groups were made up of 4-5 students who met each week for three months and worked on group assignments. At the end of the last group meeting, participants rated themselves, their peers and were also rated by all of their peers in terms of positivity of perceived traits in a 'round-

robin' format. These ratings were used to decompose self-perception into: 1) 'perceiver effects', how participants saw themselves and others, and 2) 'target effects', how participants saw themselves and were seen by others. Then, the two components were used to mathematically isolate the unique self-deceptive component. Self and other perceptions of overall performance on the task were also measured at the end of the interactions on a scale ranging from zero to 100, with higher scores indicating better task performance. Kwan et al (2004) found that the self-deception index correlated negatively with overall ratings of task performance in the group made by other members in the group's last meeting (*r*=-.42).

Leising, Locke, Kurzius and Zimmermann (2016) designed a simplified, less time consuming method to estimate the self-deceptive index, and developed aspects of the index's calculation to produce a more refined measure. They also correlated their self-deceptive index with a German translation of the BIDR SDE and IM scales along with a number of other self-report questionnaires and three intelligence tests. In Leising et al's approach each participant's behavior was video recorded and rated by four trained judges across a range of tasks, such as describing a recent successful experience. The participant also watched their own video and rated their behaviour as well as recordings of four other people performing the same tasks. To calculate their index Leising et al weighted the trait judgements by ratings of the trait's estimated desirability value rather than assuming all traits carried equal weight as Kwan had done. Then the participant's self-judgements were regressed onto both standard perceiver and standard target judgements to produce residual scores which were independent of them both. The resulting residuals served as their index.

When tested against the main two BIDR scales, Leising et al's self-deceptive index correlated (*r*=.34) with SDE and (*r*=.24) with IM. This result suggests that although it was 'general' in nature, it was more closely aligned with the egoistic bias than the moralistic bias. Leising et al also tested whether item desirability moderated the relationship between the index and self-ratings across a range of 157 self-report questionnaire items taken from seven self-report questionnaires. Correlation coefficients of the relationship between the index and participants self-ratings on the questionnaire items themselves correlated with the desirability of the items. The 157 correlations ranged from (*r*=.31) for desirable items to (*r*=-.32) for undesirable items, with evaluatively neutral items showing zero correlation. Participants scoring high on the index (self-enhancers) tended to rate themselves highly on

desirable items and lowly on undesirable items. So item desirability did moderate the influence of the index on self-ratings of personality, with the upper limit of correlations for highly desirable traits being around *r*=.30. A further analysis focused on intelligence because it is a highly desirable trait. The self-deception index score and intelligence test scores derived from three short IQ tests were regressed on to self-ratings of intelligence. This analysis indicated that the index predicted self-ratings of this desirable trait independently (β=.27) of actual IQ values (β=.27). The results suggest that participants' self-ratings of intelligence were determined by a realistic assessment of IQ with a self-deceptive enhancement component layering on top.

Leising et al's work did not involve mental imagery. However their findings, particularly those involving self-reported intelligence have parallels to the research undertaken into self-reported mental imagery described later in the thesis. Mental imagery would also appear likely to be an ability valued by individuals, and as such may be open to similar self-deceptive biases. Leising et al's analyses suggest that in low stakes testing situations: 1) self-reported mental imagery ratings may be influenced by both realistic self-assessments and self-deceptive enhancement; 2) correlations between mental imagery measures and of social desirable responding are likely to be around *r*=.30 in size.

3.9) Summary of key points

- Peoples' self-reports of their experiences are often considered to be valuable sources of information about their psychological functioning. However, while self-reports as a methodology have several positive features, they are open to a number of potential biases. Paulhus (1991, p17) defines a response bias as 'a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content' i.e. what the items were designed to measure. Response biases are traditionally classed as either 'sets' or 'styles'. Response sets are short-term biases affecting a person's responses on a questionnaire, caused by some temporary motivation or pressure in a situation. Response styles are long-term biases that consistently affect a person's responses across time and different questionnaires.

- Paulhus (2002, p.50) defines social desirable responding as "…the tendency to give overly positive self-descriptions". Paulhus (2002) conceptualizes social desirable

responding as a multifaceted construct and has proposed a two-tier model of social desirable responding. The model has a primary distinction in terms of 'personality content' between egoistic and moralistic biases. The egoistic bias refers to the tendency to claim positive social and intellectual qualities, while the moralistic bias refers to the claiming of positive moral qualities and the refutation of negative socially-deviant qualities. This distinction can be summarized as the difference between claiming to be a 'superhero' or a 'saint'. In the second tier, both the egoistic and moralistic biases are split into conscious and unconscious forms. This yields four kinds of social desirable responding: SDE, AM, SDD and IM.

- The logic behind tests of social desirable responding is that they contain items that make claims which for most people would be unlikely to be true. So when people endorse many of them, it would seem likely that their answers are not correct, and that they are exhibiting socially desirable responding. Paulhus has developed a questionnaire to assess the four different forms of social desirable responding proposed by his model called the BIDR. Paulhus (1994, 1998, 2002) provides a range of sources of support for the validity of the two most well established scales, SDE and IM. However, there is debate over aspects of the scales such as their scoring method.

- Traditionally social desirable responding has been viewed as distortion or error that should be eliminated or minimized (Domino & Domino, 2006). In terms of response style, individuals scoring high on social desirable questionnaires are argued to show a consistent tendency to overestimate their standing on desirable qualities. When applied to self-report questionnaires measuring a desirable characteristic, this tendency results in a consistent pattern of score distortion. As a consequence, self-report questionnaires that correlate highly with measures of social desirable responding are considered invalid, or at very least, suspect.

- However, the view of social desirable responding as simply reflecting distortion has been questioned. Increasing numbers of psychologists have argued that social desirable responding in some sense reflects substantive or valid responding. Several different substantive views of social desirable responding have been proposed. One substantive view of social desirable responding is that it is better viewed as a meaningful personality trait in its own right, or set of traits such as egoism and moralism, that correlates with a variety of psychological constructs and behaviours. A second way substantive responding has been viewed is to argue that, although high scores on social desirable

responding questionnaires represent a departure from reality, on a conceptual level they form an integral and inseparable part of many different constructs of personality. A third way substantive responding has been viewed is to argue that, despite their excessive nature, high scores on social desirable responding questionnaires actually reflect genuinely accurate responses and real virtues, at least when made under low demand circumstances. In relation to imagery self-report questionnaires, the second and third perspectives on social desirable responding as substance are most relevant.

- Support for the substantive position comes from two main sources: 1) findings that self-ratings made on social desirable responding questionnaires correlate with other-ratings made on the same questionnaires and also with other-ratings made on measures of the big 5 personality traits; and 2) findings that despite considerable research into social desirable responding across a wide range of contexts, few notable occurrences of statistical effects such as suppression or moderation have been found. Equally, limitations in these data sets and counter arguments have also been made. For example, the failure to observe reliable partialling effects may be due to a range of factors including the lack of precision of measurement of social desirable responding questionnaires or the lack of sophistication of the statistical techniques used. Recently, some statistical effects have been seen with SDE when using advanced statistical techniques.

- Connelly and Chang (2015) argue that discussion on the substance – distortion debate has often been polarized and framed in an all-or-nothing way, when it is likely that BIDR scores are influenced by both factors. The handful of studies that have attempted to determine the degree of influence of substance and style on the SDE and IM BIDR scale scores have supported this view. They have shown that even under anonymous, low stakes response conditions, the SDE and IM scales tap into response style i.e. a consistent personal tendency to distort responses in a positive direction. The strength of the style effect lies towards the lower end of the medium effect size range. However, the studies suggest that under these conditions, the BIDR scales are influenced to an even greater extent by substance. Results relating to the exact nature that 'substance' takes varied across the studies, though both Lönnqvist et al (2007) and Connelly and Chang found SDE to relate to other-rated neuroticism levels. The substantive effect strength lies towards the upper end of the medium effect size range. However, the estimates of substance effect size derived from these studies probably reflect an upper

bound estimate given that factors such as the influence of friendship biases on others-ratings might act to inflate estimates of substance. Further, when situational factors increase the pressure to respond desirably, the proportion of variance in social desirable responding scales loading on distortion (style and set combined) is likely to be higher than in low demand situations.

- Leite and Cooper (2010) have argued that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the characteristics of: 1) the items making up the focal scale; 2) the respondent i.e. individual differences in levels of response style; and 3) the situation that the testing takes place in. Leite and Coopers' conception of social desirable responding is supported by the available evidence and provides a useful framework for organising the different types of factors that affect whether social desirable responding will occur.

- Four recent papers (Dodaj, 2012; Galic & Jerneic, 2013; Galic et al, 2009; and Kovacic et al, 2014) have tested all four BIDR scales using a Croatian version of the BIDR using a fake-good instruction paradigm. Their results produced support for the egoistic–moralistic distinction. However, less clear-cut support was found for the unconscious–conscious distinction, although this may have been due to limits in the BIDR scales rather than the model itself. The studies found that the BIDR self-deceptive scales (SDE and SDD) were affected by situational factors, but to a lesser extent than the BIDR impression scales (AM and IM). Also, that neither AM nor IM scales appear able to effectively correct faked big 5 trait questionnaire scores using semi-partial correlations, however, they do show a degree of ability to act as moderators. Kovacic et al developed two new social desirable responding scales (the E-SDR and M-SDR) from the BIDR items which showed good performance in the one study that has tested them.

- Researchers have also attempted to develop objective indexes of departure from reality. Paulhus et al (2003) developed a measure of academic general knowledge and knowledge over-claiming called the over-claiming questionnaire (OCQ). Paulhus views the OCQ as a measure of the egoistic bias, however, responses on the questionnaire appear to be influenced by other factors such as possessing a large amount of cultural knowledge. Also Kwan et al (2004, 2008) and Leising et al (2016) have developed an index to measure the general tendency to rate oneself more favourably than one is rated by others, and more positively than one rates others. Leising at al found self-

report questionnaire item desirability moderated the influence of the index on self-ratings of personality, with the upper limit of correlations for highly desirable traits, such as intelligence, appearing to be around r=.30. Furthermore, their data suggested that participants' self-ratings of intelligence were determined by a realistic assessment of IQ with an equally influential self-deceptive enhancement component layering on top.

4) Self-report imagery questionnaires and social desirable responding

This section brings together the two previously discussed areas of self-report imagery questionnaires and social desirable responding. Factors which have the potential to make self-report imagery questionnaires open to the influence of social desirable responding are presented using Leite and Cooper's (2010) 'three-way interaction' framework. This is followed by description of the small amount of experimental and correlational research that has been conducted into the relationship between imagery questionnaire scores and social desirable responding.

Within the mental imagery literature social desirable responding has been conceptualised as distortion / error. From this perspective scores on self-report imagery questionnaires may be influenced by social desirable responding because they involve rating experiences that are subjective in nature and we may value. Self-report imagery questionnaires offer the possibility of a simple, easy and flexile way to measure differences in imagery ability. People can fill them in anywhere there is quiet and no timing is necessary. However, from the distortion perspective, they are only useful measures of imagery ability if they are not excessively confounded by social desirable responding.

Several authors have discussed the possibility that responses on theVVIQ-1 might be influenced by social desirable responding. For example, Killstrom et al (1991) collected responses from 730 American psychology students which produced a mean score of 2.08 with a standard deviation of 0.74. The distribution had a positive skew with most participants reporting that their images had been at least 'moderately clear and vivid' (a mean rating of 3). Less than 3% of their sample reported that their images were 'vague and

dim' (a mean rating of 4) or could form no images (a mean rating of 5). Killstrom et al noted that the positive skew in the VVIQ distribution can be interpreted in several ways. The most optimistic interpretation is that the skew reflects the fact that most people actually have relatively vivid visual imagery. However, the phenomena is also consistent with the possibility that participants' responses are inflated by social desirable responding. Such possibilities are lent credence by the fact that imagery experiences are typically thought of as being less vivid than perceptual experiences.

Factors that might potentially influence the occurrence of social desirable responding on self-report imagery questionnaires can be discussed within Leite and Cooper's (2010) three-way interaction framework. Leite and Cooper argued that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the characteristics of: 1) the items making up the focal scale; 2) the respondent i.e. individual differences in levels of response style; and 3) the situation that the testing takes place in.

4.1) Focal scale items

Several characteristics of self-report imagery questionnaires might encourage social desirable responding. In relation to questionnaire items, it would appear likely that some imagery properties are viewed as valued abilities, such as the ability to generate vivid visual images or the ability to generate or maintain them with ease. Participants might value such imagery characteristics in their own right, or because of the associations they have with them. For example, participants might associate vivid imagery with valued perceptual processes, traits such as intelligence or creativity, or with the utility of vivid images for memory or learning.

Aside from self-report imagery questionnaire items themselves, it is possible that some imagery questionnaire instructions may prime participants to think of imagery properties such as vividness as abilities that play an important role in human psychological processes and to consider where they fall on the imagery ability range. For example, Ahsen (1990) has argued that the instructions on the VVIQ-1 communicate to participants that visual imagery is an 'ability', that 'marked individual differences have been found in the strength and clarity

of reported visual imagery' and that differences in participants' vividness are 'of considerable psychological interest'.

In addition, it is possible that the subjective nature and ambiguity of rating imagery experiences may make responses on self-report imagery questionnaires more open to influence by extraneous variables such as social desirable responding. The research described previously by Chara and Verplank (1986) which investigated inter-participant consistency in assigning vividness ratings to visual percepts that varied in clarity, provides empirical verification of the difficulty of the task facing participants when they attempt to make VVIQ ratings.

Ambiguity in rating imagery properties may stem from the nature of the imagery experience itself or from ambiguity in the questionnaire rating scale used to describe the experience. For example, the visual images that participants attempt to inspect may change in vividness over time, or different parts of an image may differ in vividness at any particular moment. Equally, response scales require interpretation by respondents. For example, the Shapes questionnaire response scale is made up of numbers 1 to 9 which only have verbal labels for their lowest and highest values. This leaves the participant to have to decide what each number equates to. The VVIQ-1 has a sequence of five numbers each with its own verbal label. But even then, a participant must decide for themselves exactly what distinguishes between points on the scale, such as '2 Clear and reasonably vivid' as compared to '3 Moderately clear and vivid'.

4.2) The respondent

In terms of participant individual differences, possessing the tendency to give socially desirable responses would predispose a participant to give desirable answers. If participants answering self-report imagery questionnaires view the imagery property being tested as a valued ability, then the egoistic bias with its emphasis on claiming positive social and intellectual qualities, is most likely to be a threat to the validity of self-report imagery questionnaires. Paunonen and LeBel (2012) have suggested that these individual differences might operate in different ways. One possibility is that the egoistic bias is activated independently of true imagery ability and so has equal effects across all levels of imagery

ability. Another possibility is that the egoistic bias will primarily be stimulated in participants who are low in imagery ability, because they will experience the greatest need to engage in ego maintenance

4.3) The situation

Typical self-report imagery questionnaire populations and testing situations may also act to encourage social desirable responding by creating situations that are ego-threatening. Self-report imagery questionnaires are typically answered by groups of students in educational settings with an experimenter present in conditions of quiet or silence. Quiet is required to concentrate on forming the images required and to introspect on the imagery experiences. These aspects, quiet, group testing, the presence of an authority figure and taking part in a psychology study, may act to emphasize the test-taking or 'exam-like' aspects of the situation for participants and the extent to which their psychological functioning is being assessed. Also, for participants who may be relatively new to participation in psychological studies, such as non-psychology students or first year undergraduate psychology students, the role of the experimenter may not be fully understood. Such participants may perceive the experimenter to be an 'audience' in the sense that they may be privy to the questionnaire responses given once they are handed in. While most of these suggestions are speculative, as previously discussed, Richman et al's (1999) meta-analysis found social desirable responding to increase in group testing situations. The three types of factors that might potentially influence the occurrence of social desirable responding on self-report imagery questionnaires are summarized in Table 9.

Empirical research into the relationship between self-report imagery questionnaire responses and social desirable responding has been limited. The studies that have been conducted have taken one of two main approaches: experimental or correlational.

Table 9.

Factors that have the potential to influence the occurrence of social desirable responding on self-report imagery questionnaires framed within Leite and Coopers' (2010) conception of social desirable as a three-way interaction

Element	Details
Focal scale	Desirability of scale items, questionnaire instructions, ambiguity of imagery task
Respondent characteristics	Individual differences in levels the egoistic bias
Situation characteristics	Situational factors such as: quiet; group testing; the presence of an authority figure; taking part in a psychology study; and perception of anonymity

4.4) Experimental studies

Some researchers have manipulated the experimenter demands in a situation and observed their effect on self-report imagery questionnaire responses. In these studies three kinds of variables have been manipulated: 1) the perceived social desirability of vivid mental images; 2) the perceived difficulty of generating vivid mental images; and 3) the anonymity of responses. Some studies (Ashton & White, 1975; Ashton, White & Brown, cited in White, Ashton & Law, 1978) have found that manipulating the desirability of vivid mental imagery can affect participant's responses on Sheehan's QMI, but this effect was not observed by Durndell and Wetherick (1975). McKelvie (1979) manipulated participants' perceptions of the difficulty of generating mental images and found that reported imagery vividness on the VVIQ was reduced when it was suggested that generating mental images was difficult. Rinaldo and Okada (1993) found no relationship between participant anonymity and responses on the VVIQ-1. While these studies are limited in scope and their findings are not always consistent, taken overall, they do suggest that participants' beliefs about imagery characteristics can be affected by information they receive prior to answering questionnaires and that this information can influence the responses they give.

4.5) Correlational studies

Other researchers have correlated responses on self-report imagery questionnaires with psychometric measures of socially desirable responding. The MC social desirability scale has been used most in these kinds of studies. White, Sheehan and Ashton (1977) reviewed the studies that had investigated the relationship between two imagery self-report scales, Sheehan's QMI and Gordon's (TVIC: Gordon, 1949), and the MC scale. They found evidence for a relationship to be mixed, sometimes results supported an association with the MC but other times it did not. Where significant relationships were seen, they showed greater mental imagery ability was associated with greater socially desirable responding, but they were typically small in size. For example, the first study of this type, DiVesta, Ingersoll and Sunshine (1971, cited in Durndell & Wetherick, 1975), found a correlation between Sheehan's QMI and the MC of $r=.29$.

McKelvie (1994) proposed a set of criteria for judging the levels of 'contamination' from socially desirable responding in imagery self-report questionnaires. These were based on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by other authors. He suggested that a non-significant correlation r equated to 'no contamination', a correlation of $r=.15$ or less equated to 'inconsequential contamination', a correlation of $r=.16$ to $.25$ equated to 'some contamination, but of an acceptable level', and a correlation of above $r=.25$ indicated 'unacceptable contamination' and so should be the 'cut-off' point for acceptability.

McKelvie's (1995a, 1995b) review papers on Marks's VVIQ-1 used meta-analytical techniques to assess the size of its relationship with measures of social desirable responding, and then applied his criteria to evaluate its importance. McKelvie (1995b) made two estimates of the degree of relationship between the VVIQ and socially desirable responding. The first estimate was based on correlations with the MC scale alone. Data from 10 studies lead to a reported an r of $.19$ with 95% CI [.08 to .30], with more vivid imagery being associated with greater socially desirable responding. The second estimate added a further 6 correlations to McKelvie's data pool from three other tests of socially desirable responding used in a study by Anderson and Buyer (1994). Inclusion of this extra data produced a

revised *r* of .08 with 95% CI [-.04 to .11], indicating a much weaker effect whose confidence interval overlapped zero. Based on these statistics, McKelvie concluded that "…the VVIQ is not seriously contaminated by socially desirable responding, and is perhaps not contaminated at all" (1995b, p. 209).

Thus, the empirical evidence available prior to the five papers to be presented next in the thesis suggested that if there was an effect of social desirable responding on self-report imagery tests such as the VVIQ-1, the 'contamination' it created was within acceptable limits.

The starting point for the PhD series of papers

However, starting with paper 1 which was published in 2006, the papers in the thesis have drawn together findings from the self-report imagery questionnaire and social desirable responding literatures to question McKelvie's (1995b) conclusion. They argued that his research can be criticised firstly for treating socially desirable responding as a unitary phenomenon, and secondly for its predominant reliance on the MC scale. Past research, particularly that by Paulhus (2002), has provided strong evidence that social desirable responding can occur in at least two different forms, the egoistic bias and the moralistic bias. Also, research has suggested that the MC scale items primarily tap the moralistic bias (Paulhus, 1984). This is problematical as participants answering self-report imagery questionnaires will most likely view imagery as a valued ability that is being tested. As such the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, is more likely to be a threat to the validity of self-report imagery questionnaires.

Prior to paper 1 the results of two studies which included measures of egoistic and moralistic social desirable responding, McLemore (1976) and Anderson and Buyer (1994), offered some support for this critique. Both found significant correlations between self-report imagery questionnaires and social desirable responding questionnaires that load on the egoistic bias, but no correlations with questionnaires that load on the moralistic bias. However, both studies had limitations. Neither paper discussed their results within the framework of the egoistic–moralistic distinction, and the egoistic social desirable responding measures they used predated the BIDR and included items relating to poor adjustment and

psychopathology. Further, while McLemore (1976) found that greater vividness and control of imagery were associated with greater socially desirable responding, Anderson and Buyer (1994) found the opposite. Therefore further research using the BIDR scales appeared necessary.

This critique of McKelvie's (1995b) conclusion and the nature of the empirical findings presented by McLemore and Anderson and Buyer lead to the ideas being tested in paper 1 using the VVIQ-1 and the VMIQ imagery questionnaires. Following this the subsequent papers making up the thesis systematically explored further aspects of the relationship between scores on self-report imagery questionnaires and measures, such as whether relationships seen with the egoistic bias would generalize to ratings of other imagery properties and other sensory modalities.

The organization of the critical review

The critical review which follows is split into two sections. The first section presents a micro-level analysis of the five papers presented one at a time. The ordering of the papers follows publication year except for the last paper. This is so that papers 1-4 focus primarily on stylistic aspects of social desirable responding, while paper 5 focuses on substance. The coverage of each paper includes an introduction followed by a discussion of its strengths and limitations. The second section presents a macro-level analysis of the research which summarises the findings as a whole and then discusses methodological considerations, contributions made by the work, publication outcomes, and suggestions for future research.

Critical review

Micro analysis of the five research papers

Paper 1: Allbutt, Shafiullah and Ling (2006)

Reference: Allbutt, J., Shafiullah, M. & Ling, J. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, *30,* 1&2, 1-20.

Aims of the paper

Paper 1 used the original version of the Vividness of Visual Imagery Questionnaire (VVIQ-1) and the original version of the Vividness of movement Questionnaire (VMIQ), along with the most well established of the BIDR scales, SDE and IM. The VVIQ-1 measures predominantly static visual imagery and the VMIQ measures movement visual imagery.

The paper had four aims following from an overarching aim to clarify the nature of social desirable responding in self-reported imagery ability. The primary aim of the paper was to investigate whether scores on self-report imagery questionnaires measuring vividness of visual imagery and movement visual imagery, correlated with the egoistic bias and the moralistic bias as measured by the BIDR scales. Then, if they did, to establish which relationship was stronger, whether correlations with SDE and IM were independent or whether they shared variance, and to discover whether the sizes of the relationships were within the limits of acceptable 'contamination' suggested by McKelvie (1994).

The second aim was, if the VVIQ and the VMIQ were found to correlate together, to conduct a partial correlation analysis to test the extent to which this correlation was independent of socially desirable responding.

The third aim was to investigate the size of the correlations between the individual items of the imagery scales and the socially desirable responding measures. This was done to test for

items which might be deemed to be a threat to the content validity of the imagery questionnaires.

The fourth aim was to test whether the correlations between self-report imagery questionnaires and the BIDR were affected by the counterbalancing of questionnaire order. In Study 1, the order of presentation of the questionnaires was held constant at VVIQ, VMIQ and BIDR for all participants while in Study 2 it was counterbalanced across participants.

Note the Journal of Mental Imagery is no longer publishing. As such, no copywrite information is given here.

The Relationship Between Self-Report Imagery Questionnaire Scores and Subtypes of Socially Desirable Responding: Visual and Movement Imagery

JOHN ALLBUTT
University of Salford

MOHAMMED SHAFIULLAH
Leeds Metropolitan University

JONATHAN LING
University of Teesside

The relationship between scores on self-report imagery questionnaires and subtypes of socially desirable responding was investigated across two studies. The self-report imagery questionnaires used were the Vividness of Visual Imagery Questionnaire (Marks, 1973) and the Vividness of Movement Imagery Questionnaire (Isaac, Marks & Russell, 1986). The measure of socially desirable responding was the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1994, 2002). The BIDR is made up of two main subscales: the self-deceptive enhancement (SDE) scale, which is a measure of egoistic bias (the tendency to claim positive social and intellectual qualities), and the IM scale, a measure of moralistic bias (the claiming of positive moral qualities and the refutation of negative socially-deviant qualities). Samples of 98 and 109 undergraduate psychology students were tested under anonymous conditions. Results indicated that SDE correlated consistently and at a higher level with the self-report imagery questionnaires than IM. SDE correlations ranged from –0.23 to –0.33 with greater image vividness being associated with greater SDE. The research shows the importance of treating socially desirable responding as a multifactorial phenomenon, and that previous studies may have underestimated the degree of relationship between imagery scales and socially desirable responding by using inferior measures of egoistic bias.

Introduction

The reliability and validity of self-report imagery questionnaires have been the focus of much research and debate (McKelvie, 1995a, 1995b;

White, Sheehan, & Ashton, 1977). A concern in their use has been the extent to which participants' answers are influenced by socially desirable responding. Paulhus (2002) defines socially desirable responding as "the tendency to give overly positive self descriptions" (p. 50).

Imagery researchers concerned with social desirability have taken one of two approaches. The first approach is to manipulate the experimenter demands of a situation and observe their effect on imagery questionnaire responses. For example, Ashton and White (1975) found that manipulating the desirability of vivid mental imagery influenced participants' responses on Sheehan's Questionnaire Upon Mental Imagery (Sheehan's QMI; Sheehan, 1967).

The second approach is to correlate responses on self-report imagery questionnaires to psychometric measures of socially desirable responding. The Marlowe-Crowne social desirability scale (M-C; Crowne & Marlowe, 1960) has been used in most studies. The scale was developed by Marlowe and Crowne in the 1950s as a measure of socially desirable responding in self-reports using items that were not associated with poor adjustment or psychopathology and so was suitable for use with non-pathological samples. Over time, Marlowe and Crowne's conceptualization of what their scale was measuring evolved to a more general motive of "need for approval" (Crowne & Marlowe, 1964) and then an "avoidance of disapproval" (Crowne, 1979).

White, Sheehan and Ashton (1977) reviewed the studies that had investigated the relationship between two imagery self-report scales, Sheehan's QMI and the Gordon Controllability of Visual Imagery questionnaire (CVI; Gordon, 1950), and the M-C scale. They found evidence to be mixed, sometimes supporting an association with the M-C and other times not. Where significant relationships were seen, they were typically small in size and greater mental imagery ability was associated with greater socially desirable responding.

Review articles of the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) by McKelvie (1995a, 1995b), provide meta-analytical information about its relationship with the M-C scale. McKelvie (1994) set the following criteria for the evaluation of levels of contamination from socially desirable responding: Nonsignificant correlation (no contamination), a correlation of 0.15 or less (inconsequential contamination), a correlation of 0.16 to 0.25 (some contamination, but acceptable) and a correlation of above 0.25 (unacceptable contamination). McKelvie (1995b) made two estimates of the degree of relationship between the VVIQ and socially desirable responding. The first estimate was based on correlations with the M-C scale only. Data from 10 studies lead to a reported r of 0.189, with

more vivid imagery being associated with greater socially desirable responding. The second estimate added a further six correlations to McKelvie's data pool from three other tests of socially desirable responding used in a study by Anderson and Buyer (1994). Inclusion this extra data produced a revised r of 0.075, indicating a much weaker effect.

Taken together, the evidence from the two approaches tends to suggest that there is an effect, but that contamination is within acceptable limits. However, this research can be criticized, first, for treating socially desirable responding as a unitary phenomenon, and second, for its predominant reliance on the M-C scale.

Several researchers (e.g. Edwards, Diers & Walker, 1962; Wiggins, 1959), prompted by observations of low intercorrelations between different measures, have suggested that socially desirable responding may occur in more than one form. Probably, the most influential theorist into the different types of socially desirable responding has been Delroy Paulhus. In his earliest work, Paulhus (1984), suggested that there are two forms of socially desirable responding, distinguished by level of conscious awareness: "self-deception" (SD), reflecting an honest, but overly positive self-presentation, and "impression management" (IM), reflecting self-presentation tailored to an audience. But in his most recent work (Paulhus, 2002), he has shifted his position to propose a two-tier model of socially desirable responding which has a primary distinction in terms of "personality content" (Paulhus, 2002, p. 63), and a secondary distinction in terms of level of consciousness (see Figure 1).

Figure 1. Paulhus's Two-Tier Model (modified from Paulhus, 2002).

In the first tier there is a division of socially desirable responding into egoistic and moralistic biases. The egoistic bias refers to the tendency to claim positive social and intellectual qualities, while the moralistic bias refers to the claiming of positive moral qualities and the refutation of negative socially-deviant qualities. In the second tier, both the egoistic and moralistic biases are split into conscious and unconscious forms. The egoistic bias becomes self-deceptive enhancement (SDE) in its unconscious form and agency management (AM) in its conscious form. The moralistic bias becomes Self-deceptive denial (SDD) in its unconscious form and communion management (CM) in its conscious form. One implication of Paulhus's model is that different kinds of socially desirable responding may have independent or even additive effects on self-report imagery questionnaires. Therefore, estimates of effect sizes should be made separately for different socially desirable responding scales or at least different kinds of socially desirable responding.

Since its construction, a number of weaknesses have been identified in the M-C scale. Critics suggest that some of the items on the test may now be out of date (Stöber, 2001) and that the scale confounds the type of item, attribution of positive attributes and denial of negative attributes, with direction of item scoring (Helmes, 2000). Data from Paulhus (1984) shows a loading of 0.68 on a factor labelled "IM" and 0.40 on a factor labelled "SD." A reinterpretation of the meaning of these factors in terms of Paulhus's two-tier model would suggest that the M-C scale is a moderate to strong measure of moralistic bias, but only a moderate to weak measure of egoistic bias. This is of concern because the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, would seem more of a threat to the validity of self-report imagery questionnaires.

Only two studies, McLemore (1976) and Anderson and Buyer (1994), have investigated self-report imagery questionnaires using social desirability questionnaires which load more highly on the egoistic bias than the M-C scale does. McLemore (1976) administered the Edward Social Desirability Scale (Edwards, 1957), the CVI questionnaire and parts of Sheehan's QMI along with a series of other psychometric tests to participants. Edward's scale appears to be primarily a measure of egoistic bias as it is reported to load highly on the SD form of socially desirable responding (Paulhus, 1991). McLemore observed the following correlations with Edward's scale: +0.08 (QMI visual), +0.24 (QMI proprioceptive), +0.30 (QMI kinesthetic), and +0.44 (CVI), with the last three reaching statistical significance. In all cases, greater vividness of imagery and greater control of imagery were associated with greater scores on the Edward's scale (i.e., greater socially desirable responding). Anderson and Buyer (1994) admin-

istered a battery of nine imagery and social desirability tests to participants. The five imagery tests were: Sheehan's QMI, the Betts QMI (Betts, 1909), Anderson and Buyer's revised Betts QMI, the CVI questionnaire and the VVIQ. The four social desirability tests were: the M-C scale, Minnesota Multiphasic Personality Inventory (MMPI) K-Scale (Meehl & Hathaway, 1946), the Jacobson-Kellogg Social Desirability Scale (Jacobson, Kellogg, Cauce & Slavin, 1977) and the Edwards (1957) scale. From the point of view of the two-tier model, both egoistic and moralistic biases appear to have been assessed by this study, as Paulhus (1991) describes the K-Scale and Edwards Scale as loading primarily on SD while we have seen previously that the M-C loads more highly on IM. No corresponding factor loading information is available for the Jacobson-Kellogg Social Desirability (SD) Scale. However, the authors constructed the scale to be a multidimensional measure of the "need for approval" and as such it should overlap with the M-C scale. This interpretation is supported by the pattern of correlations between the four social desirability scales. Anderson and Buyer found the Edward's scale and the MMPI K-Scale to correlate at +0.68 and the M-C scale and the Jacobson-Kellogg scale to correlate at +0.56. No other correlations between socially desirable responding scales were significant. They went on to correlate the four social desirability scales with the self-report imagery questionnaires and found only two of the 20 correlations between them were statistically significant. These were a correlation of −0.32 between the Edward's scale and the VVIQ eyes closed, and a correlation of −0.23 between the MMPI K-Scale and the VVIQ eyes closed. Both correlations involved social desirability questionnaires, which appear to load primarily into the egoistic bias but were in the direction of *less* vivid imagery being associated with greater socially desirable responding.

The results of these two studies are important as they suggest that the relationship between self-report imagery questionnaires and socially desirable responding measures may be greatest for measures of egoistic bias. However, it is unclear why the direction of this relationship differed between the two studies. Anderson and Buyer do not discuss possible reasons for the counterintuitive pattern they observed or why their results differed from those of McLemore's study. One possibility is that Anderson and Buyer's results are simply the result of chance effects. This is a plausible explanation given that the correlations were small in size and only two of 20 achieved statistical significance. But whatever the reason for the difference, the findings of the two studies need to be treated with caution. This is because both the measures of the egoistic bias used, the K-Scale and the Edwards 1957 scale, were constructed some time ago and they include

items relating to poor adjustment and psychopathology. As such, they may confound psychopathology with socially desirable responding and so are not ideally suited for use with non-pathological student samples.

The current research is composed of two studies which correlated scores on two self-report imagery questionnaires: the Vividness of Visual Imagery Questionnaire (Marks, 1973) and the Vividness of Movement Imagery Questionnaire (VMIQ; Isaac, Marks, & Russell, 1986), with the SDE and IM subscales of the BIDR. The VVIQ measures vividness of visual imagery, while the VMIQ measures vividness of movement imagery. The choice of these two imagery questionnaires was based on the fact that vividness has been the property of mental imagery most often tested for in self-report imagery questionnaires, and the VVIQ has probably been the most often used of such questionnaires. The VVIQ and VMIQ typically show moderate to high correlations with each other: +0.60 (Eton, Gliner & Munz, 1998), +0.78 (Unpublished study cited in Hall & Martin, 1997), +0.35 (Isaac & Marks, 1994) and +0.81, +0.75, +0.45 and +0.65 (Isaac, Marks, & Russell, 1986) and as such appear to be assessing two related but different imagery domains. Making use of the BIDR as the measure of socially desirable responding allowed us to overcome some of the problems of previous studies. The BIDR acknowledges that socially desirable responding is a multidimensional phenomenon. No items on the BIDR are associated with poor adjustment or psychopathology and so the BIDR is suitable for use with non-pathological samples. The BIDR scales load highly on the respective forms of socially desirable responding and so offer the capacity for assessing them separately. Also, each scale is made up of 20 items, with equal numbers of positively and negatively keyed items, which militates against agreement and acceptance response sets. In order to keep the demands placed on participants to a manageable level, we used only two of the four BIDR scales. The SDE and IM scales were selected for use because they were the most well established of the four scales and their use allowed a comparison of the relationship between self-report imagery questionnaires and both egoistic and moralistic forms of socially desirable responding.

The main aim of the current research was to investigate whether self-report imagery questionnaires correlate with the egoistic bias and the moralistic bias as measured by SDE and IM. If they do, we wished to establish which relationship is stronger, whether correlations with SDE and IM are independent or whether they share variance, and to discover whether the sizes of the relationships are within the limits of acceptable contamination suggested by McKelvie (1995a). Following from this, if the VVIQ and the VMIQ are found to correlate, we wished to test the extent to which

this correlation was independent of socially desirable responding. Given that responses on the VVIQ and the VMIQ could be influenced by socially desirable responding, we asked whether it is possible that the shared variance from correlations with this variable could inflate the size of the correlation between the two imagery scales.

Also, we wanted to investigate the nature of the correlations between the individual items of the imagery scales and socially desirable responding. King and Bruner (2000) argue for the need to correlate scores on questionnaire items with measures of socially desirable responding during the development phase of questionnaires to detect items that are significantly correlated with socially desirable responding measures. Such items are a threat to the content validity of questionnaires. To our knowledge, this kind of analysis has not been carried out with either the items of the VVIQ or the VMIQ. Helmes (2000) suggests comparing the size of the correlation between the items and the measure of socially desirable responding, with the size of the correlation between the items and the total score of the questionnaire. If the socially desirable responding correlation is larger for an item, then it should be removed or revised.

Finally, we wished to test whether the correlations between self-report imagery questionnaires and the BIDR are affected by the counterbalancing of questionnaire order. In Study 1, the order of presentation of the three questionnaires was held constant at VVIQ, VMIQ and BIDR for all participants while in Study 2 it was counterbalanced across participants. Normal good practice in psychological work is to counterbalance the order of conditions in within-subjects designs to attempt to militate against order effects. However, having some participants answer the BIDR as one of their first questionnaires might make guessing the purpose of the study easier for them and so increase the chance of demand characteristics affecting their results. If this was true, then correlations between the self-report imagery questionnaires and measures of socially desirable responding might be less in Study 2. A comparison of the results of Study 1 and Study 2 would allow an assessment of the necessity to use counterbalancing in this kind of study. If counterbalancing had no effect on the pattern of results seen, then the data from the two studies could be combined to produce a larger data set on which to carry out the other analyses.

These issues can be summarized in the form of four research questions:

Research question 1: Are the correlations between self-report imagery questionnaires and the BIDR affected by the counterbalancing of order of questionnaires?

Research question 2: Do the self-report imagery questionnaires correlate

with the egoistic bias and the moralistic bias as measured by SDE and IM?

Research question 3: If the VVIQ and the VMIQ are found to correlate by the present study, to what extent is this correlation independent of socially desirable responding?

Research question 4: What is the nature of the correlations between the individual items of the imagery scales and socially desirable responding?

Method

Participants

In Study 1, 97 undergraduate psychology students took part (16 males, 81 females). Ages ranged from 17 to 52 years with a mean age of 23.8 years. In Study 2, 109 undergraduate psychology students took part (28 males, 81 females). Ages ranged from 18 to 55 years with a mean age of 26.8 years. Combining the data from the two studies produced a sample of 206 participants (44 males, 162 females) with ages ranging from 17 to 55, and a mean age of 25.4 years. In both studies participants were volunteers and no inducements were paid to them.

Materials

Materials were identical for Studies 1 and 2.

Ethical rights form. A form was used that informed participants of the rights they possessed under the British Psychological Society's code of ethics.

Vividness of Visual Imagery Scale (VVIQ; Marks, 1973). The VVIQ aims to assesses vividness of visual imagery using 16 items. Participants form a series of images of a friend or relative's face, the rising sun, a shop, and a country scene. Ratings of image vividness are made on a 5-point scale ranging from 1 ("Perfectly clear and vivid as normal vision") to 5 ("No image at all, you only 'know' you are thinking of the object"). Traditionally, participants make 16 ratings with their eyes open, then 16 ratings with their eyes closed and the two sets of scores are added together. However, McKelvie's (1995a) meta-analytical review found no difference in the ratings made with eyes open and eyes closed. Thus in the two studies, mention of imaging with eyes open was edited out of the VVIQ instructions and the participants were only asked to complete the questionnaire once with their eyes closed. The range of possible scores on the VVIQ with this procedure is 16 to 80, with lower scores indicating more vivid imagery.

Vividness of Movement Imagery Questionnaire (VMIQ; Isaac, Marks & Russell, 1986). The VMIQ aims to assesses vividness of movement imagery using 48 items. The questionnaire has two parts. First, respondents rate the vividness with which they can image movement in someone else carrying

out a series of 24 activities such as basic body movements and movements controlling an object. Second, respondents rate the vividness with which they can image kinesthetic sensations within themselves across the same 24 activities. Ratings of image vividness are made on the same 5-point scale as on the VVIQ. Each of the 48 images are made once with the eyes closed. The range of possible scores on the VMIQ with this procedure is 48 to 240, with lower scores indicating more vivid imagery.

The first 24 items dealing with ratings of movement in others and the second 24 items dealing with kinesthetic sensations in participants themselves were not treated as formal subscales by Isaac et al. (1986). However, they differ in the type of image being formed (i.e., visual vs. sensation of movement), and in the focus of the image (i.e., other vs. self). Thus it is possible that they are tapping into different imagery processes. Because of this we decided to treat these two sets of scores as subscales in addition to calculating a total score. We will refer to these "subscales" as "VMIQ Other" and "VMIQ Self" and total scores as "VMIQ Tot". Data from Studies 1 and 2 reported here showed the Cronbach's Alpha values of the VMIQ Other scale to be 0.96 and 0.95, and of the VMIQ Self scale to be 0.96 and 096. Thus both scales show sufficient internal consistency of items to be treated as subscales.

The Balanced Inventory of Desirable Responding Version 8 (BIDR-8; Paulhus, 1994, 2002). The BIDR aims to assess socially desirable responding. Use was made of the two main 20 item subscales, SDE and IM. SDE is a measure of unconscious egoistic bias. The subscale has statements about the respondent's abilities across a range of situations, half of which are positively phrased and half negatively phrased. For example "I am a completely rational person" (positive). Although a 5-point scale has sometimes been used, responses are more usually made on a 7-point scale ranging from 1 ("Not true") to 7 ("Very true"), so this was the scale we used. Scores on the negative items are reversed during scoring. IM is a measure of conscious moralistic bias and this subscale is organized in a similar fashion to that of the SDE. It has statements about the respondent's ethical behavior across a range of situations, half of which are positively phrased and half negatively phrased. For example "I never swear" (positive). Responses are made on the same 7-point scale as the SDE and responses to negative are items reversed.

There are two ways to score the subscales. The first method is to recode responses to negative items and them simply sum scores on the 7-point scale across items. The second method is to score only extreme scores. With this method, negative items are recoded and then responses of either 6 or 7 on an item are counted as 1 point, while responses of 1 to 5 do not

score. The points are totalled for each subscale separately. The range of possible scores with this procedure is 0 to 20, with lower scores indicating less SDE or IM, and higher scores indicating more SDE or IM. Although there is debate about which scoring method is most appropriate (see e.g. Stöber, Dette and Musch, 2002), the current study used the second scoring method. This approach was taken as Paulhus (1991, 1994) suggests that the second method has the advantage of only scoring clearly exaggerated or managed responses.

Whichever scoring method is used it is also possible to combine scores from the two subscales together to produce an over all BIDR score. This, however, was not done here as a measure of global socially desirable responding was not relevant to any of our research questions.

Additional instructions were added to the start of the BIDR in order to make two items from the SDE scale and one item from the IM scale relevant to all participants. The items from the SDE scale were: item 8, "I am not a safe driver when I exceed the speed limit," and item 14, "My parents are not always fair when they punish me." The item from the IM scale was: item 33, "I sometimes drive faster than the sped limit". Items 8 and 33 are problematic for participants who do not drive, so they were asked to answer the items by imagining what they would be like if they were to drive. Item 14 is problematic for older participants and so participants were asked to remember back to when they did live at home and comment on how they felt at the time. This approach was preferred to the omission of these items as it preserved the balanced nature of the BIDR scales.

The ethics form and three questionnaires were combined into a booklet. The booklet started with the ethics form, followed by the VVIQ, VMIQ and ending with the BIDR. In the BIDR the SDE scale appears before the IM scale. The VVIQ was titled "Questionnaire 1: Vividness of Visual Imagery Questionnaire"; the VMIQ was titled "Questionnaire 2: Vividness of Movement Imagery Questionnaire"; and the BIDR was titled only "Questionnaire 3." The title of the BIDR was not used to reduce the likelihood of participants guessing the purpose of the study.

Procedure

In Study 1, participants answered the three questionnaires in the last hour of a 2-hour laboratory class. They were asked to seat themselves as far apart as possible in the classrooms. Responses were made in silence and participants were told not to put their name anywhere in the booklet. Participants took between 30 and 40 minutes to complete the questionnaires. When finished, participants got up quietly and left the room. The

following week they were given a debriefing sheet which explained the purpose of the study and gave the contact details of the experimenters in case of queries.

In Study 2 the same procedure was used as in Study 1 except that the order in which the participants completed the three questionnaires was counterbalanced. Within each laboratory class participants were grouped into sixes and allocated to one of the six possible orders in which the questionnaires could be filled in. Any extra participants who were left over were randomly allocated to one of the six possible orders.

Results

The presentation of the results for Study 1 and Study 2 is organized around the series of four research questions listed at the end of the introduction.

The main statistical test used to analyze the data was the Pearson's Product Moment correlation test. A prospective power analysis using tables for Pearson's Product Moment correlation showed that for a two tailed test a sample size of 90 would be required to detect a medium effect size (Clark-Carter, 1997). The sample sizes from Studies 1 and 2 both exceeded 90 after data screening and thus were sufficient for this purpose. Also, in order to be open to any possible form the results might take, non-directional hypotheses were used throughout.

Q1: To what extent are correlations between self-report imagery questionnaires and the BIDR affected by the counterbalancing of order of questionnaires?

A series of Pearson Product Moment correlations between the VVIQ, VMIQ (Other scale, Self scale and Total), and BIDR (SDE and IM subscales) were carried out to investigate the correlations between the imagery scales and the socially desirable responding scales. Prior to the correlations being conducted, the variables were screened for univariate and bivariate outliers. Univariate outliers were identified by being three or more standard deviations from the mean. Normality of distribution was screened for using Z-ratios for skew and kurtosis at the 5% significance level. Multivariate outliers were identified by having standardized regression residuals of three or greater. Data screening reduced the Study 1 sample size to 91 and the Study 2 sample size to 105. In both studies SDE and IM scales were positively skewed and square root transformations were applied to make their distributions normal.

The rationale behind the inclusion of Study 2 was to investigate if counterbalancing the order of presentation of the VVIQ, VMIQ and BIDR had

an impact on the correlations between these questionnaires. A comparison of Study 1 and Study 2 would allow an assessment of the necessity to use counterbalancing in this kind of study, and if counterbalancing had no effect on the pattern of results seen then the data from the two studies could be combined together.

The results of a Pearson's correlational analysis between the imagery and socially desirable responding scales for both studies, and the two studies combined can be seen in Table 1.

Table 1
Correlation Matrix Showing Pearson's Correlations for Study 1 and Study 2.

		VVIQ	VMIQ Other	VMIQ Self	VMIQ Tot	SDE
Study 1	VVIQ					
	VMIQ Other	+0.66**				
	VMIQ Self	+0.48**	+0.53**			
	VMIQ Tot	+0.64**	+0.86**	+0.89**		
	SDE	-0.24*	-0.25*	-0.46**	-0.41**	
	IM	-0.03	-0.11	-0.10	-0.12	+0.19
Study 2	VVIQ					
	VMIQ Other	+0.48**				
	VMIQ Self	+0.50**	+0.63**			
	VMIQ Tot	+0.54**	+0.88**	+0.92**		
	SDE	-0.27**	-0.21*	-0.20*	-0.23*	
	IM	-0.26**	-0.06	-0.10	-0.09	+0.57**
Study 1& 2 combined	VVIQ					
	VMIQ Other	+0.48**				
	VMIQ Self	+0.46**	+0.60**			
	VMIQ Tot	+0.52**	+0.87**	+0.91**		
	SDE	-0.25**	-0.23*	-0.33*	-0.31*	
	IM	-0.20**	-0.11	-0.15*	-0.14*	+0.42**

Note   * p<0.05 ** p<0.01   (2-tailed)

In order to compare the correlations from Studies 1 and 2, differences in sizes of corresponding correlations from the two studies were calculated and tested for significance using Fisher's transformation to Z-scores analysis. Of the15 comparisons made, only the correlations between VMIQ Self and SDE, and between SDE and IM showed a significant difference between Studies 1 and 2. The correlation between VMIQ Self and SDE was lower in Study 2 and consistent with the possibility that a randomized presentation order might be more open to the effects of demand characteristics. But it is not clear why the other self-report imagery scale correlations with the BIDR scales were not similarly affected. Further, it is unclear why the correlation between SDE and IM would be significantly different given that their order was the same across the two studies as the

SDE and IM scales are always given in the same order in the BIDR. Two significant results from 15 is slightly more than would be expected by chance alone. However, given that the differences across the two studies either do not appear to be part of an overall consistent trend, or readily explicable, it would seem appropriate to collapse the data from the two studies together.

The new combined data set was screened for univariate and bivariate outliers reducing the sample size from 207 to 193. Again SDE and IM scales were found to be positively skewed and so square root transformations were applied to make their distributions normal. Because of its superior size, from this point onwards only the data from the combined data set will be used in analyses.

Q2: Do the imagery scales correlate with SD and IM?

With respect to socially desirable responding, a notable trend in the results was for the correlations between SDE and the imagery scales to be larger in size than those between IM and the imagery scales. Significant low to moderate sized correlations were observed between the imagery scales and SDE, with more vivid imagery being associated with greater SDE. The strongest of these associations was with the VMIQ self score. All of the correlations between IM and the imagery scales were negative, and most reached significance, but all were low in size. Thus, the imagery scales appear to correlate primarily with SDE rather than IM.

The correlations between the VVIQ and the VMIQ and its subscales were all moderate in size and positive, and so comparable to the kinds of correlations seen in previous studies. The correlations between VMIQ Other and VMIQ Self was positive and moderate in size suggesting a degree of association, but not so great as to undermine the decision to treat them as separate subscales. The correlation between SDE and IM was moderate in size and positive.

To explore whether the correlations between the self-report imagery scales and the two forms of socially desirable responding were independent a series of hierarchical multiple regressions were carried out for each scale. SDE and IM were treated as predictor variables and self-report imagery scales as criterion variables. SDE was entered into the regression equation on step 1, and then IM added on step 2 to test whether the addition of IM added significantly to the ability to predict self-report imagery scale scores. For the VVIQ regression R^2 Change with the addition of IM was 0.01, and for all three VMIQ regressions R^2 Change with the addition of IM was 0. In all cases these statistics were not significant, indicating that IM did not add significantly to the ability of SDE to predict imagery scores.

This suggests that the correlations between the self-report imagery scales and IM that were significant are due to shared variance between IM and SDE rather than variance unique to IM.

Q3: Are intercorrelations between imagery scales independent of socially desirable responding?

In order to assess whether the intercorrelations between the imagery scales were independent of socially desirable responding, they were correlated together with SDE and IM partialled out. Partialling reduced the correlation coefficients by between 0.01 and 0.06. A Fisher's transformation to Z-scores analysis comparing original correlation coefficients with those after partialling showed that none of these reductions were significant. Thus, the correlations between the imagery scales appear to be independent of their correlations with SDE and IM.

Q4: What is the nature of the correlations between the items of the imagery scales and socially desirable responding?

Following from Helmes (2000), scores on the items of the VVIQ and VMIQ were correlated with SDE and also total scores on the relevant self-report imagery questionnaire. If the socially desirable responding correlation for an item is larger than the correlation with the total score, then the item should be removed from the questionnaire or the item revised. VMIQ Self and Other subscales were analyzed separately, and SDE was taken as the measure of socially desirable responding as it had shown the strongest relationship with the self-report imagery scale scores in the previous analyses.

For the VVIQ, the size of the 16-item SDE correlations ranged from 0.02 to -0.28 with a mean size of –0.13. Nine of these correlations were significant (items 2, 3, 5, 6, 11, 13, 14, 15 and 16). The size of the 16-item VVIQ total score correlations ranged from 0.38 to 0.64 with a mean size of 0.55. All were significant but for no item did the correlation with SDE exceed the correlation with the VVIQ total score.

For the VMIQ Other subscale, the size of the 24-item SDE correlations ranged from 0.12 to -0.23 with a mean size of –0.15. Fourteen of these correlations were significant (items 2, 4, 5, 6, 7, 8, 13, 14, 15, 17, 18, 19, 20 and 23). The size of the 24-item VMIQ Other total score correlations ranged from 0.56 to 0.76 with a mean size of 0.66. All were significant but for no item did the correlation with SDE exceed the correlation with the VMIQ Other total score.

For the VMIQ Self subscale, the size of the 24-item SDE correlations

ranged from -0.14 to -0.29 with a mean size of –0.22. Of these correlations, 23 were significant (items 1-5 and items 7-24). The size of the 24-item VMIQ Self total score correlations ranged from 0.62 to 0.78 with a mean size of 0.69. All were sign ficant but for no item did the correlation with SDE exceed the correlation with the VMIQ Other total score.

Based on these analyses no items should be removed from the VVIQ or VMIQ.

Discussion

The main issue our research sought to address was whether the self-report imagery questionnaires showed relationships with measures of egoistic and moralistic biases. The data shows a clear relationship with SDE for both the VVIQ and VMIQ which is stable across the two studies. Reports of more vivid imagery were consistently associated with greater SDE. In contrast, correlations between the imagery scales and IM were smaller, less consistent and in a regression analysis IM did not add to the ability of SDE to predict self-report imagery questionnaire scores. This pattern of results supports the argument that socially desirable responding should be treated as a multidimensional entity, and that the relevance of different forms of socially desirable responding may vary across situations. Interpreted within the framework of the two-tier model (Paulhus, 2002), the pattern of results is consistent with the position that the greater threat to responses on self-report imagery questionnaires from socially desirable responding is from an egoistic bias rather than moralistic bias. This interpretation seems plausible given that participants answering a self-report imagery questionnaire may well view imagery as a skill that is being tested and all the significant correlations between imagery measures and SDE were positive in direction. However, given that only two of the four suggested types of socially desirable responding were assessed, and the SDE and IM scales are also thought to differ in terms of unconscious verses conscious processes, a second possibility is that level of consciousness is the key variable.

The pattern of results obtained in our studies are in line with the findings of McLemore's (1976) study, but not Anderson and Buyer (1994) who found mostly non-significant correlations and two correlations where more vivid imagery was associated with lesser SDE. The reason for this difference is not clear, although it is possible that Anderson and Buyer's two correlations may simply be chance effects.

A concern expressed in the introduction over previous studies was that the M-C scale of socially desirable responding, while tapping the egoistic bias to a degree, appears to be more of a measure of moralistic bias. Given

that the egoistic bias appears to be a greater threat to the validity of answers on self-report imagery questionnaires, use of the M-C scale might lead to an underestimate of the impact of socially desirable responding on these tests. The combined data from the two studies produced correlations between SDE and the imagery scales of: VVIQ, -0.25; VMIQ Other, -0.23; VMIQ Self, -0.33; and VMIQ Total, -0.31. These correlations are small in size, but bigger than the two estimates made by McKelvie (1995b) of 0.189 (based on the M-C scale only) and 0.075 (including correlations from other social desirability scales). For two of the self-report imagery scales that were tested, the estimate of the influence of socially desirable responding based on correlations with SDE exceeded the 0.25 level for acceptable contamination set by McKelvie (1994) and one matches it. McKelvie's suggested cut-off point was based on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by previous authors. As such it is a useful contribution in indicating generally where a cut-off point might lie. However, supplying an answer to the question "How much 'contamination' by socially desirable responding can we accept on self-report imagery questionnaires?" is not a straightforward task. In addition, rejection of self-report imagery questionnaires as unacceptably contaminated by socially desirable responding would seem premature given the support for their validity (e.g., McKelvie, 1995b). The results do suggest that the BIDR is a more appropriate measure in this context than the M-C, which appears to have correlated with self-report imagery questionnaires in the past because some of its items tap into the egoistic bias.

Investigators who view socially desirable responding as a personality characteristic or response set are divided in their opinion as to whether socially desirable responding should be treated as error variance. Our position on this issue is similar to that of Helmes (2000). We are agnostic as to whether socially desirable responding is best viewed as a personality characteristic, response set or mixture of both, but believe that its control would appear in most cases to aid more precise measurement of the construct of interest. Indeed in the current context, socially desirable responding would appear to be conceptually distinct from mental imagery, and as such its control should lead to more precise measurement. Possible intervention strategies include explicitly asking participants to respond honestly, reminding participants that mental imagery is generally thought to be less vivid than its corresponding sensory experience, rejecting data from participants whose scores pass a predetermined value on relevant social desirability scales, or using statistical techniques such as partial correlation to

statistically control for the influence of socially desirable responding. Direct appeals to participants to respond honestly and 'realistically' are likely to have most impact on conscious and deliberate forms of socially desirable responding. Thus in principle, AM forms of the egoistic bias should be more responsive to such appeals than SDE.

However, Rinaldo and Okada (1993) have suggested that Ahsen's work on the unvividness paradox (e.g., Ahsen, 1985, 1988, 1990; Hochman, 1994) shows the existence of nonconscious self-inhibitory forces which act to reduce the vividness of images and appear to form part of the imagery system. Ahsen (1990) asked participants to keep either their mother or father in mind while answering the items of the VVIQ and found that reported imagery vividness was less when fathers were kept in mind. Rinaldo and Okada suggest that this phenomenon might be viewed as a form of self-deception as participants may be erroneously making estimates of low vividness, presumably because of the influence of emotionally charged associations with their fathers. If these self-inhibitory processes form part of the normal imagery system, then the situation as to whether SDE is to be viewed as conceptually distinct from the imagery system is less clear.

For the remaining research questions we found the following. First, a comparison of the correlations from the two studies suggests that in general they did not differ. Only 2 correlations differed significantly out of 15. The correlation between VMIQ Self and SDE was lower in Study 2 and consistent with the possibility that a randomized presentation order might be more open to the effects of demand characteristics. However, it is not clear why the other self-report imagery scale correlations with the BIDR scales were not similarly affected and so this result does not appear to be part of an overall consistent trend. Thus, it would seem that counterbalancing the order the questionnaires is not a crucial factor influencing the pattern of the results obtained. Second, the partial correlation analysis showed that the correlations between imagery scales were independent of socially desirable responding. Finally, the analysis of self-report imagery questionnaire items showed that although a number of items from the self-report imagery questionnaires showed statistically significant correlations with SDE, the size of these correlations was never greater than between the item and the total score. Thus, there were no grounds on which to remove any items from these questionnaires.

Further work is needed to establish whether the pattern of results obtained extends to other kinds of self-report imagery questionnaires, for example, those measuring imagery in different modalities and those measuring style of thinking such as Paivio's (1971) Individual Differences Questionnaire. Also, given that the strongest relationship observed in the

current study was for egoistic bias, it is important that future work should assess both SDE and AM at the same time. D. Paulhus (personal communication, October 23rd, 2003) has informed us that in one data set the correlation between SDE and AM in an honest response condition was 0.39 and in a fake good condition was 0.56. Given these low/moderate sized correlations, it is possible that AM might correlate with self-report imagery questionnaires independently from SDE and that the influence of socially desirable responding could be greater than the estimate from this study which is based on SDE alone.

References

Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery, 9*(3), 1-18.

Ahsen, A. (1988). Imagery, unvividness paradox and the paradigm of control. *Journal of Mental Imagery, 12*(3&4), 1-44.

Ahsen, A. (1990). AA-VVIQ and imagery paradigm: Vividness and unvividness issue in VVIQ research programmes. *Journal of Mental Imagery, 14*(3&4), 1-58.

Anderson, D.C., & Buyer, L.S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *American Journal of Psychology, 107*(2), 207-222.

Ashton, R. & White, K. (1975). The effects of instructions on subjects' imagery questionnaire scores. *Social Behavior and Personality, 3*(1), 41-43.

Betts, G.H. (1909). *The distributions and functions of mental imagery.* New York: Teachers College, Columbia University.

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove: Psychology Press.

Crowne, D.P. (1979). *The experimental study of personality.* Hillsdale, NJ: Erlbaum.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24,* 349-354.

Crowne, D.P. & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Edwards, A.L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Edwards, A.L., Diers, C.J., & Walker, J.N. (1962). Response sets and factors loadings on sixty-one personality scales. *Journal of Applied Psychology, 46,* 220-225.

Eton, D. T., Gliner, F. H. & Munz, D. C. (1998). The measurement of imagery vividness: A test of the reliability and validity of the Vividness of Visual Imagery Questionnaire and the Vividness of Movement Questionnaire. *Journal or Mental Imagery, 22*(3&4) 125-136.

Gordon, R.A. (1950). An experiment correlating the nature of imagery with performance on a test of reversal perspective. *British Journal of Psychology, 41,* 63-67.

Hall, C.R. & Martin, K.A. (1997). Measuring movement imagery abilities: A revision of the Movement Imagery Questionnaire. *Journal of Mental Imagery, 21*(1&2), 143-154.

Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In Goffin, R. D. and Helmes, E. (Eds.), *Problems and solutions in human assessment: Honouring Douglas N. Jackson at seventy* (pp21-40). Boston: Kluwer.

Hochman, J. (1994). Ahsen's image psychology. *Journal of Mental Imagery, 18*(3&4) 1-118.

Isaac, A., & Marks, D.F (1994). Individual differences in mental imagery experience: Developmental changes and specialization. *British Journal of Psychology, 85,* 479-500.

Isaac, A., Marks, D.F., & Russell, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery, 10*(4), 23-30.

Jacobson, L.I., Kellogg, R.W., Cauce, A.M., & Slavin, R.S. (1977). A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society, 9,* 109-110.

King, M. F., & Brunner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing, 17*(2), 79-103.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17-24.

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*(3&4), 153-166.

McKelvie, S. J. (1994). Psychometric guidelines for imagery questionnaires. *Perceptual and Motor Skills, 79,* 1219-1231.

McKelvie, S. J. (1995a). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 1-106.

McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 197-251.

McLemore, C. W. (1976) Factorial validity of imagery measures. *Behavioural Research and Therapy, 14,* 399-408.

Meehl, P. E. & Hathaway, S. R. (1946). The K factor as a suppresser variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30,* 525-564.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp17-59). San Diego: Academic Press.

Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6.* Unpublished Manual, University of British Columbia, Vancouver, Canada.

Paulhus, D. L. (1999). BIDR Self-Deceptive Denial (SDD). neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm. Retrieved from the World Wide Web on 1 August 2003.

Paulhus, D. L. (2002). Socially Desirable Responding: The evolution of a construct. In Braun, H. I., Jackson, D. N. & Wiley, D. E. (Eds.), *The role of constructs in psychological and educational measurement* (pp49-69). Mahwah: Lawrence Erlbaum Associates.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Rinehart and Winston.

Rinaldo, R., & Okada, R. (1993). A test of the relationship between imagery vividness and social desirability across subject selection procedures. *Journal of Mental Imagery, 17*(3&4), 171-180.

Sheehan, P.W. (1967). A shortened version of Betts' Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology, 23,* 386-389.

Stöber, J. (2001). The Social Desirability Scale—17 (SDS-17): Convergent validity, discriminant validity and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222-232.

Stöber, J., Dette, D. E., & Musch, T. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78*(2), 370-389.

White, K.D., Ashton, R., & Law, H. (1978). The measurement of imagery vividness: Effects of format and order on the Betts' Questionnaire Upon Mental Imagery. *Canadian Journal of the Behavioural Sciences, 10*(1), 68-78.

White, K., Sheehan, P.W. and Ashton, R. (1977). Imagery assessment: A survey of self-report measures. *Journal of Mental Imagery, 1,* 145-170.

Wiggins, J.S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23,* 419-427.

Evaluation of the paper

Strengths

Paper 1 opened up new ways to conceptualise the area, helped make sense of past findings, advanced methodology and generated new findings.

Empirical contributions

Paper 1 made several important contributions, for example, it generated new knowledge in the area. Its primary aim was to investigate whether scores on self-report imagery questionnaires correlated with the egoistic bias and the moralistic bias when measured by scales that that were designed to be pure measures of each bias. The results showed that SDE correlated consistently and at a higher level with the self-report visual imagery than IM. This pattern was seen regardless of whether the imagery questionnaire measured visual or movement / kinesthetic imagery. SDE correlations ranged from $r=-0.23$ to -0.33 with greater image vividness being associated with greater SDE. The largest correlation involved the VMIQ self-ratings. Regression analyses showed that IM did not add to the ability of SDE to predict self-report imagery questionnaire scores. The implications of these findings were:

1) The MC scale is not the most appropriate measure of socially desirable responding to use in this situation;
2) Previous studies may have underestimated the degree of relationship between imagery scales and socially desirable responding.

Past research had produced inconsistent results. McLemore (1976) had found that greater vividness of imagery and greater control of imagery were both associated with higher levels of egoistic bias. Anderson and Buyer (1994) had also found a relationship with scales measuring the egoistic bias, but these were in the direction of *less* vivid imagery being associated with higher levels of egoistic bias. Our findings added support for the position that the relationship was positive in direction, as have our subsequent papers, and for the argument that the direction of Anderson and Buyer's two correlations were atypical and perhaps due to chance effects.

The partial correlation analysis which examined the relationship between the VVIQ-1 and the VMIQ scales reported that the correlation between them was not significantly affected by the partialling of the SDE or IM scores. Lönnqvist et al (2007) describe how if social desirable responding is truly distortion then it should have predictable statistical effects when its influence is controlled using partial correlation techniques. Where social desirable responding contributes variance to both variables in a correlation, it should operate as confounding common response variable, and removal of its influence should reduce the size of correlation. This situation might occur where both variables are self-report measures of constructs with desirability implications. In paper 1 we found that partialling out SDE reduced the correlations between VVIQ-1 and the VMIQ scales by between .01 and .06. It is notable that all six partial correlation analyses reduced the original correlations in size, however, none did so by significant amounts. This result could be interpreted to mean the relationship between the self-report imagery questionnaires was largely independent of social desirable responding. Equally, though it could be attributed to the lack of precision of measurement of social desirable responding questionnaires or the statistical techniques used in the analysis. For example, Watson et al (2006) found the MIRM approach produced notable reduction in sizes of correlations between self-efficacy and health behaviors in a simulated data set when the variance associated with social desirable responding was removed.

Theoretical contributions

Paper 1 advanced conceptual understanding by being the first paper to apply Paulhus's (2002) model and its view of social desirable responding as a multi-dimensional phenomena, to the topic of the relationship between self-report imagery scale questionnaires and social desirable responding. This created several original insights:

1) Application of the model to McKelvie's (1995b) meta-analytical estimation of the size of the relationship between the VVIQ-1 and social desirable responding, showed the limitations of his approach i.e. treating socially desirable responding as a unitary phenomenon, and its predominant reliance on the MC scale. Furthermore, because the MC scale appears to load primarily on the moralistic bias, McKelvie's estimate of the size of the relationship between the VVIQ-1 and social desirable responding could well be an underestimate.

2) It also allowed the pattern of correlations seen between self-report imagery scales and measures of social desirable responding in McLemore (1976) and Anderson and Buyer (1994) to be interpreted in terms of the egoistic bias.

Other conceptual advances included that the paper was the first to argue that participants answering self-report imagery questionnaires will most likely view imagery as a valued ability being tested. As such the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, is more likely to be a threat to the validity of self-report imagery questionnaires than the moralistic bias.

Paper 1 also reflected on the cut-off criterion for 'contamination' suggested by Mckelvie (1994). Working from a distortion view of social desirable responding, McKelvie proposed a set of criteria for judging the levels of contamination from socially desirable responding in self-report imagery questionnaires. These were based on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by other authors. He suggested a correlation of above r=.25 indicated 'unacceptable contamination' and the 'cut-off' point for acceptability. The estimate of the influence of socially desirable responding for the VMIQ based on correlations with SDE exceeded the .25 level for acceptable contamination set by McKelvie and for the VVIQ-1 matched it. However, in paper 1 we argued that, while being a useful contribution, McKelvie's criterion of r=.25 (6.25% variance) is an arbitrary cut off point. Supplying an answer to the question "How much 'contamination' by socially desirable responding can we accept on self-report imagery questionnaires?" is not a straightforward task. In addition, rejection of self-report imagery questionnaires as unacceptably contaminated by socially desirable responding would seem premature given the support for their validity. For example, McKelvie (1995b) reports that reliable relationships exist between the VVIQ-1 and performance on perceptual tasks such as scan path consistency during perception with an effect size of r=.44 and 95% CI [.31 to .56]. Thus, even though the VVIQ-1 appears to correlate with SDE, this does not stop it from correlating with performance on these tasks, though it may have reduced the effect sizes seen.

The paper discussed possible interpretations of the nature of the relationship between self-

report imagery scales and the egoistic bias. Past imagery papers had viewed correlations between imagery questionnaires and measures of social desirable responding as indicating contamination. Paper 1 was one of the first papers in the area of mental imagery to discuss the possibility of such relationships being substantive in nature and to offer a position on this issue.

The view of social desirable responding as simply reflecting distortion has been questioned. Increasing numbers of psychologists have argued that social desirable responding in some sense reflects substantive or valid responding. Several different substantive views of social desirable responding have been proposed: social desirable responding as a as departure from reality, but a meaningful personality trait in its own right; social desirable responding as a departure from reality, but an integral part of many different aspects of personality, and social desirable responding as honest responding. These were previously discussed in the literature review. In relation to imagery self-report questionnaires, the second and third perspectives are most relevant. The issue of whether social desirable responding forms a meaningful personality trait in its own right, or a set of traits, is not of major relevance to imagery work. Instead, key issues are whether it can be demonstrated that social desirable responding measures actually do index departures from reality, and that if this is possible, that such self-deceptive processes do not form play an integral and inseparable part of the normal functioning imagery system.

Previously, the literature review described how Paulhus (e.g. 1994) has provided a range of sources of support for the validity of the SDE scale and the BIDR scales in general. However, the studies described in the literature review that attempted to determine the degree of influence of substance and style on the SDE and IM BIDR scale scores (Connelly & Chang, 2015; Lönnqvist et al, 2007 and Pauls & Stemmler, 2003), found that under the kind of anonymous, low stakes response conditions used in papers 1-5 BIDR scores are influenced by both factors.

Paper 1 noted that Rinaldo and Okada (1993) have suggested that the nonconscious self-inhibitory forces reported by Ahsen (e.g. Ahsen, 1985), and that form the suppressor mechanism in Hishitani et al's (2011) imagery vividness model, could overlap with the unconscious self-deceptive processes thought to underlie responses on the SDE scale. In this

proposal 'substance' takes the form of social desirable responding being a departure from reality, but an integral part of the normal functioning imagery system. There is evidence which supports this proposal and suggests that the two processes are associated. For example, Paulhus (1991) reports that SDE scores are positively related to measures of defense and coping with anxiety, the largest of which is with the use of a repressive coping style (r=.51). Equally though, there appear to be differences between the two mechanisms. While the SDE bias focuses on ego enhancement, Ahsen's nonconscious self-inhibitory processes focus on ego defense. In line with this, psychophysiological studies carried out since paper 1 was published have found associations between the two processes and different brain regions. Barrios et al (2008) conducted a TMS study which found links between SDE and processing in anterior parts of the medial prefrontal cortex, an area located at the very front of the frontal lobe. In contrast, Motoyama et al (2010, cited in Hishitani et al, 2011) linked the suppressor mechanism to the left posterior cingultate gyrus, a semi-circular fold of brain tissue that covers the corpus callosum and forms part of the limbic system. Motoyama et al based this assertion on the observation of greater fMRI activity during negative imagery in the cingultate gyrus, and that the activity was highly correlated with imagery vividness (r=.82).

Furthermore, the hypothesized imagery mechanisms of Ahsen and Hishitani et al have been discussed primarily in the context of within participant variation in imagery vividness and different mechanisms may be responsible for more stable patterns of individual differences in imagery vividness. Hishitani et al suggested that such individual differences might occur for two possible reasons: 1) vivid imagers have greater capacity in the information channel responsible for transporting visual information from LTM to working memory; 2) vivid imagers possess greater visual information in their LTM. Hishitani et al argue that the first possibility is more likely because interpersonal differences in vividness still exist when images are formed of very familiar objects and scenes such as those used in the VVIQ.

In paper 1 we also offered a position on the distortion-substantive issue which drew on the work of Helmes (2000). We argued that we were agnostic as to whether socially desirable responding is best viewed as a personality characteristic, response set or mixture of both, but believed that its control would appear in most cases to aid more precise measurement of the construct of interest. Indeed in the current context, socially desirable responding

would appear to be conceptually distinct from mental imagery, and as such its control should lead to more precise measurement. Given this, we went on to suggest possible intervention strategies including explicitly asking participants to respond honestly and reminding participants that mental imagery is generally thought to be less vivid than its corresponding sensory experience. We argued that direct appeals to participants to respond honestly and realistically were likely to have most impact on conscious and deliberate forms of socially desirable responding. Thus in principle, AM forms of the egoistic bias should be more responsive to such appeals than SDE.

Methodological contributions

Paper 1 advanced methodology in this area through its use of two scales from the BIDR for the first time. The SDE and IM scales were superior to the measures of social desirable responding used by McLemore (1976) and Anderson and Buyer (1994). The BIDR scales had a balance of both positively and negatively phrased items, and did not include items relating to poor adjustment and psychopathology. The article also found that the comparison of correlations from the two studies showed that the ordering of questionnaires used did not appear to have an effect on the pattern of results seen. Futhermore, the analysis of self-report imagery questionnaire items showed that there were no grounds on which to remove any items from either the VVIQ-1 or the VMIQ.

Limitations

Despite having strengths, the first paper also had several limitations. These will now be discussed.

The self-report imagery questionnaires used were both measures of vividness and designed to assess visual or motor imagery. At this stage of the research it was unclear to what extent the correlations seen with the SDE scale would generalize to ratings of other imagery properties and other sensory modalities. This was addressed in papers 2 and 3.

Although a significant relationship had been found between two self-report measures of imagery and the SDE scale, no data existed to help determine the nature of the underlying

cause or causes of these relationships, and more specifically, whether those causes reflected distortion or substantive factors. This was addressed in papers 2, 3 and 5.

Only two of Paulhus's four BIDR scales had been used in the study, the SDE and IM BIDR scales. Although these two measures were the most well tested of the scales and cover both egoistic and moralistic biases, according to Paulhus (2002), they also differed in that SDE is best suited to measuring unconscious processes while IM is best suited to measuring deliberate, conscious processes. As such, differences in patterns of results between the two scales could be due either to differences in type of social desirable responding (egoistic vs moralistic) or level of consciousness of the social desirable response. Equally though, more recent research by Croatian researchers have questioned the extent to which the four BIDR scales capture the unconscious – conscious dimension posited by Paulhus's (2002) model. This was addressed in paper 4.

If the self-report imagery scales' correlations with SDE seen in paper 1 reflect the egoistic bias, it was unclear whether Pauhus's two BIDR egoistic scales, SDE and AM might show independent and additive relationships with self-report imagery questionnaires or not. If they did, it was possible that the relationship with socially desirable responding could be greater than the estimate from this study which was based on SDE alone. This was addressed in papers 4 and 5.

Our criticism of past studies use of the MC had received support from the pattern of correlations seen in paper 1. However, because the study had not included the MC test, we had no direct support for the validity of this critique. This was addressed in paper 4.

Only dichotomous scoring of the BIDR scales was used in this study as recommended by Paulhus (1994). But it is possible that superior results might have been obtained if we had used continuous scoring. This was addressed in papers 3 and 4.

Finally, in paper 1 the first author had approached the research topic primarily from the perspective of an imagery researcher, indeed, the first paper was submitted to a mental imagery journal. As such, much of how the topic was thought about was framed in terms of the knowledge present at the time in this subject area. Thus correlations between self-

report imagery questionnaires and measures of social desirable responding were conceptualised primarily in terms of distortion, threat to validity, acceptable and unacceptable levels of 'contamination', and measures that an experimenter might take to reduce the impact of social desirable responding. Also, the paper offered only limited suggestions as to what exactly might underlie a substantive relationship. This was addressed in papers 5.

Thus, the papers that followed this first piece of research attempted to work systematically to address these issues.

Paper 2: Allbutt, Ling and Shafiullah (2006)

Reference: Allbutt, J., Ling, J. & Shafiullah, M. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery. *Imagination, Cognition and Personality, 25*, 4, 337-353.

https://doi.org/10.2190/Y067-0726-M45W-37W8

Aims of the paper

The primary aim of paper 2 was to investigate whether the pattern of findings found in paper 1 for ratings of vividness, generalised to the ratings of other imagery properties such as the ease with which an image can be generated. Paper 2 tested one self-report imagery questionnaire, the Shapes Questionnaire (Dean & Morris, 1991, 2003), along with the SDE and IM scales of the BIDR. The shapes questionnaire assesses a wide range of self-reported visual imagery properties including vividness, but also assesses 17 other aspects of the imagery experience such as the ease of evoking an image and the amount of detail in an image.

The rationale for using the Shapes Questionnaire stems from critiques of the use of vividness to measure self-reported imagery. Some researchers e.g. Lacey and Lawson (2013), have argued that vividness may not be the best measure of subjective imagery experience that can be used and that questionnaires based on theoretical models of the underlying nature of mental imagery may be more appropriate. Lacey and Lawson argue that vividness is only weakly related to such models, and using only vividness as the measure of imagery ignores the fact that imagery may be a multifaceted phenomenon.

The Shapes Questionnaire draws on the work of Kosslyn (1980, 1994) in suggesting that mental imagery may be best viewed as a collection of relatively independent sub-processes that should be assessed separately. Kosslyn's model emphasizes four key imagery processes: image generation, image maintenance, image inspection and image transformation / manipulation. Kosslyn at al (2004) provided neuroimaging evidence that imagery subprocesses show evidence of localization in different brain areas. Thus the Shapes test represents an attempt to go beyond ratings of what has been the most popular measure of

the imagery experience i.e. vividness. Furthermore, the authors have argued that the inclusion of measures such as the ease of generation and ease of maintenance may prove to have more functional significance than pictoral aspects.

The second aim was to investigate whether the size of correlations seen between the different Shapes Questionnaire ratings and the BIDR scales showed a relationship with perceptions of whether the ratings measure imagery ability. Some of the ratings made on the Shapes Questionnaire, such as vividness or ease of generation, may be perceived by respondents as capable of indicating a person's ability level at visual imagery. In contrast other ratings, such as whether the rotation of the image is continuous or discrete, may be perceived as unconnected to ability level and more to do with cognitive style. If respondents perceive the ratings in this way, then the size of any correlations with measures of socially desirable responding may vary as a function of these perceptions. Testing this had the potential to shed light on whether the relationship between self-report imagery questionnaires and the egoistic bias was due to distortion or substantive factors. If the size of correlations between the Shapes ratings and SDE varied in line with participants' perceptions of the extent to which the ratings measure imagery ability, this would be suggestive that the relationship reflected distortion.

The Relationship Between Self-Report Imagery Questionnaire Scores and Sub-types of Socially Desirable Responding: Components of Visual Imagery

Abstract

Allbutt, Shafiullah and Ling [1] found that scores on self-report measures of visual and movement imagery vividness correlate primarily with an egoistic form of socially desirable responding rather than a moralistic form. The current study investigated whether the pattern of findings generalises to the ratings of other imagery properties such as the ease with which an image can be generated. Participants completed the Shapes Questionnaire [2] and the Balanced Inventory of Desirable Responding [3]. Several of the Shapes Questionnaire ratings correlated significantly with the egoistic form of socially desirable responding, while correlations with the moralistic form were rare. This shows the pattern of findings generalises to the ratings of properties of the imagery experience other than vividness.

Introduction

Paulhus defines socially desirable responding as "…the tendency to give overly positive self-descriptions" [3, p. 50]. Research into the relationship between socially desirable responding and responses to self-report imagery questionnaires has involved either experimental manipulation of experimenter demands or the correlation of responses on imagery questionnaires with psychometric measures of socially desirable responding. The experimental approach has shown that the manipulation of demands, such as the desirability of vivid mental imagery, can influence responses on self-report imagery questionnaires [e.g. 4], although the effects of manipulations have not always been shown to be effective [e.g. 5]. The correlational approach has proved useful in allowing estimates of the size of the influence of socially desirable responding on self-report imagery questionnaire scores. McKelvie [6, 7], provides meta-analytical information on the relationship between the most widely used psychometric measures of visual imagery and socially desirable responding, the Vividness of Visual Imagery Questionnaire (VVIQ) [8] and the Marlowe-Crowne social desirability scale (M-C) [9]. McKelvie [7] made two estimates of the degree of relationship between the VVIQ and socially desirable responding. The first estimate was based on correlations with the M-C scale only. Data from 10 studies lead to a reported r of 0.189, with more vivid imagery being associated with greater socially desirable responding. The second estimate added a further six correlations to McKelvie's data pool from three other tests of socially desirable responding used in a study by Anderson and Buyer [10]. The inclusion of this extra data produced a revised r of 0.075, indicating a much weaker effect of socially desirable responding. McKelvie [11] proposed that for self-report imagery questionnaires, correlations up to 0.25 in size indicate either no relationship or an acceptable degree of relationship depending on size, and correlations in excess of 0.25 indicate unacceptable contamination. McKelvie based his suggestion for a cut-off point on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made other authors. Thus from his analyses McKelvie concluded that "…the VVIQ is not seriously contaminated by socially desirable responding, and is perhaps not contaminated at all" [7, p. 209].

In the late 1950s and early 1960s a number of researchers [e.g. 12, 13], prompted by observations of low intercorrelations between different measures, suggested that socially desirable responding may occur in more than one form. Since then this position has gained increasing acceptance. Work by Delroy Paulhus and his colleagues into the different types of socially desirable responding has been of particular importance. In his earliest work, Paulhus [14] suggested that there are two forms of socially desirable responding: 'self-deception' (SD), reflecting an honest, but overly positive self-

presentation, and 'impression management' (IM) reflecting self-presentation tailored to an audience. The two forms were viewed as simply differing in terms of level of conscious awareness with the former being unconscious and the later conscious. In his most recent work, Paulhus [3], has shifted his position to propose a two-tier model of socially desirable responding which has a primary distinction in terms of 'personality content' [p63], and a secondary distinction in terms of level of consciousness (see Figure 1).

Figure 1. Paulhus's Two-Tier Model (modified from Paulhus, 2002).

In the first tier there is a separation of socially desirable responding into egoistic and moralistic biases. The egoistic bias refers to the tendency to claim positive social and intellectual qualities, while the moralistic bias refers to the claiming of positive moral qualities and the refutation of negative socially-deviant qualities. In the second tier, both the egoistic and moralistic biases are divided into conscious and unconscious forms. The egoistic bias becomes self-deceptive enhancement (SDE) in its unconscious form and agency management (AM) in its conscious form. The moralistic bias becomes self-deceptive denial (SDD) in its unconscious form and communion management (CM) in its conscious form. Support for the model comes from two kinds of studies. First, Paulhus and Notareschi [15] who carried out a series of studies varying the kind of 'fake good' instructions given to participants to elicit socially desirable responding. Second, Paulhus and John [16] and John and Paulhus [17] carried out factor analytic studies factoring residual scores derived from differences between self and other estimates of personality and intelligence together with measures of socially desirable responding.

Participants answering self-report imagery questionnaires will most likely view imagery as an ability being tested. Given this, the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, would seem to be more of a threat to the validity of responses than the moralistic bias. However, Paulhus [14] reported the M-C scale had a loading of 0.68 on a factor labelled 'IM' and 0.40 on a factor labelled 'SD'. In terms of the two-tier model, this suggests that the M-C scale is a moderate to strong measure of moralistic bias, but only a moderate to weak measure of egoistic bias. As both of McKelvie's [7] estimates of the degree of relationship between the VVIQ and socially desirable responding drew either exclusively or predominately on data from the M-C scale it is possible that he may have underestimated the size of the relationship.

A measure of socially desirable responding which may be better at assessing the degree of relationship between self-report imagery questionnaires and socially desirable responding is Paulhus's Balanced Inventory of Desirable Responding (BIDR). Paulhus [14] developed the BIDR in an attempt to assess the different forms of socially desirable responding. The first version of the BIDR had two scales corresponding to SD (later renamed as SDE) and IM. In the sixth version [18], an extra optional scale measuring SDD was added, and in the eighth version a forth scale measuring AM was added [3]. In this latest version of the BIDR, the SDE, AM and SDD scales map straight onto the corresponding terms of the two-tier model, while IM scale is viewed as a measure of CM. The BIDR has several properties which make it superior to the M-C and other existing measures of socially desirable responding. First, the questionnaire includes scales which measure egoistic and moralistic bias and so offers the capacity for assessing them separately. Second, each scale consists of 20 items with equal numbers of positively and negatively keyed items. The balanced nature of the scales militates against agreement and acceptance response sets. Third, no items on the BIDR are associated with poor adjustment or psychopathology and so the BIDR is suitable for use with non-pathological samples.

Allbutt, Shafiullah, and Ling [1] made use of the SDE and IM scales of BIDR in an attempt to study the relationship between two self-report imagery questionnaires and socially desirable responding. The SDE and IM scales were selected for use because they were the most well established of the BIDR scales and they included one measure of egoistic bias and one measure of moralistic bias. Including a measure of egoistic bias and a measure of moralisitic bias allowed a comparison of the relationship between self-report imagery questionnaires and these two broad types of socially desirable responding. The measures of visual imagery used were the VVIQ [8] and the Vividness of Movement Imagery Questionnaire (VMIQ) [19]. The VMIQ is a measure of the vividness of movement imagery and has two halves: imaging movement in others and imaging kinesthetic sensations within the imager themselves. Allbutt et al. analysed scores from each half of the questionnaire separately as well as when combined into a total score and referred to these as VMIQ Other, VMIQ Self and VMIQ Tot respectively. A correlational analysis found that the self-report imagery questionnaires correlated primarily with SDE rather than IM, with greater image vividness being associated with greater SDE. Allbutt et al. reported the correlations between SDE and the imagery scales: VVIQ, -0.25; VMIQ Other, -0.23; VMIQ Self, -0.33; and VMIQ Total, -0.31, these correlations are negative in direction as greater vividness of imagery on the VVIQ and VMIQ are indicated by low scores. These results suggest that scores on self-report imagery questionnaires may well be affected by socially desirable responding and to a degree in excess of the 0.25 criterion set by McKelvie [11]. They also support the suggestion that the relationship appears to be strongest with the egoistic form of socially desirable responding.

Allbutt et al.'s study was useful in showing a consistent pattern of results across two questionnaires aiming to measure the vividness of visual and movement imagery. Studying the relationship between self-reports of the vividness of mental images and socially desirable responding is an important starting point as this has been the property of mental images most studied by researchers interested in imagery self-reports. However, it remains to be seen whether the same pattern of results would be obtained if the self-reports of properties other than vividness were to be tested.

Deane and Morris [2, 20] have developed a self-report imagery questionnaire called the 'Shapes Questionnaire'. The questionnaire aims to assess a wide range of self-reported visual imagery properties including vividness, but also assesses 17 other aspects of the imagery experience such as the ease of evoking an image and the amount of detail in an image. The rationale behind the development of the test draws on the work of Kosslyn [e.g. 21, 22] in suggesting that mental imagery may be best viewed as a collection of abilities that should be assessed separately. The items used in the questionnaire are of the kind used on tests of spatial ability such as the 'block shapes' used in the Vandenberg test of mental rotation [23]. Respondents are required to look at the shapes, generate images of them, maintain these images, and finally rotate the images.

Deane and Morris [2, 20] have used two versions of their questionnaire. In Deane and Morris's 1991 paper, the Shapes Questionnaire was made up of ratings of five imaged items. In the 2003 paper only two of these items were retained, one a two dimensional shape, and the other a three dimensional shape, with analyses performed separately for each shape. In their work Deane and Morris have investigated why scores on self-report imagery questionnaires such as the VVIQ rarely correlate with scores on more objective spatial tests. In both papers they found scores on their questionnaire to show a stronger relationship with scores on spatial tests than other self-report questionnaires such as the VVIQ. They suggest that this may be because of the better match in item type and / or type of imagery between the Shapes Questionnaire and spatial tests, and the limits of only measuring imagery in terms of vividness which they argue may only be an overall measure of imagery processes.

In the current study we aimed to extend our earlier research with self-report imagery questionnaires and the BIDR to include a study of the Shapes Questionnaire. Its use allows the earlier research to be extended in two ways:

1) To investigate the pattern of correlations between self-report imagery questionnaires and measures of socially desirable responding for ratings of a range of different imagery properties in addition to vividness.
2) To investigate whether the size of correlations seen between the different Shapes Questionnaire ratings and the BIDR scales show a relationship with perceptions of whether the ratings measure imagery ability. Some of the ratings made on the Shapes Questionnaire, such as ease of generation or vividness, may be perceived by respondents as capable of indicating a person's skill level at visual imagery. Whereas other ratings, such as whether the rotation of the image is continuous or discrete, may be perceived as unconnected to skill level and more to do with cognitive style. If respondents perceive the ratings in this way, then the size of any correlations with measures of socially desirable responding may vary as a function of these perceptions.

For the study, data collection occurred in two phases. In the first phase a sample of psychology students answered the Shapes Questionnaire and the SDE and IM scales of the BIDR. We used the 1991 version of Deane and Morris's Shapes Questionnaire because it gave us the option of pulling out the data from the two shapes used in the 2003 version for analysis. In order to keep the demands placed on participants to a manageable level we used only two of the four BIDR scales. In addition, Paulhus himself [24] notes several reasons that make the use of the SDD scale unappealing: Some of the SDD scale items are considered offensive by ethical review boards, the scale has been found to correlate highly with IM under many conditions and its additional 20 items can make the BIDR overly long. Also the newest of the BIDR scales, AM, is a relatively untested scale and has as yet no published data on its psychometric properties.

The second data collection phase involved different groups of psychology students answering a questionnaire designed to identify the extent to which ratings from the Shapes Questionnaire were perceived as measures of imagery ability. We will refer to this questionnaire as the 'reflection of imagery ability questionnaire'.

Method

Participants

The Shapes Questionnaire (1991 version) and BIDR were completed by 100 first year undergraduate psychology students (18 males and 82 females). Ages ranged from 18 to 43 with a mean age of 20.4 years. To increase the sample to an adequate size to perform a factor analysis on the Shapes Questionnaire, data from a further 20 first year psychology undergraduates who had completed this

questionnaire only were added. This increased the sample size to 120 (22 males and 98 females). Ages ranged from 18 to 58 years with a mean age of 23.6 years.

The reflection of imagery ability questionnaire was answered by a different sample of 78 psychology students. Data from seven of these participants was rejected as comments made by them at the end of their questionnaires indicated that they were unsure as to the task they were being asked to do. For the remaining 71 participants 16 were males, 55 females, the age range was 18 to 51, and the mean age 25.2 years.

All participants were volunteers and no inducements were paid to them.

Materials

Ethical rights form.
A form was used which informed participants of the rights they possessed under the British Psychological Society's code of ethics.

The Shapes Questionnaire [2].
The Shapes Questionnaire assess 18 different properties of visual images via a series of ratings, for example 'How easy is it for you to evoke this image?'. Sixteen of the ratings can be viewed as subscales of the questionnaire which produce scores suitable for statistical analysis. The names of these ratings can be seen in Table 2 in the results section. The ratings are numbered from 1 to 16. However, ratings 5 and 12 are split into 'a' and 'b' parts relating to changes in detail and clarity while either maintaining or rotating the image. When answering the questionnaire participants image in turn five shapes of the kind seen in tests of spatial ability. For each shape an image is generated, maintained and then rotated. While carrying out these procedures the images are rated for the different properties. The ratings of most properties are made on 9-point rating scales with appropriate end labels for example 1, 'Very Difficult' and 9, 'Very Easy' and then summed across the five shapes. In general, higher scores indicate more of the property being assessed for example easier, more vivid, or more of the shape seen. There are two exceptions to this pattern of scoring. Ratings of change of detail and clarity are first rated on a scale of 1 to 9, then transformed to a scale of –8 to +8 depending on whether the change was in favour of less or more detail / clarity. A higher score indicates more detail / clarity. Ratings of whether the rotation is continuous or discrete have response options of 'Continuous' and 'Discrete'. These were scored as 1 and 0 respectively with a higher score indicating more continuous rotation.

The Balanced Inventory of Desirable Responding Version 8 (BIDR-8) [3].
The BIDR aims to assess socially desirable responding. Use was made of the two main 20 item subscales, SDE and IM. SDE is a measure of unconscious egoistic bias. The subscale has statements about the respondent's abilities across a range of situations, half of which are positively phrased and half negatively phrased. For example 'My first impressions of people usually turn out to be right' (positive). Responses were made on a seven point scale ranging from 1, 'Not true', to 7, 'Very true'. IM is a measure of conscious moralistic bias and this subscale is organised in a similar fashion to that of the SDE. It has statements about the respondent's ethical behaviour across a range of situations, half of which are positively phrased and half negatively phrased. For example 'I never cover up my mistakes' (positive). Responses are made on the same 7-point scale as the SDE.

For scoring, negative items were recoded and then responses of either six or seven on an item were counted as one point, while responses of one to five did not score. The points were totalled for each subscale separately. Higher scores indicate more SDE or IM. The rationale behind the scoring method is that only clearly exaggerated or managed responses achieve scores [18].

Additional instructions were added to the start of the BIDR to make two items from the SDE scale and one item from the IM scale relevant to all participants. The items from the SDE scale were: item 8, 'I am not a safe driver when I exceed the speed limit', and item 14, 'My parents are not always fair when they punish me'. The item from the IM scale was: item 33, 'I sometimes drive faster than the speed limit'. Items 8 and 33 are problematic for participants who do not drive, so they were asked to answer the items by imagining what they would be like if they were to drive. Item 14 is problematic for older participants and so the instructions asked participants to remember back to when they lived at home and comment on how they felt at the time. This approach was preferred to the omission of these items as it preserved the balanced nature of the BIDR scales.

The ethics form and three questionnaires were combined into a booklet. The booklet started with the ethics form, followed by the Shapes Questionnaire and the BIDR.

The reflection of imagery ability questionnaire

The questionnaire introduced the concept of mental imagery and explained the nature of the Shapes Questionnaire. The idea that some kinds of ratings of imagery properties could inform us about differences in imagery ability while others could inform only about differences in cognitive style was described. A list of the ratings from the Shapes Questionnaire for example ease of evoking an image, amount of detail, was given. Respondents were asked to tick either 'not an indicator at all', 'a weak indicator of ability', 'a moderate indicator of ability' and 'a strong indicator of ability' for each of the Shapes Questionnaire ratings. For data analysis purposes these options were scored as 0, 1, 2 and 3 respectively.

Procedure
For the data from the Shapes Questionnaire and BIDR, participants answered the questionnaires in the last hour of a two hour laboratory class. They were asked to seat themselves as far apart as possible. Responses were made in silence and participants were told not to put their name anywhere in the booklet. Participants answered the questionnaires in the order in which they appeared in the booklet. Following this they were given a debriefing sheet which explained the purpose of the study and gave the contact details of the experimenters in case of queries.

For the data from the reflection of imagery ability rating questionnaire, the students answered the questionnaires at the start of a lecture. The instructions on the questionnaires were read out loud by an experimenter to aid understanding. When all participants had completed the questionnaires they were debriefed.

Results

The presentation of the results is organised into two parts: investigation of the relationship between the Shapes Questionnaire ratings and BIDR scales, and an investigation of the relationship between the Shapes Questionnaire – BIDR scales correlations and reflections of imagery ability questionnaire scores.

The relationship between the Shapes Questionnaire ratings and the BIDR scales

The relationship was investigated in two ways. First, by factor analysing the Shapes Questionnaire ratings to reveal their underlying factors and then correlating factor scores with the BIDR scales. This analysis gave an overall measure of the relationship between the Shapes Questionnaire ratings and the BIDR scales. Second, by correlating each Shapes Questionnaire rating with the BIDR scales. This analysis gave a measure of the relationship between individual Shapes Questionnaire ratings and the BIDR scales.

Prior to the data analysis being conducted, the variables were screened for univariate and multivariate outliers reducing the total sample size to 114, and the sample size for participants who had completed both the Shapes Questionnaire and the BIDR to 94. Non-normal distributions were corrected with transformations. A prospective power analysis using tables for Pearson's Product Moment correlation showed that for a two- tailed test a sample size of 90 would be required to detect a medium effect size [25].

Prior to the factor analysis the 16 ratings were correlated using Pearson Product Moment correlations. This analysis revealed a degree of multi-collinearity among variables. To overcome this, if two ratings correlated at 0.80 or greater, then one of them was removed. This criterion was applied so as to leave as many different kinds of scales as possible in the analysis and to be as consistent in omissions as possible. As a result of this procedure five ratings (ratings 2, 3, 5b, 12b, and 13) were removed from the data set. A factor analysis was then carried out to reduce the 16 ratings of the Shapes Questionnaire to their underlying factor(s). A Principle Axis factor analysis with Oblimin rotation was used.

A scree plot analysis clearly showed the presence of only one factor. The factor had an eigenvalue of 4.83 and it accounted for 43.94% of variance in the data set. It was the only factor to have an eigenvalue in excess of 1.The Shapes Questionnaire ratings loadings on this factor can be seen in the structure matrix in Table 1. Only loadings of 0.3 or greater are shown.

Table 1. Structure matrix from Shapes Questionnaire factor analysis

Item	Factor loadings
4) Ease of maintenance	0.93
8) Vividness of image	0.85
1) Ease of evoking image	0.80
16) Vividness of image during rotation	0.78
6) Amount of shape imaged	0.77
9) Ease of rotation	0.72
7) Size of image	0.42
5a) Amount of detail change during maintenance	0.39
11) Size change during rotation	0.39
12a) Amount of detail change during rotation	0.35
10) Continuous vs. discrete rotation	-0.32

The pattern of factor loadings suggest that the factor is a global measure of imagery ability taking in both static and dynamic aspects, which primarily reflects the ease and vividness of imagery processes.

The BIDR scales were next correlated with the factor scores. The Shapes Questionnaire factor scores showed a low positive significant correlation with SDE such that 'better' imagery was associated with more SDE (r=+0.25, df=92, p<0.05, two-tailed), and a low negative non-significant correlation with IM (r=-0.11, df=92, p>0.05, two-tailed). There was also a low positive significant correlation between SDE and IM (r=+0.25, df=92, p<0.05, two-tailed).

The BIDR scales were then correlated with each individual rating from the Shapes Questionnaire. The results of the correlational analysis can be seen in Table 2. For simplicity, the table shows only the correlations between the imagery ratings and the two BIDR scales[1].

Table 2. Matrix of Pearson's correlations for Shapes Questionnaire and BIDR

Ratings	SDE	IM
IM	+0.25*	
1) Ease of evoking the image	+0.22*	+0.10
2) Amount of detail	+0.26*	+0.23*
3) Clarity and sharpness	+0.25*	+0.14
4) Ease of maintenance	+0.22*	+0.05
5a) Detail change while maintaining	+0.08	-0.02
5b) Clarity change while maintaining	+0.10	+0.09
6) Amount of shape imaged at any one time	+0.18	+0.13
7) Size of image	+0.13	-0.02
8) Vividness of image	+0.20	+0.18
9) Ease of rotation	+0.16	+0.02
10) Whether rotation is continuous or discrete	+0.01	-0.13
11) Size change during rotation	+0.11	+0.15
12a) Detail change while rotating	+0.37**	+0.09
12b) Clarity change while rotating	+0.34**	-0.01
13) Amount of shape imaged while rotating	+0.20*	+0.16
16) Vividness of image while rotating	+0.30**	+0.09

Note: * $p<0.05$ ** $p<0.01$ (2-tailed)

The first four and last four ratings on the Shapes Questionnaire showed small, but significant positive correlations with SDE. For these questions greater ease, amount, clarity, detail and vividness of imagery processes were associated with greater socially desirable responding. Some of the differences between the ratings that reached significance and those that did not are surprising. For example both detail and clarity change while rotating were significant, while detail and clarity change while maintaining were non-significant and near zero in size. Overall the average size of the correlations between the 16 Shapes Questionnaire ratings and SDE was +0.20. Only one correlation with IM reached significance, this involved the amount of detail in the image.

The relationship between the Shapes Questionnaire – BIDR scales correlations and reflections of imagery ability questionnaire scores

The purpose of this part of the analysis was to investigate whether the size of the Shapes Questionnaire - BIDR scales correlations showed a relationship with perceptions of whether the Shapes Questionnaire ratings measure imagery ability. First the agreement of participants' ratings of their perceptions of the likelihood of each imagery dimension reflecting imagery ability was assessed using a two-way random effects model intraclass correlation (ICC) analysis. Then the Shapes Questionnaire - BIDR scale correlations were correlated against the reflection of imagery ability ratings. If the size of the Shapes Questionnaire – SDE correlations is influenced by the extent to which participants judge the ratings to reflect imagery as an ability then there should be a significant positive correlation for the Shapes Questionnaire - SDE correlations only. The reason for the inclusion of the Shapes Questionnaire – IM correlations in the analysis was to act as a control as this correlation should not be significant. The ICC analysis showed that there was good agreement in ratings across students (ICC=0.88 for consistency and 0.84 for absolute agreement). The data for the correlations was screened for parametric data assumptions, no univariate or multivariate outliers were found and all distributions were normal. The correlation between Shapes Questionnaire – SDE correlations and the reflection of imagery ability ratings was positive in direction and significant (r=+0.53, df=14, p<0.05, two-tailed). The direction of the correlation was consistent with the larger Shapes Questionnaire - SDE correlations being associated with

participant's judgements that a Shapes Questionnaire rating reflects imagery ability. The correlation between Shapes Questionnaire – IM correlations and the reflection of imagery ability ratings was non-significant (r=+0.41, df=14, p>0.05, two-tailed).

Discussion

The main aim of the research was to investigate the pattern of correlations between the Shapes Questionnaire and socially desirable responding as measured by the SDE and IM scales of the BIDR. Previous research by Allbutt et al. [1] found that scores on self-report imagery questionnaires which measured the vividness of visual and movement imagery correlated primarily with the SDE rather than IM. The current paper aimed to use the Shapes Questionnaire to investigate whether this pattern of results would extend to ratings of properties of the imagery experience other than vividness.

The results show that, overall, the data from the Shapes Questionnaire showed a similar pattern of correlations to those observed by Allbutt et al. The correlation between the factor scores from the factor analysis of the Shapes Questionnaire only reached significance with SDE, and for the individual Shapes Questionnaire ratings half of the correlations with SDE reached significance while only one correlation with IM reached significance. The correlations that reached significance included processes such as ease of generation and amount of detail, and so were not restricted to just self-reports of imagery vividness. All of these correlations were positive in direction such that 'better' imagery, for example greater ease of imagery processes, were associated with higher social desirability scores. Interpretation of the meaning of the difference in the pattern of correlations is complicated by the fact that the SDE and IM scales differ in terms of type of socially desirable responding measured (egoistic versus moralistic biases) and level of consciousness of processes (conscious versus unconscious). However, two features of the findings make it more plausible that type of socially desirable responding underlies the difference. First, the correlations between the Shapes Questionnaire and SDE were all positive in direction i.e. higher SDE was associated with 'better' imagery. Second, the correlation between the Shapes Questionnaire – SDE correlations and the reflection of imagery ability ratings was positive in direction and significant, while the correlation correlation between the Shapes Questionnaire – IM correlations and the reflection of imagery ability ratings was not significant. Both of these patterns of results would be expected if participants answering the Shapes Questionnaire are viewing imagery as a skill that is being tested and the results were caused by differences in type of socially desirable responding. These patterns of results would not be expected if the results were caused by differences in levels of consciousness.

Only some of the Shapes Questionnaire ratings correlated significantly with SDE. The significant positive correlation between the Shapes Questionnaire – SDE correlations and the reflection of imagery ability ratings suggests in general that the differences between the imagery ratings that did and did not correlate significantly reflect the extent to which participants perceived a rating as indicating a person's skill at visual imagery. However, some of the differences between ratings that did correlate and those that did not are surprising. For example both detail and clarity change while rotating were significant, while detail and clarity change while maintaining were non-significant and near zero in size. Beyond the possible effect of chance variation in our data we can see no obvious reason why differences such as these could have occurred.

McKelvie [11] proposed that correlations up to 0.25 in size between self-report imagery questionnaires and measures of socially desirable responding would indicate either no relationship or an acceptable relationship, while correlations in excess of 0.25 would indicate an unacceptable degree of contamination. In our study the statistic which probably produces the best overall assessment of the degree of association between the Shapes Questionnaire ratings and socially desirable responding is the correlation between the factor scores from the factor analysis and SDE. This correlation reached a level of 0.25. In addition several of the correlations between the Shapes Questionnaire ratings exceeded 0.25 in size. McKelvie's suggested cut-off

point of 0.25 (6.25% variance) was based on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by previous authors. As such it is a useful contribution in indicating generally where a cut-off point might lie. However, we suggest that, like all cut-off points, the exact value chosen is somewhat arbitrary, and judgements of the validity of imagery questionnaires needs to be made in light of data on other aspects of validity as well.

The different pattern of correlations seen in this study for the SDE and IM scales shows the importance of treating socially desirable responding as a multidimensional entity, and suggests that the BIDR is the most appropriate measure of socially desirable responding to use to attempt to assess the relationship between self-report imagery questionnaires and socially desirable responding. The main weakness of the M-C scale for assessing this relationship is that it is a single scale which loads more highly on the moralistic form of socially desirable responding than the egoistic form. In contrast, the BIDR has separate scales to measure egoistic and moralistic biases, and both scales load highly on their respective form of socially desirable responding. The different questionnaires used may explain why the results from Allbutt et al. and the current study suggests there may be a stronger relationship between self-report imagery questionnaire scores and socially desirable responding than the estimate made by McKelvie [7] based on studies using the M-C scale.

Theorists have conceptualised socially desirable responding as either a personality characteristic or response style and thus differed in whether they view it as a meaningful source of variance or not. Our position on this issue is similar to that of Helmes [26] who argues that in practice being able determine the exact nature of socially desirable responding is not crucial. This is because in most cases it is likely that socially desirable responding, however it is conceptualised, will be distinct from the phenomena under investigation. In the current context, socially desirable responding would appear to be conceptually distinct from mental imagery, and so its control should lead to more accurate measurement. Possible strategies to reduce the impact of an egoistic bias include: requests to participants to respond honestly or realistically (scores on questionnaires like the VVIQ are often very high yet mental imagery is generally thought to be less vivid than its corresponding sensory experience Chara and Hamm [27]), removing data from participants who show extreme social desirability scores, or using statistical techniques to partial out the influence of socially desirable responding. We suspect that direct appeals to participants to respond honestly or 'realistically' are likely to have most impact on conscious and deliberate forms of socially desirable responding. Thus AM forms of the egoistic bias should be more responsive to such appeals than SDE.

However, Ahsen's work on the unvividness paradox [e.g. 28-31], shows the existence of nonconscious self-inhibitory forces that affect the vividness of images, and which may form part of the imagery system. For example, Ahsen [30] asked participants to think about one of their parents while answering the VVIQ and found that the reported vividness of images was less when fathers were thought about than mothers. Rinaldo and Okada [32] suggest that such phenomena show a form of self-deception as participants may be incorrectly making estimates of low vividness because of the influence of emotionally-charged associations with their father. If these self-deceptive inhibitory processes form part of the normal imagery system, then it could also be true that that the self-enhancing processes tapped by the SDE scale are also part of the system. Ahsen's work shows the potential for emotional associations linked to stimuli to influence ratings of vividness and would be of particular concern for self-report imagery questionnaires such as the VVIQ which ask participants to image stimuli such as a well know friend or relative's face. Responses made to questionnaires such as the Shapes Questionnaire may be less open to these kinds of influences because of the abstract nature of their items.

Future work might attempt to establish whether the pattern of results seen in this and our previous paper extends to other kinds of self-report imagery questionnaires, for example those measuring imagery in different modalities and those measuring style of thinking such as Pavio's [33] Individual Differences Questionnaire. Also when details of the psychometric properties of the AM are published it would be helpful for future work to assess both SDE and AM at the same time given that the strongest relationship observed in the current study was for the measure of egoistic bias. D. L. Paulhus (personal communication, October 23[rd,] 2003) has informed us that in an unpublished study the correlation between SDE and AM was 0.56 in a fake good condition and 0.39 in an honest response condition. The moderate size of these correlations raises the possibility that AM might correlate with self-report imagery questionnaires independently from SDE and that the influence of socially desirable responding could be greater than the estimate made from our studies which were based on SDE alone.

References

1. Allbutt, J., Shafiullah, M. & Ling, J. (2006, in press). The relationship between self-report imagery questionnaire scores and sub-types of socially desirable responding: Visual and Movement Imagery. *Journal of Mental Imagery*.

2. Deane, G. & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology, 94*, 245-273.

3. Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun, D.N. Jackson & D.E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp49-69). Mahwah: Lawrence Erlbaum Associates.

4. Ashton, R. & White, K. (1975). The effects of instructions on subjects' imagery questionnaire scores. *Social Behaviour and Personality, 3*, 41-43.

5. Durndell, A.J. & Wetherick, N.E. (1975). Reported imagery and social desirability. *Perceptual & Motor Skills, 41*, 987-992.

6. McKelvie, S. J. (1995a). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*, 1-106.

7. McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*, 197-251.

8. Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

9. Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

10. Anderson, D.C. & Buyer, L.S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *American Journal of Psychology, 107*, 207-222.

11. McKelvie, S. J. (1994). Psychometric guidelines for imagery questionnaires. *Perceptual and Motor Skills, 79*, 1219-1231.

12. Edwards, A.L., Diers, C.J. & Walker, J.N. (1962). Response sets and factors loadings on sixty-one personality scales. *Journal of Applied Psychology, 46*, 220-225.

13. Wiggins, J.S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419-427.

14. Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Psychology, 46*, 598-609.

15. Paulhus, D.L. & Notareschi, R.F. (1993). *Varieties of faking manipulations*. Unpublished manuscript, University of British Columbia.

16. Paulhus, D.L. & John, O.P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1024-1060.

17. John, O, P. & Paulhus, D.L. (2000). *The structure of self enhancement*. Unpublished manuscript, University of California, Berkley.

18. Paulhus, D. L. (1994). Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6. Unpublished Manual, University of British Columbia, Vancouver, Canada.

19. Isaac, A., Marks, D.F. & Russell, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery, 10*, 23-30.

20. Deane, G. & Morris, P. E. (1991). Imagery and spatial ability: When introspective reports predict performance. In R. H. Logie and M. Denis (Eds.), *Mental images in human cognition* (pp331-347). Amsterdam: Elsevier Science.

21. Kosslyn, S. M. (1980). *Image and mind*. London: Harvard University Press.

22. Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.

23. Vandenberg, S.G. & Kuse. A.R. (1978). Mental rotations: A group test of three-dimensional spatial visualisation. *Perceptual & Motor Skills*, *47,* 599-604.

24. Paulhus, D. L. (1999). BIDR Self-Deceptive Denial (SDD). neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm. Retrieved from the World Wide Web on 1 August 2003.

25. Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove: Psychology Press.

26. Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R.D. Goffin and E. Helmes (Eds.), *Problems and solutions in human assessment: Honouring Douglas N. Jackson at seventy* (pp21-40). Boston: Kluwer Academic Pub.

27. Chara, P. J. and Hamm, D. A. (1989). An inquiry into the construct validity of the vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 69*, 127-136.

28. Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery, 9*, 1-18.

29. Ahsen, A. (1988). Imagery, unvividness paradox and the paradigm of control. *Journal of Mental Imagery*, *12*, 1-44.

30. Ahsen, A. (1990). AA-VVIQ and imagery paradigm: Vividness and unvividness issue in VVIQ research programmes. *Journal of Mental Imagery, 14*, 1-58.

31. Hockman, J. (1994). Ahsen's image psychology. *Journal of Mental Imagery, 18,* 1-118.

32. Rinaldo, R. & Okada, R. (1993). A test of the relationship between imagery vividness and social desirability across subject selection procedures. *Journal of Mental Imagery*, *17*, 171-180.

33. Pavio, A. (1971). *Imagery and verbal processes*. New York: Rinehart and Winston.

Evaluation of the paper

Strengths

Paper 2 extended the work began in paper 1 by generating two main empirical findings:

1) The data from the Shapes Questionnaire showed a similar pattern of correlations to those observed in paper 1. The Shapes Questionnaire ratings correlated primarily with SDE. When judged using the global measure of factor scores on the single underlying Shapes Questionnaire factor, the correlation between the Shapes Questionnaire and SDE was of a similar size, $r=.25$, to the correlations seen for the imagery scales in paper 1. The analysis showed that the correlations with SDE were not restricted solely to self-reports of image vividness, or to images of naturalistic scenes. Instead they extended to other pictorial aspects such as the amount of detail in an image, and to the ratings of imagery processes such as ease of generation and ease of maintenance. All correlations with Shapes questionnaire ratings were positive in direction such that 'better' imagery, for example greater ease of imagery processes, were associated with higher social desirability scores.

2) The significant positive correlation between the Shapes Questionnaire–SDE correlations and the reflection of imagery ability ratings suggests that the correlation reflects the extent to which participants perceived an imagery rating as indicating a person's ability at visual imagery. For example, ratings of image clarity and sharpness correlated $r=.25$ with SDE and were rated highly by participants as indicating ability, whereas whether rotation was continuous or discrete correlated $r=.01$ and was rated lowly as indicating ability. This finding is more consistent with the self-report imagery questionnaire and SDE correlations reflecting distortion rather than substantive factors. Paper 2 was the first to explore participants' perceptions of the extent to which different imagery properties might reflect imagery ability.

The paper also explored the factor structure of the Shapes Questionnaire. The factor analysis suggested that the Shapes Questionnaire items, when administered as a 5 item test, load on just one factor. The factor was a global measure of imagery ability taking in both static and

dynamic aspects, which primarily reflected the ease and vividness of imagery processes and accounted for 43.94% of variance in the data set.

In their work Dean and Morris (1991, 2003) have compared the size of the correlations seen between the Shapes Questionnaire and objective spatial tests compared to other self-report imagery questionnaires such as the VVIQ and spatial tests, in an attempt to understand why such self-report imagery questionnaires rarely correlate with spatial tests. In both papers they found scores on their questionnaire to show a stronger relationship with scores on spatial tests than other self-report imagery questionnaires and report correlations as high as $r=.51$ (ease of maintenance) using their test. They suggest that this may be because of the better match in item type and / or type of imagery between the Shapes Questionnaire and spatial tests, and the limits of only measuring imagery in terms of vividness which they suggest may only be an overall measure of imagery processes.

It is interesting to note that the highest correlation seen across Dean and Morris's studies involved the ease of maintenance scale. In paper 2, this scale showed a correlation of $r=.22$ with SDE. Thus the Shapes Questionnaire's relationship with SDE does not appear to prevent it from correlating with objective measures of imagery ability.

Limitations

The second paper had some of the same limitations as paper 1. For example, only two of Paulhus's four BIDR scales were used in the study, the SDE and IM. The limitations of this approach have been previously discussed in paper 1.

The data in paper 2 was collected across two samples. The first sample of 120 participants answered the Shapes Questionnaire, of whom 100 also answered the SDE and IM scales of the BIDR. The second sample of 78 participants answered the reflection of imagery ability questionnaire. However, correlations comparing the sizes of the Imagery scale–SDE correlations with the ratings of imagery ability were made across samples. This situation opens up the results to the influence of gross individual differences between samples. Interpretation of the results would have been more straightforward if all of the data had been collected from the same participants. Also, it would also have been preferable if the 20

participants who only answered the Shapes Questionnaire had also answered the BIDR scales.

While the sample size of 120 for the exploratory factor analysis was within acceptable limits, ideally it would had been larger still. Field (2013) describes samples of over 300 as 'good'. Guadagnoli and Velicer (1988) report that factors with four or more loadings greater than .6 are reliable regardless of the sample size. In line with this, the single factor which emerged for the Shapes Questionnaire in paper 2 had six items which loaded higher than .6. Futhermore, a more rigorous test of the factor structure of the questionnaire would have been achieved if the theory-driven approach of CFA had been used.

Paper 3: Allbutt, Ling, Heffenan and Shafiullah (2008)

Reference: Allbutt, J., Ling, J., Heffernan, T., M., & Shafiullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and visual thinking style. *Journal of Individual Differences, 29,* 4, 181-188. https://doi.org/10.1027/1614-0001.29.4.181

Aims of the paper

The primary aim of paper 3 was to investigate whether the pattern of findings observed in the previous papers generalised to the ratings of vividness of auditory imagery i.e. a different sensory modality, the second version of the VVIQ which had been claimed by Marks to be an advance on its predecessor, and a measure of visual thinking style which had been suggested to be less open to social desirable responding effects than scales based on properties of imagery such as ratings of its vividness. Paper 3 reported results from three self-report imagery questionnaires, the Auditory Imagery Scale (AIS) which measures the vividness of auditory imagery, the second version of the VVIQ called the VVIQ version 2 (VVIQ-2) and a measure of habitual visual thinking style the Individual Differences Questionnaire – Imagery Habit Scale (IDQ-IHS), together with the SDE and IM scales of the BIDR. None of the three self-report imagery questionnaires had been correlated before with a measure of social desirable responding.

The second aim was to investigate whether the method used to score the BIDR scales, dichotomous versus continuous, had an impact on the pattern of results seen.

The third aim was to investigate whether the size of correlations seen between the different self-report imagery questionnaire ratings and the BIDR scales showed a relationship with participants' ratings of desire to possess different aspects of imagery i.e. how much they were valued. Testing this had the potential to shed light on whether the relationship seen between self-report imagery questionnaires and the egoistic bias was due to distortion or substantive factors. If the size of correlations between the imagery questionnaires and SDE varied in line with participants' ratings of desire to possess different aspects of imagery, this would be suggestive that the relationship reflected distortion.

The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and thinking style

Abstract

Allbutt, Ling and Shafiullah (2006a) and Allbutt, Shafiullah and Ling (2006b) found that scores on self-report measures of visual imagery vividness and other imagery properties such as ease of generation, correlate primarily with the egoistic form of social desirable responding (the tendency to claim positive social and intellectual qualities) rather than the moralistic form (the claiming of positive moral qualities and the refutation of negative socially-deviant qualities). Here, three studies are reported which investigate whether this pattern of findings generalises to the ratings of imagery vividness in the auditory modality, a new version of the Vividness of Visual Imagery Questionnaire (Marks, 1995), and reports of visual thinking style. Samples of undergraduate psychology students were tested. The measure of social desirable responding used was the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002). Continuous and dichotomous scoring methods of the BIDR were also compared. Correlational analyses found the following: 1) The VVIQ-2 correlated with the egoistic bias as its predecessor had done in Allbutt *et al.*'s previous work; 2) Vividness of auditory imagery and visual thinking style were not significantly related to BIDR scores; and 3) Both BIDR scoring methods produced similar patterns of results. The results are interpreted in terms of our assertion that properties of imagery that are perceived to be associated with imagery as a skill and / or are more valued by respondents are more likely to correlate with measures of social desirable responding.

Introduction

Mental images are quasi-perceptual mental representations. In the visual modality mental imagery would refer to the experience of mentally visualising the appearance of something, usually without it being present. In the auditory modality it would refer to the experience of 'hearing' the sound made by the object in your mind under the same circumstances. A commonly used approach to assessing individual differences in imagery ability has been the use of subjective self-report imagery tests. In these tests participants are asked to introspect on their imagery experience and comment on selected aspects such as the vividness of their imagery, the ability to control their images or the extent to which they use imagery in certain specified situations. Since the first use of such tests by Francis Galton (1883), numerous self-report tests have been developed with almost all of them focusing on visual aspects of mental imagery. Self-report imagery questionnaires have the advantage of being quick and easy to use and for some aspects of imagery experience, such as image vividness, appear to be the only measurement option available. However, self-report visual imagery questionnaires typically only show only weak correlations with objective behavioural spatial tests (e.g. Poltrock & Brown, 1984) which are assumed to also involve mental imagery. One factor causing this may be differences in the nature of the items to be imaged, as Dean and Morris (1991) found correlations between both types of tests as high as +0.51 when the items to be imaged on the self-report test were of the kind used in spatial tests. Another factor may be that different kinds of imagery are involved in the two tests with self-report questionnaires involving visual representations while spatial tests involve more amodal spatial representations. This is plausible as psychophysiological research has shown brain activity associated with visual imagery tasks such as visual image generation, which is greater in amplitude for high scorers on self-report visual imagery questionnaires, to peak over left occipital areas (e.g. Allbutt, 2000; Farah & Peronnet, 1989). In contrast, brain activity associated with typical spatial task manipulations such as mental rotation peak over right parietal areas (e.g. Rösler, Schumacher & Sojka, 1990; Peronnet & Farah, 1989).

Social Desirable Responding: Distortion and error or valid responding?

A concern over the use of self-report imagery questionnaires has been the extent to which responses on them might be 'contaminated' by social desirable responding. Paulhus (2002) defines social desirable responding as "…the tendency to give overly positive self-descriptions" (p50). Traditionally social desirable responding has been viewed as distortion or error that should be eliminated or minimised. This does not necessarily mean a person scoring high on a social desirability questionnaire is lying, they may simply have a distorted view of themselves, or a high need to be valued and accepted by others. But whatever the cause the end result is the same, a departure from reality. However, more recently this view point has come to be questioned. Many psychologists now believe that social desirable responding is better viewed as a personality trait in its own right that is related to other positive traits such as psychological adjustment or conscientiousness, and so should be treated as substantive or valid responding. Findings by Piedmont, McCrae, Riemann and Angleitner (2000) and McCrae and Costa (1983) that that controlling social desirable responding usually had either no effect on correlations between self assessments on the 'big five' factors of personality and external criterions (parallel spouse or peer ratings) or actually reduced the size of these correlations, have usually been viewed as supporting this position.

Social Desirable Responding and self-report imagery questionnaires

Within the imagery literature social desirable responding has been viewed as distortion or error, and research into it has taken two forms. The first kind of research has either experimentally manipulated the social pressure in a situation or the perceived desirability of imagery ability and shown that such factors can have an influence on imagery test responses (e.g. Ashton & White, 1975). The second kind has correlated responses on imagery tests with responses on psychometric measures of social desirable responding to assess the size of the relationship between these variables. McKelvie (1995a) drew on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by other authors, to suggest that correlations in excess of 0.25 in this context would indicate unacceptable 'contamination' of imagery questionnaires by social desirable responding.

McKelvie's (1995b) meta-analysis is an important starting point for attempting to gauge the size of the relationship between self-report imagery scales and social desirable responding. This is because he attempted to estimate the size of correlation between probably the most widely used imagery scale, the VVIQ-1 (Vividness of Visual Imagery Questionnaire version 1, Marks 1973), and most widely used test of social desirable responding the Marlowe-Crowne test (MC; Crowne & Marlowe, 1960). McKelvie (1995b) made two estimates of the size of this relationship. The first estimate was based on correlations with the MC test only. Data from ten studies lead to a reported r of 0.189, with more vivid imagery being associated with greater social desirable responding. The second estimate added a further six correlations to McKelvie's data pool from three other tests of social desirable responding used in a study by Anderson and Buyer (1994). Inclusion of this extra data produced a revised r of 0.075, indicating a much weaker relationship with social desirable responding. Working from a view of social desirable responding as distortion or error McKelvie (1995b) concluded that "…the VVIQ is not seriously contaminated by social desirable responding, and is perhaps not contaminated at all" (p. 209). However, in previous articles, (Allbutt, Ling & Shafiullah, 1996a; Allbutt, Shafiullah, & Ling, 1996b), we have argued that McKelvie's estimate of the size of the relationship between the VVIQ-1 and social desirable responding may be an underestimate because previous research is limited in that it: 1) theorises social desirable responding as a unitary construct, and 2) because it relies predominantly on the MC test as its measure of social desirable responding which may not be the most appropriate measure for this particular context.

Forms of Social Desirable Responding

Several researchers, prompted by observations of low intercorrelations between different measures, have argued that social desirable responding may occur in more than one form (e.g. Edwards, Diers &

Walker, 1962; Wiggins, 1959). Work by Paulhus into the types of social desirable responding has been particularly influential, not least because it has led to the development of a questionnaire to measure different forms of social desirable responding called the Balanced Inventory of Desirable Responding Version 8 (BIDR, Paulhus, 2002). In his latest theory, Paulhus (2002) has proposed a two-tier model of social desirable responding which has a primary distinction in terms of 'personality content' (Paulhus, 2002, p63), and a secondary distinction in terms of level of consciousness (see Figure 1).
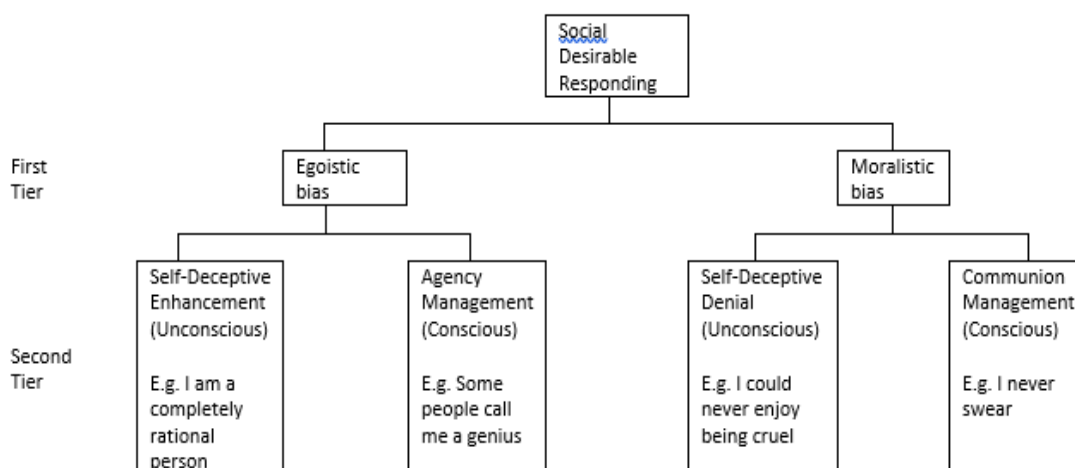

Figure 1. Paulhus's Two-Tier Model (modified from Paulhus, 2002).

In the first tier there is a separation of social desirable responding into egoistic and moralistic biases. The egoistic bias refers to the tendency to claim positive social and intellectual qualities, while the moralistic bias refers to the claiming of positive moral qualities and the refutation of negative socially-deviant qualities. In the second tier, both the egoistic and moralistic biases are divided into conscious and unconscious forms. The egoistic bias separates into self-deceptive enhancement (SDE) in its unconscious form and agency management (AM) in its conscious form. The moralistic bias separates into self-deceptive denial (SDD) in its unconscious form and communion management (CM) in its conscious form. These four types of social desirable responding map straightforwardly onto the four scales of the BIDR which bear the same names, except for CM which is referred to as the impression management scale (IM) rather than the CM scale. Paulhus and his colleagues have offered support for this model from studies that have varied the kind of 'fake good' instructions given to participants to elicit social desirable responding (Paulhus & Notareschi,1993), and studies that have factored residual scores derived from differences between self and other estimates of personality and intelligence together with scores from measures of social desirable responding (John & Paulhus, 2000; Paulhus & John, 1998).

<div style="text-align:center">Problems with using the Marlowe-Crowne Test</div>

Both of McKelvie's (1995b) estimates of the degree of relationship between the VVIQ-1 and social desirable responding drew either exclusively or predominately on data from the MC scale. The MC test assumes that social desirable responding is a unitary construct. Further, it would seem likely that respondents answering self-report imagery questionnaires will view imagery as an ability being tested and will value responses they perceive to indicate 'better' imagery ability. Given this, the egoistic bias with its emphasis on claiming positive social and intellectual qualities, would seem to be more likely to correlate with self-report imagery questionnaires than the moralistic bias. However, Paulhus (1984) found the MC test to load 0.68 on a factor labelled as 'impression management' (corresponding to CM in Paulhus's 2002 model) and 0.40 on a factor labelled as 'self-deception' (corresponding to SDE in Paulhus's 2002 model). Thus the MC test would seem to be more a measure of moralistic bias than egoistic bias. As such the estimates made by McKelvie may have

underestimated the size of the relationship between the VVIQ-1 and social desirable responding.

<div align="center">Our previous research</div>

To date we have conducted a series of studies to investigate the relationship between a variety of self-report imagery tests and social desirable responding (Allbutt *et al.*, 2006b; Allbutt, *et al.*, 2006a), the results of which support our critique of the previous literature and its reliance on data from the MC test. In our research we have used the SDE and IM scales of the BIDR as our measures of social desirable responding. The SDE and IM scales were selected because they are the most well established of the BIDR scales; the scales include one measure of egoistic bias and one measure of moralistic bias. Only these two scales of the BIDR were used for several reasons. First, in order to keep the demands placed on participants to a manageable level. Second, and as noted by Paulhus (1999) himself, there are several features of the SDD scale that make its use at the present time unappealing: some of the SDD scale items are considered offensive by ethical review boards, the scale has been found to correlate highly with IM under many conditions and its additional 20 items can make the BIDR overly long. Third, the newest of the BIDR scales, AM, is a relatively untested scale and has as yet no published data on its psychometric properties. The results of both of our studies found the correlations with self-report imagery tests and SDE to be higher than those with IM, and 'greater' imagery ability was always associated with higher social desirable responding. Allbutt *et al.* (2006b) used two imagery tests which attempt to measure the vividness of either visual imagery or movement imagery: the VVIQ-1 and the Vividness of Movement Imagery Questionnaire (VMIQ; Isaac, Marks & Russell, 1986). The correlations between SDE and the imagery scales were found to be: VVIQ -0.25; VMIQ -0.31. Allbutt *et al.*, (2006a) used a further imagery test: The Shapes Questionnaire (Dean & Morris, 1991; 2003). This questionnaire aims to assess a wide range of self-reported visual imagery properties including vividness, but also assesses 17 other aspects of the imagery experience such as the ease of evoking an image and the amount of detail in an image. This imagery test was used as it allowed us to investigate whether the pattern of results found by the first article would generalise beyond ratings of imagery vividness. A factor analysis of the questionnaire's 16 continuously scored items found one global imagery factor that accounted for 43.94% of variance. The correlation between factor scores from this factor and SDE was 0.25. In addition, eight of the 16 items from the questionnaire correlated significantly with SDE, and the size of these correlations correlated positively with the extent to which participants rated the properties as measuring imagery ability. These results suggested that the pattern seen in our first article did indeed generalise to ratings of other aspects of visual imagery. Further, the correlations between the imagery scales and SDE from the two articles suggest a stronger relationship between self-report imagery measures and social desirable responding than suggested by McKelvie's (1995b) two estimates.

Our position on the 'substance vs style' issue is as follows. The results of Piedmont *et al.* (2000) and McCrae and Costa (1983) would at first seem to imply that correlations between self-report imagery questionnaires and SDE reflect a substantive relationship. Following from these attempts to control for SDE might also have either no effect, or reduce correlations between self-report imagery questionnaires and other variables. Indeed, Pauls and Stemmler (2003) have replicated these results using shortened German versions of the SDE and IM BIDR scales, and Paulhus (1994) himself states that under low demand conditions (e.g. anonymity) SDE correlations are almost always substantive.

However, Pauls and Stemmler (2003) have questioned Piedmont *et al.* and McCrae and Costa's conclusions that their results necessarily imply substantive relations. They note that there are both similarities and differences between 'self' and 'other' ratings made on the big five personality traits. They argue that similarities might occur for other reasons than substantive ones. High social desirable responding scorers may believe the distorted view they have of themselves, project the view to others and strive to behave as consistently with it, thus causing others to share this distorted view too to a substantial, but not to a complete degree. This could contribute to why controlling for social desirable responding has no beneficial effect. Pauls and Stemmler also draw on observations of trends in their and other's data, and further analysis not conducted by Piedmont *et al.* and McCrae and Costa, to highlight important differences between self and other ratings. They note a trend in

studies of this type for scores on measures of social desirable responding to correlate more highly with self ratings than other ratings. They also calculated self rating inflation scores by for big five traits by regressing self ratings on other ratings, and then correlated these residual scores with the BIDR scales. Their analysis showed that both SDE and IM were correlated with relevant inflation scores. In light of this, Pauls and Stemmler suggest that SDE and IM scales assess both valid personality traits and distortion, and that at least under anonymous conditions they are good indicators of distortion. However, attempting to control for social desirable responding does not appear to be an effective way to handle bias in the area of personality assessment.

In addition to the points made by Pauls and Stemmler (2003), we would argue that these findings relate to the effects of controlling social desirable responding on correlations between personality questionnaires. It may be the case that even if social desirable responding reflects a substantive effect, it may still be an independent construct from mental imagery and so its control would aid more precise measurement of mental imagery processes. Indeed Helmes (2000) argues that 'Whether variance associated with social desirability is due to a response style or to a meaningful construct, the net result is the same. The variance that is associated with social desirability is almost always different at a conceptual level from the variable(s) in question' (p35-36). Data of some relevance to this issue comes from Allbutt *et al.* (1996b). There we found that partialing out SDE and IM reduced the size of correlations between the VVIQ-1 and a measure of movement imagery (the Vividness of Movement Imagery Questionnaire; Isaac, Marks, & Russell, 1986) by only between 0.01 and 0.06. Further work is needed though, in order to examine whether this pattern of results can be extended to other types of criterion e.g. perceptual and visual memory tests before firm conclusions can be drawn. If the relationship does prove to be substantive and mental imagery is not an independent construct, charting the relationship between types of social desirable responding and types of self-report imagery questionnaires would advance our understanding of factors affecting responses on such questionnaires.

<div align="center">The current paper</div>

In this paper we report the findings of three studies which attempt to extend our previous work by investigating whether the pattern of findings observed in our previous articles will be seen with: 1) a measure of vividness of auditory imagery i.e. a different sensory modality; 2) a new version of the VVIQ which has been claimed to be an advance on its predecessor; 3) a measure of habitual thinking style which has been suggested to be less open to social desirable responding effects (e.g. Richardson, 1977a) than scales based on properties of imagery such as ratings of its vividness; and 4) to investigate whether the method used to score the BIDR scales has an impact on the pattern of results seen. In keeping with our previous two articles and for the same reasons stated previously, the research reported here made use of only the SDE and IM scales of the BIDR.

<div align="center">Study 1: Auditory imagery, visual imagery and social desirable responding</div>

In Study 1 participants answered a measure of vividness of auditory imagery, the Auditory Imagery Scale (AIS; Gissurarson 1992), as well as a new version of the VVIQ called the VVIQ version 2 (VVIQ-2; Marks, 1995). The AIS has never been correlated with a measure of social desirable responding. It is possible that the value participants place on the ability to form vivid images in different modalities might vary. In particular, because of its importance to us as a sensory modality, vision (and as a consequence visual imagery), might be valued above other sensory modalities. If this is true we might expect to observe a larger relationship between social desirable responding and responses to self-report tests of visual imagery than with auditory imagery and as a consequence see higher correlations with SDE for such tests. Marks developed the VVIQ-2 in response to criticisms of the VVIQ-1. Compared to the original VVIQ the VVIQ-2 has 16 additional items so that the scale samples a wider range of imagery ratings, a reversal of its scoring of responses such that greatest vividness of imagery is now scored '5' rather than '1' and all ratings are made only once with eyes closed. We used the VVIQ-2 because we were interested in whether the new version of the VVIQ would correlate with SDE as its predecessor had.

We also investigated the impact of the scoring method used for the BIDR on the pattern of results obtained. There are two ways to score the subscales. The 'continuous method' involves scoring all answers on a continuous scale. The 'dichotomous method' is to score only extreme responses (after negative items are recoded responses of either six or seven on an item are counted as one point while any other response does not score). In our previous investigations into the relationship between self-report imagery scales and social desirable responding we have only used the dichotomous scoring method. Paulhus (1991, 1994) favours this method as it scores only clearly exaggerated or managed responses, and for the SDE he argued that dichotomous scoring 'emphasises the distinction between typical responses and ones indicating extreme claims of confidence. Thus this scoring provides some assurance that over-confidence rather than confidence is being tapped' (Paulhus, 1994, p21). We propose a minor revision to this assessment in suggesting that the dichotomous scoring method most likely reflects over-confidence, while the continuous scoring method most likely reflects a combination of over-confidence, confidence and also under-confidence. Stöber, Dette and Musch (2002) attempted to compare the effectiveness of the two scoring methods using a German version of the BIDR. They argued that their data found support for the continuous method as it showed: 1) higher Cronbach's alphas; 2) higher convergent correlations with other measures of social desirable responding; 3) more consistent effects with self-presentation instructions (fake-good vs fake-bad instructions); and 4) SDE scores showed higher correlations with those traits of the Five-Factor model or personality for which substantial correlations were expected i.e. neuroticism, extraversion and conscientiousness. Further, Stöber *et al*. found correlation coefficients between SDE scored dichotomously and continuously across three studies to be +0.79, +0.69 and +0.67, and between IM scored dichotomously and continuously to be +0.84, +0.82 and +0.79. These patterns of results support the view that the two scoring methods may be assessing different constructs, particularly for the SDE scale. Given Stöber *et al*.'s findings it seems prudent to consider the impact of the BIDR scoring method on our results.

Method

Participants

There were 113 undergraduate psychology students (31 males, 82 females). Ages ranged from ages 18 to 51 years with a mean age of 25.2 years. All participants were volunteers and no inducements were paid to them.

Materials

Ethical rights form. A form was used which notified participants of the rights they possessed under the British Psychological Society's code of ethics.

Auditory Imagery Scale (AIS; Guissurason, 1992). Assesses vividness of auditory imagery using 7 items. Participants form a series of images of the sounds made by a car, telephone, footsteps, water dripping, snapping twigs, people talking, and music. Ratings of image vividness are made on a 4-point scale ranging from 1, (*Very clear sound / noise*), to 4, (*No sound / noise at all*). Lower scores indicate more vivid imagery. Guissurason (1992) reported a Cronbach's alpha of 0.80 and in a factor analysis found all items to load on one common factor. He also found the AIS to correlate 0.48 with the VVIQ and –0.23 with Gordon's Test of Visual Imagery control (GTVIC; Gordon, 1949), indicating that self-reported vivid auditory imagery was positively associated with vivid visual imagery and better control of visual imagery. The moderate size of the first correlation and low size of the second suggests that the AIS is measuring a related but distinct construct to those measured by the VVIQ-1 and GTVIC. The test-retest reliability of the AIS has not yet been tested.

Vividness of Visual Imagery Scale Version 2 (VVIQ-2; Marks, 1995). Assesses vividness of visual imagery using 32 items. Participants form a series of images of a friend or relative's face, the rising sun, a shop, a country scene, being driven in a car, a beach, a railway station, and a garden with lawns. Ratings of image vividness are made on a 5-point scale ranging from 5, (*Perfectly clear and vivid as normal vision*), to 1, (*No image at all, you only 'know' you are thinking of the object*). The ratings are made once with

the eyes closed. In contrast to the way the original VVIQ was scored, the new version of the VVIQ is scored so that higher scores indicate more vivid imagery. No reliability or validity data has been published on the VVIQ-2, although in the current study Cronbach's alpha was found to be 0.93. Also, because the first 16 items of the VVIQ-2 are the same as the 16 items that form the VVIQ-1, results on the VVIQ-1 suggest the likely pattern for the VVIQ-2. McKelvie (1995b) reviewed over 100 articles relevant to the VVIQ-1 and concluded it had good internal consistency, acceptable test-retest reliability and found support for its validity from tests of content and criterion-related validity.

The Balanced Inventory of Desirable Responding Version 8 (BIDR; Paulhus, 2002). Assesses social desirable responding. The two main 20 item subscales, SDE and IM, were used. SDE is a measure of unconscious egoistic bias. The subscale has statements about the respondent's abilities across a range of situations, half of which are positively phrased and half negatively phrased. For example 'My first impressions of people usually turn out to be right' (positive). Responses were made on a seven point scale ranging from 1, (*Not true*), to 7, (*Very true*). IM is a measure of conscious moralistic bias and this subscale is organised in a similar fashion to that of the SDE. It has statements about the respondent's ethical behaviour across a range of situations, half of which are positively phrased and half negatively phrased. For example 'I never cover up my mistakes' (positive). Responses are made on the same 7 point scale as the SDE.

The subscales were scored using the continuous and the dichotomous methods to allow a comparison of their effects. The 'continuous method' is to recode responses to negative items and then simply sum scores on the 7-point scale across items. The 'dichotomous method' is to score only extreme scores. Negative items are recoded and then responses of either six or seven on an item are counted as one point, while responses of one to five do not score. The points are totaled for each subscale separately. For both scoring methods higher scores indicate more SDE or IM.

Additional instructions were added to the start of the BIDR to make two items from the SDE scale and one item from the IM scale relevant to all participants. The items from the SDE scale were: item 8, "I am not a safe driver when I exceed the speed limit", and item 14, "My parents are not always fair when they punish me". The item from the IM scale was: item 33, I sometimes drive faster than the sped limit". Items 8 and 33 are problematic for participants who do not drive, so they were asked to answer the items by imagining what they would be like if they were to drive. Item 14 is problematic for older participants and so the instructions asked participants to remember back to when they lived at home and comment on how they felt at the time. This approach was preferred to the omission of these items as it preserved the balanced nature of the BIDR scales.

Paulhus (1994) reported Cronbach's alphas for SDE scored dichotomously to range from 0.65 to 0.75 and scored continuously to range from 0.70 to 0.82. Test-retest reliability over a five week interval scored dichotomously was 0.69. Cronbach's alphas for IM scored dichotomously range from 0.75 to 0.80 and scored continuously to range from 0.80 to 0.86. Test-retest reliability over a five week interval scored dichotomously was 0.77. Paulhus reports a number of studies that support the validity of the BIDR, including the results of factor analyses and evidence supporting the convergent and discriminant validity of the two scales.

The ethics form and three questionnaires were combined together into a booklet, starting with the ethics form, followed by the AIS, VVIQ-2 and ending with the BIDR. In the BIDR the SDE scale appears before the IM scale.

Procedure

Participants answered the questionnaires in the last hour of a two hour laboratory class. They were asked to seat themselves as far apart as possible. Responses were made in silence and participants were told not to put their name anywhere in the booklet. Participants answered the questionnaires in the order in which they appeared in the booklet. The following week they were given a debriefing sheet which explained the purpose of the study and gave the contact details of the experimenters in

case of queries.

Results

A series of Pearson Product Moment correlations between the AIS, VVIQ-2, and BIDR SDE and IM subscales were carried out to investigate the correlations between the imagery scales and the social desirable responding scales. Prior to analysis the variables were screened for univariate and bivariate outliers. Screening revealed two univariate outliers reducing the sample size to 111. Non-normal distributions were corrected with transformations such that AIS received a logarithmic transformation and dichotomously scores SDE a square root transformation. A prospective power analysis using tables for Pearson's Product Moment correlation showed that for a two tailed test a sample size of 90 would be required to detect a medium effect size (Clark-Carter, 1997). Thus the sample size was sufficient for this purpose. The results of the Pearson's correlational analysis can be seen in Table 1. Continuous scoring of BIDR scales is indicated by the suffix 'c' and dichotomous scoring by the suffix 'd'. Because the AIS and VVIQ-2 differ considerably in their number of items (7 versus 32 items), the correlations between the AIS and BIDR scales were corrected for test length using the Spearman-Brown correction formula taken from Thorndike (1982). The correction left the correlation between AIS and IM-c unaltered, and increased the size of the three other AIS – BIDR scale correlations by 0.01.

Table1.
Matrix for Pearson's correlations between measures for Study 1

	AIS	VVIQ-2	SDE-c	IM-c	SDE-d
AIS					
VVIQ-2	-0.35**				
SDE-c	-0.09	+0.31**			
IM-c	-0.05	+0.11	+0.15		
SDE-d	-0.12	+0.31**	+0.76**	-0.17	
IM-d	-0.09	+0.18	+0.22*	+0.77**	+0.24*

Note   * $p<0.05$ ** $p<0.01$ (2-tailed)

The AIS and VVIQ-2 self-report imagery questionnaires showed a significant low negative correlation with each other indicating that high auditory and visual vividness of imagery were weakly associated.

With respect to the relationship between self-report imagery questionnaire scores and social desirable responding, although correlations with SDE were slightly higher than with IM, neither reached statistical significance whichever scoring method was used. In contrast, the VVIQ-2 showed a significant low positive correlation with SDE that was identical in size for both scoring methods with more vivid imagery being associated with greater SDE. No correlations between the VVIQ-2 and IM were significant. Thus only the visual form of imagery appeared to be associated with social desirable responding, and only with SDE.

The correlations between SDE and IM were low and positive, but only reached significance for the dichotomous scoring method. This suggests that for this scoring method, high levels of both forms of social desirable responding were weakly associated. The correlations between SDE continuous and dichotomous scoring methods, and between IM continuous and dichotomous scoring methods, were both positive, similarly large in size and significant. This shows that the scores obtained by both methods were highly related.

Discussion

The main result from Study 1 was that scores from the visual form of imagery appeared to be associated with social desirable responding and only with the measure of the egoistic bias. The correlation seen between the VVIQ-2 and SDE was 0.31 for both scoring methods and so comparable in size to that seen for the original VVIQ (0.25) in our first study (Allbutt *et al.*, 2006b).

The fact that only vividness of visual imagery correlated significantly with social desirable responding is consistent with the position that visual imagery might be valued to a greater extent by participants than imagery ability in other modalities. The AIS differs from the VVIQ-2 in several respects other than being a measure of auditory imagery in that it has: 1) a 4 point response scale rather than 5 point response scale; 2) a scoring system whereby low scores indicate greater imagery ability rather than high scores; and 3) the AIS has only seven items compared to the VVIQ-2's 32 items. However, none of these differences would seem able to account for the difference in the pattern of results for the two modalities.

Study 2: Ratings of the relative desire to possess vivid mental imagery in auditory and visual sensory modalities

From the results of Study 1 it was apparent that vividness of visual imagery correlated significantly with social desirable responding while vividness of auditory imagery did not, and we speculated that this might be because, while vividness in both modalities might be viewed as skills, vividness in the visual modality was valued to a greater extent by participants than imagery ability in other modalities. However, we had no direct evidence that vividness of visual imagery was valued more greatly than vividness of auditory imagery. Thus, in Study 2 participants were asked to rate their relative desire to possess vivid mental imagery in auditory and visual sensory modalities. They were asked to express this separately for each modality on a short questionnaire using four point ratings scales.

Method

Participants
There were 54 undergraduate psychology students (7 males, 47 females). Ages ranged from ages 19 to 54 years with a mean age of 23.4 years.

Materials
Ethical rights form. As per study 1.

A questionnaire for rating the relative desire to possess vivid mental imagery in auditory and visual sensory modalities. Two short questions of the same format to assess the relative importance a person places on having vivid auditory and visual imagery. For example the statement for the auditory modality reads: 'I would like to have vivid auditory imagery'. Response ratings are made on 4-point scales ranging from 0, (*No, does not matter to me*), to 3, (*Yes, I would like to very much*). Higher scores indicate a greater desire to have vivid imagery in a given modality.

Procedure
The procedure was similar to that used in Study 1 except that the data was collected at the start of a lecture.

Results

Data screening revealed no outliers but the distribution of visual imagery scores was positively skewed so the data was analysed using a Wilcoxon signed ranks test for matching pairs. A prospective power analysis showed a sample size of 42 would be required to detect a medium effect size (Clark-Carter, 1997). The median rating for auditory imagery was 2 with an inter-quartile range of 2, while the median rating for visual imagery was 3 with an inter-quartile range of 0.25. The Wilcoxon test showed that vividness of visual imagery was desired significantly more than vividness

of auditory imagery (Z=-4.72, p<0.0005, 2-tailed, N=54).

Discussion

The results of Study 2 supported the hypothesis that people in general value vividness of visual imagery over vividness of auditory imagery. This adds further support to our proposed explanation for why in Study 1 the measure of vividness of visual imagery correlated with SDE while the measure of vividness of auditory imagery did not.

Study 3: Thinking style and social desirable responding

In Study 3 we sought to explore the relationship between measures of thinking style and social desirable responding. Participants answered two measures of habitual thinking style the Verbilizer-Visualizer Questionnaire (VVQ; Richardson, 1977a) and the Individual Differences Questionnaire – Imagery Habit Scale (IDQ-IHS; Cohen & Saslona, 1990) together with the BIDR, although only the results for the IDQ-IHS are reported here because research has questioned aspects of the reliability and validity of the VVQ (e.g. Antonietti and Giorgetti, 1998).

The IDQ-HIS was developed from a subset of items contained in a previous questionnaire, Paivio's Individual Differences Questionnaire (IDQ; Paivio, 1971), which aimed to assess habitual visual and verbal thinking styles via imagery and verbal sub-scales. However, Paivio and Harshman (1983) carried out a factor analysis on the IDQ and found a pattern repeated across two samples each made up of more than 300 respondents, that the imagery scale contained a mixture of factors. Three of these factors were of a visual nature and were interpreted as habitual use of imagery, use of images to solve problems and vividness of dreams, daydreams and imagination. In an attempt to produce a questionnaire that was a purer measure of visual thinking style Cohen and Saslona (1990) developed the IDQ-HIS from 13 items that loaded on a factor Paivio and Harshman's first imagery factor.

Although the IDQ-HIS has never been correlated with a measure of social desirable responding before, the IDQ has been in two studies. Richardson (1977b) reported correlations between the IDQ visual scale and MC separately for males as –0.05 and for females as –0.06, while Hiscock (1978) reported the correlation between the IDQ and MC for both genders combined as 0.10. While both these studies used the MC test and so are open to the same limitations described previously in our article, their results raise the possibility that measures attempting to assess habitual styles of thinking might correlate to a lesser degree with social desirable responding than measures based on properties of imagery such as ratings of its vividness. The possibility seems further to be supported by the fact that when compared to questionnaires such as the VVIQ, most of the IDQ and IDQ-IHS items appear less obviously connected to imagery as a skill e.g. 'My thinking often consists of mental pictures or images', although some items of both IDQ questionnaires could be interpreted as assessing skill aspects of imagery e.g. 'When someone describes something that happens to them, I sometimes find myself vividly imagining the events that happened'. Three items of the IDQ-HIS (items 2, 5, and 8) refer to the ease generating an image and two items refer to the vividness of imagery processes (items 3 and 6) and so might be interpreted by respondents as assessing a skill aspect of imagery, while the remaining eight items relate to the frequency of use of imagery in different situations and so do not.

Thus this study aimed to investigate the relationship between the IDQ-HIS and the scales of the BIDR, and the study also investigated the impact of BIDR scoring method on the pattern of results seen as in Study 1.

Method

Participants

There were 102 undergraduate first year psychology students (79 female, 23 male). Ages ranged from 17 to 74 with a mean age of 26.7 years. Participants were volunteers and no inducements were paid to them.

Materials
Ethical rights form. As per study 1.

Individual Differences Questionnaire (IDQ; Cohen & Saslona, 1990). Assesses the extent to which people prefer a visual style of thinking using 13 items. Responses are made on a 5-point response scale ranging from 1 (strongly agree) to 5 (strongly disagree). Seven items are 'positively' phrased and six 'negatively' phrased. An example item is 'Listening to someone recount their experiences does not usually arouse mental pictures of the incidents being described' (negative). Higher scores indicate a greater preference for a visual style of thinking. Alstedt (1988, cited in Cohen & Saslona, 1990) reported a Cronbach's alpha of 0.85 and a test-retest reliability coefficient after a 6 to 8 week interval of 0.76. Cohen and Saslona found higher scores on the scale associated with better performance on an intentional memory test for pictured objects and on an incidental memory test for their colours.

The BIDR. SDE and IM scales as per Study 1.

The ethics form and three questionnaires were combined together into a booklet. The booklet started with the ethics form, followed by the VVQ, IDQ-IHS and ended with the BIDR.

Procedure
The procedure was identical to that used in Study 1.

<div align="center">Results</div>

A series of Pearson Product Moment correlations between the IDQ-IHS and BIDR (SDE and IM subscales scored using continuous and dichotomous methods) were carried out to investigate the correlations between the imagery scales and the social desirable responding scales. The procedures followed were identical to those used in Study 1. Data screening revealed 12 outliers reducing the sample size to 90. Non-normal distributions were corrected with transformations such that the IDQ-IHS received a logarithmic transformation and both BIDR subscales a square root transformation. As per Study 1 a prospective power analysis showed a sample size of 90 would be required to detect a medium effect size (Clark-Carter, 1997). The results of the Pearson's correlational analysis can be seen in Table 2. Continuous scoring of BIDR scales is indicated by the suffix 'c' and dichotomous scoring by the suffix 'd'.

Table 2.
Matrix for Pearson's correlations between measures for Study 3

	IDQ-IHS	SDE-c	IM-c	SDE-d
IDQ-IHS				
SDE-c	+0.15			
IM-c	+0.04	+0.03		
SDE-d	+0.13	+0.60**	+0.07	
IM-d	+0.03	+0.02*	+0.99**	+0.01

Note　*　$p<0.05$　　**　$p<0.01$　　(2-tailed)

The IDQ-IHS did not correlate significantly with SDE, but its correlations with SDE were positive in direction, similar in size and approached significance for both scoring methods while those with IM were much nearer zero. None of the intercorrelations between SDE and IM for either scoring method reached significance. The correlations between SDE continuous and dichotomous scoring methods, and between IM continuous and dichotomous scoring methods, were both positive and large in size, however, the degree of association was greater for IM than for SDE. This shows that the scores obtained by both methods were highly related but more so for the IM scale in this sample.

Discussion

The IDQ-IHS showed no significant correlations with SDE, however, its correlations with SDE approached significance for both methods of scoring while its correlations with IM were much more clearly near zero. It is possible that with a larger sample size and the accompanying increase in statistical power, a significant relationship with SDE might have been detected. Therefore, it seems that the IDQ-HIS, and possibly other questionnaires that attempt to measure thinking style, correlate to a lesser degree with social desirable responding than tests such as the VVIQ, and that this is probably because fewer of the IDQ-HIS items relate to imagery as a skill.

General discussion

The correlations between imagery self-report questionnaires in Study 1 and Study 3 show a similar pattern to that seen before in that the strongest correlations are always with SDE rather than IM and that the correlations with SDE are always positive in direction, that is, higher SDE was associated with 'better' imagery. However, while the VVIQ-2 correlated significantly with SDE as its predecessor had done in our previous work, vividness of auditory imagery and visual thinking style were not significantly related to BIDR scores. The patterns of results were consistent across the dichotomous and continuous scoring methods of the BIDR scales, and as such would appear to be due to the over-confidence elements of SDE which is likely to be tapped by both scoring methods. Taken overall, the results suggest that properties of imagery that are perceived to be associated with imagery as a skill and / or are more valued by respondents are more likely to correlate with egoistic measures of social desirable responding. The results also support our assertion that the BIDR is a superior measure of social desirable responding to the MC test because of its ability to treat social desirable responding as a multifactorial phenomenon.

In our introduction, we noted that there is debate over whether correlations with social desirable responding should best be viewed as distortion and error or valid responding. If social desirable responding is best viewed as distortion or best viewed as different on a conceptual level then McKelvie's (1995a) suggestion that correlations with measures of social desirable responding in excess of 0.25 (6.25% overlap of variance) would indicate an unacceptable degree of contamination becomes relevant. Certainly the correlations from Study 1 involving the VVIQ-2 exceed this value. However, we would suggest that, like all cut-off points, the exact value chosen is somewhat arbitrary, and even assuming social desirable responding reflects distortion, judgements of the validity of imagery questionnaires would need to be made in light of data on other aspects of validity as well.

But if social desirable responding is best viewed as valid responding in its own right the crucial question becomes less how to control the influence of SDE, and more to why SDE should correlate with the VVIQ-2 and the other scales from our previous articles? One possible answer relates to Ahsen's work on the unvividness paradox (e.g. Ahsen, 1985, 1988, 1990; Hochman, 1994) which shows the existence of nonconscious self-inhibitory forces which act to reduce the vividness of images. For example, Ahsen (1990) asked participants to keep either their mother or father in mind while answering the items of the VVIQ-1 and found that reported imagery vividness was less when fathers were kept in mind. Rinaldo and Okada (1993) suggest that this phenomenon might be viewed as a form of self-deception as participants may be erroneously making estimates of low vividness, presumably because of the influence of emotionally charged associations with their fathers. This interpretation is given support by the fact that Paulhus (1994) reports the highest correlates of the SDE scale to be trait anxiety (-0.52) and self esteem (+0.50) and Davies, French and Keogh (1998) report a correlation between the SDE and Eysenck's neuroticism scale from the Eysenck Personality Questionnaire-Revised (Eysenck & Eysenck, 1991) to be -0.53.

With respect to limitations of our study it should be noted that interpretation of the meaning of the difference in pattern of correlations between SDE and IM scales is not straightforward as the scales differ not only in terms of egoistic versus moralistic biases but also conscious versus unconscious processes. However, we view the difference in type of bias to be most likely the key variable underlying the differences in results we have obtained.

Possible directions for future work include the following. Testing the assertion that social desirable responding is an independent construct from mental imagery processes and so its control would aid the assessment of self-reported imagery should be a priority for future work. McKelvie's (1995b) meta-analytical review reported relationships between the VVIQ-1 and other self-report imagery questionnaires, perceptual measures and memory measures that were argued to provide evidence to support the construct validity of the VVIQ. Repeating some of these studies and observing the effect that controlling social desirable responding has would provide valuable data to help resolve this issue. Also collecting data on the BIDR and MC test in the same study would allow a more direct comparison of findings between the two tests. The measure of over-claiming developed by Paulhus, Harms, Bruce and Lysy (2003), which is a concrete operalisation of self-enhancement, could be used in an attempt to replicate our results using a measure of the egoistic bias other than the BIDR. Given that the strongest relationship observed in the current study was for egoistic bias, when details of the psychometric properties of the AM are published it would be desirable to assess both SDE and AM at the same time. It is possible that AM correlates with self-report imagery questionnaires independently from SDE and that the relationship with social desirable responding could be greater than the estimate from this study which is based on SDE alone.

References

Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery, 9*(3), 1-18.

Ahsen, A. (1988). Imagery, unvividness paradox and the paradigm of control. *Journal of Mental Imagery*, *12*(3&4), 1-44.

Ahsen, A. (1990). AA-VVIQ and imagery paradigm: Vividness and unvividness issue in VVIQ research programmes. *Journal of Mental Imagery, 14*(3&4), 1-58.

Allbutt, J. (2000). A study of visual image generation using a combined ERP and introspective questionnaire approach. *Journal of Psychophysiology, 14*, 3, 184.

Allbutt, J., Ling, J. and Shafiullah, M. (2006a). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery. *Imagination, Cognition and Personality, 25*, 4, 337-353.

Allbutt, J., Shafiullah, M. & Ling, J. (2006b). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, *30,* 1&2, 1-20.

Anderson, D.C. & Buyer, L.S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *American Journal of Psychology, 107*(2), 207-222.

Antonietti, A. & Giorgetti, M. (1998). The Verbalizer-Visualizer Questionnaire: A review. *Perceptual and Motor skills, 86,* 227-239.

Ashton, R. & White, K. (1975). The effects of instructions on subjects' imagery questionnaire scores. *Social Behaviour and Personality, 3*(1), 41-43.

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove: Psychology Press.

Cohen, B. H., & Saslona, M. (1990). The advantage of being an habitual visualiser. *Journal of Mental Imagery*, *14*(3-4), 101-112.

Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Davies, M. F., French, C. C. and Keogh, E. (1998). Self-deceptive enhancement and impression management correlates of EPQ-R dimensions. Journal of Psychology, 132, 4, 401-406.

Dean, G. & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology, 94*, 245-273.

Dean, G. & Morris, P. E. (1991). Imagery and spatial ability: When introspective reports predict performance. In R. H. Logie and M. Denis (Eds.), *Mental images in human cognition* (pp331-347). Amsterdam: Elsevier Science.

Edwards, A.L., Diers, C.J. & Walker, J.N. (1962). Response sets and factors loadings on sixty-one personality scales. *Journal of Applied Psychology, 46*, 220-225.

Eysenck, H. J. & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult).* London: Hodder & Stoughton.

Farah, M. J. & Peronnet, F. (1989). Event-related potentials in the study of mental imagery. *Journal of Psychophysiology, 3*, 2, 99-109.

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

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology, 39*, 156-167.

Guissurason, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery, 16*(3&4), 117-122.

Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R.D. Goffin and E. Helmes (Eds.), *Problems and solutions in human assessment: Honouring Douglas N. Jackson at seventy* (pp21-40). Boston: Kluwer.

Hiscock, M. (1978). Imagery assessment through self-report: What do imagery questionnaires measure? *Journal of Consulting and Clinical Psychology, 46*, 2, 223-230.

Hochman, J. (1994). Ahsen's image psychology. *Journal of Mental Imagery, 18*(3&4) 1-118.

Isaac, A., Marks, D.F., & Russell, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery, 10*(4), 23-30.

John, O.P. & Paulhus, D.L. (2000). *The structure of self enhancement*. Unpublished manuscript, University of California, Berkley.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*(3&4), 153-166.

McCrae, R. R. & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting Clinical Psychology, 51*, 6, 882-888.

McKelvie, S. J. (1995a). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 1-106.

McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 197-251.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In Robinson, J. P., Shaver, P. R. and Wrightsman, L. S. (Eds.), *Measures of Personality and Social Psychological Attitudes* (pp17-59). San Diego: Academic Press.

Paulhus, D. L. (1994). Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6. Unpublished Manual, University of British Columbia, Vancouver, Canada.

Paulhus, D. L. (1999). BIDR Self-Deceptive Denial (SDD). neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm. Retrieved from the World Wide Web on 1 August 2003.

Paulhus, D. L. (2002). Social Desirable Responding: The evolution of a construct. In Braun, H. I., Jackson, D. N. & Wiley, D. E. (Eds.), *The role of constructs in psychological and educational measurement* (pp49-69). Mahwah: Lawrence Erlbaum Associates.

Paulhus, D.L., Harms, P.D., Bruce, M.N. and Lysy, D.C. (2003). The over claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 4, 890-904.

Paulhus, D.L. & John, O.P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1024-1060.

Paulhus, D.L. & Notareschi, R.F. (1993). *Varieties of faking manipulations*. Unpublished manuscript, University of British Columbia, Vancouver, Canada.

Pauls, C. A. & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*, 2, 263-275.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rhinehart & Winston.

Paivio, A. and Harshman, R. (1983). Factor analysis of a questionnaire on imagery and verbal habits and skills. *Canadian Journal of Psychology, 37*, 4, 461-483.

Peronnet, F. & Farah, M. J. (1989). Mental rotation: An event-related potential study with a validated mental rotation task. *Brain and Cognition, 9*, 2, 279-288.

Piedmont, R. L., McCrae, R. R., Riemann, R. & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78,* 3, 582-593.

Poltrock, S. E. & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence, 8*, 2, 93-138.

Richardson, A. (1977a). Verbaliser – Visualiser: A cognitive style dimension. *Journal of Mental Imagery*, *1,* 109-125.

Richardson, A. (1977b). The meaning and measurement of memory imagery. *British Journal of Psychology, 68*, 1, 29-43.

Rinaldo, R. & Okada, R. (1993). A test of the relationship between imagery vividness and social desirability across subject selection procedures. *Journal of Mental Imagery*, *17,* 3&4, 171-180.

Rösler, F., Schumacher, G. & Sojka, B. (1990). What the brain reveals when it thinks: Event-related potentials during mental rotation and mental arithmetic. *The German Journal of Psychology, 14*, 3, 185-203.

Sheehan, P.W. (1967). A shortened version of Bett's Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology, 23*, 386-389.

Stöber, J., Dette, D. E. & Musch, T. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78*(2), 370-389.

Thorndike, R. L. (1982). *Applied Psychometrics*. Boston: Houghtn Mifflin.

Vandenberg, S. G. & Kuse. A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualisation. *Perceptual & Motor Skills*, *47,* 599-604.

Wiggins, J.S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419-427.

Evaluation of paper

Strengths

Paper 3 produced four main findings:

1) The correlations between self-report imagery questionnaires in Studies 1 and 2 show a similar pattern to that seen before in papers 1 and 2 in that the strongest correlations are always with SDE rather than IM and that the correlations with SDE are always positive in direction, that is, higher SDE was associated with 'better' imagery.

2) Of the correlations with SDE, the correlation with the VVIQ-2 was largest and significant, the correlation with the IDQ-HIS was next largest and approached significance and the correlation with the AIS was the smallest and not significant. The VVIQ-2 showed a similar size correlation with SDE as had the VVQ-1 in paper 1 despite its innovations such as the addition of four new items. Claims that questionnaires measuring the habitual use of imagery as a thinking style are less likely to correlate with social desirable responding were supported by the data.

3) The patterns of results were consistent across the dichotomous and continuous scoring methods of the BIDR scales, and as such would appear to be due to the over-confidence elements of SDE which is likely to be tapped by both scoring methods. It was concluded that the method of scoring used for the BIDR scales had little effect on the results and so neither scoring method appeared superior to the other.

4) The participants' ratings of desire to possess the three aspects of imagery suggests that imagery in the visual modality is valued above the auditory modality, and that within the visual modality participants value vivid imagery more than a visual thinking style. Overall, imagery in the visual modality appeared to be valued more than imagery in the auditory modality because participants believed it would enhance remembering. Within the visual modality vivid imagery appeared to be valued more than a thinking style because more participants believed it would enhance remembering, and because some participants believed that vividness was just more 'important' or 'useful' than a thinking style without having a more specific reason for why this was so. The relative desire to possess the three kinds of imagery mapped directly onto the size of their correlations with SDE. As with the findings of paper 2, this result is more consistent with the self-report imagery

questionnaire and SDE correlation reflecting distortion rather than a substantive relationship.

Paper 3 was the first to explore the relationship between vividness of auditory imagery and social desirable responding. It was also the first to explore participants' relative desire to possess different forms of imagery. The collection of qualitative data alongside the ratings, gave insights into why participants held these different values.

When these results are taken together with those from paper 2, they suggest that the pattern of correlations between the self-report imagery questionnaires and SDE was seen because properties of imagery that are perceived to be associated with imagery as an ability, or are more desired and valued by respondents, are more likely to correlate with egoistic measures of social desirable responding. Moreover, the fact that the sizes of the correlation of the self-report imagery scales were in line with ratings of imagery ability and value, provided evidence to suggest that self-report imagery questionnaire–SDE correlations are likely to reflect distortion rather than substance.

Research on the 'hierarchy of the senses' i.e. the importance of the five human senses for human experience by philosophers, linguists and biologists, is in agreement with our finding that visual imagery was valued more highly by our student participants than auditory imagery. The Greek philosopher Aristotle (cited in Jutte, 2005) argued for an ordering of vision, hearing, smell, taste and touch, with vision being the most important. More recently linguists, for example San Roque et al (2015), have tested the hierarchy by exploring the nature of sensory related words across different world languages. They have studied features such as the number of linguistic terms present in languages to express experience in each sensory modality and the universality of such terms. Their research suggests that vision is dominant in all languages and that there is a strong trend for hearing to be placed second, but that the position of the remaining three senses vary considerably across languages.

Paper 3 also explored the sensitivity of the two methods of scoring the BIDR scale within an imagery questionnaire context. There are two ways to score the subscales: continuous scoring and dichotomous scoring, and opinion is divided on which method is preferable. Paulhus (1991, 1994) prefers the dichotomous method for the BIDR scales because it has the

advantage of only scoring clearly exaggerated or managed responses, thus it involves not just an endorsement of desirable features and rejection of undesirable features, but rather a departure from reality. Most use of the BIDR follows Paulhus's preference, however opinion on its use is split. For example, Stöber *et al.* and Cervellione *et al.* have both criticized the dichotomous scoring system for treating social desirable responding as an 'all or nothing' phenomena rather than one that occurs in degrees and argued that the dichotomization of the measurement scale leads to a loss of information which may add to the error in measurement. Further, they and other authors have found empirical support for the continuous scoring method for example, superior psychometric properties. However, in paper 3 both scoring methods produced similar results, so neither scoring method appeared superior to the other.

Limitations

The third paper had one of the same limitations as papers 1 and 2 i.e. only two of Paulhus's four BIDR scales had been used in the study, the SDE and IM. The limitations of this approach have been previously discussed.

The data in paper 3 was collected across three samples. As with paper 2, it would have been preferable if all of the data had been collected from the same participants.

Paper 4: Allbutt, Ling, Rowley and Shafiullah (2011)

Reference: Allbutt, J., Ling, J., Rowley, M. & Shafiullah, M. (2011) Vividness of visual imagery and social desirable responding: Correlations of the Vividness of Visual imagery Questionnaire with the Balanced Inventory of Desirable Responding and the Marlowe-Crowne. Behavior Research Methods, 43, 3, 791-799. https://doi.org/10.3758/s13428-011-0086-8

Aims of the paper

Papers 2 and 3 had investigated the SDE correlation with self-reported imagery questionnaires in different modalities and with ratings of imagery properties other than vividness, and collected evidence that might help unpick the underlying nature of the imagery scale–SDE correlations. Paper 4 returned to the most widely used imagery questionnaire, the VVIQ-1, and ratings of visual imagery vividness to address some important issues that had remained outstanding from papers 1-3, such as whether our critique of the use of the MC scale by previous authors was correct.

The primary aim of paper 4 was to make a direct comparison of how the VVIQ-1 correlated with the four scales of the BIDR and the MC scale in a single sample of participants. The previous three papers had argued that McKelvie (1995b) had underestimated the degree of relationship between the VVIQ-1 and social desirable responding because of a reliance on data from the MC scale, which appears primarily be a measure of moralistic bias. However, the validity of this argument had not been tested directly using the MC scale. Also, the previous three papers had only used the SDE and IM scales of the BIDR. They had been chosen because they were the most well established of Paulhus's scales. However, according to Paulhus's (2002) model, SDE and IM differ not only in terms of the egoistic and moralistic biases, but also in their suitability to measure unconscious and conscious levels of social desirable responding. As such, it was not possible to say for sure that the reason why the self-report imagery questionnaires correlated with SDE, but not with IM, was because SDE is a measure of the egoistic bias and IM a measure of the moralistic bias.

Other aims focused on methodology and were to investigate whether the VVIQ-1 correlated

significantly and independently with both BIDR egoistic scales, SDE and AM. Also, to investigate further the effect of the BIDR scoring system (continuous versus dichotomous) on the pattern of results obtained. As previously discussed, Paulhus and Trapnell (2008) state that the AM items are extreme enough that they have low endorsement rates in private test conditions so that this leaves 'room' for the scale to sample consciously enhanced responses that might be made by participants under public conditions. In paper 1 we described how D. Paulhus (personal communication, October 23^{rd,} 2003) had informed us that in honest response conditions the correlation between SDE and AM could be as low e.g. r=.39. As such, it was possible that AM could correlate with self-report imagery questionnaires independently from SDE, and that the relationship with socially desirable responding could be greater than estimates based on SDE alone. Additionally, although testing in papers 1, 2 and 3 took place in relatively low demand anonymous conditions, it is still possible that situational factors, such as the presence of an experimenter and group testing, might create sufficient pressure to lead to more conscious forms of the egoistic bias which might be better picked up by the AM scale than the SDE scale.

Vividness of Visual Imagery and Social Desirable Responding: Correlations of the Vividness of Visual Imagery Questionnaire with the Balanced Inventory of Desirable Responding and the Marlowe-Crowne Scale

Abstract

Correlational research investigating the relationship between scores on self-report imagery questionnaires and measures of social desirable responding has shown only a weak association. However, researchers have argued that this research may have underestimated the size of the relationship as it relied primarily on the Marlowe-Crowne scale (MC; Crowne & Marlowe, 1960), which loads primarily on the least relevant form of social desirable responding for this particular context, the moralistic bias. Here we report the analysis of data correlating the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973), with the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002) and the MC scales under anonymous testing conditions. The VVIQ correlated significantly with the self-deceptive enhancement (SDE) and Agency Management (AM) BIDR subscales and the MC. The largest correlation was with SDE. The ability of SDE to predict VVIQ scores was not significantly enhanced by adding either AM or MC. Correlations between the VVIQ and BIDR egoistic scales were larger when the BIDR was continuously rather than dichotomously scored. This analysis indicates the relationship between self-reported imagery and social desirable responding is likely to be stronger than previously thought.

Introduction

Mental images are quasi-perceptual mental representations that can occur in all sensory modalities. Visual imagery has received the most attention from psychologists and refers to the experience of mentally visualising the appearance of something usually without it being present. Individual differences in imagery experience, such as its vividness, are often reported by individuals. Psychologists have commonly assessed these differences using subjective self-report imagery scales which request participants to form mental images and then rate some aspect of their imagery experience. Self-report imagery questionnaires have appeal because they are quick and easy to use and, for some aspects of imagery experience, such as image vividness, appear to be the only measurement option available.

An important concern over the results obtained from self-report imagery questionnaires has been the extent to which they might be influenced by social desirable responding. Paulhus (2002, p.50) defines social desirable responding as "…the tendency to give overly positive self-descriptions". There has been much debate over how best to conceptualise social desirable responding. Domino and Domino (2006) state that there are differing opinions over whether social desirable responding should be thought of as distortion or error that should be eliminated or minimised, or whether it is best thought of as a personality trait which is related to other positive traits such as psychological adjustment or conscientiousness, and should be treated as substantive or valid responding. Leite and Cooper (2010) argue that there is evidence (Li & Bagger, 2006; Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist & Verkaslo, 2007; & Ones, Visweswaran, & Reiss, 1996) that whether a participant's response to a focal scale item is affected by social desirable responding or not depends on a three way interaction between the nature of the items making up the focal scale, the respondent's characteristics and the situation that the testing takes place in.

Within the mental imagery literature social desirable responding has been conceptualised as distortion / error. McKelvie (1995a) has drawn on considerations such as factors that affect reliability and validity, to suggest that correlations in excess of .25 between imagery questionnaires and

measures of social desirable responding would indicate unacceptable 'contamination'. In a meta-analytic review paper McKelvie's (1995b) estimated the size of correlation between the most widely used imagery scale, the VVIQ (Vividness of Visual Imagery Questionnaire version 1, Marks 1973), and social desirable responding. McKelvie made two estimates of the size of this relationship. The analysis which produced the largest estimate drew only on data from the most widely used test of social desirable responding, the Marlowe-Crowne scale (MC; Crowne & Marlowe, 1960). The analysis estimated the size of this relationship to have a mean r of .189 with 95% confidence intervals of .078 and .296 respectively based on data from 10 studies. McKelvie concluded that the VVIQ was not seriously affected by social desirable responding, if affected at all. However, in previous work (e.g. Allbutt, Ling, Shafiullah & Heffenan, 2008; Allbutt, Shafiullah, & Ling, 2006a and Allbutt, Ling & Shafiullah, 2006b), we have argued that McKelvie's estimate of the size of the relationship between the VVIQ and social desirable responding may have been limited because previous research: 1) theorised social desirable responding to be a unitary construct, and 2) has predominantly used the MC scale as its measure of social desirable responding which may not be the most appropriate measure to use with self-report imagery tests.

Early research into social desirable responding conceptualised the phenomena as a unitary construct. However, observations of low intercorrelations between different measures and the results of factor analysis led several researchers to argue that social desirable responding could occur in several forms (e.g. Edwards, Diers & Walker, 1962). Work by Paulhus (e.g. 1984) has been particularly influential, and led to the development of a 'two-tier model' of the different forms of social desirable responding and a questionnaire to measure them called the Balanced Inventory of Desirable Responding Version 8 (BIDR, Paulhus, 2002, see Figure 1).
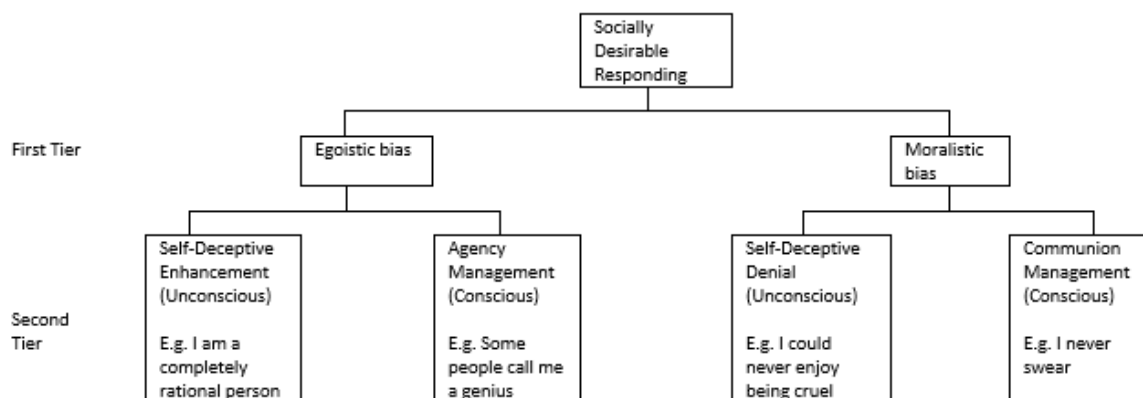
Figure 1. Paulhus's Two-Tier Model (based on Paulhus, 2002).

In the first tier of the model, social desirable responding is divided into egoistic and moralistic biases. Egoistic bias refers to the tendency to claim positive social and intellectual qualities, while moralistic bias is the claiming of positive moral qualities and the refutation of negative socially-deviant qualities. In the second tier, both biases are split into unconscious and conscious forms. The egoistic bias becomes self-deceptive enhancement (SDE) and agency management (AM), while the moralistic bias becomes self-deceptive denial (SDD) and communion management (CM). The four scales of the BIDR map straightforwardly onto the model apart from the fact that in the BIDR CM is referred to by a different term, impression management (IM, which we will use throughout). The nature of the differentiation of unconscious from conscious levels differs between the two kinds of biases. For the egoistic bias, SDE scale items differ from AM items in the degree of exaggeration made about ability claims such that AM claims are very extreme. In contrast, for the moralistic bias, SDD scale items cover more sensitive and emotionally charged themes than IM items. These include the admission of negative emotions such as anger, jealousy or sexual feelings, and have a defensive tone similar to

that of psychoanalytic denial, while IM items cover compliance with norms relating to socially acceptable behaviour such as not telling lies or not covering up one's mistakes.

Paulhus (2002) views SDE and SDD as best suited to measuring response styles, long lasting and consistent biases across time and questionnaires, while AM and IM are best suited to measuring responses to instructional sets which are more short-lived biases caused by a temporary factors such as the desire to impress a particular audience. Paulhus's conceptualisation of social desirable responding, and consequently the BIDR, has evolved over time. Early versions of the BIDR only included the SDE and IM scales and the addition of the AM scale has been the most recent development to it. Research has supported the validity of Paulhus's two-tier model model (e.g. John & Paulhus, 2000).

It is possible to conceptualise Paulhus's model in light of the most recent thinking on social desirable responding, that is that social desirable responding is best thought of as a three way interaction (between the nature of the items making up the focal scale, the respondent's characteristics and the situation the testing takes place in), that it might be distortion / error, substantive or a combination of both. In this view, correlations between a focal scale and either of the measures of unconscious bias (SDE or SDD) would equate to 3-way interactions, but with respondent characteristics having greater weight in the interaction than situational factors. In contrast, correlations between either of the measures of conscious bias (AM or IM) would also equate to 3-way interactions, but with situational factors having equal or greater weight in the interaction than respondent characteristics. 'Respondent characteristics' would correspond to the expression of traits that lead to distortion / error, substantive responding or a combination of both. Situational influences would lead to distortion / error. Research by Lönnqvist et al. (2007) on Finnish versions of the SDE and IM scales offers some support for this. They found that neither scale was a 'pure' measure and that scores on both were affected by 'style', 'set' and substantive individual differences but to differing degrees.

When published in 1960 the MC scale was intended by its authors to be a measure of social desirable responding in self-reports. However, Crowne and Marlowe (1964) later suggested that the scale taps a more general motive of a person's need for approval from others, which Crowne (1979) refined further as a drive to avoid disapproval from others. Paulhus (1991, p.29) notes the scale sustains 'a dual existence as a social desirable responding scale and a measure of the approval-dependent personality'. McKelvie's (1995b) conclusion that the VVIQ was not seriously affected by social desirable responding was derived mainly from data involving the MC scale.

A number of specific weaknesses have been identified in the MC scale. Critics suggest that some of the items on the scale are out of date (Stöber, 2001), that the scale may confound the type of item, attribution of positive attributes and denial of negative attributes, with direction of item scoring (Helmes, 2002) and that confirmatory factor analysis has not supported either single or two factor models for the scale (Leite & Beretvas, 2005). Further, the MC scale was originally developed on the assumption that social desirable responding is a unitary construct; a problematic assumption given evidence suggesting that social desirable responding occurs in more than one form.

There is also evidence that McKelvie underestimated the size of the relationship between the VVIQ and social desirable responding because the MC scale is predominantly a measure of moralistic bias. We would argue that respondents answering self-report imagery questionnaires will most likely view imagery as an ability being tested, rather than a 'value neutral' aspect of their cognition and therefore will value responses they perceive to indicate 'better' imagery ability. This would be particularly true when data is collected in settings which may emphasise ability and competition such as data collected from groups of students in academic settings with the presence of a tutor. Given this, the egoistic bias with its emphasis on claiming positive social and intellectual qualities, would seem to be more likely to correlate with self-report imagery questionnaires than the moralistic bias. Paulhus (1984) factor analysed scores from several measures of social desirable responding and found the MC scale to load .68 on a factor labelled as 'impression management' and .40 on a factor

labelled as 'self-deception'. Interpreted in terms of Paulhus's two-tier model it is not possible to know for sure if the MC loaded more highly on the impression management factor because it is predominantly a measure of conscious forms of social desirable responding, the moralistic bias or a mixture of both, as the factors differed in terms of content (moralistic bias vs. egoistic bias) and level of consciousness (conscious vs. unconscious). However, from studying the items of the MC and noting that they do not appear to tap into the emotionally charged themes characteristic of SDD, we would argue that it is most likely that the key distinguishing feature between the factors is content and that the MC scale is predominantly a measure of the moralistic bias.

Previous work investigating the relationship between a variety of self-report imagery scales and social desirable responding (Allbutt, *et al.*, 2006a, Allbutt *et al.*, 2006b, Allbutt *et al.* 2008) supports our critique of the reliance on data from the MC scale. In this research the SDE and IM scales of the BIDR have been used as measures of social desirable responding. The SDE and IM scales were selected because they are the best established of the BIDR scales and the scales include one measure of egoistic bias and one measure of moralistic bias. Together these studies found the following: 1) 'greater' imagery ability was always associated with higher social desirable responding; 2) correlations with self-report imagery scales and SDE were always higher than those with IM; 3) the magnitude of correlations between imagery scales and SDE sometimes exceeded McKelvie's (1995a) .25 criterion; 4) significant correlations between imagery properties and SDE were seen for scales measuring vividness of visual imagery and for some other visual properties of images such as ease of image generation and ease of image maintenance, but not for measures of vividness of visual thinking style and auditory imagery ; 5) the size of imagery scale – SDE correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which the imagery property was valued; and 6) the form of BIDR sub-scale scoring did not appear to greatly affect the pattern of results obtained with SDE and IM scales.

In the current study, participants completed the VVIQ, all four scales of the BIDR and the MC. The study had three main aims which extend previous work:

1) A direct comparison of how the VVIQ correlates with the scales of the BIDR and the MC scale in a single sample of participants. Earlier work has argued that McKelvie (1995b) underestimated the degree of relationship between the VVIQ and social desirable responding because of a reliance on data from the MC scale which may primarily be a measure of the moralistic bias. However, we have not directly tested the validity of our argument using the MC scale in our previous work. If our critique is correct we would expect the VVIQ to correlate more highly with BIDR egoistic scales than with the MC, and for the MC scale to correlate more highly with BIDR moralistic scales than egoistic scales.

2) Investigate whether the VVIQ correlates significantly and independently with both BIDR egoistic scales. Our previous work has shown a consistent relationship between the VVIQ and SDE, however we have not yet tested for a relationship between the VVIQ and AM. We would normally expect anonymous testing conditions to generate minimal social desirable situational pressures. However, data collected from groups of students as part of an experiment carried out in an academic setting and with a tutor present acting as the experimenter potentially could create situational demands to respond consciously in a socially desirable way. If so, AM might correlate with self-report imagery questionnaires independently from SDE and that the relationship with socially desirable responding seen under our testing conditions could be greater than the estimate made from our studies which were based on SDE alone.

3) Finally, investigate the effect of BIDR scoring system on the pattern of results obtained. There are two ways to score the BIDR subscales, continuously and dichotomously. After recoding of negative items, in the continuous approach the scores are summed to create a total for each subscale, while in the dichotomous approach, recoded extreme scores (either six or seven on an item), are counted as one point and are added to create a total for each subscale. Paulhus (1994) favours the dichotomous approach to scoring arguing that it has the advantage of only scoring

clearly exaggerated or managed high responses. Other authors, however, have challenged this assertion based on findings with the SDE and IM subscales observing better performance using continuous scoring (higher Cronbach's alphas and higher convergence correlations) (Stöber, Dette & Musch, 2002) and better fit for the data and their student sample (Cervellione, Lee & Bonanno, 2009). Both Stöber *et al.* and Cervellione *et al.* suggest that one reason why performance may be superior with continuous scoring is that, unlike dichotomous scoring, it may also tap into exaggerated low responses as well as exaggerated high responses. Allbutt *et al.* (2008) tested the effect of scoring method of the SDE and IM subscales for three self-report imagery subscales and found little difference in pattern of results obtained, however AM and SDD subscales remain yet to be tested in this way.

<div align="center">Method</div>

Participants

Participants were 123 undergraduate psychology students (23 males, 100 females). Ages ranged from 18 to 54 years with a mean age of 19.6 years. Students received course credit for participation, although they had the option to choose which studies they took part in.

Materials

Vividness of Visual Imagery Scale (VVIQ; Marks, 1973). The VVIQ assesses vividness of visual imagery using 16 items. Participants form a series of images of a friend or relative's face, the rising sun, a shop, and a country scene. Ratings of image vividness are made on a five point scale ranging from 1, 'Perfectly clear and vivid as normal vision', to 5, 'No image at all, you only 'know' you are thinking of the object'. Traditionally participants make 16 ratings with their eyes open, then 16 ratings with their eyes closed and the two sets of scores are added together. However, McKelvie's (1995b) meta-analytical review found no difference in the ratings made with eyes open and eyes closed. Thus in the two studies mention of imaging with eyes open was edited out of the VVIQ instructions and the participants were only asked to complete the questionnaire once with their eyes closed. Lower total scores indicate more vivid imagery. McKelvie's (1995b) meta analytical review estimated internal consistency, as measured by Cronbach's alpha, to be 0.88. Cronbach's alpha from the current study was 0.82.

The Balanced Inventory of Desirable Responding Version 8 (BIDR; Paulhus, 2002). The BIDR has four 20-item subscales (SDE, AM, SDD, IM) which assess four kinds of social desirable responding. All scales are made up of alternating positively and negatively phrased statements, 10 of each type. Responses are made on a seven point scale ranging from 1 (*Not true*) to 7 (*Very true*). Both egoistic scales have statements about the respondent's abilities across a range of situations. The four subscales were scored separately and all were scored both continuously and dichotomously. Paulhus (1994) recommends omitting items from the BIDR scales if they are not appropriate or cause confusion for a particular sample e.g. where an item refers to being a car driver and yet most participants in a sample are likely to not own a car. Because of this SDE items 8 and 14, and IM item 13 were omitted. Paulhus (1999) also notes that some of the items of the SDD scale make its use ethically problematic and may be offensive to participants e.g. 'I have never felt like I wanted to kill someone'. Because of this SDD items 6, 12 and 13 were also omitted. Paulhus (1994) reported Cronbach's alphas for SDE scored dichotomously to range from 0.65 to 0.75, and alphas for IM scored dichotomously to range from 0.75 to 0.80. We are not aware of any currently published reliability data for the AM or SDD scales. In the current study Cronbachs alpha for the SDE, AM, SDD and IM subscales scored continuously were: 0.75, 0.82, 0.75, 0.72; and scored dichotomously were: 0.70, 0.69, 0.64, 0.60.

The Marlowe-Crowne scale (MC; Crowne & Marlowe, 1960). This scale assesses social desirable responding, specifically the 'need' to avoid the disapproval of others. It consists of 33 statements to which respondents reply 'true' or 'false'. Eighteen statements are positively phrased and 15 negatively, which are randomly ordered across the questionnaire. Positively phrased item responses of 'true' score one point, while negatively phrased item responses of 'false' score one point. Higher

total scores indicate greater social desirable responding / need for approval. Item 27 'I never make a long trip without checking the safety of my car' was omitted for similar reasons as described above for BIDR item omission. Paulhus (1991) reports Cronbach's alphas to range from 0.73 to 0.88. Cronbach's alpha from the current study was 0.77.

Ethical rights form. A form was used which notified participants of the rights they possessed under the British Psychological Society's code of ethics.

The ethics form and three questionnaires were combined together into a booklet, starting with the ethics form, followed by the VVIQ, then BIDR and ending with the MC.

Procedure
Participants answered the questionnaires in groups of about 20 with a tutor present who acted as the experimenter. They were asked to seat themselves as far apart as possible in the room. Responses were made in silence and participants were told not to put their name anywhere in the booklet. Participants answered the questionnaires in the order in which they appeared in the booklet.

Results
Descriptive statistics were calculated for all variables and can be seen in Table 1. The VVIQ, all BIDR subscales scored dichotomously and MC scales were found to be mildly positively skewed and so median and inter-quartile range (IQR) descriptive are presented as well as mean and standard deviation (SD). Note that throughout all the results tables continuous scoring of BIDR scales is indicated by the suffix 'c' and dichotomous scoring by the suffix 'd'.

Table1.
Descriptive statistics for scales

Variable	Number of scale items	Mean	SD	Median	IQR	Max	Min
VVIQ	16	38.52	10.13	38	15	18	66
SDEc	18	66.68	13.13	65	16	107	40
SDEd	18	3.37	2.79	3	4	13	0
AMc	20	59.87	15.16	59	19	93	30
AMd	20	2.19	2.30	1	3	9	0
SDDc	17	59.15	13.92	59	18	107	27
SDDd	17	3.46	2.53	3	3	13	0
IMc	19	68.29	14.26	67	17	112	22
IMd	19	4.81	2.65	4	3	14	0
MC	32	12.72	5.01	12	6	26	2

Note: N=123

One feature of note shown by the descriptive statistics is the effect the scoring system had on the spread of AM scores. When scored continuously, AM scores showed the largest spread of scores of all the BIDR subscales, however, when scored dichotomously it showed the smallest spread.

A series of Pearson Product Moment correlations between the VVIQ, BIDR subscales (scored continuously and dichotomously) and MC were carried out. Prior to carrying out the correlations the positive skews seen in the VVIQ and BIDR subscales scored dichotomously and the MC were corrected with square root transformations to make their distributions normal. Four multivariate outliers were removed reducing the sample size to 119. The results of the correlations are shown in Table 2. The reader should note that a large number of statistical tests have been carried out and as such some correlations may only be due to chance. However, the fact that most significant correlations observed indicated large effect sizes, the pattern of correlations is meaningful when

viewed in terms of Paulhus's (2002) model, and the degree of consistency with the results seen in our previous three papers, argues against the effects only having been caused by chance alone.

Table2.
Matrix for Pearson's correlations between measures

	VVIQ	SDEc	SDEd	AMc	AMd	SDDc	SDDd	IMc	IMd
VVIQ									
SDEc	-.35**								
SDEd	-.25**	.76**							
AMc	-.27**	.34**	.29**						
AMd	-.14	.25**	.34**	.74**					
SDDc	-.05	.31**	.13	-.09	-.15				
SDDd	-.03	.20*	.13	-.17	-.13	.84**			
IMc	-.11	.38**	.21*	.15	.04	.66**	.51**		
IMd	-.05	.29**	.17	.07	.03	.60**	.53**	.86**	
MC	-.18*	.38**	.30**	.15	.12	.66**	.58**	.68**	.62**

Note: * p<.05; ** p<.01 (2-tailed)

The VVIQ showed significant low negative correlations (reflecting the opposite keying of the VVIQ to the other scales) with some of the social desirability scales indicating that more vivid imagery was associated with greater social desirable responding. VVIQ correlations were typically largest with SDE and when BIDR scales were continuously scored. The VVIQ correlated significantly with both SDE and AM when they were scored continuously, and with SDE only when scored dichotomously. The VVIQ also correlated with the MC but to a lesser extent than the BIDR egoistic scales (apart from AM dichotomously scored) and showed no significant correlations with the BIDR moralistic subscales.

The pattern of correlations between the BIDR social desirable scales in general supported the two-tier model, though with continuous scoring the SDE subscale did not clearly correlate to a higher level with AM than with the BIDR moralistic subscales. SDD and IM subscales were more highly correlated than SDE and AM. Correlations between the same BIDR subscales when scored continuously and dichotomously were high and ranged from .74 for AM to .86 for IM. The MC correlated significantly with SDE, SDD and IM but notably more highly with the BIDR moralistic scales than SDE, but only marginally more highly with IM than with SDD. Intercorrelations between BIDR subscales tended to be larger with continuous scoring, and BIDR subscales correlations with the MC also tended to be slightly larger with continuous scoring.

While SDE, AM and the MC all correlated significantly with the VVIQ, considerations of Paulhus's (2002) model together with the pattern of correlations observed (SDE, AM and MC showed only small sized correlations between them, and the MC correlated most highly with the BIDR moralistic scales) raise the possibility that these scales might share unique variance with the VVIQ. Thus two stepwise multiple regressions were conducted to investigate the degree of independence of SDE, AM and the MC correlations with the VVIQ. In both regressions the VVIQ acted as the criterion variable. In the first regression the predictor variables were SDE scored continuously, AM scored continuously and the MC. In the second regression the predictor variables were SDE scored dichotomously and the MC. The results of the regressions are shown in Table 3. Neither regression proceeded beyond step one and both had only SDE in the equation, indicating that addition of variables beyond SDE did not significantly improve prediction of VVIQ scores, although in the first regression AM came close to satisfying entry requirements at step 2.

Table 3.
Results of stepwise regressions

Regression	Variable	B	SE B	β	R²	T	Sig
Regression 1 **Step 1**							
Variables in the equation	Constant	7.61	.67			20.62	.00
	SDEc	-.02	.00	-.35	.12	-4.02	.00
Variables not in the equation	AMc			-.17		-1.86	.06
	MC			-.06		-.64	.52
Regression 2 **step 1**							
Variables in the equation	Constant	6.52	.15			43.55	.00
	SDEd	-.23	.08	-.25	.06	-2.77	.01
Variables not in the equation	MC			-.12		-1.29	.20

F-to-enter (d) = .05, F-to-remove (e) = .10.

Finally, further Pearsons's correlations tested the possibility that dichotomously scored BIDR subscales might underestimate the size of the correlations with the VVIQ by not tapping into exaggerated low responses. To do this new analysis, SDE and AM scores were calculated in an analogue of the dichotomous scoring method but transposed to the bottom end of the scoring range with item responses of 1 or 2 scored as 1 point and all other responses scored zero. Distributions of these new variables were normally distributed and so did not require transformation. The resultant new SDE and AM scores were correlated with the VVIQ. Both VVIQ correlations were not significant, for VVIQ – SDE (r=-.05, p=.62, df=117, 2-tailed), for VVIQ – AM (r=-.10, p=.30, df=117, 2-tailed).

Discussion

This study was the first time a direct comparison had been made between the BIDR subscales and the MC. Previously we have argued that the MC scale is most likely primarily a measure of the moralistic bias and so is not ideally suited to assessing the relationship between scores on the VVIQ and social desirable responding, and that the BIDR is a superior measure of social desirable responding to the MC scale because of its ability to treat social desirable responding as a multifactorial phenomenon. The VVIQ correlation with the MC scale was lower than with the BIDR egoistic scales (apart from AM scored dichotomously). The coefficient value of .18 for the VVIQ – MC correlation matches McKelvie's (1995b) meta-analysis estimate from 10 studies. The MC correlated significantly with SDE, SDD and IM but notably more highly with the BIDR moralistic scales than SDE, but only marginally more highly with IM than with SDD. The results showed several significant correlations between the VVIQ and the measures of social desirable responding in which greater imagery ability was always associated with higher social desirable responding. The largest of these correlations were with the BIDR measures of egoistic bias, particularly SDE. This pattern of results supports previous research (e.g. Allbutt et al., 2008).

Regression analyses investigated the relationship between the VVIQ, the BIDR egoistic subscales and the MC. We were concerned that, despite the use of anonymous responding, factors such as participation in an experiment, group testing in an academic setting, and presence of a tutor in the role of the experimenter might create the conditions for the VVIQ to correlate with the AM independently of SDE. However, the stepwise regression analyses showed no conclusive evidence

that correlations between the VVIQ and SDE were enhanced by adding other social desirable responding scales.

The continuous and dichotomous BIDR scoring methods were also compared. Allbutt et al. (2008) found no major differences with BIDR scoring method in studies investigating the relationship between a new, longer version of the VVIQ called the VVIQ version 2 (Marks, 1995), the Individual Differences Questionnaire – Imagery Habit Scale (Cohen & Saslona, 1990), the Auditory Imagery Scale (Gissurarson 1992) and the SDE and IM subscales of the BIDR. However, several notable differences in patterns of results were seen in this study with the two scoring methods. VVIQ correlations were higher with SDE and AM when continuously scored. The MC correlated more highly with BIDR subscales when they were continuously scored. Also when continuously scored BIDR subscales showed higher Cronbach's alphas, were normally distributed as opposed to positively skewed, and showed higher correlations with each other. This pattern of results is in agreement with those of Stöber, et al. (2002) and Cervellione et al, (2009).

In contrast, the pattern of correlations seen with dichotomous BIDR scoring showed Paulhus's (2002) two-tier structure more clearly. The fact that the correlations between each BIDR subscale scored using the two scoring methods were around .7- .8, and the difference in distribution normality between the scoring approaches suggests that they measure related but not identical constructs. The possibility that dichotomous scoring might underestimate the size of the correlations with the VVIQ by not tapping into exaggerated low responses was explored by scoring SDE and AM scales using a 'mirror image' analogue of the dichotomous approach. However, no significant correlations were found between the VVIQ and the BIDR egoistic subscales scored in this way which suggests that no relationship between VVIQ and exaggerated low responses existed. The scoring method may also have had an effect on VVIQ correlations with the AM subscale. The VVIQ – AM correlation reached significance with AM scored continuously but not when scored dichotomously. AM scale items differ from SDE items in being particularly strong egoistic claims. Under our anonymous testing conditions descriptive statistics showed that AM dichotomous average scores were low and not widely spread. This restricted spread of AM scores when scored dichotomously may have contributed to the difference in VVIQ – AM correlations when scored using continuous and dichotomous approaches.

In our introduction, we noted that there is debate over whether correlations with social desirable responding should best be viewed as distortion / error, substantive responding or a mixture of both. We have argued that respondents answering self-report imagery questionnaires will most likely view imagery as an ability being tested and will value responses they perceive to indicate 'better' imagery ability. The results of previous studies have provided some support for this in that the size of imagery scale – SDE correlations were correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which the imagery property was valued. If social desirable responding is best viewed as distortion / error or as different on a conceptual level from mental imagery processes then McKelvie's (1995a) suggestion that correlations with measures of social desirable responding in excess of .25 (6.25% overlap of variance) would indicate an unacceptable degree of contamination becomes relevant. The correlations between the VVIQ and SDE either exceeded this criterion when SDE was scored continuously (r=-.35), or matched the criterion when SDE was scored dichotomously (r=-.25). However, like all cut-off points, the exact value chosen is somewhat arbitrary, and even assuming social desirable responding reflects distortion, judgements of the validity of imagery questionnaires would also need to be made in light of data on other aspects of validity.

If social desirable responding is best viewed as valid substantive responding in its own right the crucial question becomes less what is an unacceptable level of contamination and how can we control the influence of SDE, and more why the VVIQ and other scales from our previous work should correlate with the egoistic bias. One possible answer is that the relationship between the VVIQ and the egoistic bias is mediated by anxiety levels. Although no study has directly correlated VVIQ total scores with anxiety levels, Bent and Wick (2005) measured participants' vividness of visual imagery

using Ahsen's adapted VVIQ (Ahsen,1985) when imaging their mothers and fathers, and when imaging other VVIQ scenes while keeping either their mother or father in mind. Anxiety levels were also measured while doing this using Speilberger's State-Trait Anxiety Inventory (Speilberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). Bent and Wick (2006) found vividness difference scores for imaging mothers and fathers correlated significantly with trait anxiety difference scores (r=.27), and the vividness difference scores for VVIQ scenes while keeping mothers and fathers in mind correlated significantly with state anxiety (r=.43) and with trait anxiety (r=.60). Similarly, Paulhus (1994) reports the highest correlates of the SDE scale to be trait anxiety (r=-.52) and self esteem (r=+.50) and Davies, French and Keogh (1998) report a correlation between the SDE and Eysenck's neuroticism scale from the Eysenck Personality Questionnaire-Revised (Eysenck & Eysenck, 1991) to be -.53.

A priority for future work should be testing the assertion that social desirable responding is an independent construct from mental imagery processes and so its control would aid the assessment of self-reported imagery. Testing the relationship between VVIQ scores, egoistic bias scores and anxiety levels would help determine the extent to which VVIQ – egoistic bias correlations might be mediated by anxiety. Also, McKelvie's (1995b) meta-analytical review reported relationships between the VVIQ-1 and other self-report imagery questionnaires, perceptual measures and memory measures that were argued to provide evidence to support the construct validity of the VVIQ. Repeating some of these studies and observing the effect of controlling social desirable responding would provide valuable data to help resolve this issue. Meta-analytic reviews (Li & Bagger, 2006; Ones, Viswesvaran & Reiss, 1996) have found partial correlation approaches to be unable to effectively control for social desirability effects. However, more recently, approaches based on confirmatory factor analysis that have the capacity to control for item measurement error but that require large sample sizes and specialist statistical software, have shown greater promise and suggest that social desirable responding can act as both a suppressor and a moderator variable (Konstabel, Aavik & Allik, 2006; Leite & Cooper, 2010; Ziegler, Buehner, 2009).

References

Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery, 9*(3), 1-18.

Allbutt, J., Shafiullah, M. & Ling, J. (2006a). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, *30,* 1&2, 1-20.

Allbutt, J., Ling, J. & Shafiullah, M. (2006b). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery. *Imagination, Cognition and Personality, 25*, 4, 337-353.

Allbutt, J., Ling, J., Heffenan, T. and Shafiullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and visual thinking style. *Journal of Individual Differences, 29*. 4, 181-188.

Bent, N. A. & Wick, E. (2006). Beyond vividness: Parental filters as moderators in mental imagery and measured anxiety level. *Journal of Mental Imagery*, *30*(1&2), 21-38.

Cohen, B. H., & Saslona, M. (1990). The advantage of being an habitual visualiser. *Journal of Mental Imagery*, *14*(3&4), 101-112.

Crowne, D.P. (1979). *The experimental study of personality.* Hillsdale, NJ: Erlbaum.

Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.

Crowne, D.P. & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Davies, M. F., French, C. C. & Keogh, E. (1998). Self-deceptive enhancement and impression management correlates of EPQ-R dimensions. *Journal of Psychology, 132*, 4, 401-406.

Domino, G. & Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge: Cambridge University Press.

Edwards, A.L., Diers, C.J. & Walker, J.N. (1962). Response sets and factors loadings on sixty-one personality scales. *Journal of Applied Psychology, 46*, 220-225.

Eysenck, H. J. & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult).* London: Hodder & Stoughton.

Guissurason, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery*, *16*(3&4), 117-122.

John, O.P. & Paulhus, D.L. (2000). *The structure of self enhancement*. Unpublished manuscript, University of California, Berkley.

Konstabel, K., Aavik, T. & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20*, 7, 549-566.

Li, A. & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*, 2, 131-141.

Leite, W. L. & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioural Research, 45*, 2, 271-293.

Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J.& Verkaslo, M, (2007). Substance and style in social desirable responding. *Journal of Personality, 75*, 2, 291-322.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*(3&4), 153-166.

McKelvie, S. J. (1995a). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 1-106.

McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*(3&4), 197-251.

Ones, D. S. Visweswaran, C. & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 6, 660-679.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver and L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp17-59). San Diego: Academic Press.

Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6*. Unpublished Manual, University of British Columbia, Vancouver, Canada.

Paulhus, D. L. (1999). *BIDR Self-Deceptive Denial (SDD)*. neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm. Retrieved from the World Wide Web on 1 August 2003.

Paulhus, D. L. (2002). Social Desirable Responding: The evolution of a construct. In Braun, H. I., Jackson, D. N. & Wiley, D. E. (Eds.), *The role of constructs in psychological and educational measurement* (pp49-69). Mahwah: Lawrence Erlbaum Associates.

Spielberger, C.D., Gorsuch, R., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Stöber, J., Dette, D. E. & Musch, T. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78*(2), 370-389.

Ziegler, M. & Buehner, M, (2009). Modelling social desirable responding and its effects. *Educational & Psychological Measurement, 69*, 4, 548-565.

Evaluation of paper

Strengths

Paper 4 made contributions to extend the work began in papers 1, 2 and 3 by generating several new insights. It was the first time we had made a direct comparison of how the VVIQ-1 correlated with all four of the scales of the BIDR and the MC scale in a single sample of participants. As such it promised to be the most comprehensive and direct test of the central ideas behind the research. Paper 4 produced several important empirical findings:

1) The results showed several significant correlations between the VVIQ-1 and the measures of social desirable responding in which greater imagery ability was always associated with higher social desirable responding. As was seen in the past three papers, the largest of these correlations were with the BIDR measures of egoistic bias, particularly SDE.

2) The VVIQ-1 correlation with the MC scale was lower than its correlation with the BIDR egoistic scales (apart from AM scored dichotomously). This provided support for the critique of McKelvie's (1995b) meta-analysis which was based primarily on data from the MC questionnaire.

3) The MC correlated significantly with SDE, SDD, and IM, but notably more highly with the BIDR moralistic scales than with SDE, and only marginally more highly with IM than with SDD. This provided further data showing that it is primarily a measure of the moralistic bias.

The pattern of results seen in paper 4 provided strong support for the critique which had started the series of papers. Thus it does appear that past research has been limited by its use of the MC scale. Due to the findings from Paper 4, we can also say with more confidence that the VVIQ-1 correlates with SDE but not with IM because it is a measure of the egoistic bias, rather than because one scale is more effective at tapping unconscious social desirable responding processes while the other is better at tapping more conscious, deliberate processes.

In the regression analyses involving the VVIQ-1, SDE and AM scales, in the first stepwise regression, AM scored continuously came close to satisfying the entry requirements at step 2

with *p*=.06. This points to the possibility that the continuously scored AM scale might tap additional aspects of the egoistic bias to that sampled by the SDE scale. Because this result only approached significance at the 5% level, for the time being it can only be taken as indicating a possible effect that needs replication by future data sets. If the effect were confirmed and it has a non-substantive basis, several explanations are possible. For example, the effect might reflect the greater sensitivity of the more extreme AM items to tap into more conscious forms of the egoistic bias. Alternatively, it might reflect item content differences between the BIDR egoistic scales. The AM scale has notably more items sampling self-confidence in relation to other people than the SDE scale, eight compared to three, making it more able to sample this aspect of self-enhancement than the SDE scale. However, until this effect is replicated, at present we can only conclude that the past studies do not appear to have greatly underestimated the relationship with the egoistic bias by omitting the AM scale.

Paper 4 also explored the scoring method for the BIDR further. Paper 4 was not able to resolve conclusively the issue of which BIDR scoring method is to be preferred. However, it did replicate some of the past findings and produced new observations on the issue. The findings that related to this included:

1) The fact that the correlations between each BIDR subscale scored using the two scoring methods were around *r*=.8, and the differences in distribution normality between the scoring approaches, suggests that the different scoring methods measure highly related but not necessarily identical constructs. Alternatively, this pattern of results may reflect the loss of information when continuous scores are converted to dichotomous scores.

2) Also, some differences in the patterns of results were seen between the two scoring methods. Results which favoured the continuous scoring method included: BIDR continuously scored subscales showed higher Cronbach's alphas and their distributions were normally distributed as opposed to positively skewed; VVIQ-1 correlations were higher with SDE and AM when continuously scored; and the MC correlated more highly with the BIDR subscales when they were continuously scored. These findings are in line with those of researchers such as Stöber, Dette and Musch (2002), but differ from the results of our Paper 3 where no notable differences were found between the two scoring methods.

3) In contrast, dichotomous BIDR scoring showed the distinction between egoistic and moralistic biases from Paulhus's (2002) two-tier model more clearly in terms of the inter-correlations between the scales. This result is in line with Leite and Beretvas (2005) and Gignac (2013) who tested the two scoring methods on the SDE and IM scales using CFA and found that the dichotomous scoring system was a better fit for their data. Equally though, these results may have reflected the tendency for the dichotomously scored BIDR scales to produce lower correlations overall.

Paper 4 also developed thinking around possible substantive mechanisms underlying the relationship between self-report imagery questionnaires and SDE further by describing studies that showed a relationship between visual imagery vividness and anxiety, and between SDE and anxiety. This issue will be discussed more fully in the introduction to paper 5.

Subsequent analyses on the paper 4 data

As described previously, the Croatian researchers Kovacic et al (2014) attempted to overcome problems seen with the full four BIDR scales performance by developing two new scales from the BIDR items, the E-SDR and M-SDR. Of most relevance in the current context is the E-SDR scale, which was made up of four items from the SDE scale and six items from the AM scale. When the E-SDR was answered in honest conditions it most strongly correlated with self-peer personality trait discrepancy scores, while when answered in a relevant simulated job applicant condition it most strongly correlated with honest self–applicant personality trait discrepancy scores. Further, the E-SDR showed patterns of results which indicated that it was most sensitive to changes in the egoism content domain, and importantly, E-SDR correlations with personality trait discrepancy measures were higher than with the trait peer-ratings. If these 10 BIDR items are 'pulled out' from the SDE and AM scales used in paper 4 to form the E-SDR scale, they correlate $r_s=.23, p<.01$ with the VVIQ-1 when scored dichotomously, and $r=.40, p<.001$ scored continuously. This additional analysis provides further support that the VVIQ–egoism correlations seen in paper 4 were at least in part reflecting distortion rather than simply substance.

Limitations

Paper 4 attempted to advance conceptual understanding by incorporating insights from Leite and Cooper (2010). Leite and Cooper have argued that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the nature of the items making up the focal scale, the respondent's characteristics and the situation that the testing takes place in. In the paper an initial expression of how Leite and Cooper's ideas might apply to Paulhus's (2002) two-tier model was made by suggesting that the four BIDR scales mapped straightforwardly onto their respective component of the model. However, the conception of the meaning of focal scale - BIDR scale correlations presented in paper 4 could have been described more fully.

Several studies discussed in the literature review, such as the Croatian studies, have questioned how well the BIDR scales exclusively tap into unconscious and conscious social desirable responding processes. For example, the studies found that the BIDR self-deceptive scales (SDE and SDD) were affected by situational factors, but to a lesser extent than the BIDR impression scales (AM and IM). Also, Paulhus and Trapnell (2008) state that the AM and IM scales are 'not especially useful as individual difference measures. In private administrations, much of the individual difference variance represents actual content differences in positive qualities' (p504). Thus, a more sophisticated conception of the relationship between the BIDR scales and the two-tier model is needed.

Pulling this information together, it would seem plausible that under conditions of low situational pressure to give socially desirable responses, a combination of the nature of the items making up the focal scale and the respondent's characteristics would be most influential for responses on the focal scale. Under conditions of high pressure, situational factors would also become influential. While all four BIDR scales could possibly pick up effects in either situation, the SDE and SDD scales would be most likely to do so in low demand situations, while the IM and AM scales would be most likely to do so in high demand situations. This is because, for example in the egoistic domain, in low demand situations the less extreme items of the SDE scale would be most likely to produce wider range of scores than the AM scale, while under high

demand situations the reverse would be true. In addition to this, significant correlations could also reflect substantive factors or a mixture of substance and distortion.

Up to this point in papers 1-4 had produced a considerable amount of data which supported the basic ideas behind the research. However, as yet the research had not attempted to explore the relationship between the self-report imagery scale-egoistic relationship and other constructs such as state and trait anxiety. Paper 5 attempted to address is omission.

Paper 5: Allbutt and Ling (2009)

Reference: Allbutt, J. & Ling, J. (2009) The relationship between self-report imagery questionnaire scores, the egoistic sub-type of social desirable responding and state-trait anxiety. *British Psychological Society Conference,* Brighton.

Aims of the paper

Papers 2 and 3 had found that self-report imagery questionnaires and items perceived by participants as tapping valued aspects of imagery were more likely to correlate with SDE. These findings support a distortion view of the self-report imagery questionnaire-SDE relationship. However, research discussed previously, e.g. Connelly and Chang (2015), has found that under anonymous, low stakes response conditions SDE scores are influenced by a mixture of style and substance. This raises the possibility that, despite the results of papers 2 and 3, self-report imagery questionnaire correlations with SDE might result from a combination of substance and style factors. Up to this point, no possible substantive explanations for the relationship between self-report imagery questionnaires and SDE had been directly explored. Paper 5 constituted a first attempted to address this omission.

One construct that is a candidate for a third, unobserved confounding variable underlying the relationship between imagery vividness and the egoistic bias is anxiety. Anxiety is an emotion experienced when facing 'uncertain existential threat' (Lazarus, 2000, p. 230). Cox (2012) describes how anxiety is multidimensional in nature. First, it can be divided into state and trait components. State anxiety is an emotional state and refers to feelings of apprehension, tension, nervousness, and worry felt at the present moment in time, while trait anxiety is a general and long-standing predisposition to perceive environmental stressors as threatening and to respond to them with anxiety. State and trait anxiety can both be further subdivided into a mental component (cognitive anxiety) reflecting fears and worries such as the fear of being negatively evaluated by others, and a physical component (somatic anxiety) reflecting the perception of bodily responses indicating fear such as increased heart rate and sweating.

In papers 1-4 we discussed how the nonconscious self-inhibitory forces reported by Ahsen (e.g. Ahsen, 1985) and that form the suppressor mechanism in Hishitani et al's (2011) imagery vividness model, might offer a substantive explanation for the relationship between self-report imagery questionnaire scores and the egoistic bias. Rinaldo and Okada (1993) have suggested that these forces might overlap with the unconscious self-deceptive processes thought to underlie responses on the SDE scale. This substantive possibility was discussed previously in the coverage of paper 1.

There is also psychometric data which links both imagery vividness and SDE to anxiety. Although no study has directly correlated VVIQ-1 total scores with anxiety levels, Euse and Haney (1975) tested the relationship between visual imagery properties (clarity, controllability and associated affect), state-trait anxiety, neuroticism and extraversion in 109 undergraduate participants. Euse and Haney reported that high anxiety and neuroticism, but low extraversion, were all associated with poor clarity and controllability of visual images (imagery control correlated $r=-.31$ with neuroticism, .21 with extraversion, -.52 with state anxiety and -.52 with trait anxiety, while imagery clarity correlated -.36 with neuroticism, .26 with extraversion, -.58 with state anxiety and -.51 with trait anxiety). They interpreted their findings to suggest that anxiety acts as a mediating variable for the clarity and controllability of visual images. Further, Bent and Wick (2006) tested the effect of keeping either their mother or father in mind on participants' visual imagery vividness ratings using Ahsen's adapted VVIQ (Ahsen,1985), vividness ratings while imaging their mother or father and also on anxiety levels measured using the Speilberger State-Trait Anxiety Inventory (Speilberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). Bent and Wick found relationships between vividness and both state and trait anxiety scores, for example, the vividness difference scores for VVIQ scenes while keeping mothers and fathers in mind correlated significantly with both state ($r=.43$) and trait ($r=.60$) anxiety. Equally, Paulhus (1994) reports the highest correlates of SDE to be trait anxiety ($r=-.52$) and neuroticism ($r=-.53$). So, because anxiety has a relationship with both imagery properties such as clarity and controllability and the SDE, it is possible that it could operate as a 'third variable' underlying the relationship between vividness and SDE. Furthermore, the need to explore the potential substantial role of anxiety / neuroticism is lent further credence by the results of Lönnqvist et al (2007) and Connelly and Chang (2015), who both observed self-rated SDE to correlate with

other-rated neuroticism levels.

The studies from cognitive and clinical psychology that have explored the relationship between working memory capacity, anxiety and performance provide data that suggest how the cognitive processes underlying imagery vividness and anxiety might interact. For example, Moran (2016) conducted a recent meta-analysis of 177 data samples which found that self-reported anxiety was related to poorer performance on measures of working memory capacity (g=-.334). In addition, he conducted a narrative review which found that both self-reported anxiety and experimentally induced anxiety were related to poorer task performance. When considered separately, trait anxiety had a uniformly negative relationship with performance, while state anxiety had an inverse 'U' shape relationship, although its relationship with performance on complex tasks was negative. Moran described how most cognitive theories of the relationship between anxiety and working memory suggest that anxiety competes with task related processes for limited cognitive resources, and so anxiety processes can be thought of as a dual, or secondary task that reduces working memory's ability to retain information on relevant stimuli and to ignore irrelevant stimuli. Furthermore, the fact that anxiety processes reflect the activation of an important survival mechanism that prepares an individual for dealing with a potential threat, means that these processes will often take precedence over task related processes. So performance on a task will decline when cognitive resources such as executive function, which are important for conducting the task, are taken up by anxiety processes. If the logic behind these effects is extended to anticipate the likely effects of anxiety on self-reports of imagery experiences such as vividness, it can be predicted that anxiety processes would act to reduce the vividness of experienced imagery because of their consumption of limited cognitive resources.

Taken together, the literature on this topic suggests the possibility that correlations between imagery properties such as the vividness of visual imagery and SDE might reflect a shared general relationship with state and / or trait anxiety. Furthermore, that the direction of the relationship between imagery vividness and anxiety would be predicted to be such that lower vividness of imagery would be associated with higher levels of anxiety. Thus the fifth paper

collected participant responses to the VVIQ-1, the BIDR egoistic scales (SDE and AM) and Speilberger et al's anxiety scales to investigate:

1) The extent to which scores on a self-report imagery questionnaire, the VVIQ-1, correlated with state and trait anxiety;

2) The extent to which the correlation between the imagery questionnaire and the egoistic bias was independent of correlations with state and trait anxiety.

A correlation between the VVIQ-1 and state anxiety would suggest an association between image vividness and short-term anxious feelings felt at the time of completing the questionnaires, while a correlation with trait anxiety would suggest an association between image vividness and a tendency towards more a general and long-standing predisposition to perceive environmental stressors as threatening and to respond to them with anxiety. Directional hypotheses were used with associated one-tailed significance tests because past research had shown trends for higher vividness to be associated with higher egoistic scale scores and lower anxiety scores, and because of the findings from the anxiety and working memory literature. The BIDR scales were scored dichotomously because the paper was presented in 2009 after our third paper which had shown no notable effect of scoring method.

The relationship between self-report imagery questionnaire scores, the egoistic sub-type of social desirable responding and state-trait anxiety

John Allbutt[1] & Jonathan Ling[2]

1 University of Salford, Directorate of Psychology, University of Salford, Frederick Road, Salford, M6 6PU, United Kingdom. j.d.allbutt@salford.ac.uk

2 Department of Psychology, Keele University, Keele, Staffordshire, ST5 5BG.

Background

- Research question = Are responses on self-report imagery questionnaires related to socially desirable responding (SDR)?
- If so, what kind of SDR and to what extent?

Self-report imagery questionnaires

- Mental images are quasi-perceptual mental representations.
- Subjective self-report imagery tests = a commonly used approach to assessing individual differences in imagery ability in which participants introspect on their imagery experience and comment on aspects such as vividness of their imagery.

2

Social Desirable Responding: Distortion and error or valid responding?

- A concern with self-report imagery questionnaires is the extent to which responses might be 'contaminated' by SDR.
- Paulhus (2002, p.50) defines SDR as "…the tendency to give overly positive self-descriptions".
- Traditionally SDR has been viewed as distortion or error that should be eliminated or minimised.

- More recently this viewpoint has been questioned….
- Many psychologists now believe that SDR is better viewed as a personality trait in its own right that is related to other positive traits such as psychological adjustment, and so should be treated as substantive or valid responding.

- Within the imagery literature SDR has been viewed as distortion or error, and research into it has taken two forms:
  1) Experimental 2) Correlational

Both approaches suggest there is a relationship with SDR.   3

How much 'contamination' is acceptable?

- McKelvie (1995a) set the following criteria for the evaluation of levels of 'contamination' based on studies that calculated the average size of correlations between large numbers of psychological variables:

| Size of correlation	Conclusion
Non-significant	No contamination
0.15 or less	Inconsequential contamination
0.16 to 0.25	Some contamination (acceptable)
Above 0.25	Unacceptable contamination

4

What size of relationship do studies show?

- McKelvie (1995a, 1995b) meta-analysed the relationship between the most widely used tests:

1) Vividness of Visual Imagery Questionnaire (VVIQ: Marks 1973)

2) Marlowe-Crowne test of SDR (MC; Crowne & Marlowe, 1960).

- 2 estimates were made:

1) $r = 0.189$ (MC scale only).

2) $r = 0.075$ (MC scale + data from 3 other SDR tests).

- McKelvie (1995b) concluded that:

- "…the VVIQ is not seriously contaminated by social desirable responding, and is perhaps not contaminated at all" (p. 209).

- We question this estimate of the size of the relationship because….

5

Is SDR a unitary construct?

- The MC test assumes that SDR is a unitary construct.
- But, correlations between SDR scales are often low….
- Paulhus (2002) has proposed a model of SDR with 4 types of SDR.
- The Balanced Inventory of Desirable Responding Version 8 (BIDR. Paulhus 2002) measures these forms of SDR.

Figure 1. Paulhus's Two-Tier Model (modified from Paulhus, 2002).
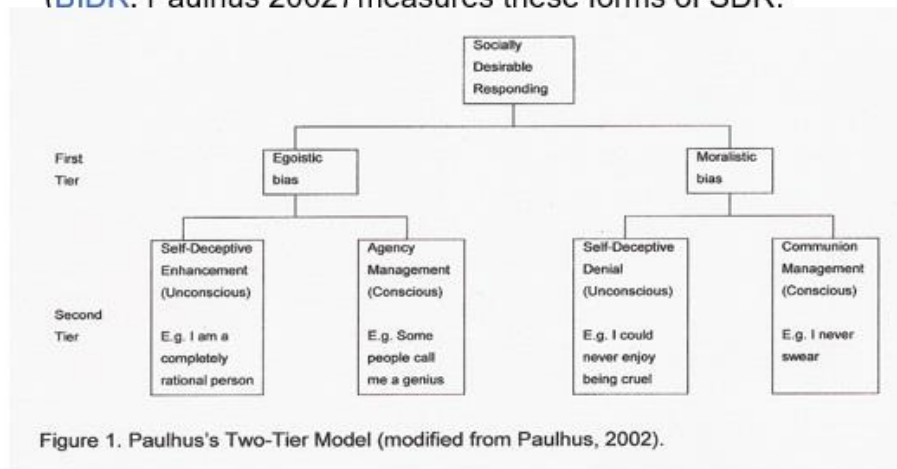
6

Is the most relevant form of SDR being assessed?

- No….

- Paulhus (1984) found the MC test to load:

1) 0.68 on a factor labelled 'impression management' (moralistic bias).

2) 0.40 on a factor labelled 'self-deception' (egoistic bias).

- So the MC test is more a measure of moralistic bias.

- But participants answering imagery questionnaires will most likely view imagery as an ability being tested.

- So, the egoistic bias is more likely to correlate with responses on self-report imagery scales.

- Both of McKelvie's (1995b) estimates drew either exclusively or predominately on data from the MC scale so he may have underestimated the size of the relationship.

7

Evidence from our previous research

- 5 papers have investigate the relationship between a variety of imagery tests and the SDE and IM scales of the BIDR.

- Results:

1) Correlations with the egoistic bias are always higher than with the moralistic bias, and 'greater' imagery ability is associated with higher SDR.

2) Our estimates tend to suggest a stronger relationship than McKelive (1995a,b) suggests.
   - e.g. Allbutt, et al. (2006a) used VVIQ and Vividness of Movement Imagery Questionnaire (VMIQ; Isaac, Marks & Russell, 1986).
   - Correlations with SDE were: VVIQ -0.25; VMIQ Total -0.31

3) Imagery questionnaires or questionnaire items perceived by participants as tapping ability aspects of imagery or are more valued are more likely to correlate with SDE.

8

Our position on the 'substance vs style' issue

☐ The third finding listed above seems to support an error / distortion view of the imagery questionnaire-SDE relationship.

☐ However, the relationship may also be partly substantive in nature. Paulhus (1994) reports the highest correlates of the SDE scale to be trait anxiety (-0.52) and self esteem (0.50) and Davies et al., (1998) report a correlation between the SDE and Eysenck's neuroticism scale from the Eysenck Personality Questionnaire-Revised (Eysenck & Eysenck, 1991) to be -0.53.

☐ These phenomena are possible candidates for a substantive relationship.

9

The Current Study

☐ The current study investigated:

1) The extent to which scores on a self-report imagery questionnaire, the VVIQ, correlated with state and trait anxiety.

2) The extent to which the correlation between the imagery questionnaire and the egoistic bias was independent of correlations with trait and state anxiety.

10

Method

- 53 undergraduate students, 47 female, 6 male, mean age = 22.8, sd = 6.3.
- The following questionnaires were used:

1) *Vividness of Visual Imagery Scale* (VVIQ; Marks, 1973).

- Assess vividness of visual imagery.
- 16 items. 5 point response scale: 1, ('Perfectly clear and vivid as normal vision'), to 5, ('No image at all, you only 'know' you are thinking of the object').
- Ratings were made once with the eyes closed.
- *Lower scores = more vivid imagery.

11

2) *The Balanced Inventory of Desirable Responding Version 8* (BIDR; Paulhus, 2002)

- Only 2 egoistic sub-scales used:

a) Self-deceptive enhancement (SDE)
  - Items cover: Control own thoughts, behaviour & life, self-knowledge, quality of judgements, e.g. 'My first impressions of people usually turn out to be right'.

b) Agency Management (AM)
  - Items cover: Persuasive powers, intellect, problem solving, leadership, interpersonal forcefulness, decisions & judgement, etc., (More exaggerated claims than SDE). E.g. 'My persuasive powers are impossible to resist'.

- 20 items and a 7 point response scale: 'Not true' to 'very true'.
- Scales were scored 'dichotomously' (after recoding of negative items 6 & 7 = 1 point).

12

3) *State – Trait Anxiety Index (Speilberger et al., 1983)*

☐ 2 scales: state anxiety and trait anxiety.

a) State anxiety
- – Items cover: Feelings of apprehension, tension, nervousness, and worry felt at the present moment in time e.g. 'I feel calm.

b) Trait anxiety
- – Items cover: More general and long-standing feelings of anxiety. e.g. 'I am a steady person.

☐ Both scales have 20 items and a 4 point response scale: 'Not at all' to 'very much so'.

13

Results

Correlations

		wiq	sde_dsq	AM_dsq	state_sq	Trait anxiety
wiq	Pearson Correlation	1	-.259*	-.180	.171	.250*
	Sig. (1-tailed)		.039	.113	.125	.045
	N	47	47	47	47	47
sde_dsq	Pearson Correlation	-.259*	1	.272*	-.281*	-.365**
	Sig. (1-tailed)	.039		.032	.028	.006
	N	47	47	47	47	47
AM_dsq	Pearson Correlation	-.180	.272*	1	-.180	-.209
	Sig. (1-tailed)	.113	.032		.113	.080
	N	47	47	47	47	47
state_sq	Pearson Correlation	.171	-.281*	-.180	1	.729**
	Sig. (1-tailed)	.125	.028	.113		.000
	N	47	47	47	47	47
Trait anxiety	Pearson Correlation	.250*	-.365**	-.209	.729**	1
	Sig. (1-tailed)	.045	.006	.080	.000	
	N	47	47	47	47	47

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

14

▢ Correlational analyses found the following main findings:

1) The VVIQ correlated significantly with SDE and trait anxiety.

2) Further analysis using partial correlation found that the correlation between the VVIQ and the egoistic bias was largely unaffected by partialing out trait anxiety:

▢ VVIQ - SDE correlation before partialing = r=-0.259, after partialing = r=-0.186.

▢ Thus the correlations between the VVIQ and the egoistic bias, and the VVIQ and trait anxiety, appear to be largely independent.

15

Conclusions

▢ The correlations between the VVIQ and the egoistic bias, and the VVIQ and trait anxiety, appear to be largely independent.
▢ The correlations between self-report imagery scales and the egoistic bias in past research does not seem to be due to trait anxiety.

16

References

▫ Allbutt, J., Shafiullah, M., Ling, J., & Anderson, J. (2005). The Relationship Between Self-Report Measures Of Auditory And Visual Imagery And Four Sub-Types Of Social Desirable Responding. *9th European Congress of Psychology*, Granada.

▫ Allbutt, J., Shafiullah, M. & Ling, J. (2006a). The Relationship Between Self-Report Imagery Questionnaire Scores and Sub-types of Socially Desirable Responding: Visual and Movement Imagery. *Journal of Mental Imagery, 30, 1&2, 1-20.*

▫ Allbutt, J., Ling, J. & Shafiullah, M. (2006b). The Relationship Between Self-Report Imagery Questionnaire Scores and Sub-types of Socially Desirable Responding: Components of Visual Imagery. *Imagination, Cognition and Personality, 25, 4, 337-353.*

▫ Allbutt, J., Ling, J., Anderson, J. & Dean, G. (2007). The Relationship Between the Vividness of Visual Imagery Questionnaire, the Balanced Inventory of Socially Desirable Responding and the Marlowe-Crowne test. *10th European Congress of Psychology, Prague.*

▫ Allbutt, J., Ling, J., Heffernan, T. & Shafiullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and thinking style. *Journal of Individual Differences, 29, 4, 181-188.*

▫ Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24, 349-354.*

▫ Davies, M.F., French, C.C. & Keogh, E. (1998). Self-deceptive enhancement and impression management correlates of EPQ-R dimensions. Journal of Psychology, 132, 4, 401-406.

▫ Eysenck, H.J. & Eysenck, S.B.G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult).* London: Hodder & Stoughton.

17

▫ Isaac, A., Marks, D.F. & Russell, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery, 10, 23-30.*

▫ Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64, 17-24.*

▫ Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19, 3 and 4, 153-166.*

▫ McKelvie, S. J. (1995a). The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19, 3 and 4, 1-106.*

▫ McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond, vividness and Its measurement. *Journal of Mental Imagery, 19, 3 and 4, 197-251.*

▫ Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46, 598-609.*

▫ Paulhus, D. L. (1994). Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6. Unpublished Manual, University of British Columbia, Vancouver, Canada.

▫ Palhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In Braun, H.I., Jackson, D.N. and Wiley, D.E. *The role of constructs in psychological and educational measurement.* New Jersey: Lawrence Erlbaum Associates.

▫ Spielberger, C.D., Gorsuch, R., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Consulting Psychologists Press.

18

Evaluation of paper

Strengths

Paper 5 introduced a new direction to the research in correlating the VVIQ-1 and the BIDR egoistic scales with a construct which could offer a substantive explanation for their relationship, anxiety. The key findings of the correlational analysis were:

1) The results showed significant correlations between the VVIQ-1 and the SDE measure of social desirable responding (*r*=-.26), in which greater imagery vividness was associated with higher social desirable responding. This result was consistent with the results of papers 1-4.

2) The VVIQ-1 also correlated significantly with trait anxiety (*r*=.25) with greater vividness associated with lower levels of anxiety. However, the VVIQ-1 correlation with state anxiety (*r*=.17) did not reach significance, although this may have been due to the small sample size. These correlations were in the same direction, but notably smaller in size than those seen by Euse and Haney. The limitations of the small sample used, and the fact that state and trait anxiety are often correlated (Moran, 2016), mean it is not possible to say with certainty whether the VVIQ's relationship is exclusive to trait anxiety, or whether it extends to also include state anxiety, or variance shared by both state and trait anxiety.

3) Further analysis using partial correlation found that the correlation between the VVIQ-1 and the egoistic bias was largely unaffected by partialling out trait anxiety, the VVIQ–SDE correlation fell to *r*=.19 after trait anxiety was removed.

The VVIQ-1 involves imaging life-like images that are likely to have limited negative emotional value (a friend or relative's face, the rising sun, a familiar shop and a country scene). As such it would appear unlikely that content related imagery moderation mechanisms such as the hypothesized suppressor mechanism from Hishitani et al's model were engaged in this study. However, the association between lower image vividness and higher trait anxiety is consistent with an effect whereby anxiety process acts to reduce the vividness of experienced imagery because of their consumption of limited cognitive resources in working memory. Also, the fact that the VVIQ-1 correlations with state and trait anxiety were not bigger than with SDE and the

outcome of the partial correlation analysis, suggests that in the self-report imagery questionnaire–SDE correlations seen in papers 1-4, SDE was not acting as a proxy for either state or trait anxiety.

Limitations

Paper 5 had some limitations. It was a small study of limited scope whose findings need replicating with a larger sample of participants. The poster format meant that only limited coverage was given to some aspects of the study.

The paper only presented the results from the BIDR egoistic scales when scored dichotomously. Analysis of the BIDR scales scored continuously showed a broadly similar pattern to the dichotomously scored scales though were smaller in size. The correlation between the VVIQ-1 and SDE was $r=-.19$, and with AM was $r=-.14$. One aspect of these results worthy of note is that in contrast to paper 4, here higher correlations with the VVIQ-1 were seen with the BIDR scales scored dichotomously i.e. the reverse pattern to that see in paper 4.

Aside from anxiety, it is possible that other substantive factors might underlie the self-report imagery scale–egoistic bias relationship. For example, Li and Bagger's (2006) meta-analysis found SDE to correlate with extraversion at $r=.31.$ Further, several studies have found results that suggest a relationship exists between extraversion-introversion and imagery vividness. As previously discussed, Euse and Haney (1975) found extraversion to be positively related to visual imagery clarity and controllability. Other studies have also found similar findings e.g. McDougall and Pfeifer (2012). They selected groups of extraverts (n=15) and introverts (n=15) using Eysenck's Personality Inventory from a larger sample (N=54) and tested their VVIQ-1 scores and memory for concrete and abstract nouns. Extraverts reported more vivid imagery than introverts ($r=-.43$ when controlled for neuroticism) but this did not translate into better recall, even for concrete stimuli. Morris and Gale (1973) suggested the relationship could reflect the fact that extraverts are 'stimulus hungry' and seek arousal through their imagery experiences. In contrast McDougall and Pfeifer suggested that extraverts' higher vividness scores might be more reflective of response distortion stemming from the fact that extraverts

are more impulsive than introverts and so less cautious when making vividness judgements rather than necessarily having more vivid imagery.

As previously discussed, using the NEO-PI-R, Holden and Passey (2010) found SDE to correlate primarily with big 5 facet correlations that share a common thread of confidence in abilities: C1 competence ($r=.32$), N6 vulnerability ($r=-.31$) and E3 assertiveness ($r=.31$). This finding suggests that SDE correlates with big 5 traits primarily because they contain elements of confidence / over-confidence in abilities. Given this, it is possible that confidence / overconfidence underlies the big 5 traits correlations with imagery self-report questionnaire scores.

Finally, it should also be born in mind that just as scores on the BIDR egoism scales appear to be a mixture of substance and distortion, so self-ratings on tests of anxiety or personality traits may well themselves also be a mixture of substance and response distortion. So the overlap in the VVIQ-1 variance accounted for by SDE and trait anxiety is not straightforward to interpret.

Overall reflection on the research ('macro' perspective)

Overview

The macro level reflection is organised into the following sections:

- Summary of findings;
- Overall contribution the papers have made to the discipline and theoretical base;
- Methodological considerations;
- Publication strategy and measures of impact;
- Suggestions for future research.

1) Summary of findings

The thesis describes a research project carried out across five papers, four that were published journal articles in peer reviewed journals, and one that was a conference paper. The research investigated the relationship between scores on self-report imagery questionnaires and measures of sub-types of socially desirable responding. Researchers have typically argued that self-report imagery questionnaires are not heavily influenced by social desirable responding, e.g. McKelvie (1995b). However, this thesis argues that past research on the relationship between self-report imagery questionnaires and social desirable responding is limited. The five papers making up the thesis systematically explored the relationship between scores on self-report imagery questionnaires and measures of social desirable responding in a series of samples of psychology students tested anonymously in classroom settings. Their results provide empirical data to support the following:

1) Social desirable responding should be conceptualised as a multifactorial phenomenon with the need for distinction between egoistic and moralistic biases being crucial (Paulhus, 2002; Paulhus & Trapnell, 2008);

2) Respondents answering self-report imagery questionnaires most likely view imagery as a

valued ability being tested rather than as a 'value neutral' aspect of their cognition, and therefore will value responses they perceive to indicate 'better' imagery ability. As such egoistic forms of social desirable responding are more of a 'threat' to the validity of self-report imagery questionnaires;

3) Past studies have underestimated the size of the relationship between self-report imagery questionnaires and social desirable responding because they have used inappropriate measures of social desirable responding that primarily measure moralistic forms of bias. The research conducted in the thesis papers made use of the Balanced Inventory of Desirable Responding (BIDR; Paulhus 1994, 1998, 2002) which has separate scales to measure the two biases and so does not have this limitation. Also, most imagery researchers have conceptualised social desirable responding as distortion and overlooked the possibility that it might reflect substantive relationships with imagery;

4) The series of five research papers reported in this thesis have produced evidence that responses to self-report imagery questionnaires show small but consistent positive correlations with the egoistic form of social desirable responding in the approximate range of r=.20 to .35, with 'greater' imagery ability associated with higher levels of the egoistic bias;

5) While there is evidence e.g. Connelly and Chang (2015) that questionnaire measures of social desirable responding are not pure measures and tap a combination of substance and distortion, the findings of the research reported here suggest that the self-report imagery questionnaire–egoistic bias correlations mostly reflect distortion. Specifically: 1) the size of imagery scale–egoistic bias correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which an imagery property was desired; and 2) the finding that the correlation between the Vividness of Visual Imagery Questionnaire (Marks, 1974) and the egoistic bias was found to be largely independent of its correlation with trait anxiety;

6) However, while the thesis data suggests that the relationship between self-report imagery questionnaire responses of valued imagery properties and measures of social desirable responding is larger than previously thought, the size of this relationship does not appear to reach to the extent where it impacts notably on imagery questionnaire performance. Despite the relationship with social desirable responding being larger than previously

thought, self-report imagery questionnaires do still correlate with other related measures. Furthermore, in the thesis, the results of a partial correlational analysis reported in paper 1 suggests that the removing the variance associated with social desirable responding from the relationship between imagery questionnaires largely leaves the effect sizes unaffected. However, this conclusion needs to be confirmed using more advanced statistical techniques such as Multiple Item Response Modelling (MIRM) which have the capacity to weight items and take account of cross-loadings across the items of different questionnaires.

It is argued that the five papers making up the thesis represent original peer reviewed research that has increased our understanding of the relationship between self-report imagery questionnaires and social desirable responding and extended the forefront of the mental imagery questionnaire literature.

2) Overall contribution the papers have made to the discipline and theoretical base

It is argued that the five papers make empirical, theoretical, methodological and practical contributions to the study of the relationship between self-report imagery questionnaire responses and social desirable responding.

2.1) Empirical contributions

Papers 1-5 made several empirical contributions. First, two aspects of the results suggest that at least a major part of the self-report imagery questionnaire correlations with SDE reflects self-enhancing egoistic distortion. These were: 1) the size of imagery scale–SDE correlations were themselves correlated positively with the extent to which participants rated imagery properties as measuring imagery ability and the extent to which an imagery property was desired; and 2) the finding that the correlation between the VVIQ-1 and SDE was found to be largely independent of its correlation with trait anxiety. A further analysis reported in the PhD thesis on the paper 4 data set also provided support for this view. The analysis showed that the 10 items of Kovacic et al's (2014) E-SDR scale, which showed higher correlations with personality trait self-peer discrepancy measures than with the trait peer-ratings, correlated r_s=.23, p<.01 with

the VVIQ-1 when scored dichotomously, and *r*=.40, *p*<.001 scored continuously.

Second, the size of the correlation between self-report imagery questionnaires and social desirable responding is larger than was previously thought. McKelvie (1995b) made two estimates of the degree of relationship between the VVIQ and socially desirable responding. An estimate of *r* of .19 with 95% CI [.08 to .30] based on correlations with the MC scale alone, and an estimate of *r* of .08 with 95% CI [-.04 to .11] based on all available data. Based on these statistics, McKelvie concluded that "…the VVIQ is not seriously contaminated by socially desirable responding, and is perhaps not contaminated at all" (1995b, p. 209). McKelvie (1994) suggested that a correlation of above *r*=.25 indicated 'unacceptable contamination' and so should be the 'cut-off' point for acceptability. This was based on the findings of studies that have calculated the average size of correlations between large numbers of psychological variables, considerations of factors that affect reliability and validity and the suggestions for cut-off points made by other authors. The results from the five thesis papers show that the magnitude of correlations between self-report imagery scales and SDE sometimes exceeded McKelvie's (1994) *r*=.25 criterion. For example, in paper 3 the correlation between the VVIQ-2 and SDE was *r*=-.31.

The size of these correlations is consistent with work by Leising et al (2016) who have developed an index to measure the general tendency to rate oneself more favourably than one is rated by others, and more positively than one rates others. Leising at al found self-report questionnaire item desirability moderated the influence of the index on self-ratings of personality, with the upper limit of correlations for highly desirable traits, such as intelligence, being around *r*=.30. Thus, the size of correlations found by the five papers in this thesis are in line with Leising at al's findings.

In addition to these findings, other empirical contributions included:

1) The SDE correlations were shown to generalized beyond ratings of vividness and to the kind of items used in spatial tests;

2) The Shapes Questionnaire items load on one global imagery factor;

3) The VVIQ-2 shows a similar size correlation with SDE as the VVIQ-1;

4) Participants value visual imagery over auditory imagery, and vividness of visual imagery over a visual thinking style;

5) Some aspects of visual imagery, such as vividness, appear to be valued by participants because they are believed to have the potential to enhance remembering.

Further empirical findings from papers 1-5 will now be discussed as is appropriate in the following sections.

2.2) Theoretical contributions

The papers also made five theoretical contributions to the mental imagery literature. First, before the publication of paper 1, researchers active in the mental imagery area had primarily conceptualised social desirable responding as a unitary construct. However, the five thesis papers have shown that it is important to treat socially desirable responding as a multifactorial phenomenon, and that Paulhus's two-tier model (Paulhus, 2002) of social desirable responding gives an insightful framework for reflecting on the imagery literature and the results of the five papers.

Second, paper 1 was the first to argue that participants answering self-report imagery questionnaires will most likely view imagery as a valued ability being tested. As such the egoistic bias, with its emphasis on claiming positive social and intellectual qualities, is more likely to be a threat to the validity of self-report imagery questionnaires than the moralistic bias.

Third, papers 1-5 argued that, while being a useful contribution McKelvie's criterion of $r=.25$ (6.25% variance), is a somewhat arbitrary cut off point. So rejection of self-report imagery

questionnaires as unacceptably contaminated by socially desirable responding would seem premature given the support for their validity. For example, McKelvie (1995b) reports that reliable relationships exist between the VVIQ-1 and performance on perceptual tasks such as scan path consistency during perception. Thus, even though the VVIQ-1 appears to correlate with SDE, this does not stop it from correlating with performance on these tasks.

A potential explanation for the limited impact of social desirable responding on focal scale tests has been proposed by Paunonen and LeBel (2012). They used Monte Carlo simulation methods to model the effects of social desirable responding on the predictive validity of personality trait scores for performance. Their analysis of computer generated samples for models of social desirable responding suggested that at the individual level the construct validity of a personality test was compromised by social desirable responding such that it became a mixed measure of the trait and desirable responding. However, the test's predictive validity, as measured by correlation coefficients at the group level, were only noticeably affected when social desirable responding reached large effect sizes. Given the relationships seen between self-report imagery questionnaires and performance measures, it would seem likely that this situation also holds true for imagery self-report questionnaires answered under typical experimental conditions. Social desirable responding in the form of the egoistic bias appears likely to add error variance to the measurement of effects but not to the extent that they are notably affected or will not be seen.

Fourth, mental imagery researchers have traditionally conceptualized social desirable responding as distortion or error that should be eliminated or minimized. The five thesis papers have advanced debate in the area by drawing attention to the 'substance versus style' issue and have offered a more informed discussion of the factors that could underlie a relationship between a self-report mental imagery questionnaire and a social desirable responding measure. Possible substantive factors were discussed for example that nonconscious self-inhibitory forces which act to reduce the vividness of images could overlap with self-deceptive enhancement processes.

Fifth, paper 4 drew on insights from Leite and Cooper (2010) to suggest that whether or not a participant's response to a focal scale item is affected by social desirable responding depends on a three-way interaction between the nature of the items making up the focal scale, the respondent's characteristics and the situation that the testing takes place in. In the paper an initial expression of how Leite and Cooper's ideas might apply to Paulhus's (2002) two-tier model was made, while in the thesis a fuller description was presented. The thesis also made use of the three-way interaction framework to organize the description of potential factors that might encourage the occurrence of social desirable responding on self-report imagery questionnaires.

2.3) Methodological contributions

The papers have advanced methodology in this area in five ways. First, they have shown that the MC scale is not appropriate to use in this context because it loads primarily on the moralistic bias. In contrast the BIDR has scales to measure both egoistic and moralistic forms of social desirable responding. However, the BIDR scales themselves have limitations, for example, they appear to tap both substance and style making interpretation of effects complex.

Second, papers 2 and 3 advanced methodology by measuring the extent to which participants rated imagery properties as measuring imagery ability and the extent to which an imagery property was desired, and then relating these measures to the size of imagery scale–SDE correlations. This approach provided evidence that the imagery scale–SDE correlations were at least partially caused by distortion.

Third, papers 4 and 5 found no convincing evidence that the BIDR AM scale significantly improves the ability of SDE to predict VVIQ-1 scores. So it appears that papers 1-3 do not appear to have greatly underestimated the relationship with the egoistic bias by omitting the AM scale.

Fourth, papers 3, 4 and 5 produced evidence relevant to an evaluation of the BIDR dichotomous and continuous scoring methods. The results obtained showed that in an imagery context, neither scoring method showed a clear advantage over the other.

Fifth, paper 1 found that counterbalancing the order of self-report imagery and BIDR questionnaires is not crucial to the results obtained.

2.4) Practical contributions

Self-report imagery questionnaires have been used in a range of applied settings such as clinical psychology (Pearson et al., 2013), sports psychology (Munzert & Lorey, 2013) and medical surgeon training (Sevdalis, Moran & Arora, 2013). The results of the five thesis papers suggests that the use of imagery questionnaires in such settings is not problematical. In terms of experimental technique though, researchers should take care to use standardized approaches when collecting data. Past research has shown that participants' responses on self-report imagery questionnaires can be affected by communicating that vivid mental imagery is desirable or difficult (e.g. Ashton & White, 1975; Ashton, White & Brown, cited in White, Ashton & Law, 1978; and McKelvie, 1979). If such care is not taken, then the conclusions reached in the thesis that imagery questionnaire data is not notably affected by the egoistic bias may not apply.

3) Methodological considerations

The primary research method used in the five papers was the use of standardised psychometric questionnaires. Samples were volunteer undergraduate British psychology students tested under anonymous university classroom conditions by lecturers. The papers used a range of statistical methods, but the main core of the statistical analysis involved correlations or used methods based on correlation.

3.1) Use of the BIDR scales

The main results of the five papers centred around the SDE egoistic scale of the BIDR. Previously, the literature review described how Paulhus (e.g. 1994) has provided a range of sources of support for the validity of the SDE scale and the BIDR scales in general. However, research by Connelly and Chang (2015), Lönnqvist et al (2007) and Pauls and Stemmler (2003) has suggested that under low demand conditions, BIDR scores are influenced by both substance and style. Thus, high SDE scores appear to reflect a mixture of some participants self-enhancing on items and other participants responding honestly. The fact that the BIDR scales tap into both substance and style means that effects observed with the scales are open to more than one interpretation. Results relating to the exact nature that 'substance' takes varied across the studies, though both Lönnqvist et al and Connelly and Chang found SDE to relate to other-rated neuroticism levels.

Furthermore, four recent papers (Dodaj, 2012; Galic & Jerneic, 2013; Galic et al, 2009; and Kovacic et al, 2014) have tested either all four BIDR scales or only the AM and IM scales, using a Croatian version of the BIDR using fake-good instructions in simulated job or student application situations designed to illicit either egoistic or moralistic biases. Their results have produced strong support for the egoistic–moralistic distinction. However, less clear cut support was found for the unconscious–conscious distinction, although this may be due to limits of the BIDR scales rather than the model itself.

3.2) Use of a correlational approach

The five research papers used a range of statistical methods: data screening, transformation, correlation, regression, factor analysis, intra-class correlation, non-parametric ANOVA. However, the main core of the statistical analysis involved correlations or used methods based on correlation.

Correlation coefficients and techniques based on correlation, can be affected by a range of methodological factors such as: multivariate outliers, gaps in distributions, restricted range of

scores, non-linearity and heterocedasticity (Coolican, 2015; Field 2013, & Howell, 2010). In the literature review when discussing the VVIQ-1, it was noted that Killstrom et al (1991) collected responses from 730 American psychology students and found that the distribution had a positive skew with most participants reporting that their images had been at least 'moderately clear and vivid'. Killstrom et al point out that these properties of the VVIQ mean that in a small sample the spread of scores may well not be large and this would decrease the chances of obtaining a correlation between VVIQ scores and another variable. Equally, the distribution of the BIDR scales, especially when scored with the dichotomous scoring, method may lead to low mean scores, skew and a restricted range of scores, particularly in low demand situations. These factors could lead to reduced correlation coefficients and the underestimation of effect sizes. However, in our data, skewed distributions were transformed when they occurred which should have limited the impact of restricted range on correlations. Equally, univariate and multivariate distributions were screened to check for the potential problems described above.

The term 'correlation' can also be used in relation to a type of design. Coolican (2015) explains that in a true experiment an independent variable is manipulated and then a dependent variable is measured while all other variables are kept constant. To do this the researcher has to: 1) be able to control the levels of the independent variable, and 2) randomly allocate participants to groups (if it is an independent groups design) or to a random order of conditions (if it is a repeated measures design). Because the changes in the independent variable come first we can infer that the changes in the independent variable cause the changes in the dependent variable. In a correlational study there is no independent variable or dependent variable, the researcher only measures values on two or more pre-existing variables and no independent variable is manipulated. This makes it more difficult to interpret the results of correlations, because the direction of causality cannot be straightforwardly assumed as it can be in a true experiment. Usually we can only say that a relationship exists between the two variables and speculate as to the direction and nature of causality. One direction of causality can be more likely than another if one variable is *prior* to the other e.g. if height is correlated with success at a relevant sport. This is because it is not possible for success to cause height, so that direction of causality is ruled out. However, it is still possible that a third variable such as a genetic influence could underlie the relationship. These interpretation issues apply to the studies in all the five papers. So, although a

strong and replicable pattern was seen for self-report imagery questionnaire scores to correlate with the egoistic bias, the nature and direction of causality mechanisms behind the relationship can only be inferred.

A final correlational issue that should be considered is the possible effect of common method variance on the results. Common method variance is the 'systematic error variance shared among variables measured with, and introduced as a function of, the same method and/or source' and can stem from various causes (Richardson, Simmering, & Sturman, 2009, p. 763). In the context of the five thesis papers it relates to the use of a questionnaire methodology. It is possible that correlations between different questionnaires could in principle stem from the common measurement method used. For example, if participants adopted extreme or inattentive response sets that were systematic and independent of questionnaire content, this could potentially create spurious correlations. However, one aspect of the thesis results which argues against this having happened is the diverging patterns of imagery questionnaire correlations seen with the egoistic and moralistic BIDR scales.

3.3) Sample and experimental issues

All samples in the five papers were made up of undergraduate psychology students. Student samples have been criticized for lacking generalizability, for example Henrich, Heine and Norenzayan (2010) characterize student participants as 'westernized, educated people from industrialized, rich democracies' (p. 7). In the current context aspects of student characteristics that may act to limit the external validity of the findings include: student participants may not have fully formed their sense of self and they may have a stronger need to be viewed positively by their peers (Sears, 1986). Furthermore, all participants in the research were psychology students and this may have influenced their behaviour in the studies. For example, in paper 3 where participants were asked to describe why they valued aspects of mental imagery, reference to visual imagery's ability to act as a mnemonic may have been influenced by their educationally derived knowledge of aids to memory. Equally, the fact that participants were predominantly female may have been influential. Earlier in the thesis sex differences were described in relation to responses made to the VVIQ-1 and the egoistic and moralistic scales of the BIDR. This raises

the possibility that the sex of the participants might have acted to confound the nature of the results in some way. It is possible that the over-representation of females in the samples coupled with the lower value they attach to egoistic characteristics might have acted to reduce the size of the correlations observed with SDE seen throughout the papers. However, supplementary analyses conducted on the data, but not reported in the published papers, suggest that the findings were not substantially affected by gender. For example, a further analysis of the correlations from paper 4 between the VVIQ-1 and the measures of social desirable responding used (the four BIDR scales and the MC) that controlled for gender, showed that no correlations changed by more than $r=.02$ when gender was taken into account. The matrix from this analysis can be seen in Appendix 7. So, in conclusion it can be said that beyond the impact of sex, it remains to be established whether the results obtained generalize to other samples.

The results were collected under strictly controlled experimental conditions e.g. participants were directed using standardized instructions, to attempt to combat the possibility of experimenter effects. This provides support for the validity of the results obtained. However, it would be helpful for all patterns of results observed in the thesis papers to be retested to confirm that they were reliable effects.

4) Publication strategy and measures of impact

4.1) Publication strategy

The findings of the five papers were published in four journals and one conference proceedings. Publication details can be seen in Table 10. The first two papers were published in imagery journals, the third paper in a personality journal, and the fourth paper in a research methods journal. All four journal articles were blind reviewed by two or more reviewers, while the conference poster was also reviewed prior to being accepted. During the review process for each paper I was the corresponding author. The choice of journals helped to develop our thinking about the topic through gaining feedback from expert reviewers in different fields (mental imagery, individual differences and research methods). For example, a reviewer for

paper 4 brought our attention to the recent application of FMM to the topic of social desirable responding.

Table 10.

Details of the publication strategy

Paper number	Place of publication	Details	Rationale for choice	Journal impact factor
Paper 1	Journal of Mental Imagery	The journal is an interdisciplinary journal established in 1977 that focuses on the phenomenon of mental imagery.	Up to this point, many of the key papers on this topic had been published in this journal thus, it appeared to be a suitable starting point for publication.	No impact factor information currently exists for this journal.
Paper 2	Imagination, Cognition and Personality	The journal is a psychology journal established in 1981, which focuses on the phenomenon of conscious experience, including mental imagery.	Publishing in a second imagery journal had the possibility of widening the audience for the research.	No impact factor information currently exists for this journal.
Paper 3	Journal of Individual Differences	The journal is a psychology journal which focuses on individual differences in all their forms in humans and animals i.e. in cognition, emotion, behavior, and developmental aspects.	This publication broadened the audience for our work to researchers working in the area of personality.	The journal's impact factor for 2014 is 1.222, its five-year impact factor is 2.378.
Paper 4	Behavior Research Methods	The journal is a psychology journal which focuses on methods, techniques and instrumentation used in experimental psychology research.	This publication broadened the audience for our work to researchers working in the area of psychological research methods.	The journal's impact factor for 2014 is 2.928, its five-year impact factor is 3.900.
Paper 5	2009 British Psychological Society (BPS) Conference, Brighton	The BPS is the representative body for psychology and psychologists in the UK i.e. does things that promote psychology & help psychologists. The BPS was formed in 1901 and has more than 45,000 members. The BPS conference is an annual event and is the	Paper 5 presented initial findings from the first study to explore possible substantive explanations for the self-deceptive enhancement – egoistic bias correlation.	NA

organisation's flagship
conference.

4.2) Measures of impact

The impact of the papers can be judged by a range of metrics including: journal impact factor, citation reports, number of downloads, Research Gate interest in the work, requests for copies of the papers sent to the first author by email and invitations to the first author to act as a reviewer for journals. Details of citation reports conducted on the 1st May 2018 using Web of Science, Scopus, and Google Scholar are shown in Table 11 and total 22 papers. The reports show a modest degree of impact suggesting that the research on self-report imagery questionnaires is a niche area. The papers were primarily cited by researchers working within the area of imagery self-report questionnaires themselves. Impact of the papers is also shown by the fact that publishing in this field lead me to act as a reviewer for Behavior Research Methods and Perceptual and Motor Skills. Thus editors of these journals noted our work and sought our input when related articles were submitted to them. Since registering on Research Gate on the 30th May 2016 to the 19th July, 2018 my research has been read by over 400 visitors (see Appendix 6).

Table 11.

Details of papers citing the research

Thesis paper	Citing paper
1	1. Campos, A. (2011) Internal consistency and construct validity of two versions of the revised vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 113,* 2, 454-460. 2. Jacobs, K.E. & Roodenburg, J. (2014) The development and validation of the Self-Report Measure of Cognitive Abilities: A multitrait-multimethod study. *Intelligence, 42*, 1, 5-21.

2	1.	Campos, A. (2011) Internal consistency and construct validity of two versions of the revised vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 113,* 2, 454-460.
	2.	Young, R. L. (2011) *Predictors of client responsiveness to the Bonny Method of Guided Imagery and Music (BMGIM)*. Unpublished PhD thesis, Temple University.
3	1.	Campos, A. & Jose Perez-Fabello, M. (2009) Psychometric quality of a revised vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 108*, 3, 798-802.
	2.	Willander, J. & Baraldi, S. (2010) Development of a new Clarity of Auditory Imagery Scale. *Behavior Research Methods, 42*, 3, 785-790.
	3.	Hubbard, T. L. (2010) Auditory Imagery: Empirical Findings. *Psychological Bulletin, 136*, 2, 302-329.
	4.	Campos, A. (2011) Internal consistency and construct validity of two versions of the revised vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 113,* 2, 454-460.
	5.	Young, R. L. (2011) *Predictors of client responsiveness to the Bonny Method of Guided Imagery and Music (BMGIM)*. Unpublished PhD thesis, Temple University.
	6.	Goins, R. R. (2012) *An examination of guided imagery and its relationship to self-efficacy and the implications for employability in college graduates seeking employment*. Unpublished PhD thesis Harold Abel School of Social and Behavioral Sciences.
	7.	Hubbard, T.L. (2013) Auditory aspects of auditory imagery. In: S. Lacey, & R. Lawson (Eds.), *Multisensory Imagery* (pp. 271-283). New York: Springer.
	8.	Runge, M., Bakhilau, V., Omer, F. & D'Angiulli, A. (2015) Trial-by-trial vividness self-reports versus VVIQ: a meta-analytic comparison of behavioral, cognitive and neurological correlations*. Imagination, Cognition and Personality: Consciousness in Theory, Research & Clinical Practice, 35*, 2, 137-165.
4	1.	Lacey, S. & Lawson, R. (2013) Imagery Questionnaires: Vividness and beyond. In: S. Lacey, & R. Lawson (Eds.), *Multisensory Imagery* (pp. 271-283). New York: Springer.
	2.	Madan, C. R. & Singhal, A. (2014) Improving the TAMI for use with athletes. *Journal of Sports Sciences, 32*, 14, 1351-1356.
	3.	Madan, C. R. & Singhal, A. (2014) Introducing TAMI: an objective test of ability in movement imagery. *Journal of Motor Behavior, 45*, 2, 153-166.
	4.	Madan, C. R. & Singhal, A. (2014) Improving the TAMI for use with athletes. *Journal of Sports Sciences, 32*, 14, 1351-1356.
	5.	Spiller, M. J., Jonas, C. N. & Simner, J.(2015) Beyond visual imagery: How modality-specific is enhanced mental imagery in synesthesia? *Consciousness and Cognition, 31*, 73-85.
	6.	Runge, M., Bakhilau, V., Omer, F. & D'Angiulli, A. (2015) Trial-by-trial vividness self-reports versus VVIQ: a meta-analytic comparison of behavioral, cognitive and neurological correlations*. Imagination, Cognition and Personality: Consciousness in Theory, Research & Clinical Practice, 35*, 2, 137-165.
	7.	Sawczak, C. M. (2015) *Mechanisms underlying the prosocial construction effect: detail and self-referential processing*. Unpublished PhD thesis University of Toronto.
	8.	Runge, M., Bakhilau, V., Omer, F. & D'Angiulli, A. (2015) Trial-by-trial vividness self-reports versus VVIQ: a meta-analytic comparison of behavioral, cognitive and neurological correlations*. Imagination, Cognition and Personality: Consciousness in Theory, Research & Clinical Practice, 35*, 2, 137-165.

9. Bartels, R. M., Harkins, L. & , Beech, A. R. (2017) The Influence of Fantasy Proneness, Dissociation, and Vividness of Mental Imagery on Male's Aggressive Sexual Fantasies. *Journal of Interpersonal Violence*. Advanced online publication.

10. Wilson, A.C., Schwannauer, M. ,McLaughlin, A., Ashworth, F. & Chan, S. W. Y. (2018) Vividness of positive mental imagery predicts positive emotional response to visually presented Project Soothe pictures. *British Journal of Psychology, 109*, 2, 259-276.

5) Suggestions for future research

Suggested directions for future work include:

1) Several authors have suggested that it may be possible to statistically control for social desirable responding using advanced statistical techniques. They have argued that social desirable responding may be able to create statistical effects such as suppression or moderation, but that approaches based on advanced statistical techniques are needed to show these effects. It would be important to test techniques such as multidimensional item response modeling (MIRM) and Factor Mixture Modelling (FMM) in the self-reported imagery questionnaire domain to confirm the conclusions made from the thesis data. Such studies would require very large sample sizes. If effective, FMM for example, would identify the sub-group of participants whose focal scale responses were affected by the egoistic bias, and removal of these participants would serve to increase the size of the relationship between a self-report imagery questionnaire and a criterion variable.

2) Replicate paper 4 using one or more social desirable responding indexes of departure from reality. Potential indexes that might be used are listed in Table 12 with brief details describing them. If the argument presented in the thesis that relationship between the self-report imagery questionnaires and the egoistic bias seen in the five thesis papers primarily reflects distortion is correct, then we would expect that the indexes of departure from reality would show similar relationships imagery questionnaires to the BIDR SDE scale.

Our research group has made a start on this project. Allbutt and Ling (2008) presented a conference poster of early pilot work using the VVIQ-1, BIDR egoistic scales and the OCQ with a sample of 60 participants. This paper has not been included in the body of the thesis because only

the abstract was published in the conference proceedings and so the paper does not qualify for inclusion in the thesis as published paper under the University of Salford regulations. The OCQ was scored with a bias index made up of unweighted hits plus false alarms which produced a correlation with the VVIQ-1 of only $r(58)=-.12$, $p=.188$, 1-tailed. Since then we have collected a larger (n=171) unpublished data set and conducted a more advanced analysis on it. The OCQ bias score was calculated by first converting hits and false alarms to Z scores so that each carried equal weight, and the false alarms measure was also included in the analysis. Initial analysis of this larger data set showed significant, correlations between the VVIQ-1 and several egoistic bias measures: SDE dichotomously scored $r_s=-.39$, $p<.001$; SDE continuously scored $r_s=-.34$, $p<.001$; AM continuously scored $r_s=-.19$, $p<.05$; OCQ bias score $r_s=-.21$, $p<.01$; and OCQ false alarms $r_s=-.24$, $p<.001$. Greater vividness of imagery was always associated with greater overclaiming. The correlation matrix for this data analysis can be seen in Appendix 8. Also, though not an index of departure from reality, it would also be valuable to correlate self-report imagery questionnaire scores with Connelly and Chang's (2015) index of the egoistic bias as common method variance on relevant big 5 trait scales.

Table 12.

Social desirable responding indices of departure from reality

Measure	Details
Residual scores derived from a regression of other- onto self-ratings of egoistic scores	Several authors have used a regression 'residual approach' to attempt to isolate the part of the variance in SDE scores which indicate a departure from 'reality' as indexed by other-ratings (e.g. Pauls & Stemmler, 2003). The rationale behind this approach is that the estimates of knowledgeable others, such as peers and family, act as objective standards to which self estimates can be compared and so the residuals reflect departure from social reality. Deriving the index though, requires sampling other- as well as self-ratings.
Paulhus et al's (2003) measure of academic general knowledge and knowledge over-claiming, the over-claiming questionnaire (OCQ)	The OCQ is a 150 item measure where respondents rate their familiarity with items from 10 academically oriented general knowledge domains, for example historical names and events, fine arts, language, books and poems. Paulhus et al. (2003) has presented the OCQ over-claiming indexes as measuring unconscious egoistic bias.

Kwan et al (2004, 2008) and Leising et al's (2016) index	Kwan et al (2004, 2008) and Leising et al (2016) have developed an index to measure the general tendency to rate oneself more favourably than one is rated by others, and more positively than one rates others. Deriving the index though, even using Leising et al's streamlined video methodology, is a lengthy process.

3) Replicate paper 5 but with a larger sample and more extensive range of questionnaires. These might include: the Speilberger State-Trait Anxiety Inventory (Speilberger et al, 1983) to measure both state and trait anxiety; and the NEO Personality Inventory-Revised (NEO PI-R) (Costa & McCrae, 1992). The NEO PI-R measures the big five personality traits and 6 facets for each of the personality traits. Exploration of the VVIQ-1's relationship with the big 5 facets could reveal the fine grained detail needed to understand exactly which factors underlie its relationship with the big 5 personality traits. An ideal study would include not only self-ratings on these scales, but also other-ratings. The expectations for a data set of this kind would be that self-report imagery questionnaires would generate two types of effects: 1) correlations with personality traits that actually reflect egoistic distortion, and 2) correlations with personality traits that are substantive in nature. For example, self-report imagery questionnaires would be expected to correlate with self-reports of big five traits that include facets reflecting confidence in abilities (conscientiousness, emotional stability, extraversion and also openness if it is operationalized to emphasize intellectual and creative aspects). This is because these facets would tap into the egoistic bias. Substantive relationships might include relationships with both self- and other-reported trait and state anxiety, and neuroticism as suggested by the results of paper 5. Also, imagery questionnaires might correlate with self- and other reported artistic aspects of openness because artistic individuals may genuinely possess enhanced imagery ability.

4) Test for social desirable responding suppressor and moderation effects in the self-reported imagery questionnaire domain. McKelvie's (1995b) meta-analytical review reported relationships between the VVIQ-1 and other self-report imagery questionnaires, perceptual measures and memory measures that were argued to provide evidence to support the construct validity of the VVIQ. Also strong relationships with the VVIQ-1 have been observed

from psychophysiological recording techniques. Repeating some of these studies and testing for suppressor and moderation effects using the BIDR egoistic scales, the OCQ or other indexes of departure from reality listed in Table 8 would provide valuable data. However, following from the findings of researchers such as Paunonen and LeBel (2012), we might only expect such statistical effects where there is a strong relationship between the self-report imagery tests and criterion measure and where advanced statistical tests are used.

5) Explore whether combining egoistic bias measures, such as the BIDR SDE and AM scales with the OCQ measures, through the creation of factor sores might capture the egoistic bias more effectively than single scales alone.

6) Test the BIDR scoring method (continuous versus dichotomous) against one or more social desirable responding indexes of departure from reality described earlier. Paulhus (1991, 1994) advocated use of the dichotomous method for the BIDR scales arguing that it has the advantage of only scoring clearly exaggerated or managed responses, thus it involves not just an endorsement of desirable features and rejection of undesirable features, but it involves a departure from reality. If Paulhus is correct, then when the BIDR egoistic scales are scored dichotomously they should show significantly larger correlations with the indexes of departure from reality than when continuously scored.

7) Modern theory driven self-report imagery questionnaires are an important development in the imagery questionnaire field. Paper 2 tested the Shapes Questionnaire's relationship with the BIDR scales. In the future it would be informative to test other theory driven questionnaires such as Blajenkova and Kozhevnikov's (2009) Object-Spatial Imagery and Verbal Questionnaire (OSIVQ) and Blajenkova's (2016) Vividness of Object and Spatial Imagery Questionnaire (VOSI). Here it would be predicted that, because these questionnaires appear likely to tap into aspects of imagery valued by participants, that they should also show significant correlations with measures of the egoistic bias.

8) Paper 3 observed that the size of the self-report imagery questionnaire–SDE correlation was associated with the extent to which the imagery property was desired. A further and more

stringent test of this effect would be to use Shehan's QMI or Andrade et al's (2014) Plymouth Sensory Imagery Questionnaire (Psi-Q), which both measure vividness of imagery across seven sensory modalities, but this time in a single sample. For this study, following from research on the hierarchy of the senses by e.g. San Roque et al (2015), it would be predicted that the most valued aspects of imagery would show the largest relationships with the egoistic bias and the least valued the lowest i.e. visual imagery would show the largest relationship, auditory imagery the next largest, and the other sensory modalities the smallest relationships.

9) It would be valuable to replicate the effects seen in the thesis papers with a more diverse and representative sample than undergraduate psychology students. It is possible that even other groups of students, might possess a different profile of imagery sensory values which might affect the pattern of results seen. For example, students studying sports-related subjects might value kinesthetic imagery more highly than students studying other academic subjects, and this could affect the pattern of results seen with the egoistic bias.

10) The research project could broaden its scope to look at other threats to self-report imagery questionnaire validity such as inattentive responding. Some imagery questionnaires such as the Shapes Questionnaire are quite lengthy to complete on their own, or when combined with other questionnaires. This raises the possibility that some participants might not give the questionnaires their full attention. McKibben and Silvia (2015) used latent class analysis (an advanced statistical procedure similar in nature to FMM) to identify a subgroup of participants (24% of the sample) who scored highly on multiple measures of inattentive responding when answering batteries of self-report scales of creativity and artistic knowledge presented on-line and towards the end of an academic semester. When these participants were removed from the data set, correlations between openness, creativity and artistic knowledge questionnaires increased by approximately $r=.10$. Here we might predict a similar effect to occur between imagery self-report questionnaires and a criterion variable, particularly where a large battery of questionnaires is used in a study and / or when student participants are tested at a time when they are simultaneously under pressure to meet academic assessment deadlines.

11) Finally, the approach used here to explore social desirable responding might be extended to the self-assessment of other cognitive abilities. Recently Jacobs and Roodenburg (2014) took the first steps in developing a self-report measure of cognitive abilities. They argued that such a measure could be useful in providing a practical way to screen the full range of cognitive abilities to identify possible areas of weakness which could then be followed up with more conventional targeted objective tests. However, given the findings presented in this thesis, it would also seem prudent to test such a self-report measure for the presence of relationships with measures of the egoistic bias. Given that most cognitive abilities are likely to reflect abilities participants will value, it would be expected that self-report measures of these abilities would also correlate with the egoistic bias.

6) End note

In conclusion it is argued that this thesis meets the HEQ descriptors for qualifications at doctoral level, and the BPS (2008) criteria for assessing PhD theses in psychology. The HEQ states that work at doctoral level should generate new knowledge through conducting original research. Such work should generate a substantial body of knowledge which is at the forefront of the academic discipline. The BPS states that for a PhD by publication 'Normally, a minimum of three separate publications will be required and that the publications should constitute a coherent programme of research and make a significant contribution to knowledge.' (p55). It is argued that the five papers presented in this thesis have met these criteria.

References

Abele, A. and Wojciszke, B. (2018). The big two of agency and communion as an overarching framework in psychology. In A. E. Abele & B. Wojciszke (Eds.), *Agency and communion in social psychology (pp. 1-12)*. London: Routledge.

Ahsen, A. (1985). Unvividness paradox. *Journal of Mental Imagery, 9,* 3, 1-18.

Ahsen, A. (1986). Prologue to unvividness paradox. *Journal of Mental Imagery,* 10, 1, 1-8.

Ahsen, A. (1987). Epilogue to unvividness paradox. *Journal of Mental Imagery, 11, 1,* 13-60.

Ahsen, A. (1988). Imagery, unvividness paradox and the paradigm of control. *Journal of Mental Imagery*, *12*, 3&4, 1-44.

Ahsen, A. (1990). AA-VVIQ and imagery paradigm: Vividness and unvividness issue in VVIQ research programmes. *Journal of Mental Imagery, 14*, 3&4, 1-58.

Allbutt, J., Shafiullah, M. & Ling, J. (2006a). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, *30,* 1&2, 1-20.

Allbutt, J., Ling, J. & Shafiullah, M. (2006b). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery. *Imagination, Cognition and Personality, 25*, 4, 337-353. http://dx.doi.org/10.2190/Y067-0726-M45W-37W8

Allbutt, J., Ling, J., Heffernan, T., M., & Shafiullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and visual thinking style. *Journal of Individual Differences, 29,* 4, 181-188. http://dx.doi.org/10.1027/1614-0001.29.4.181

Allbutt, J. & Ling, J. (2009) The relationship between self-report imagery questionnaire scores, the egoistic sub-type of social desirable responding and state-trait anxiety. *British Psychological Society Conference,* Brighton.

Allbutt, J., Ling, J., Rowley, M. & Shafiullah, M. (2011) Vividness of visual imagery and social desirable responding: Correlations of the Vividness of Visual Imagery Questionnaire with the Balanced Inventory of Desirable Responding and the Marlowe-Crowne. Behavior Research Methods, 43, 3, 791-799. http://dx.doi.org/10.3758/s13428-011-0086-8

Amati, F., Oh, H., Kwan, V. S. Y., Jordan, K. & Keenan, J. P. (2010) Overclaiming and the medial prefrontal cortex: A transcranial magnetic stimulation. *Cognitive Neuroscience, 1*, 4, 268–276. https://doi.org/10.1080/17588928.2010.493971

Anderson, D.C. & Buyer, L.S. (1994). Is imagery a functional component of the "bizarre imagery" phenomenon? *American Journal of Psychology, 107*, 207-222. http://dx.doi.org/10.2307/1423037

Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014) Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. *British Journal of Psychology, 105*, 4, 547-563. http://dx.doi.org/10.1111/bjop.12050

Antonietti, A. & Giorgetti, M. (1998). The Verbalizer-Visualizer Questionnaire: A review. Perceptual and Motor skills, 86, 227-239. http://dx.doi.org/10.2466/pms.1998.86.1.227

Ashton, R. & White, K. (1975). The effects of instructions on subjects' imagery questionnaire scores. *Social Behaviour and Personality, 3,* 41-43. http://dx.doi.org/10.2224/sbp.1975.3.1.41

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man.* Boston: Beacon Press.

Barrios, V., Kwan, V. S. Y., Ganis, G., Gorman, J., Romanowski, J. & Keenan, J. P. (2008) Elucidating the neural correlates of egoistic and moralistic self-enhancement. *Consciousness and cognition, 17*, 2, 451-456. https://doi.org/10.1016/j.concog.2008.03.006

Belardinelli, O. M., Palmiero, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., D'Ausilio, A., Ferretti, A.,Del Gratta, C. & Romani, G. L. (2009). An fMRI investigation on image generation in different sensory modalities: The influence of vividness. *Acta Psychologica*, *132*, 190-200. http://dx.doi.org/10.1016/j.actpsy.2009.06.009

Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1974). *Manual for the Differential Aptitude Tests, Form S and T*. New York, NY: The Psychological Corporation.

Bent, N. A. & Wick, E. (2006). Beyond vividness: Parental filters as moderators in mental imagery and measured anxiety level. *Journal of Mental Imagery*, *30*, 1&2, 21-38.

Beretvas, S. N., Meyers, J., L. & Leite, W. L. (2002) A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Education and Psychological Measurement, 62*, 570-589. http://dx.doi.org/10.1177/0013164402062004003

Berry, C. M., Page, R. C. & Sackett, P. R. (2007) Effects of self-deceptive enhancement on personality – job performance relationships. *International Journal of Selection and Assessment, 15*, 1, 94-109. https://doi.org/10.1111/j.1468-2389.2007.00374.x

Betts, G. H. (1909). *The distributions and functions of mental imagery*. New York, NY: Teachers College. Columbia University.

Blajenkova, O. (2016) Vividness of object and spatial imagery. *Perceptual and Motor Skills, 122*, 2, 490-508. http://dx.doi.org/10.1177/0031512516639431

Blajenkova, O. & Kozhevnikov, M. (2009) The new object-spatial-verbal cognitive style model: theory and measurement. *Applied Cognitive Psychology, 23*, 5, 638-663. http://dx.doi.org/10.1002/acp.1473

Bonett, D. G., & Price R. M. (2005). Inferential methods for the tetrachoric correlation coefficient. *Journal of Educational and Behavioral Statistics, 30,* 213. http://dx.doi.org/10.3102/10769986030002213

Borkenau, P. & Ostendorf, F. (1989) Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality, 3,* 31-45. http://dx.doi.org/10.1002/per.2410030105

British Psychological Society (2008) *Guidelines for Assessment of the PhD in Psychology and Related Disciplines, Revised version*. Leicester: British Psychological Society.

Callow, N., & Roberts, R (2010). Imagery research: An investigation of three issues. *Psychology of Sport & Exercise, 11,* 4, 325-329. https://doi.org/10.1016/j.psychsport.2010.03.002

Callow, N., Roberts, R. J., Hardy, L. J., Jiang, D., & Edwards, M. G. (2013). Performance improvements from imagery: evidence that internal visual imagery is superior to external visual imagery for slalom performance. *Frontiers in Human Neuroscience*, *7,* 697. https://doi.org/10.3389/fnhum.2013.00697

Campos, A. (2014) Gender differences in imagery. Personality and Individual Differences, 59, 107-111. http://dx.doi.org/10.1016/j.paid.2013.12.010

Campos, A., and Sueiro, E. (1993). Sex and age differences in visual imagery vividness. *Journal of Mental Imagery, 17,* 3&4, 91-94.

Carmichael, L. Hogan, H.P. & Walter. A.A. (1932) An experimental study of the effect of language on the production of visually perceived form. *Journal of Experimental psychology, 15*, 73-86. http://dx.doi.org/10.1037/h0072671

Cattell, R.B. & Scheier, I.H. (1961). Extension of meaning of objective test personality factors: Especially into anxiety, neuroticism, questionnaire, and physical factors. *Journal of General Psychology*, *61*, 287-315. http://dx.doi.org/10.1080/00221309.1959.9710264

Cervellione, K.L., Lee, Y-S., & Bonanno, G. A. (2009). Rasch modelling of the Self-Deception Scale of the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement, 69*, 3, 438-458. http://dx.doi.org/10.1177/0013164408322020

Chara, P. J. and Hamm, D. A. (1989). An inquiry into the construct validity of the vividness of visual imagery questionnaire. *Perceptual and Motor Skills, 69*, 127-136. http://dx.doi.org/10.2466/pms.1989.69.1.127

Chara, P.J. & Verplanck, W.S. (1986). The Imagery Questionnaire: An Investigation of its Validity. *Perceptual and Motor Skills*, *63*, 915-920. http://dx.doi.org/10.2466/pms.1986.63.2.915

Clark-Carter, D. (1997). *Doing quantitative psychological research: From design to report.* Hove: Psychology Press.

Cohen, B. H., & Saslona, M. (1990). The advantage of being an habitual visualiser. *Journal of Mental Imagery*, *14*, 3&4, 101-112.

Comer, R. J. (2015) *Abnormal Psychology*. 9th edition. New York, NY: Worth Publishers.

Conger, A. J. and Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and psychological measurement, 32*, 579-599. http://dx.doi.org/10.1177/001316447203200303

Connelly, B. S. & Chang, L. (2015) A meta-analytic multirater separation of substance and style in social desirability scales. *Journal of Personality, 84*, 3, 319-334. https://doi.org/10.1111/jopy.12161

Coolican, H. (2015) *Research Methods and Statistics in Psychology*. London: Hodder and Stoughton.

Cornoldi, C., De Beni, R., Giusberti, F., Marucci, F., Massironi, M., & Mazzoni, G. (1991). The study of vividness of imagery. In R.H. Logie & M. Denis (Eds.), *Mental images in cognition* (pp. 305-313). North Holland: Elsevier Science Publishers B.V.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor (NEO-FFI) Inventory professional manual*. Odessa, FL: PAR.

Cox, R. H. (2012). *Sport psychology: concepts and applications.* 7th ed. New York, NY: McGraw-Hill.

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements*. New York, NY: Wiley.

Crowne, D.P. (1979). *The experimental study of personality.* Hillsdale, NJ: Erlbaum.

Crowne, D.P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354. http://dx.doi.org/10.1037/h0047358

Crowne, D.P. & Marlowe, D. (1964). *The approval motive.* New York, NY: Wiley.

Cui, X., Yang, C. B., Montague, P. R., & Eagleman, D. M. (2007) Vividness of mental imagery: individual variability can be measured objectively. *Vision Research, 47*, 474-478. http://dx.doi.org/10.1016/j.visres.2006.11.013

Cunningham, M. R., Wong, D. T. & Barbee, A. P. (1994) Self-presentation dynamics on overt integrity tests: experimental studies of the Reid Report. *Journal of Applied Psychology, 79*, 5, 643-658. https://doi.org/10.1037/0021-9010.79.5.643

Davies, M. F., French, C. C. & Keogh, E. (1998). Self-deceptive enhancement and impression management correlates of EPQ-R dimensions. *Journal of Psychology, 132*, 4, 401-406. http://dx.doi.org/10.1080/00223989809599274

Davis, D. & Loftus, E. (2014). Repressed memories. In R. L. Cautin & S. L. Lilienfeld (Eds.), The Encyclopedia of Clinical Psychology (pp. 1–3). Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781118625392.wbecp270

de Vries, R. E., Zettler, I & Hilbig, B. E. (2014) Rethinking trait conceptions of social desirability scales: impression management as an expression of honesty-humility. *Assessment, 21*, 3, 286-299. https://doi.org/10.1177/1073191113504619

Dean, G. & Morris, P. E. (2003). The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology, 94*, 245-273. http://dx.doi.org/10.1348/000712603321661912

Dean, G. & Morris, P. E. (1991). Imagery and spatial ability: When introspective reports predict performance. In R. H. Logie and M. Denis (Eds.), *Mental images in human cognition* (pp. 331-347). Amsterdam: Elsevier Science. http://dx.doi.org/10.1016/S0166-4115(08)60522-1

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 6, 1246-1256. http://dx.doi.org/10.1037/0022-3514.73.6.1246

Djikic, M., Peterson, J. B. & Zelazo, P. D. (2005) Attentional biases and memory distortions in self-enhancers. *Personality and Individual Differences, 38*, 3, 559-568. https://doi.org/10.1016/j.paid.2004.05.010

Dodaj, A. (2012) Social desirability and self-reports: testing a content and response-style model of socially desirable responding. *Europe's Journal of Psychology, 8*, 4, 651-666. http://dx.doi.org/10.5964/ejop.v8i4.462

Domino, G. & Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511813757

Dunlop, P. D., Bourdage, J. S., de Vries R. E., Hilbig, B. E., Zettler, I. & Ludeke, S. G. (2017) Openness to (reporting) experiences that one never had: Overclaiming as an outcome of the knowledge accumulated through a proclivity for cognitive and aesthetic exploration. *Journal of Personality & Social Psychology, 113*, 5, 810-834. https://doi.org/10.1037/pspp0000110

Durndell, A.J. & Wetherick, N.E. (1975). Reported imagery and social desirability. *Perceptual & Motor Skills, 41*, 987-992. http://dx.doi.org/10.2466/pms.1975.41.3.987

Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.

Edwards, L. K. & Clark, C. L. (1987). Social desirability values and dispersions for items in two SD scales: The MMPI SD scale and an experimental SD scale. *Psychological Reports, 60*, 1083-1086. http://dx.doi.org/10.2466/pr0.1987.60.3c.1083

Edwards, A.L., Diers, C.J. & Walker, J.N. (1962). Response sets and factors loadings on sixty-one personality scales. *Journal of Applied Psychology, 46*, 220-225. http://dx.doi.org/10.1037/h0040280

Embretson, Susan E.; Reise, Steven P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Eton, D. T., Gliner, F. H. & Munz, D. C. (1998). The measurement of imagery vividness: A test of the reliability and validity of the Vividness of Visual Imagery Questionnaire and the Vividness of Movement Questionnaire. *Journal or Mental Imagery*, 22, 3&4, 125-136.

Euse, F. J., & Haney, J. N. (1975). Clarity, controllability, and emotional intensity of image: Correlations with introversion, neuroticism, and subjective anxiety. *Perceptual Motor Skills*, *40*, 443–447. http://dx.doi.org/10.2466/pms.1975.40.2.443

Eysenck, H. J. & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult).* London: Hodder & Stoughton.

Eysenck, M. W. (2012). *Fundamentals of cognition*. New York, NY: Psychology Press.

Fan, J., Wong, C. C., Carroll, S. A., & Lopez, F. J. (2008). An empirical investigation of the influence of social desirability on the factor structure of the Chinese 16PF. *Personality and Individual Differences, 45,* 8, 790-795. https://doi.org/10.1016/j.paid.2008.08.008

Farah, M. J., Peronnet, F., Weisberg, L. L. & Monheit, M. (1988) Brain activty underlying mental imagery: event-related potentials during image generation. *Journal of Cognitive Neuroscience, 1*, 4, 302-316. http://dx.doi.org/10.1162/jocn.1989.1.4.302

Farah, M. J., Steinhauer, S. R., Lewicki, M. S. & Zubin, J. & Peronnet, F. (1988) Invdidual differences in the vvidness of mental imagery. *Psychophysiology, 25*, 444-445.

Finke, R.A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.

Gale, A., Morris, P. E., Lucas, B., & Richardson, A. (1972). Types of imagery and imagery types: An EEG study. *British Journal of Psychology*, *63*, 4, 523–531. http://dx.doi.org/10.1111/j.2044-8295.1972.tb01302.x

Galic, Z. & Jerneic, Z. (2013) Measuring faking on the five-factor personality questionnaires: the usefulness of communal and agentic management scales. *Journal or Personnel Psychology, 12*, 3, 115-123. http://dx.doi.org/10.1027/1866-5888/a000087

Galton, F. (1883). *Inquiries into Human Faculty and its Development*. London: Macmillan. http://dx.doi.org/10.1037/14178-000

Ganis, G. (2013) Visual mental imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory imagery*. (pp. 9-28). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-5879-1_2

Ganster, D.C., Hennessey, H.W. & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal, 26*, 321-331. http://dx.doi.org/10.2307/255979

Gignac, G. E. (2013) Modelling the Balanced Inventory of Desirable Responding: Evidence in favor of a revised model of Socially Desirable Responding. *Journal of Personality Assessment, 95*, 6, 645-656. http://dx.doi.org/10.1080/00223891.2013.816717

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology, 39*, 156-167. http://dx.doi.org/10.1111/j.2044-8295.1949.tb00215.x

Gralton, M. A., Hayes, Y. A. & Richardson, T. E. (1979) Introversion-extraversion and mental imagery. *Journal of mental imagery, 3*, 1-10.

Griffith, R. L., & Peterson, M. H. (2008) The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 3, 308-311. https://doi.org/10.1111/j.1754-9434.2008.00053.x

Guissurason, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery*, *16*, 3&4, 117-122.

Hall, C. R. & Martin, K. A. (1997). Measuring movement imagery abilities: A revision of the Movement Imagery Questionnaire. *Journal of Mental Imagery, 21*, 1-2, 143-154.

Hart, C. M., Ritchie, T., Hepper, E. G. & Gebauer, J. E. (2015) The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open, 5*, 4, 1-9. https://doi.org/10.1177/2158244015621113

Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 21-40). Boston, MA: Kluwer. http://dx.doi.org/10.1007/978-1-4615-4397-8_2

Henrich, J. Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 2-3, 61-83. https://doi.org/10.1017/S0140525X0999152X

Hirsch, C. R. (2003) Self-images can play a causal role in social phobia. *Behaviour Research and Therapy 41,* 8, 909–92. https://doi.org/10.1016/S0005-7967(02)00103-1

Hiscock, M. (1978). Imagery assessment through self-report: What do imagery questionnaires measure? *Journal of Consulting and Clinical Psychology, 46*, 2, 223-230. http://dx.doi.org/10.1037/0022-006X.46.2.223

Hishitani, S., Miyazaki, T. & Motojama, H. (2011) Some mechanisms responsible for the vividness of mental imagery: Suppressor, closer and other functions. *Journal of Mental Imagery, 35*, 1&2, 5-32.

Hochman, J. (1994). Ahsen's image psychology. *Journal of Mental Imagery, 18,* 1-118.

Holden, R. (2010). Social desirability. In B. Weiner & W. E. Craighead, W. E. (Eds), *The Corsini Encyclopedia of Psychology, Volume 1*, 4th Edition (pp. 912-913). New York, NY: Wiley. http://dx.doi.org/10.1002/9780470479216.corpsy0889

Holden, R. R. & Passey, J. (2010) Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49*, 5, 446-450. https://doi.org/10.1016/j.paid.2010.04.015

Holden, R. R., & Passey, J. (2009). Social desirability. In M. R. Leary, & R. H. Hoyle (Eds), Handbook of Individual Differences in Social Behavior (pp. 441-454). New York, NY: Guilford Press.

Holden, R. R. (2007) Socially desirable responding does moderate scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science 39*, 3, 184-201. http://dx.doi.org/10.1037/cjbs2007015

Holtgraves , T. (2004) Social desirability and self-reports: Testing models of social desirable responding. *Personality and Social Psychology Bulletin, 30*, 161-172. http://dx.doi.org/10.1177/0146167203259930

Howell, D. C. (2010) *Statistical methods for psychology*. Belmont, CA: Cengage Wadsworth.

Howitt D, Cramer D (2011). *Introduction to Research Methods in Psychology*. Harlow: Pearson.

Huckabee, M. W. (1974) Introversion-extraversion and imagery. *Psychological Reports, 34*, 453-454. http://dx.doi.org/10.2466/pr0.1974.34.2.453

Isaac, A., Marks, D.F. & Russell, D.G. (1986). An instrument for assessing imagery of movement: The Vividness of Movement Imagery Questionnaire. *Journal of Mental Imagery, 10*, 23-30.

Jackson, D.N. & Messick, S. (1962). Response styles on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 65, 285-299. http://dx.doi.org/10.1037/h0045340

Jacobson, L.I., Kellogg, R.W., Cauce, A.M., & Slavin, R.S. (1977). A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society, 9*, 109-110. http://dx.doi.org/10.3758/BF03336944

Jiang, D., Edwards, M. G., Mullins, P., & Callow, N. (2015). The neural substrates for the different modalities of movement imagery. *Brain and Cognition*, *97*, 22-31. https://doi: 10.1016/j.bandc.2015.04.005

John, O.P. & Paulhus, D.L. (2000). *The structure of self enhancement*. Unpublished manuscript, University of California, Berkeley.

Jutte, R. (2005) *A history of the senses: From antiquity to cyberspace*. Malden: Polity Press.

Katz, A.N. (1983). What does it mean to be a high imager? In: J.C. Yuille (Ed.), *Imagery, memory and cognition* (pp. 39-63). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kihlstrom, J.F., Glisky, M.L., Peterson, M.A., Harvey, E.M., & Rose, P.M. (1991). Vividness and control of mental imagery: A psychometric analysis. *Journal of Mental Imagery*, *15*, 133-142.

King, M. F. & Brunner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing, 17*, 2, 79-103. http://dx.doi.org/10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0

Konstabel, K., Aavik, T. & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20*, 7, 549-566. http://dx.doi.org/10.1002/per.593

Kosslyn, S. M. (1980). *Image and mind*. London: Harvard University Press.

Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press. http://dx.doi.org/10.1093/brain/117.5.1055

Kosslyn, S. M., Thompson. W.L., Shephard, J. M., Ganis, G., Bell, D., Donovitch, J., Wittenberg, L. A. & Alpert, N. M. (2004) Brain rCBF and performance in visual imagery tasks: Common and distinct processes. *European Journal of Cognitive Processes, 16,* 696-716. http://dx.doi.org/10.1080/09541440340000475

Kovacic, M. P., Galic, Z. & Jermeic, Z. (2014) Social desirability scales as indicators of self-enhancement and impression management. *Journal of Personality Assessment, 96*, 5, 532-543. http://dx.doi.org/10.1080/00223891.2014.916714

Kreuger, J. (1998) Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin, 24*, 5, 505-516. https://doi.org/10.1177/0146167298245006

Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H. & Robins, R. W. (2004) Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychological Review, 111*, 1, 94-110. https://doi.org/10.1037/0033-295X.111.1.94

Kwan, V. S., John, O. P., Robins, R. W. & Kuang, L. L. (2008) Conceptualizing and Assessing Self-Enhancement Bias: A Componential Approach. *Journal of Personality and Social Psychology, 94*, 6, 1062–1077. https://doi.org/10.1037/0022-3514.94.6.1062

Kwekkeboom, K. L. (2000) Measuring imaging ability: Psychometric testing of the imaging ability questionnaire. *Research in Nursing & Health, 23,* 301-309. http://dx.doi.org/10.1002/1098-240X(200008)23:4<301::AID-NUR6>3.0.CO;2-0

Lacey, S. & Lawson, R. (2013) Imagery Questionnaires: Vividness and beyond. In: S. Lacey, & R. Lawson (Eds.), *Multisensory Imagery* (pp. 271-283). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-5879-1_14

Laing, J., Morland, T. & Fornells-Ambrojo, M. (2016) The phenomenology and generation of positive mental imagery in early psychosis. *Clinical Psychology & Psychotherapy, 23*, 6, 521-532. https://doi.org/10.1002/cpp.1976

Lanyon, R. I. & Carle, A. C. (2007) Internal and external validity scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement, 67*, 5, 859-876. http://dx.doi.org/10.1177/0013164406299104

Lazarus, R. S. (2000). How emotions influence performance in competitive sports. *The Sport Psychologist, 14*, 3, 229-252. https://doi.org/10.1123/tsp.14.3.229

Leahy, T.M. (2013) *A history of psychology*. Harlow: Pearson.

Lee, K. & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39,* 2, 329-358. https://doi.org/10.1207/s15327906mbr3902_8

Leising, D., Locke, K. D., Kurzius, E. & Zimmermann, J. (2016) Quantifying the association of self-enhancement bias with self-ratings of personality and life satisfaction. *Assessment, 23*, 5, 588–602. https://doi.org/10.1177/1073191115590852

Leite, W. L. & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement, 65*, 1, 140-154. http://dx.doi.org/10.1177/0013164404267285

Leite, W. L. & Cooper, L. A. (2010). Detecting social desirability bias using factor mixture models. *Multivariate Behavioural Research, 45*, 2, 271-293. http://dx.doi.org/10.1080/00273171003680245

Li, A. & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14*, 2, 131-141. http://dx.doi.org/10.1111/j.1468-2389.2006.00339.x

Li, A. & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement, 67*, 525-544. http://dx.doi.org/10.1177/0013164406292087

Li, F. & Li, Y. (2008) The Balanced Inventory of Desirable Responding (BIDR): A factor analysis. *Psychological reports, 103*, 727-731. http://dx.doi.org/10.2466/pr0.103.7.727-731

Lönnqvist, J. E., Paunonen, S., Tuulio-Henriksson, A., Lönnqvist, J., & Verkaslo, M, (2007). Substance and style in social desirable responding. *Journal of Personality, 75*, 2, 291-322. http://dx.doi.org/10.1111/j.1467-6494.2006.00440.x

Ludeke, S. G. & Makransky, G. (2016) Does the Over-Claiming Questionnaire measure overclaiming? Absent convergent validity in a large community sample. *Psychological Assessment, 28*, 6, 765-774. http://dx.doi.org/10.1037/pas0000211

McCrae, R. R. & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 6, 882-888. http://dx.doi.org/10.1037/0022-006X.51.6.882

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24. http://dx.doi.org/10.1111/j.2044-8295.1973.tb01322.x

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery, 19*, 3&4, 153-166.

Marks, D. F. (1998) The VVIQ: A measure of mental contents, state, or trait? *Journal of Mental Imagery, 17,* 1&2, 337-346.

McDougall, S. & Pfeifer, G. (2012) Personality differences in mental imagery and the effects on verbal memory. *British Journal of Psychology, 103*, 4, 556-573. http://dx.doi.org/10.1111/j.2044-8295.2011.02094.x

McKelvie, S.J. (1979). Effects of instructions and format on reported visual imagery. *Perceptual & Motor Skills*, *49*, 567-571. http://dx.doi.org/10.2466/pms.1979.49.2.567

McKelvie, S. J. (1994). Psychometric guidelines for imagery questionnaires. *Perceptual and Motor Skills, 79*, 1219-1231. http://dx.doi.org/10.2466/pms.1994.79.3.1219

McKelvie, S. J. (1995a). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery, 19*, 3&4, 1-106.

McKelvie, S. J. (1995b). Response to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery, 19*, 3&4, 197-251.

McKibben, W. B. & Silvia, P. J. (2015) Evaluating the distorting effects of inattentive responding and social desirability on self-report scales creativity and the arts. *Journal of Creative Behavior, 51*, 1, 57-69. https://doi.org/10.1002/jocb.86

McLemore, C. W. (1976) Factorial validity of imagery measures. *Behavioural Research and Therapy, 14*, 399-408. http://dx.doi.org/10.1016/0005-7967(76)90086-3

Meehl, P. E. & Hathaway, S. R. (1946). The K factor as a suppresser variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30*, 525-564. http://dx.doi.org/10.1037/h0053634

Messick, S. (1962). Response style and content measures from personality inventories. *Educational and Psychological Measurement*, 22, 41-56. http://dx.doi.org/10.1007/978-1-4614-5879-1_17

Moran, T. P. (2016) Anxiety and Working Memory Capacity: A Meta-Analysis and Narrative Review. *Psychological Bulletin, 142*, 8, 831–864. https://doi.org/10.1037/bul0000051

Morris, P. E. & Gale, A. (1973) A correlational study of variables related to imagery. *Perceptual and Motor Skills, 38*, 659-665. http://dx.doi.org/10.2466/pms.1974.38.2.659

Munroe, K. J., Giacobbi, P., Hall, C. & Weinberg, R. (2000). The four W's of imagery use: where, when, why and what. *Sport Psychologist, 14*, 2, 119-137. https://doi.org/10.1123/tsp.14.2.119

Munzert, J. & Lorey, B. (2013). Motor and visual imagery in sports. In S. Lacey & R. Lawson. (Eds.), *Multisensory imagery* (pp. 319-341). New York, NY: Springer.

Nichols, D. S. & Greene, R. L. (1980) *Adaptive or defensive: An evaluation of Paulhus's two-factor model of social desirability responding in the MMPI*. Unpublished manuscript, Dammasch State Hospital, Wilsonville, OR.

Oh, I-S., Wang, G. & Mount, M. K. (2011) Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied psychology, 96*, 4, 762-773. https://doi.org/10.1037/a0021832

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 6, 660-679. http://dx.doi.org/10.1037/0021-9010.81.6.660

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rhinehart & Winston.

Paivio, A. & Harshman, R. (1983). Factor analysis of a questionnaire on imagery and verbal habits and skills. *Canadian Journal of Psychology, 37*, 4, 461-483. http://dx.doi.org/10.1037/h0080749

Palmer, S. E. (1999) *Vision science: Photons to phenomenology*. Cambridge, MA: Bradford Books.

Palmiero, M., Belardinelli, M.O., Nardo, D., Sestieri, C., Di Matteo, R., D'Ausilio, A. & Romani, G.L. (2009) Mental imagery generation in different modalities activates sensory-motor areas. *Cognitive Processes, 10*, Supplement 2, S268-71. http://dx.doi.org/10.1007/s10339-009-0324-5

Parmac Kovacic, M. (2015) Saints and superheroes, or overall perfect: Testing the possibility to differentiate egoistic from moralistic socially desirable responding. *Psihologijske Teme, 24*, 2, 259-277.

Paulhus, D. L. (1984). Two-component models of social desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609. http://dx.doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6*. Unpublished Manual, University of British Columbia, Vancouver, Canada.

Paulhus, D.L. (1998). *Manual for the Balanced Inventory of Desirable Responding: Version 7*. Toronto/Buffalo: Multi-Health Systems.

Paulhus, D. L. (1999). *BIDR Self-Deceptive Denial (SDD)*. Retrieved 1 August 2003 from http://neuron4.psych.ubc.ca/~dpaulhus/research/sdd.htm.

Paulhus, D. L. (2002). Social Desirable Responding: The evolution of a construct. In H. I. Braun, D.N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49-69). Mahwah, NJ: Erlbaum.

Paulhus, D. L. (2006) A comprehensive Inventory of Desirable Responding (CIDR). Poster presented at the meeting of the Association for Research in Personality, New Orleans.

Paulhus, D.L., Harms, P. D., Bruce, M.N., & Lysy, D.C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology, 84*, 890-904. http://dx.doi.org/10.1037/0022-3514.84.4.890

Paulhus, D. L. & Holden, R. R. (2010) Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew, D. E. Carlson, W. G. Graziano & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 227-246). New York, NY: Oxford University Press.

Paulhus, D.L. & John, O.P. (1998). Egoistic and moralistic bias in self-perceptions: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1024-1060. http://dx.doi.org/10.1111/1467-6494.00041

Paulhus, D.L. & Notareschi, R.F. (1993). *Varieties of faking manipulations*. Unpublished manuscript, University of British Columbia.

Paulhus, D. L. & Reid, D. B. (1991) Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60,* 307-317. http://dx.doi.org/10.1037/0022-3514.60.2.307

Paulhus, D. P., Robins, R. W., Trzesniewski, K. H. & Tracy, J. L. (2004) Two replicable suppressor situations in personality research. *Multivariate Behavioural Research, 39*, 2, 303-328. https://doi.org/10.1207/s15327906mbr3902_7

Paulhus, D. L. & Vazire, S. (2007). The self-report method. Robins, R. W.,Fraley, R. C., & Krueger, R. F. (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York, NY: Guilford Press.

Pauls, C. A. & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*, 2, 263-275. http://dx.doi.org/10.1016/S0191-8869(02)00187-3

Paunonen, S. V. & LeBel, E. P. (2012) Socially desirable responding and its elusive effects on the validity of personality assessments. *Journal of Personality and Social Psychology, 103,* 1, 158-175. https://doi.org/10.1037/a0028165

Pearson, G. G., Deeprose, C., Wallace-Hadrill, S. M. A. Burnett Hayes & Holmes, E. A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review 33*, 1-23. http://dx.doi.org/10.1016/j.cpr.2012.09.001

Pearson, J., Naselaris, T., Holmes, E. A. & Kosslyn, S. M. (2015) Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends in Cognitive Sciences, 19*, 10, 590-602. https://doi.org/10.1016/j.tics.2015.08.003

Piedmont, R. L., McCrae, R. R., Riemann, R. & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78,* 3, 582-593. http://dx.doi.org/10.1037/0022-3514.78.3.582

Poltrock, S. E. & Brown, P. (1984) Individual differences in imagery and spatial ability. *Intelligence*, 8, 93-138. http://dx.doi.org/10.1016/0160-2896(84)90019-9

Pylyshyn, Z.W. (1973). What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin*, *80,* 1-25. http://dx.doi.org/10.1037/h0034650

Rademaker, R. L. & Pearson, J. (2012) Training Visual Imagery: Improvements of Metacognition, but not Imagery Strength. *Frontiers in Psychology, 10*, 3, 23-33. http://dx.doi.org/10.3389/fpsyg.2012.00224

Reynolds, C.R. and Paget, K.D. (1981) Factor Analysis of the Revised Children's Manifest Anxiety Scale for Blacks, Whites, Males and Females with a National Normative Sample. *Journal of Consulting and Clinical Psychology, 49,* 3, 352-359. https://doi.org/10.1037/0022-006X.49.3.352

Richardson, A. (1964). Has mental practice any relevance to physiotherapy? *Physiotherapy*, *50*, 148-151.

Richardson, A. (1977a). Verbaliser – Visualiser: A cognitive style dimension. *Journal of Mental Imagery*, *1,* 109-125.

Richardson, A. (1977b). The meaning and measurement of memory imagery. *British Journal of Psychology, 68*, 1, 29-43. http://dx.doi.org/10.1111/j.2044-8295.1977.tb01556.x

Richardson, H.A., Simmering, M.J. & Sturman, M.C. (2009) A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods, 12***,** 4, 762–800. https://doi.org/10.1177/1094428109332834

Richardson, J.T.E. (1980). *Mental Imagery and Human Memory*. London: Macmillan. http://dx.doi.org/10.1007/978-1-349-16354-0

Richardson, J. T. E. (1988) Vividness and unvividness: reliability, consistency, and validity of subjective imagery ratings. *Journal of Mental Imagery*, *12*, 3&4, 115-122.

Richardson, J. T. E. (1997) *Imagery*. Hove: Psychology Press.

Richman, W. L., Kiesler, S., Weisband, S. & Drasgow, F. (1999) A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires and interviews. *Journal of Applied Psychology, 84,* 754-777. http://dx.doi.org/10.1037/0021-9010.84.5.754

Riding, R. J., & Dyer, V. A. (1980). The relationship between extraversion and verbal-imagery learning style in twelve-year-old children. *Personality and Individual differences*, *1*, 273–279. http://dx.doi.org/10.1016/0191-8869(80)90059-8

Rinaldo, R. & Okada, R. (1993). A test of the relationship between imagery vividness and social desirability across subject selection procedures. *Journal of Mental Imagery*, *17,* 3&4, 171-180.

Roberts, R., Callow, N., Hardy, L., Markland, D. & Bringer, J. (2008) Movement imagery ability: development and assessment of a revised version of the Vividness of Movement Imagery Questionnaire. *Journal of Sport Exercise Psychology, 30*, 200-221. http://dx.doi.org/10.1123/jsep.30.2.200

San Roque. L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, N. J., Floyd, S., Hammond, J., Rossi, G., Tufvesson, S., van Putten, S., Majid, A. (2015) Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics, 26*, 1, 31-60. https://doi.org/10.1515/cog-2014-0089

Sears, D.O. (1986) College sophomores in the lab: Influences of a narrow database on psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 3, 515-530. https://doi.org/10.1037/0022-3514.51.3.515

Shane, M. S. & Peterson, J. B. (2004) Self-induced memory distortions and the allocation of processing resources at encoding and retrieval. *Cognition and Emotion, 18*, 4, 533-558. http://dx.doi.org/10.1080/02699930341000437

Sheehan, P.W. (1967). A shortened version of Bett's Questionnaire Upon Mental Imagery. *Journal of Clinical Psychology, 23*, 386-389. http://dx.doi.org/10.1002/1097-4679(196707)23:3<386::AID-JCLP2270230328>3.0.CO;2-S

Sinha, R. R. & Kreuger, J. (1998) Idiographic self-evaluation and bias. *Journal of Research in Personality, 32*, 2, 131-155. https://doi.org/10.1006/jrpe.1997.2211

Spielberger, C.D., Gorsuch, R., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Stigler, S. M. (1986) *The history of statistics: The measurement of uncertainty before 1900.* Cambridge, MA: Belknap Press.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment,* 17, 3, 222-232. http://dx.doi.org/10.1027//1015-5759.17.3.222

Stöber, J., Dette, D. E. & Musch, T. (2002). Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78,* 2, 370-389. http://dx.doi.org/10.1207/S15327752JPA7802_10

Strahan, R. & Gerbasi, C. G. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 2, 191-193. http://dx.doi.org/10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G

Strelow, B. R. & Davidson, W. B. (2002) Introversion-extraversion, tempo and guided imagery. *Psychological Reports, 90*, 619-626. http://dx.doi.org/10.2466/pr0.2002.90.2.619

Sevdalis, N., Moran, A. & Arora, S. (2013). Mental imagery and mental practice applications in surgery: State of the art and future directions. In S. Lacey & R. Lawson. (Eds.), *Multisensory imagery* (pp. 343-363). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-5879-1_18

Thomas, N.J.T. (2003). Mental Imagery, Philosophical Issues About. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science* Volume 2 (pp. 1147–1153). London: Nature Publishing/Macmillan.

Tonkovic, M., Galic, Z. & Jerneic, Z. (2011) The construct validity of over-claiming as a measure of egoistic enhancement. *Review of Psychology, 18*, 1, 13-21.

Tracey, T. J. G. (2016) A note of socially desirable responding. *Journal of Counselling Psychology, 62*, 2, 224-232. https://doi.org/10.1037/cou0000135

Uziel, L. (2010) Rethinking social desirability scales: From Impression Management to Interpersonally Oriented Self-Control. *Perspectives on Psychological Science, 5*, 3, 243-262. https://doi.org/10.1177/1745691610369465

Van den Hout, M. A. (2012) Tones inferior to eye movements in the EMDR treatment of PTSD. *Behaviour Research and Therapy, 50,* 5, 275-279. https://doi.org/10.1016/j.brat.2012.02.001

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualisation. *Perceptual and Motor Skills, 47,* 599-604. http://dx.doi.org/10.2466/pms.1978.47.2.599

Vecchione, M. & Alessandri, G. (2014) Egoistic and moralistic self-enhancement in the eye of the beholder: A cross-informant study. *Journal of Personality, 82,* 5, 432-439. https://doi.org/10.1111/jopy.12073

Vecchione, M. & Alessandri, G. (2013) Distinguishing trait from state components in the assessment of egoistic and moralistic self-enhancement. *Personality and Individual Differences, 54*, 8, 884-889. https://doi.org/10.1016/j.paid.2012.12.027

Vecchione, M., Alessandri, G. & Barbaranelli, C. (2013) Measurement and application of egoistic and moralistic self-enhancement. *International Journal of Selection and Assessment, 21*, 2, 170-182. https://doi.org/10.1111/ijsa.12027

Watson, K., Baranowski, T., Thopmson, D., Jago, R., Baranowski, J. and Klesges, L. M. (2006). Innovative application of a multidimensional item response model in assessing the influence of social desirability on the pseudo-relationship between behavior and self-efficacy and behavior. *Health Education Research Theory and Practice Special Edition, 21*, i85-i97. http://dx.doi.org/10.1093/her/cyl137

White, K.D., Ashton, R. & Law, H. (1978). The measurement of imagery vividness: Effects of format and order on the Bett's Questionnaire Upon Mental Imagery. *Canadian Journal of the Behavioural Sciences, 10*, 1, 68-78. http://dx.doi.org/10.1037/h0081537

White, K., Sheehan, P.W. & Ashton, R. (1977). Imagery assessment: A survey of self-report measures. *Journal of Mental Imagery*, *1*, 145-170.

Wiggins, J.S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology, 23*, 419-427. http://dx.doi.org/10.1037/h0047823

Wright, C., Hogard, E., Ellis, R., Smith, D. & Kelly, C. (2008) Effect of the PETTLEP imagery training on performance of nursing skills: pilot study. *Journal of Advanced Nursing, 63,* 3, 259-265. https://doi.org/10.1111/j.1365-2648.2008.04706.x

Zerbe, W. & Paulhus, P. (1987). Social desirable responding in organizational behavior: A recognition. *Academy of management Review, 12*, 2, 250-264. http://dx.doi.org/10.5465/AMR.1987.4307820

Zettler, I., Hilbig, B. E., Moshagen, M. & de Vires, R. E. (2015) Dishonest responding or true virtue? A behavioral test of impression management. *Personality and individual differences, 81*, 107-111. https://doi.org/10.1016/j.paid.2014.10.007

Ziegler, M. & Buehner, M, (2009). Modelling social desirable responding and its effects. *Educational & Psychological Measurement, 69*, 4, 548-565. http://dx.doi.org/10.1177/0013164408324469

Appendices

Appendix 1: Paper 5 screenshot of conference proceedings

Appendix 2: Abstracts of the five papers

Appendix 3: Email confirmation of contributions from authors

Appendix 4: Emails applying for permission to include articles in the thesis

Appendix 5: Main questionnaires used in the five papers

Appendix 6: Research Gate read statistic screenshot 30th July, 2018

Appendix 7: Additional correlational analysis on paper 4 data controlling for gender

Appendix 8: Additional correlational analysis involving the VVIQ-1, BIDR egoistic scales and the over-claiming questionnaire

Appendix 1: Paper 5 screenshot of conference proceedings

http://abstracts.bps.org.uk/index.cfm?&ResultsType=Abstracts&ResultSet_ID=4502&FormDisplayMode=view&frmShowSelected=true&localAction=details

A 93 word abstract and 288 longer 'body' of the paper were submitted to the conference. The conference organizers published the body in their conference proceedings (see Figure 2).

Figure 3. Paper 5 screenshot of conference proceedings

Appendix 2: Abstracts of the five papers

Paper 1 abstract

The relationship between scores on self-report imagery questionnaires and sub-types of socially desirable responding was investigated across two studies. The self-report imagery questionnaires used were the Vividness of Visual Imagery Questionnaire (Marks, 1973) and the Vividness of Movement Imagery Questionnaire (Isaac, Marks & Russell, 1986). The measure of socially desirable responding was the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1994, 2002). The BIDR is made up of two main sub-scales: the self-deceptive enhancement (SDE) scale, which is a measure of egoistic bias (the tendency to claim positive social and intellectual qualities), and the IM scale, a measure of moralistic bias (the claiming of positive moral qualities and the refutation of negative socially-deviant qualities). Samples of 98 and 109 undergraduate psychology students were tested under anonymous conditions. Results indicated that SDE correlated consistently and at a higher level with the self-report imagery questionnaires than IM. SDE correlations ranged from –0.23 to –0.33 with greater image vividness being associated with greater SDE. The research shows the importance of treating socially desirable responding as a multifactorial phenomenon, and that previous studies may have underestimated the degree of relationship between imagery scales and socially desirable responding by using inferior measures of egoistic bias.

Paper 2 abstract

Allbutt, Shafiullah and Ling [1] found that scores on self-report measures of visual and movement imagery vividness correlate primarily with an egoistic form of socially desirable responding rather than a moralistic form. The current study investigated whether the pattern of findings generalises to the ratings of other imagery properties such as the ease with which an image can be generated. Participants completed the Shapes Questionnaire [2] and the Balanced Inventory of Desirable Responding [3]. Several of the Shapes Questionnaire ratings correlated significantly with the egoistic form of socially desirable responding, while correlations with the

moralistic form were rare. This shows the pattern of findings generalises to the ratings of properties of the imagery experience other than vividness.

Paper 3 abstract

Allbutt, Ling and Shafiullah (2006a) and Allbutt, Shafiullah and Ling (2006b) found that scores on self-report measures of visual imagery experience correlate primarily with the egoistic form of social desirable responding. Here, three studies are reported which investigated whether this pattern of findings generalised to the ratings of imagery vividness in the auditory modality, a new version of the Vividness of Visual Imagery Questionnaire (Marks, 1995), and reports of visual thinking style. The measure of social desirable responding used was the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002). Correlational analysis replicated the pattern seen in our earlier work and of the correlations with the egoistic bias, the correlation with vividness of visual imagery was largest and significant, the correlation with visual thinking style next largest and approached significance and the correlation with vividness of auditory imagery was the smallest and not significant. The size of these correlations mirrored the extent to which the three aspects of imagery were valued by participants.

Paper 4 abstract

Correlational research investigating the relationship between scores on self-report imagery questionnaires and measures of social desirable responding has shown only a weak association. However, researchers have argued that this research may have underestimated the size of the relationship because it relied primarily on the Marlowe-Crowne scale (MC; Crowne & Marlowe, 1960), which loads primarily on the least relevant form of social desirable responding for this particular context, the moralistic bias. Here we report the analysis of data correlating the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973), with the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002) and the MC scale under anonymous testing conditions. The VVIQ correlated significantly with the self-deceptive enhancement (SDE) and Agency Management (AM) BIDR subscales and with the MC. The largest correlation was with SDE. The ability of SDE to predict VVIQ scores was not significantly enhanced by adding either

AM or MC. Correlations between the VVIQ and BIDR egoistic scales were larger when the BIDR was continuously rather than dichotomously scored. This analysis indicates that the relationship between self-reported imagery and social desirable responding is likely to be stronger than previously thought.

Paper 5 abstract

The study tested the relationship between self-report imagery questionnaire scores, egoistic social desirable responding and state-trait anxiety. Measures used were the Vividness of Visual Imagery Questionnaire (VVIQ), the two egoistic scales of the Balanced Inventory of Desirable Responding (BIDR) and the State – Trait Anxiety Index. Correlational analyses found the VVIQ correlated with trait anxiety, and the relationship between the VVIQ and the egoistic bias was largely unaffected by partialling out trait anxiety. Thus the correlations between the VVIQ and the egoistic bias and the VVIQ and trait anxiety appear to be largely independent.

Appendix 3: Email confirmation of contributions from authors

14th September 2015

Dear John

Many thanks for your email. I am really pleased to hear that you are registering for a PhD by published works at Salford. I can confirm that the author contributions you have attributed to each of the papers represent accurate descriptions of the input each of your co-authors had and I am happy with your description of my role (as, I am sure, will the other authors be).

If I can be of any further assistance in supporting your already excellent case for a PhD by published work, please let me know. Best of luck with your application.

Very best wishes

Jon

Dr Jonathan Ling
Dept. of Pharmacy, Health & Wellbeing
University of Sunderland
Chester Road
Sunderland
SR1 3SD

jonathan.ling@sunderland.ac.uk

Fuse - The UKCRC Centre for Translational Research in Public Health fuse.ac.uk
14th September 2015

Hi John,

I hope you are well. The documents are fine.

Do you still do OU summer school? I lost their contact with my email account at De Montfort. If you let know then I would be grateful.

Best wishes,

Shafi

mohammed.shafiullah@unisg.ch
http://www.unisg.ch/en/personenverzeichnis/6e4b8ec8-ee55-4e09-80af-d2b5c3c0c57a.aspx

14th September 2015

This looks fine to me.
Tom

Thomas M. Heffernan Northumbria University
tom.heffernan@northumbria.ac.uk
https://www.northumbria.ac.uk/about-us/our-staff/h/tom-heffernan/

15th September 2015

Hi John. That looks fine to me.

Thanks
Martin

Martin Rowley Keele University
m.g.rowley@keele.ac.uk

Appendix 4: Emails applying for permission to include articles in the thesis

Article	Email sent	Email reply
1	jnlmentimag@aol.com Saturday 4th June 2016 Dear Dr. Hochman I hope that you are well. In 2006 your journal, the journal of mental imagery, published my paper: Allbutt, J., Shafiullah, M. & Ling, J. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Visual and movement imagery. *Journal of Mental Imagery*, *30,* 1&2, 1-20. I am in the process of preparing my PhD thesis entitled: Self-report mental imagery & social desirable responding. The degree is a PhD by published works & my paper is included in my thesis. As such I am writing to you, the owner of the paper's copywrite, to ask for permission to include my article in the thesis. I expect that I will need to produce 5 draft copies of the thesis prior to my oral viva, & if I am successful, 5 copies of the final thesis itself. I would welcome a response to this request at your earliest convenience. Best wishes, John. John Allbutt Lecturer in psychology Directorate of psychology & public health, Room L820 Allerton Building, University of Salford, Frederick Road, Salford, M6 6PU. Email j.d.allbutt@salford.ac.uk Phone +44 (0) 161 295 2156	No email reply was received from the journal.

		Monday 13th June 2016 Dear Dr. Hochman I hope that you are well. Last week I sent you the email message shown below. I would welcome a response to this request at your earliest convenience. Best wishes, John. Friday 8th July 2016 Dear Dr. Hochman I hope that you are well. A while ago I sent you the email messages shown below. I have now obtained permission for the inclusion of all the other articles in my PhD by published works, so I would welcome a response to my request at your earliest convenience. In addition to the physical copies of my thesis mentioned in my first email, I also anticipate an electronic version of my thesis being made available in the university library for interested students. Best wishes, John.	
2		sphone@lsu.edu j.cumming@bham.ac.uk; markman@ohio.edu Dear Professr Honeycutt, (copies to Prof. Markman & Prof. Cumming) I hope that you are well. In 2006 your journal, Imagination, Cognition and Personality, published my paper: Allbutt, J., Ling, J. & Shafiullah, M. (2006). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Components of visual imagery.	Hello John, Cc: Sarah Shinkle, Sage Associate Editor for Social Science Journals I am forwarding your request to Sarah Shinkle who is the associate editor of the social science journals for Sage and may be in a position to answer your query. Obviously, you will cite your paper in your dissertation, but it is not clear to me if you are using verbatim copies of the article in the dissertation including tables, narratives, etcetera. Sarah, I hope you can assist us in this request.

Imagination, Cognition and Personality, 25, 4, 337-353.

I am in the process of preparing my PhD thesis entitled: Self-report mental imagery & social desirable responding. The degree is a PhD by published works & my paper is included in my thesis. As such I am writing to you, the owner of the paper's copywrite, to ask for permission to include my article in the thesis. I expect that I will need to produce 5 draft copies of the thesis prior to my oral viva, & if I am successful, 5 copies of the final thesis itself.

I would welcome a response to this request at your earliest convenience.

Best wishes, John.

John Allbutt
Lecturer in psychology

All the best,

Jim

James Honeycutt, Ph.D.
LSU Distinguished Professor
Co-Editor: *Imagination, Cognition, and Personality*
Dept. of Communication Studies
Louisiana State University
Baton Rouge, LA 70803-3923
cell: 225-252-0032
e-mail: sphone@lsu.edu

Hi both,

I'm checking with our permissions department on this and will get back to you as soon as I hear.

Best,
Sarah

Sarah Shinkle
Associate Editor, Social Science Journals
SAGE Publishing
2455 Teller Road
Thousand Oaks, CA 91320
USA

Dear John Allbutt,

I work in the permissions department at SAGE Publications and your request was forwarded to us for handling. Thank you for your request.

You may use the published version of your article (version 3) in the printed version of your thesis. However, if you wish to post your thesis online, we ask that you use the version of your article that was accepted by the journal (version 2). Please note that this permission does not cover any 3rd party material that may be found within the work. You must properly credit the original source, *Imagination, Cognition and Personality*. Please let us know if you have further questions.

Best regards, Michelle Binur

		Rights Coordinator SAGE Publishing 2455 Teller Road Thousand Oaks, CA 91320 USA www.sagepublishing.com Los Angeles | London | New Delhi Singapore | Washington DC | Melbourne
3	beauducel@zem.uni-bonn.de juergen.hogrefe@hogrefe.com Dear Professor Dr. André Beauducel, (copy to Dr. G.-Jürgen Hogrefe) I hope that you are well. In 2008 your journal, the journal of individual differences, published my paper: Allbutt, J., Ling, J., Heffernan, T., M., & Shafiullah, M. (2008). The relationship between self-report imagery questionnaire scores and sub-types of social desirable responding: Auditory imagery, visual imagery and visual thinking style. *Journal of Individual Differences, 29*. 4, 181-188. I am in the process of preparing my PhD thesis entitled: Self-report mental imagery & social desirable responding. The degree is a PhD by published works & my paper is included in my thesis. As such I am writing to you, the owner of the paper's copywrite, to ask for permission to include my article in the thesis. I expect that I will need to produce 5 draft copies of the thesis prior to my oral viva, & if I am successful, 5 copies of the final thesis itself. I would welcome a response to this request at your earliest convenience. Best wishes, John. John Allbutt Lecturer in psychology Monday 13th June	Dear Mr. Allbutt, Thank you very much for your permission request which was forwarded to me. We are happy to grant permission to include your paper *Self-Report Imagery Questionnaire: Scores and Subtypes of Social-Desirable Responding Auditory Imagery, Visual Imagery, and Thinking* *Style* by John Allbutt, Jonathan Ling，Thomas M. Heffernan, and Mohammed Shafiullah, published on pp. 181-188 in *Journal of Individual Differences,* 2008, Vol. 29 (4). as part of your thesis as detailed in your original request, provided that you will please make sure that the following copyright line will appear: Used by permission from the *Journal of Individual Differences*, 2008, Vol. 29(4):181-188 © 2008 Hogrefe & Huber Publishers (now Hogrefe publishing)　www.hogrefe.com　DO 10.1027/1614-0001.29.4.181 you will follow the guidelines as specified below: Authors of articles in journals published by the Hogrefe Group also may: • Archive or post on their own or their institution's website or in their institutional repository a pre-print of

Dear Professor Dr. André Beauducel, & Dr. G.-Jürgen Hogrefe,

I hope that you are well.

Last week I sent you the email message shown below. I would welcome a response to this request at your earliest convenience.

Best wishes, John.

their submitted manuscript (i.e., manuscript version before peer-review) for noncommercial purposes at any time
- Archive or post on their own or their institution's website or in their institutional repository a post-print of their accepted manuscript (i.e., manuscript version after peer-review) for noncommercial purposes 12 months after publication of the respective journal issue
- Upon written request, archive on a website or in a repository mandated by their funding bodies a post-print of their manuscript (i.e., manuscript version after peer-review) (1) 12 months after publication of the respective journal issue or (2) as a result of and in accordance with legal obligations

Conditions

The following conditions apply:

- **Pre-print:** This manuscript version should be labeled with its date and a statement that the manuscript in this form has not yet been accepted for publication.
- **Post-print:** Only the final draft manuscript post-refereeing may be used, not the version of the paper prepared and released by the publisher (PDF, XML). This final draft manuscript may only be posted 12 months after the article has been published in the respective journal issue. It must link to the DOI of the published version of the article, carry the publisher's copyright notice in the form"[Journal Title], [Volume No.], [Issue No.], © [Year] by [Publisher's name]", and include the following statement: "This article may not exactly replicate the final version published in [Journal Title]. It is not the version of record and is therefore not suitable for citation."

| | | • **e-Offprint/published version of record**: The e-offprint is provided exclusively for the authors' personal use. Other uses of the e-offprint/published version of record, including but not limited to the following, are not permitted except with the express written permission of the publisher: posting the e-offprint/published version of record to a personal or institutional website or to an institutional or disciplinary repository; changing or modifying the digital file; reproducing, distributing, or licensing the article in whole or in part for commercial use.

Kind regards,

Gitta Moos
Customer Service

customerservice@hogrefe.com
+49 551 999 50 421

Hogrefe Publishing GmbH
Merkelstr. 3
37085 Göttingen
Germany
Tel. + 49 551 999 50 0
Fax + 49 551 999 50 111
www.hogrefe.com

CEO: Dr. G.-Jürgen Hogrefe
Registered: Göttingen, HRB 2224
VAT no.: DE 115303194 |

4	jonesmn@indiana.edu	Monday 6th June 1016

jonesmn@indiana.edu

Dear Professor Dr. Michael Jones,

I hope that you are well.

In 2011 your journal, Behavior Research Methods, published my paper: Allbutt, J., Ling, J., Rowley, M. & Shafiullah, M. (2011) Vividness of Visual Imagery and Social Desirable Responding: Correlations of the Vividness of Visual Imagery Questionnaire with the Balanced Inventory of Desirable Responding and the Marlowe-Crowne. Behavior Research Methods, 43, 3, 791-799.

I am in the process of preparing my PhD thesis entitled: Self-report mental imagery & social desirable responding. The degree is a PhD by published works & my paper is included in my thesis. As such I am writing to you, the owner of the paper's copywrite, to ask for permission to include my article in the thesis. I expect that I will need to produce 5 draft copies of the thesis prior to my oral viva, & if I am successful, 5 copies of the final thesis itself.

I would welcome a response to this request at your earliest convenience.

Best wishes, John.

John Allbutt
Lecturer in psychology

permissions.springer@spi-global.com

Springer Science+Business Media
Rights and Permissions
Tiergartenstrasse 17
69121 Heidelberg
GERMANY

14th June 2016

Dear Springer,

I hope you are well.

Monday 6th June 1016

Hi John,

 When you transfer copyright, it is held by the publisher (Springer in this case). I really doubt that you need to obtain permission to use your own work in a dissertation though....that is quite commonly written in as a normal permissible use of an author's work without requiring permission.

But you can find Springer's policy and request permission online here: http://www.springer.com/gp/rights-permissions/obtaining-permissions/882

Best, Michael

Dear John Allbutt,

Good day from Springer Rights and Permissions. Thank you for your e-mail.

As we have streamlined our permissions procedure, please visit "Obtaining Permissions" on http://www.springer.com/gp/rights-permissions/obtaining-permissions/882.
Here you will find detailed information about the possibilities of obtaining permission and will be guided through the process.

We hope that this information will help to meet your needs.

With all the best wishes,

Rights and Permissions
Springer Science+Business Media
Tiergartenstr. 17
69121 Heidelberg
Germany

15th June 2016

Dear John Allbutt,,

I am seeking permission to include a copy of my journal article: Allbutt, J., Ling, J., Rowley, M. & Shafiullah, M. (2011) Vividness of Visual Imagery and Social Desirable Responding: Correlations of the Vividness of Visual Imagery Questionnaire with the Balanced Inventory of Desirable Responding and the Marlowe-Crowne. Behavior Research Methods, 43, 3, 791-799, in my PhD by published works. Please see email trail below for details.

Could you please assist me in my request.

Best wishes, John.

Hello Springer,

Thank you for your message & the link to your Obtaining Permissions webpage.

I was sent this link by Professor Dr. Michael Jones (please see the email below). I followed the link to my article & clicked on the reprints & permissions link. When I did this I got the following error message which told me to contact you:

Please Contact Springer

Please contact the publisher of this content directly to obtain information regarding this reuse. One or more of the content data field values is preventing Rightslink from processing this request.

Springer contact information

Copyright © 2016 Copyright Clearance Center, Inc. All Rights Reserved.
Comments? We would like to hear from you.
E-mail us at customercare@copyright.com

By 'publisher' does the message mean:
1) Springer, 2) The journal editor i.e. Professor Dr. Michael Jones, or 3) someone else?

Could you please advise me how to proceed with my request?

Thank you, John.

17th June 2016

Hello Springer,

Thank you for sending me your form. I have done my best to fill it in – see attached.

Good day from Springer Rights and Permissions.

For this reason, please fill in the questionnaire attached and return the completed form back to permissions.springer@spi-global.com so that we can manually process and review your request. Thank you.

With all the best wishes,

Rights and Permissions
Springer Science+Business Media
Tiergartenstr. 17
69121 Heidelberg
Germany

23rd June 2016

Dear John Allbutt,

Good day from Springer Rights and Permissions.

Please see attached permission letter.

Thank you.

With all the best wishes,

Rights and Permissions
Springer Science+Business Media
Tiergartenstr. 17
69121 Heidelber
Germany

	Do you have any idea how long it will take to process this request? Best wishes, John.	
5	enquiries@bps.org.uk Dear BPS enquiries, I hope that you are well. In 2009 I presented a poster article at the BPS conference entitled: Allbutt, J. & Ling, J. (2009) The relationship between self-report imagery questionnaire scores, the egoistic sub-type of social desirable responding and state-trait anxiety. *British Psychological Society Conference,* Brighton. I am in the process of preparing my PhD thesis entitled: Self-report mental imagery & social desirable responding. The degree is a PhD by published works & my paper is included in my thesis. As such I am writing to you to ask for permission to include my article in the thesis. I expect that I will need to produce 5 draft copies of the thesis prior to my oral viva, & if I am successful, 5 copies of the final thesis itself. I am writing on the belief that you may be the owner of the copywrite on my paper. I would welcome a response to this request at your earliest convenience. Best wishes, John. John Allbutt Lecturer in psychology	Monday 6th June 1016 Hi John. We do not ask for any assignment of copyright for poster presentations, so the copyright remains with the originator of the poster. If that is you, then you are free to re-use the material in any way you like. Good luck with the thesis. Peter **Peter Dillon-Hooper** LLB | Academic Resources Manager St Andrews House, 48 Princess Road East, Leicester LE1 7DR t: +44 (0)116 252 9528 w: www.bps.org.uk/hopc

Appendix 5: The questionnaires used in the five papers

Vividness of Visual Imagery Questionnaire

Instructions

Visual imagery refers to the ability to visualise, that is, the ability to form mental pictures, or to 'see in the mind's eye'. Marked individual differences have been found in the strength and clarity of reported visual imagery and these differences are of considerable psychological interest.

The aim of this test is to determine the vividness of your visual imagery. The items of the test will possibly bring certain images to your mind. You are asked to rate the vividness of each image by reference to the five-point scale given below. For example, if your image is 'vague and dim' then give it a rating of 4. After each item write the appropriate number in the box provided. Before you turn to the items on the next page, familiarise yourself with the different categories on the rating scale. Throughout the test, refer to the rating scale when judging the vividness of each image. Try to do each item separately, independent of how you have done other items.

Please form all of the images with your eyes CLOSED.

Rating scale

1	Perfectly clear and vivid as normal vision
2	Clear and reasonably vivid
3	Moderately clear and vivid
4	Vague and dim
5	No image at all, you only 'know' that you are thinking of the object

For items 1 - 4 think of some relative or friend whom you frequently see (but who is not with you at present) and consider the picture that comes before your mind's eye.

	Item	Response
1	The exact contour of face, head, shoulders and body.	
2	Characteristic poses of the head, attitude of the body etc.	
3	The precise carriage, length of step etc. in walking.	
4	Different colours worn in some particular clothes.	

Visualise the rising sun. Consider carefully the picture that comes before your mind's eye.

	Item	Response
5	The sun is rising above the horizon into a hazy sky.	
6	The sky clears and surrounds the sun with blueness.	
7	Clouds. A storm blows up with flashes of lightning.	
8	A rainbow appears.	

Think of a shop which you go to. Consider the picture that comes before your mind's eye.

	Item	Response
9	The overall appearance of the shop from the opposite side of the road.	
10	A window display including colours, shapes and details of individual items for sale.	
11	You are near the entrance. The colour, shape and details of the door.	
12	You enter the shop and go to the counter. The counter assistant serves you. Money changes hands.	

Finally, think of a country scene which involves trees, mountains and a lake. Consider the picture that comes before your mind's eye.

	Item	Response
13	The contours of the landscape.	
14	The colour and shape of the trees.	
15	The colour and shape of the lake.	
16	A strong wind blows on the trees and on the lake, causing waves.	

Please explain below any difficulties you had answering this questionnaire or add any general comments you have to make.

Thank you for your cooperation.

Vividness of Movement Imagery Questionnaire

TOTAL SCORES

a) Other =
b) Self   =
Total a + b =

Instructions

Movement imagery refers to the ability to imagine a movement. The aim of this test is to determine the vividness of your movement imagery. The items of the test are designed to bring certain images to your mind. You are asked to rate the vividness of each item by reference to the 5-point scale. After each item, write the appropriate number in the box provided. The first box is for an image obtained watching somebody else and the second box is for an image obtained doing it yourself. Try to do each item separately, independently of how you may have done other items. Complete all items obtained watching someone else and then return to the beginning of the questionnaire and rate the image obtained doing it yourself. The two ratings for a given item may not in all cases be the same. For all items please have your eyes CLOSED.

RATING SCALE The image aroused by each item might be:

Perfectly clear and vivid as normal vision	RATING 1
Clear and reasonably vivid	RATING 2
Moderately clear and vivid	RATING 3
Vague and dim	RATING 4
No image at all you only know you are thinking of the skill	RATING 5

Think of each of the following acts, and classify the images according to the degree of clearness and vividness as shown on the RATING SCALE.

Item	Watching someone else	Doing it yourself
1) Standing		
2) Walking		
3) Running		
4) Jumping		

Item	Watching someone else	Doing it yourself
5) Reaching for something on Tiptoe		
6) Drawing a circle on paper		
7) Kicking a stone		
8) Bending to pick up a coin		

RATING SCALE The image aroused by each item might be:

Perfectly clear and vivid as normal vision	RATING 1
Clear and reasonably vivid	RATING 2
Moderately clear and vivid	RATING 3
Vague and dim	RATING 4
No image at all you only know you are thinking of the skill	RATING 5

Think of each of the following acts, and classify the images according to the degree of clearness and vividness as shown on the RATING SCALE.

Item	Watching someone else	Doing it yourself
9) Falling forwards		
10) Running up stairs		
11) Jumping sideways		
12) Slipping over backwards		

Item	Watching someone else	Doing it yourself
13) Catching a ball with two hands		
14) Throwing a stone into water		
15) Kicking a ball in the air		
16) Hitting a ball along the ground		

Item	Watching someone else	Doing it yourself
17) Running downhill		
18) Climbing over a high wall		
19) Sliding on ice		
20) Riding a bike		

Item	Watching someone else	Doing it yourself
21) Jumping into water		
22) Swinging on a rope		
23) Balancing on one leg		
24) Jumping off a high wall		

Please explain below any difficulties you had answering this questionnaire or add any general comments you have to make.

Thankyou for your cooperation.

The Shapes Questionnaire

Mental images are mental representations of the external form of an object. In the visual modality mental imagery would refer to the experience of visualising an object, such as a car, in your mind's 'eye'. The individual differences that have been found in people's ability to form metal images of items have long been of interest to psychologists. Some people have very lifelike 'mental pictures', while others only 'know' that they are thinking about things.

In this questionnaire, you will be asked to imagine several shapes. The shapes will be shown at the top of each page, underneath will be a number of questions. Once you have looked at the shape, we would like you to shut your eyes and imagine it, then answer the question underneath it (it will probably help you to imaging the shape each time you answer the question, rather than trying to answer all the questions from imaging the shape only once). If you can imagine the shape more easily with your eyes open then do so, but DO NOT LOOK AT THE SHAPE WHILST IMAGING.

The first question will always ask you how easily you can evoke an image of the shape in question, and to rate it on a scale of 1 - 9. What we are interested in here, is how easily you can evoke the image (or how easily you know you are thinking about it), and ONLY this. We are not interested in the properties of your image at this stage (e.g. how vivid or detailed your image is), questions on these properties will come later.

The questions underneath will ask you to rate the various properties and qualities of the image you have formed. Remember, we wish you to consider ONLY THE PROPERTY OR QUALITY BEING ASKED ABOUT, and to rate each question about your image SEPARATELY AND INDEPENDENTLY of all the others. This is a very important point, as you may find that your ratings of each property or quality are different.

Concentrate on imaging the actual shape rather than the property you are trying to judge. This sounds difficult, but what we mean, is that we do not wish you to change what you are imaging to improve a particular property. Rather we wish you to rate the level of that property which occurred 'automatically', when you imaged the scene. Some of the questions will ask you to rate how easily you can perform various manipulations of the image in question. Remember, we only wish you to rate the ease of the manipulation, not any aspect of the quality of the image, there are separate questions concerning this.

There is no time limit for this questionnaire, all we wish is that you think carefully about each item, and give as accurate an answer as possible. If at any time you feel yourself getting tired or loosing concentration, take a rest! REMEMBER, your accurate and honest answers are vital of this study.

You may begin when ready, if you have any doubts about what to do, please ask the experimenter now.

If at any point you have difficulty in rating something, RATE IT AS BEST YOU CAN, then write by the side what you found difficult, or why you found it difficult. The same applies if you feel that the ratings do not capture something essential about your images. However, PLEASE TRY TO MAKE A RATING IN ALL CASES, then write down what was wrong.

Shape 1: Imagine this shape
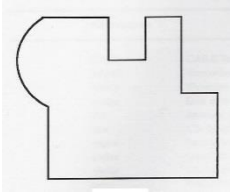
1) How easy is it for you to evoke this image?

   Very difficult   1   2   3   4   5   6   7   8   9   Very easy

2) How much detail is there in your image?

   Very little   1   2   3   4   5   6   7   8   9   A great deal

3) How clear and sharp (in pictorial terms) is you image?

   Not clear at all   1   2   3   4   5   6   7   8   9   Very sharp & clear

4) How easily can you maintain this image now that you have evoked it?

   Not at all   1   2   3   4   5   6   7   8   9   Very easily

5) How much does a) the detail, and b) the clarity of your image change when you try to maintain your image?

a) The Detail

   Not at all   1   2   3   4   5   6   7   8   9   A great deal

If there was a change was there more or less detail (please circle a response)?

More   Less

b) The Clarity

   Not at all   1   2   3   4   5   6   7   8   9   A great deal

If there was a change was the image more or less clear (please circle a response)?

More   Less

6) How much of the shape can you form an image of, at any one time?

1/10 th   1   2   3   4   5   6   7   8   9   All

Please mark on the shape on this page the parts you imagine (circle the shape if you can imagine all of it).

7) How large is your image of this shape?

Very small (As if   1   2   3   4   5   6   7   8   9   Very large (As is
seeing it from a   seeing it from
great distance)   very close up)

8) How vivid is your image of the shape?

Not at all vivid   1   2   3   4   5   6   7   8   9   Very vivid

9) Imagine the shape rotating in the plane of the paper; How easily can you perform the rotation? (Note we only wish you to rate how easy the rotation is, not how good the quality of your image is whilst you are doing it).

Not at all   1   2   3   4   5   6   7   8   9   Very easily

10) Is the rotation you can image continuous or in discrete stages (please circle a response)?

Continuous   Discrete

11) Does the size of the imaged shape (or the apparent distance at which it is imagined) change when you rotate it? If so, what way does your image of the shape change?

(As if) Much   1   2   3   4   5   6   7   8   9   (As if) Much larger
smaller

12) Does rotating the shape affect a) the detail & b) the clarity of your image in any way?

a) The detail

   Not at all   1   2   3   4   5   6   7   8   9   A great deal

If there was a change was there more or less detail (please circle a response)?

More   Less

b) The clarity

   Not at all   1   2   3   4   5   6   7   8   9   A great deal

If there was a change was the image more or less clear (please circle a response)?

More   Less

13) How much of the shape do you form an image of when rotating it?

   1/10 th   1   2   3   4   5   6   7   8   9   All

14) If you only image parts of the shape whilst rotating it, can you mark these on the shape above.

15) Are there any other ways in which your image of the shape changes when you rotate it? If there are can you say what these are, AND rate the amount of change on a 1-9 scale.

_____

_____

   Very little   1   2   3   4   5   6   7   8   9   Very much

16) Finally, how vivid is your image of the shape rotating?

   Not at all vivid   1   2   3   4   5   6   7   8   9   Very vivid

 The shapes questionnaire continues in this format requiring ratings for a further 4 shapes:
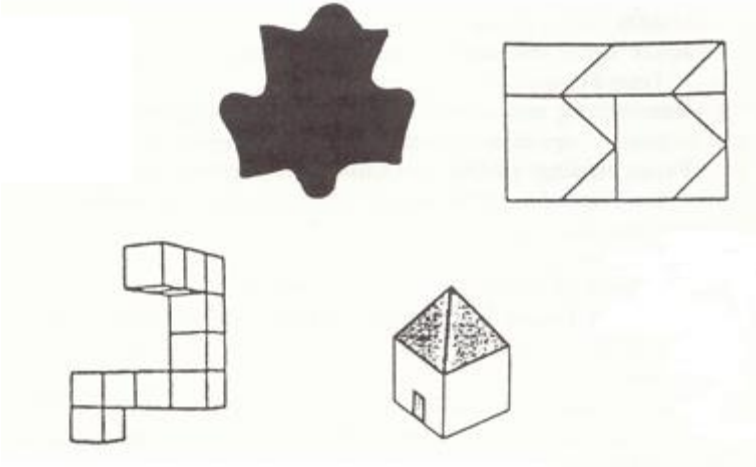
The reflection of imagery ability questionnaire

Mental images are mental representations of the external form of an object. In the visual modality mental imagery would refer to the experience of visualizing an object, such as a car, in your mind's 'eye'. The individual differences that have been found in people's ability to form mental images of items have long been of interest to psychologists. Some people have very life-like 'mental pictures', while others only 'know' that they are thinking about things.

The Shapes Questionnaire (Deane and Morris, 1991, 2003) is a new imagery questionnaire. For each of five shapes (see figure 1) respondents are asked to form an image of the shape, maintain the image, and then rotate the image. While doing this they rate the images for a number of properties.

Fig 1) The five shapes to be imaged from the Shapes Questionnaire
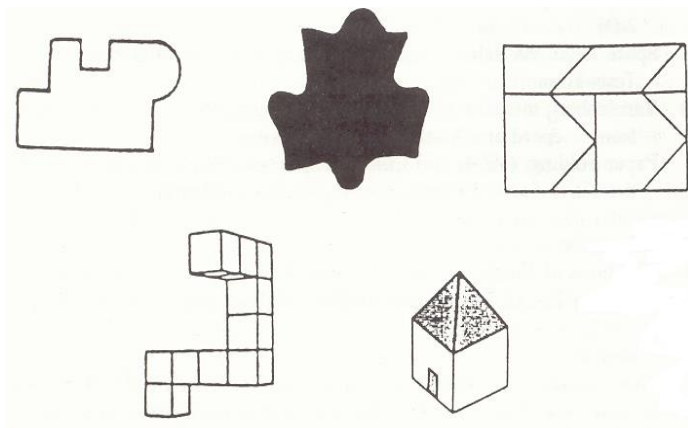
We would like you to read the list of properties the images are rated for in the table on page 2. Then, for each property in turn, indicate whether you believe the property measures 'ability' at visual imagery or not. For example, for the first property (1. Ease of evoking an image) do you think that ratings of this property inform us about a person's ability at visual imagery or not? If you think that such differences only indicate a difference in style, and not that people are in some sense better or worse at imagery, then tick the 'Not an indicator of ability at all' option. However, if you think that the differences in the ease with which people can evoke a visual image indicate a difference in imagery ability, then tick one of the remaining three responses options ('A weak indicator of ability', A moderate indicator of ability', A strong indicator of ability') to indicate how strong an indicator of ability you think the property is.

There are no right or wrong answers to what we are asking, we are just interested in your judgements about this issue.

Shapes questionnaire	Not an indicator of ability at all	A weak indicator of ability	A moderate indicator of ability	A strong indicator of ability
1. Ease of evoking the image				
2. Amount of detail				
3. Clarity and sharpness				
4. Ease of maintenance				
5a. Detail change while maintaining				
5b. Clarity change while maintaining				
6. Amount of shape imaged at any one time				
7. Size of image				
8. Vividness of image				
9. Ease of rotation				
10. Whether rotation is continuous or in discrete stages				
11. Size change during rotation				
12a. Detail change while rotating				
12b. Clarity change while rotating				
13. Amount of shape imaged while rotating				
16. Vividness of image while rotating				

Please check that you have answered all the questions and supplied information on sex and age at the start of the questionnaire. In the space below explain any difficulties you had answering these questions or add any general comments you have to make.

Thank you for answering this questionnaire.

Auditory Imagery Scale

Instructions

Consider the following items. Rate them according to how well you can manage to imagine the sound / noise.

1	Very clear sound / noise
2	Moderately clear sound / noise
3	Vague sound / noise
4	No sound / noise at all

Circle one response for each image that you form

	Item	Response			
1	Imagine the sound of a car driving in the road in front of a house	1	2	3	4
2	Imagine the monotonous beep-like sound like in a telephone	1	2	3	4
3	Imagine the sound of footsteps coming up a stair	1	2	3	4
4	Imagine the sound of water dripping	1	2	3	4
5	Imagine the sound of snapping twigs	1	2	3	4
6	Imagine the noise of a conversation as if there were a party taking place next door	1	2	3	4
7	Imagine your favourite piece of music	1	2	3	4

Please explain below any difficulties you had answering this questionnaire or add any general comments you have to make.

Thank you for your cooperation

Individual Differences Questionnaire - IHS

The statements in this questionnaire represent ways of thinking which are true for some people and not for others. Read each statement and decide whether or not it is true with respect to yourself. Then indicate your answer by ticking one of the five possible answers which range from strongly agree to strongly disagree. Answer the statements as carefully and honestly as you can. The statements are not designed to assess the goodness or badness of the way you think. They are attempts to discover the methods of thinking you use consistently in various situations. There are no right or wrong answers.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
1. Listening to someone recount their experiences does not usually arouse mental pictures of the incidents being described					
2. I can easily picture moving objects in my mind					
3. I have only vague visual impressions of scenes I have experienced					
4. I think that most people think in terms of mental pictures whether they are completely aware or not					
5. I can close my eyes and easily picture a scene I have experienced					
6. When someone describes something that happens to them, I sometimes find myself vividly imagining the events that happened					
7. I never use mental pictures or images when trying to solve problems					
8. I find it difficult to form a mental picture of anything					
9. My thinking often consists of mental pictures or images					
10. I do not form a mental picture of people or places when reading of them					
11. I enjoy the use of mental pictures to reminisce					
12. I often use mental images or pictures to help me remember things					
13. When remembering a scene, I use verbal descriptions rather than mental pictures					

If you have any difficulties in answering the questionnaire or general comments to make please write them in this space. Thankyou for answering this questionnaire.

A questionnaire for rating the relative desire to possess vivid mental imagery in auditory and visual sensory modalities and visual thinking style

Mental images are sensations of an object in our mind when the object is NOT present. They can occur in all five of our senses, for example:

1) Auditory imagery - refers to the experience of 'hearing' the sound of an object in your mind.
2) Visual imagery - refers to the experience of 'visualizing' the visual appearance an object in your mind.

Vividness / thinking style

For a long time psychologists have known that people vary in their experience of mental imagery. Some people have very vivid and life-like mental images, while others only have the vaguest of sensory sensations. Psychologists have also found that people appear to differ in whether they have a visual style of thinking. People with a visual style of thinking find it easy to think visually and do so often.

We are interested in the value you would place on having vivid images in the visual and auditory modalities, and the value you would place on having a visual thinking style. To help us assess this please answer the questions below by placing a tick in the appropriate box to indicate your response. There are no right or wrong answers to what we are asking, just what you believe.

Question / modality / thinking style	No, does not matter to me	Yes, I would like to	Yes, I would like to quite a lot	Yes, I would like to very much
1) I would like to have vivid auditory imagery				
2) I would like to have vivid visual imagery				
3) I would like to have a visual thinking style				

Please check that you have answered all the questions and supplied information on sex and age at the start of the questionnaire. In the space below explain any difficulties you had answering these questions or add any general comments you have to make.

Thank you for answering this questionnaire.

Balanced inventory of desirable responding

Self-deceptive enhancement

Using the response scale below as a guide, write a number next to each statement to indicate how much you agree with it.

1	2	3	4	5	6	7
Not true			Somewhat true			Very true

Response		Item
	1.	My first impressions of people usually turn out to be right.
	2.	It would be hard for me to break my bad habits.
	3.	I don't care to think what other people really think of me.
	4.	I have not always been honest with myself.
	5.	I always know why I like things.
	6.	When my emotions are aroused, it biases my thinking.
	7.	Once I've made up my mind, other people can seldom change my opinion.
	8.	I am not a safe driver when I exceed the speed limit.
	9.	I am fully in control of my own fate.
	10.	It's hard for me to shut off a disturbing thought.
	11.	I never regret my decisions.
	12.	I sometimes loose out on things because I can't make up my mind soon enough.
	13.	The reason I vote is because my vote can make a difference.
	14.	My parents are not always fair when they punish me.
	15.	I am a completely rational person.
	16.	I rarely appreciate criticism.
	17.	I am very confident of my judgements.
	18.	I have sometimes doubted my ability as a lover.
	19.	It's alright if some people happen to dislike me.
	20.	I don't always know the reasons why I do the things I do.

Impression management

1	2	3	4	5	6	7
Not true			Somewhat true			Very true

Response		Item
	21.	I sometimes tell lies if I have too.
	22.	I never cover up my mistakes.
	23.	There have been occasions when I have taken advantage of someone.
	24.	I never swear.
	25.	I sometimes try to get even rather forgive and forget.
	26.	I always obey laws, even if I'm unlikely to get caught.
	27.	I have said something bad about a friend behind his or her back.
	28.	When I hear people talking privately, I avoid listening.
	29.	I have received too much change from a sales person without telling him or her.
	30.	I always declare everything at customs.
	31.	When I was young I sometimes stole things.
	32.	I have never dropped litter on the street.
	33.	I sometimes drive faster than the speed limit.
	34.	I never read sexy books or magazines.
	35.	I have done things that I don't tell people about.
	36.	I never take things that don't belong to me.
	37.	I have taken sick leave from work or school even though I wasn't really sick.
	38.	I have never damaged a library book or store merchandise without reporting it.
	39.	I have some pretty awful habits.
	40.	I don't gossip about other people's business.

Self-deceptive denial

1	2	3	4	5	6	7
Not true			Somewhat true			Very true

Response		Item
	41.	I sometimes feel irritated when I don't get my own way.
	42.	I could never enjoy being cruel.
	43.	Seeing any attractive person of the opposite sex makes me think about sex.
	44.	I have never felt joy over someone else's failure.
	45.	I have gotten so angry at a friend that I felt like hitting him(her).
	46.	I have never felt like I wanted to kill someone.
	47.	There have been occasions when I was mean to someone unimportant.
	48.	I never enjoy watching sexy scenes in movies.
	49.	I enjoy it when obnoxious people get put down.
	50.	I rarely have sexual fantasies.
	51.	Once in a while I think of things too bad to talk about.
	52.	I have never wanted to rape or be raped by someone.
	53.	More than once it felt good when I heard on the news that someone had been killed.
	54.	I can't think of anyone I hate deeply.
	55.	There have been occasions when I felt like smashing things.
	56.	Few of the things I do are simply for my own gain.
	57.	I must admit that revenge can be sweet.
	58.	I never get jealous over the good fortune of others.
	59.	There have been times when I felt like rebelling against authorities, even though I knew they were right.
	60.	I have never done anything that I'm ashamed of.

Agency management

1	2	3	4	5	6	7
Not true			Somewhat true			Very true

Response		Item
	61.	My persuasive powers are impossible to resist.
	62.	My personality doesn't charm everyone.
	63.	Everyday I come up with creative, ingenious ideas.
	64.	My persuasive powers are sometimes weak.
	65.	No one can beat me in a debate.
	66.	Some problems just don't have a solution.
	67.	Some people call me a genius.
	68.	I can't win at everything.
	69.	My leadership of a group guarantees the group's success.
	70.	Some people are just impossible to persuade.
	71.	I push until things happen.
	72.	Some powerful people cannot be overcome.
	73.	I can dominate people whenever I want.
	74.	I have met people smarter than myself.
	75.	I inspire others by my success.
	76.	I sometimes need other people's help.
	77.	I have mastered every challenge put before me.
	78.	My decisions are sometimes unwise.
	79.	I am successful in everything I do.
	80.	Giving into others is sometimes necessary.

Please check that you have answered all the questions and supplied information on sex and age at the start of the questionnaire. In the space below explain any difficulties you had answering this questionnaire or add any general comments you have to make.

Thank you for your cooperation.

Appendix 6: Research Gate read statistic screenshot 30th July, 2018

Figure 4. Research Gate read statistic screenshot 30th July, 2018

Appendix 7: Additional correlational analysis on paper 4 data controlling for gender

Key

VVIQ = Vividness of visual imagery questionnaire version 1
SDE = Self-deceptive enhancement (scored continuously and dichotomously)
AM = Agency management (scored continuously and dichotomously)
SDD = Self-deceptive denial (scored continuously and dichotomously)
IM = Impression management (scored continuously and dichotomously)
MC = The Marlowe Crowne

Table 13.

Matrix showing correlational analysis involving the VVIQ-1, BIDR scales and the Marlow-Crowne scale controlling for gender

Control Variables			VVIQ_sq
sex	VVIQ_sq	Correlation	1.000
		Significance (2-tailed)	.
		df	0
	SDE_c	Correlation	-.341
		Significance (2-tailed)	.000
		df	116
	SDE_d_sq	Correlation	-.233
		Significance (2-tailed)	.011
		df	116
	AM_c	Correlation	-.260
		Significance (2-tailed)	.004
		df	116
	AM_d_sq	Correlation	-.130
		Significance (2-tailed)	.160
		df	116
	SDD_c	Correlation	-.074
		Significance (2-tailed)	.428
		df	116
	SDD_d_sq	Correlation	-.042
		Significance (2-tailed)	.654
		df	116
	IM_c	Correlation	-.114
		Significance (2-tailed)	.218
		df	116
	IM_d_sq	Correlation	-.047
		Significance (2-tailed)	.614
		df	116
	MC_sq	Correlation	-.187
		Significance (2-tailed)	.043
		df	116

Appendix 8: Additional correlational analysis involving the VVIQ-1, BIDR egoistic scales and the over-claiming questionnaire

Key

VVIQ = Vividness of visual imagery questionnaire version 1
SDE = Self-deceptive enhancement (scored continuously and dichotomously)
AM = Agency management (scored continuously and dichotomously)
OCQ_BiasZ = Over-claiming bias index calculated from adding hits Z score to false alarms Z score
OCQ_False alarms = Number of false alarms

Spearman's correlations were used because of skewed distributions across all variables.

Table 14.

Matrix showing correlational analysis involving the VVIQ-1, BIDR egoistic scales and the over-claiming questionnaire

Correlations

			wiq	SDE_c	SDE_d	AM_c	AM_d	OCQ_BiasZ	OCQ_FAs
Spearman's rho	wiq	Correlation Coefficient	1.000	-.339**	-.386**	-.191*	-.136	-.213**	-.241**
		Sig. (2-tailed)	.	.000	.000	.012	.077	.005	.001
		N	171	171	171	171	171	171	171
	SDE_c	Correlation Coefficient	-.339**	1.000	.822**	.499**	.393**	.219**	.224**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.004	.003
		N	171	171	171	171	171	171	171
	SDE_d	Correlation Coefficient	-.386**	.822**	1.000	.466**	.524**	.253**	.238**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.001	.002
		N	171	171	171	171	171	171	171
	AM_c	Correlation Coefficient	-.191*	.499**	.466**	1.000	.617**	.194*	.234**
		Sig. (2-tailed)	.012	.000	.000	.	.000	.011	.002
		N	171	171	171	171	171	171	171
	AM_d	Correlation Coefficient	-.136	.393**	.524**	.617**	1.000	.204**	.245**
		Sig. (2-tailed)	.077	.000	.000	.000	.	.007	.001
		N	171	171	171	171	171	171	171
	OCQ_BiasZ	Correlation Coefficient	-.213**	.219**	.253**	.194*	.204**	1.000	.907**
		Sig. (2-tailed)	.005	.004	.001	.011	.007	.	.000
		N	171	171	171	171	171	171	171
	OCQ_FAs	Correlation Coefficient	-.241**	.224**	.238**	.234**	.245**	.907**	1.000
		Sig. (2-tailed)	.001	.003	.002	.002	.001	.000	.
		N	171	171	171	171	171	171	171

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).