British Journal of Psychology (2003), 94, 245–273

www.bps.org.uk

The relationship between self-reports of imagery and spatial ability

Graham M. Dean[1]* and Peter E. Morris[2]
[1]FMAS, Southampton Institute, UK
[2]Department of Psychology, Lancaster University, UK

A puzzling question arising from imagery research is why no relationship has been found between self-reports of imagery and performance on spatial tests thought to require the use of imagery. To investigate this, spatial ability, measured by performance on two spatial tests, was compared with performance on the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and a newly constructed imagery questionnaire. The choice of items and ratings for the new questionnaire was based on Kosslyn's (1980, 1994) theories of the imagery system. Ratings on the new questionnaire consistently correlated significantly with performance on the spatial tests, whereas ratings from the VVIQ did not. The new ratings captured more of the imagery process than ratings of vividness alone, but the largest change depended upon the type of item imagined. Ratings of items of the same type as used on the spatial tests predicted performance on the spatial tests, whereas vividness ratings of items recalled or constructed from long-term memory did not. Participants can successfully introspect on several different properties of their images, and their ratings do predict performance on tasks thought to require imagery. The large effect of item type raises questions about the predictive value of existing models of the imagery system especially in relation to the role of phenomenological properties of our images.

Explanations of the results of mental rotation experiments often suggest that mental imagery plays a functional role (e.g. Shepard & Metzler, 1971). The linear increase in response times with increasing angular disparity between the stimulus items that are being compared provides an objective measure that supports subjective reports that mental imagery is involved in spatial manipulations. Paivio (1971) and others have suggested that spatial tests based on this paradigm can be used as indirect measures of imagery ability, and Ernest (1977) refers to such tests as 'objective' measures of imagery.

*Requests for reprints should be addressed to Graham M. Dean, Faculty of Media, Arts and Society, Southampton Institute, East Park Terrace, Southampton SO14 0YN, UK (e-mail: graham.dean@solent.ac.uk).

Theories and models of spatial ability within the psychometric and information-processing traditions (Carpenter & Just, 1986; Kyllonen, 1996; Lohman, 1996; Mumaw, Pellegrino, Kail, & Carter 1984; Pelegrino & Kail, 1982; Poltrock & Agnoli, 1986) regularly identify the existence and quality of an internal visuospatial mental representation as having a central role in the solution of spatial problems and as being one possible source of individual differences in the performance of spatial tasks.

It is puzzling, therefore, that little or no correlation has been found between measures based upon subjective reports of the conscious experiences of imagery and experimental tasks or spatial tests that are explained in terms of their use or manipulation of mental images (e.g. Danaher & Thoresen, 1972; DiVesta, Ingersoll, & Sunshine, 1971; Durndell & Wetherick, 1976a, b; Ernest, 1977; Hatakeyama, 1981, 1984; Kosslyn, Brunn, Cave, & Wallach 1984; Paivio, 1971; Poltrock & Brown, 1984; Rehm, 1973; Richardson, 1977; Sheehan & Neisser, 1969). The results seem to show that imagery measures based on introspection,[1] especially questionnaires, are psychometrically independent of spatial ability tests, often loading on orthogonal factors.

There are three plausible explanations for this lack of relationship. The first is that the phenomenological experience of a mental image either plays no functional role in spatial tasks or does not reflect the processes that do. The second is that our methods of measuring mental imagery using introspection are at fault. The third is the incorrect use of introspective measures in a way that assumes vivid imagery is better imagery and thus implicitly treats imagery as a single ability. We believe that much of the problem arises from the failure to ask theory-driven questions about 'established' introspective measures of imagery and the ways in which they are used. Difficulties arise between what is purportedly measured by self-reports and the nature of the questionnaires themselves. As Paivio observed:

> Despite its venerable history, the individual differences approach has not shown the same degree of systematic, empirical and theoretical progress as approaches based on experimental procedures and manipulation of stimulus materials . . . . Instead, the research has consisted mainly of piecemeal attempts to predict memory or some other skill using scores on a single test that is supposed to measure imagery ability. (Pavio, 1989, p. 8)

Imagery 'ability' is most frequently measured by asking participants to rate the vividness of their images. The use of vividness as a measure of imagery can be traced back to Galton's (1883) original breakfast table questionnaire. The majority of subsequent vividness questionnaires have been derived from this measure. The widely used Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) is an expanded version of the visual sub-scale of Betts' Questionnaire on Mental Imagery (QMI; Betts, 1909), which in turn was derived from Galton's original questionnaire. It requires respondents to rate the vividness of images of 16 scenes or items recalled from long-term memory. The VVIQ has been shown to correlate with a very wide array of tasks with little clear theoretical similarity, though the correlations are not consistently found (see Marks, 1989, for a bibliography, and McKelvie, 1995). However, as has been observed above, most studies show that, like other self-report measures of imagery, the VVIQ fails to correlate with performance on the spatial tests that are thought to involve the use of imagery.

Vividness is supposed to represent that most fundamental property of the subjective

---

[1] *In this paper, we use the term 'introspections' to refer to subjective reports through which participants attempt to convey their conscious experiences. We do not refer to the sophisticated techniques that some psychologists in the early days of experimental psychology (e.g. Titchener; see Morris & Hampson, 1983) sought to employ.*

experience of imagery, its actual resemblance to perceiving real objects and events (Morris & Hampson, 1983), though McKelvie (1995) provides alternative definitions. The one thing common to the many interpretations of vividness is that none of them has a particularly strong reference to theoretical models and findings concerning the nature of the imagery system. As Morris and Hampson (1983) and Richardson (1994) observed, there is no empirical or a priori reason for the choice of this factor to measure imagery.

The use of vividness as a measure of imagery carries the implicit assumptions that vivid imagery is better imagery, that imagery is a single ability (Morris & Hampson, 1983) and that the best way to characterize imagery is by measuring vividness. The work of Kosslyn and his co-workers (e.g. Kosslyn 1980, 1994) and Farah (1984) provides evidence that imagery is best characterized as a collection of abilities. Kosslyn (1980) conceptualizes an image as the end product of a collection of differentiated constructive processes. The surface properties of this image (our phenomenological experience) are a reflection of both the results of these processes and the properties of a structure he refers to as the 'visual buffer'.

The three broad classes of process that Kosslyn identifies are those of image formation, maintenance and transformation. The main surface properties of the visual buffer are its resolution and spatial extent. Experimental evidence from studies on image latency and formation time (e.g. Cocude & Denis, 1988; Morris & Reid, 1973) and image scanning (Kosslyn, 1973) seems to indicate that these processes, or properties related to them, are accessible to introspection. Considering this, is vividness the most appropriate way to measure imagery? An overall measure of imagery such as vividness is at best a reflection of the combination of the properties and processes involved. It may not fully reflect the contribution of all these properties and underlying processes to our experience of imagining, or might fail to capture their contribution at all. Likewise, treating imagery as a single ability and dividing subjects into 'good' and 'poor' imagers on the basis of vividness scores might not reflect the contribution of different processes and properties of an image (i.e. multiple abilities) used in solving a spatial task. The reasoning that measurements of vividness of imagery are different to the processes measured on spatial tasks is supported by Poltrock and Agnoli (1986) and Richardson (1999), who observe that although men tend to score more highly than women on tests of spatial ability, the reverse is true for reports of vividness of imagery. If we make the assumption that the properties of images open to introspection are used in performing spatial tests, this reversal of group differences clearly indicates different things are being measured. Baddeley and Andrade (1998) argued that vivid imagers only differ from non-vivid imagers in their strategic use of supplementary information to fill in default values in an image. They further suggest that the two groups do not differ in the amount of information they acquire. This, they reason, explains why no difference between vivid and non-vivid imagers is found on tasks where performance is determined by the information acquired.

Further evidence that image quality is not necessarily a good guide to the accuracy of imagining comes from Hinton's (1979) finding when participants imagine an antiprismatic cube. Frequently, they reported imagining such a cube fairly confidently. However, they then reported an incorrect number of visible corners and incorrect proportions for the visible surfaces. Self-reports of successfully imagining an object are, therefore, not necessarily reflections of the accuracy of the image.

The evidence suggests that the functional role of imagery in spatial ability tests is unrelated to the vividness of imagery. It follows that ratings of the properties and processes identified by theories of imagery may provide more functional measurements of imagery than vividness alone.

Reference to theories of visual processing and the imagery system also indicates a second aspect of imagery questionnaires that needs to be examined: the items. Striking differences can be found between the stimuli contained in imagery questionnaires and those used in the spatial tests with which they are compared. The first major difference is in the composition of the items. Imagery questionnaires (e.g. VVIQ) ask participants to imagine a variety of real-life scenes or objects. These often contain or require several types of motion or manipulation, and colour, and often contain several items. Conversely, spatial tests frequently require participants to manipulate abstract geometrical shapes of one type in a specific way. These shapes are black and white line drawings and are usually previously unseen. The second major difference is that the origin of the stimuli is very different. The real-life items in questionnaires are either recalled or constructed from elements stored in long-term memory. The shapes manipulated on the spatial tests are usually previously unseen. They are perceived and then held or manipulated in short-term memory. The large differences in stimuli type raise the question of how valid it is to assume that the processes and quality or performance of imagining these different types of stimuli is functionally equivalent.

If, as the findings of Kosslyn (1980, 1994) and Farah (1984) suggest, imagery is the result of a collection of processes rather than a single ability, as is incorrectly implied by the misuse of vividness as a measure of imagery, it seems logical to suggest that the content of images could place proportionally different demands on the processes that produce, manipulate and (possibly) inspect them. Cooper (1995) argued that different cognitive tasks activate different representational properties of distinguishable sub-systems. It is also possible that different content places differential demands on the structures that support the visual image. This might be reflected in the phenomenological properties of the surface image.

Problems with existing questionnaires as a measure of imagery may arise because items recalled or constructed from long-term memory could involve processes not used when imagining items that are perceived then imagined. However, Logie (1995), in presenting a modified version of the working memory model, suggested that visual input is placed in the visual or spatial short-term stores via long-term memory representations of visual form or spatial information. This would suggest that many of the operations are shared between these two modes of input to the short-term store. The difference in processes used between these two types of input may therefore not be as great as is suggested by the observation that there are two routes of input.

We suggest that ratings of everyday items do not reflect the same set of processes that are used when imagining the shapes on spatial tests. Some evidence for this was provided by Lorenz and Neisser (1985), who found that performance on two tests of spatial ability correlated with a questionnaire based on imagery in performing the tests but found no correlation with scores on standard imagery questionnaires. A similar finding was reported by Cochran and Wheatley (1989), who found a correlation between the rated difficulty of a holistic imagery strategy and performance on a space relations test. Therefore, the failure to find a relationship between self-reports of imagery and spatial tasks could be a result of existing imagery questionnaires failing to adequately capture the properties and processes of imagery that are relevant.

Selection of spatial tests

For our study, mental rotation was selected as the spatial task to be explored. Thirty years of extensive investigation into mental rotation (e.g. Shepard & Metzler, 1971) has identified it as a distinct mental operation that can be measured by both psychometric tests and experimental tasks. Considering that the previous research on imagery ability and spatial performance to which this experiment would be compared was largely based on pencil-and-paper spatial tests, it was decided to use this format to assess spatial ability.

The first test chosen was the Comprehensive Ability Battery Spacial Test (CAB-S; Hakstian & Cattell, 1976). This requires participants to perform rotation of two-dimensional (2D) shapes in the plane of the picture. The second test selected was the Vandenberg test of mental rotation (Vandenberg & Kuse, 1978), which requires participants to manipulate three-dimensional (3D) shapes of the same kind used in Shepard and Metzler's (1971) original experiment on mental rotation. Cooper (1991) has argued that reflection rotation is not part of the structural system and thus might be different from 3D mental rotation. There is evidence from the individual differences tradition (see Lohman, Pellegrino, Alderton, & Regian, 1987 for a review) that different types of rotation tasks, whilst correlated together, load on different spatial sub-factors. 2D tests of rotation involve simple left–right discrimination that is achieved as quickly as possible by a variety of strategies. These tests load on the space relations or speeded rotation sub-factor. The more complex 3D rotation tasks show stronger loadings on the most general spatial sub-factor spatial visualisation. Whilst both the CAB-S and the Vandenberg tests are tests of spatial ability, and do measure common processes, it is clear that they also measure different processes. It is likely that 3D tests of rotation place greater demands on the properties and qualities of internal mental representations. Selection of the two tests using the same operation but placing different demands on the internal representation used to solve the items therefore provides a useful comparison for the correlation of the questionnaire ratings. One shape was selected from each test as a stimulus to be imagined for the new questionnaire.

To examine whether imagery played a functional role in the two spatial tests, scores on these tests were compared with both an existing imagery questionnaire (VVIQ) and a new questionnaire constructed with reference to the issues raised above. The new questionnaire had to be designed to assess properties of imagery related to models of the imagery system other than vividness and contain relevant items.

Selection of items

The new questionnaire was constructed to ask participants to look at an item then imagine it and rate that image. The choice of the items for the new questionnaire was dictated by the choice of spatial tests.

A common thread to several theories of visual and spatial processing is the distinction between processes that deal with 2D and 3D information. Cooper (1991) distinguishes between structural and pictorial representations, Kosslyn (1980, 1994) makes a similar distinction between pictorial and spatiotopic groups of processes, and the visuospatial sketchpad of the working memory model (e.g. Baddeley, 1986; Logie, 1995) has distinct visual and spatial components. This well-defined distinction provides the opportunity to investigate whether differences in an item's spatial content affect the properties of an image that are experienced. Considerable evidence exists that item content can significantly affect vividness levels (e.g. Bat-Zion, 1986). However, the difference might not be simply quantitative. Different processes are probably involved in processing 2D and 3D stimuli. Therefore, ratings for each shape may not correlate, or may correlate selectively with 2D and 3D spatial tasks, because the selective use of these different underlying processes might be reflected in the quality of the surface image.

Selection of ratings

The aspects of imagery it was hoped to capture with the new questionnaire were the processes identified by Kosslyn (1984), the structural properties of the medium (e.g. spatial extent or resolution), and the properties and qualities of the representation itself. However, when using introspective ratings, it is difficult to separate out the properties of the representation from the qualities or properties of the image that reflect the operation of underlying processes or structures of the system.

Two of the defining features of the visual buffer (Kosslyn, 1984) are its limited resolution and spatial extent. It is possible that introspectively available properties of an image such as clarity and the amount of detail are reflections of the resolution of the visual buffer. Assessment of the spatial extent of the buffer is more problematical, as this might be reflected in the proportion of an item imagined or the size or relative distance at which an item is imagined. The logical possibility exists that there are trade-offs between the resolution and spatial extent of the visual buffer. The ratings chosen to assess the properties of the visual buffer were of detail, clarity, proportion and relative size of an image. Additionally, a rating of vividness was added so that vividness ratings could be compared between the shape stimuli on the new questionnaire and vividness ratings from the VVIQ.

The demands of transforming an image or the process of rotation might be reflected by a change in the surface properties of an image, so a comparison needs to be made between such ratings of a static shape and ratings of a shape that is being rotated. Ratings were therefore selected to assess the above properties of both a static and a rotating shape.

Three main classes of processes acting upon the visual buffer were identified: formation, maintenance and transformation. Assessment of the processes raises several problems. Evidence from image latency and formation experiments (e.g. Cocude & Denis, 1988; Morris & Reid, 1973) suggests that participants are aware of these processes, but in all probability, they cannot accurately introspect on their speed. Therefore, it is necessary to resort to judgments of the subjective ease of these processes. Three ratings were selected: how easily participants found it to evoke the image; how easily they could maintain it once formed; and how easily they could transform the image.

In order to reduce the demands on participants, it was decided to ask them to rate all the above properties sequentially. This would have the additional effect of collecting ratings concerning the same imagery event, though place demands on memory for that event. The questionnaire required participants to rate the ease of forming an image and then rate the ease of maintaining it. Following this, they would be asked to rate the pictorial properties of this same image. Then, the questionnaire asked participants to imagine the shape rotating and rate the ease of this before assessing the pictorial properties.

It is possible that the order of presentation of the tests could affect the results. It was decided to give the imagery questionnaires before the spatial tests rather than after. It could be argued that giving the questionnaire before persuades participants to use an imagery strategy in the spatial tests. However, giving the questionnaire afterwards means that participants are very familiar with the shapes and the action of rotating them. Additionally, as McKelvie (1995) notes, experimental tasks can cue VVIQ responses. If the experimental task is administered first and participants perceive their performance as good or poor, this may influence their self-rating as good or poor visualizers.

Sex differences and spatial ability

The examination of sex differences in performance on the tests and questionnaire was not the primary purpose of this experiment. However, as Richardson (1991) observes, reporting of the existence of sex differences is frequently haphazard, which both complicates and confuses the future analysis of spatial test results.

Spatial tests frequently show a male advantage in performance (Kimura, 1996). Some of the largest differences have been found using 3D mental rotation tests such as the Vandenberg test of mental rotation (Linn & Petersen, 1985; Wilson & Vandenberg, 1978). These differences are far greater than those shown for 2D rotation tasks (Collins & Kimura, 1997).

The cause of both the observed sex differences and the differential magnitude of differences across different rotation tasks is a matter of considerable debate (see Caplan, MacPherson, & Tobin, 1985; Linn & Petersen, 1985; McGee, 1979; Tapley & Bryden, 1977). For the purposes of this experiment, there are two issues arising from these observed sex differences in performance on spatial tests and the explanations given to them: first, whether the factors responsible for variation between groups are the same as those responsible for variation in scores within the entire sample (as suggested by Mumaw *et al*., 1984); second, whether there are qualitatively different ways in which the two groups use imagery in relation to spatial performance. The major concern for this investigation is whether it is valid to analyse the correlations of questionnaire ratings for the entire sample, or whether qualitative differences in males' and females' use of imagery (possibly strategy differences) would result in a differential pattern of correlations.

Horan and Rosser (1984) present evidence that there is a male advantage in transforming a 2D picture into an internal 3D representation. Collins and Kimura (1997) and Stumpf (1993) argue that performance differences are a function of task difficulty caused in part by the difficulty of the representing the information necessary to perform a rotation task. These latter two explanations would suggest that a male advantage should be found in imagining stimuli that are similar to those used on spatial tests and that it is differences in the quality of representation formed that are responsible for differences in performance on the spatial tests. This is the opposite of what has already been noted: The male advantage in performance on spatial tests is the opposite of the female advantage sometimes displayed on measures of vividness. A prerequisite of investigating this issue is that it is necessary in the first instance to see whether any relationship can be found between self-reports of imagery and performance on spatial tests.

Method

Participants

Initially, 211 undergraduates took part in this experiment. They were an opportunistic sample, volunteering to take part in the experiment after their scheduled class. Unfortunately, not all of the participants completed all the testing sessions. The data reported come from the 165 who completed both the questionnaires and at least one of the spatial tests. Fifty-one were males, and 114 were females. None had previously completed any imagery questionnaire. Participants were aged between 18 and 21 and took part voluntarily.

Materials

The two spatial tests used were the (CAB-S) Test of Mental Rotation and the Vandenberg Test of Mental Rotation (Vandenberg & Kuse, 1978). The imagery questionnaires used were the VVIQ (Marks, 1973) and the questionnaire constructed to explore the issues raised above (Dean & Morris, 1991).

The questionnaire contained the same ratings as the questionnaire used by Dean and Morris (1991). The properties that participants were required to rate are listed below, and the items participants were required to imagine are displayed in Fig. 1. These items were displayed at the top of each set of questions, and participants were instructed to look at the shapes, imagine them and then rate their images.

Participants were asked to rate the static shape on the following parameters: How easily can you evoke an image of this shape?

(1) How much detail is there in your image?
(2) How clear and sharp (in pictorial terms) is your image?
(3) How easily can you maintain this image now that you have evoked it?
(4) How much does (a) the detail and (b) the clarity of your image change when you try to maintain your image?
(5) How much of the shape can you form an image of at any one time?
(6) How large is your image of the shape?
(7) How vivid is your image of the shape?

Participants were also asked to imagine the shape rotating and rate it according to the following questions:

(1) How easily can you imagine the rotation?
(2) Is the rotation you can imagine continuous or in discrete stages?
(3) Does the size of the imagined shape (or the apparent distance at which it is imagined) change when you rotate it?
(4) If so, in what way does your image of the shape change?
(5) Does rotating the shape affect (a) the detail and (b) the clarity of your image in any way?
(6) How much of the shape do you form an image of, at any one time when rotating it?
(7) How vivid is your image of the shape rotating?

All ratings were made on a scale of 1 to 9 (1 = very poor, 9 = good) except for ratings of change in detail and clarity (rated on a scale of 1 to 9 then transformed to a scale of – 8 to + 8 according to whether a subject circled more or less below the scale), proportion (rated on a scale of 1 to 10) and whether the rotation was performed in continuous or discrete stages, which was rated 'Yes or No'. Additionally, participants were asked to mark on copies of the stimuli to be imagined what parts of the shape they imagined

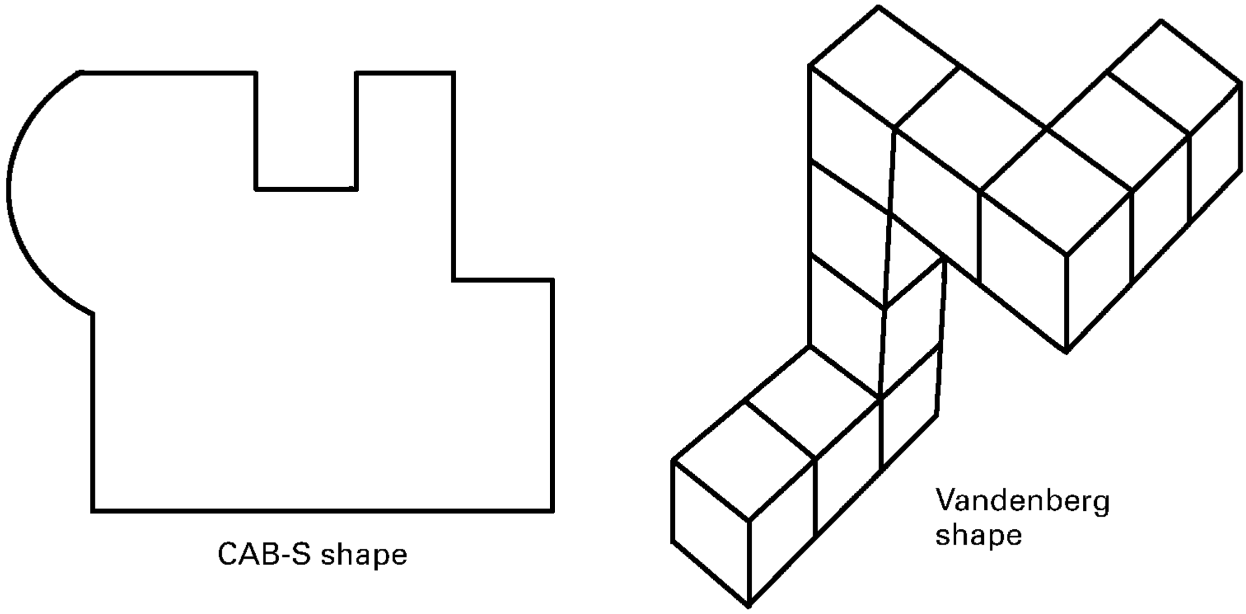

Figure 1. 2D and 3D shapes used in the questionnaire.

when static and rotated if they had imagined all the stimulus at one time. The questionnaire is reproduced in the Appendix.

Procedure

The tests were administered in four sessions (one test or questionnaire per session) with a 1-week interval between each session. Participants were tested in eight separate groups of approximately 20 people each. Sessions lasted between 12 and 20 min depending on the test being administered. The tests were administered in the following order: VVIQ, the new questionnaire, CAB-S and Vandenberg.

Participants were allowed as much time as they wished to answer both the questionnaires. The spatial tests had time limits of 3 min for CAB-S and 2 min for each section of the Vandenberg test. Before the administration of both spatial tests, participants read through the test instructions both in their own time and had them read to them by the test administrator to ensure their understanding.

Total scores for the VVIQ were calculated by adding the scores on both the eyes-closed and eyes-open administration of the VVIQ. The CAB-S test scores were the number of correct items selected. Scores on the Vandenberg test were calculated using the correction for guessing procedure outlined by Vandenberg and Kuse (1978). The rating for each item on the new questionnaire was taken as the score for that item.

Results and discussion

Abbreviations and numerical values used for each of the variables in the tables are listed in Table 1. For all analyses, missing data have been deleted pairwise. Because of the large number of possible comparisons between variables, we draw attention to patterns of intercorrelations rather than highlighting individual significant relationships.

Table 1. Abbreviations used in tables and figures

Abbreviation	Variable
CABS	CAB-S Test
VAND	Vandenberg Test
VVIQ	Vividness of Visual Imagery Questionnaire
evoke	Ease of evocation
det	Amount of detail
clar	Clarity of image
maint	Ease of maintenance
chdet	Detail change during maintenance
chclar	Clarity change during maintenance
prop	Proportion of shape imagined
size	Relative size of image
vivid	Vividness of image
Rease	Ease of rotation of image
Rcdis	Continuous or discrete stages
Rsizch	Size change during rotation
Rdetch	Detail change during rotation
Rclach	Clarity change during rotation
Rprop	Proportion of shape imagined during rotation
Rvivid	Vividness of image during rotation

Note. The prefix C indicates the 2D shape from the CAB-S test, and prefix V the 3D shape from the Vandenberg test.

Correlations of the spatial tests and VVIQ

It can be seen from the correlations of the questionnaires with the spatial tests (Table 2) that the VVIQ fails to correlate significantly with either the CAB-S test or the Vandenberg test, but the two spatial tests correlate together significantly.

Correlations of the questionnaire ratings with the spatial tests

Examination of the ratings from the new questionnaire reveals that several of the ratings for both the shapes correlate significantly with performance on both the tests.

For the CAB-S test, the significant correlations (see Table 2) for ratings of the 2D shape are: evoke, detail, clarity, proportion, vividness, ease of rotation, proportion during rotation and vividness during rotation. The significant correlations for ratings of the 3D shape are: evoke, detail and clarity change during rotation.

For the Vandenberg test, the significant correlations of ratings of the 2D shape are: maintain, proportion, vividness, proportion during rotation and vividness during rotation. The significant correlations for ratings of the 3D shape are: evoke, detail, clarity, maintain, size, ease of rotation, continuous or discrete rotation, size change during rotation, proportion during rotation and vividness during rotation.

Ratings for the 2D shape (obtained from the CAB-S test) have a greater number of significant correlations with performance on the CAB-S test than with performance on the Vandenberg test. Ratings for the 3D shape (obtained from the Vandenberg test) have an even clearer pattern of significant correlations with only three ratings correlating significantly with CAB-S scores and 10 ratings correlating significantly with Vandenberg scores. In all cases, ratings for the 2D shape that correlate significantly with CAB-S scores

Table 2. Correlations of ratings with spatial tests and VVIQ ($N = 165$)

Variable	VVIQ	CABS	VAND
VVIQ	1.00	−.01	−.04
CABS	−.01	1.00	.39**
VAND	−.04	.39**	1.00
Cevoke	−.11	.33**	.09
Cdet	−.14	.26**	.14
Cclar	−.16	.17*	.15
Cmaint	−.12	.14	.22**
Cchdet	−.08	−.11	.06
Cchcla	−.09	−.13	−.01
Cprop	−.11	.30**	.16*
Csize	−.05	.14	.06
Cvivid	−.11	.19*	.22**
CRease	−.19*	.25**	.13
CRcdis	−.23**	.09	.05
CRsizch	.07	−.01	.04
CRdetch	−.24**	−.04	.08
CRclach	−.20*	−.03	.04
CRprop	−.21*	.35**	.19*
CRvivid	−.22*	.22**	.18*
Vevoke	−.07	.16*	.23**
Vdet	−.10	.21**	.32**
Vclar	−.15	.05	.22**
Vmaint	−.26**	.15	.28**
Vdetch	−.18*	−.12	−.01
Vclach	−.19*	−.11	−.03
Vprop	−.22*	.12	.14
Vsize	−.04	.06	.18*
Vvivid	−.27**	.00	.16
VRease	−.31**	.04	.37**
VRcdis	−.26**	.08	.20*
VRsizch	.06	.06	.18*
VRdetch	−.14	−.15	.09
VRclach	−.12	−.20*	.13
VRprop	−.23**	.10	.27**
VRvivid	−.21*	.00	.33**

$*p<.05$; $**p<.01$.

account for more of the variation in performance on the CAB-S test than ratings for the 3D shape. The same is true for ratings of the 3D shape compared with ratings of the 2D shape against prediction of performance on the Vandenberg test. Care has to be taken when comparing the significance or magnitude of two correlations, as the two correlation coefficients may not be significantly different in magnitude. Evidence that these differences in the magnitude of correlation are meaningful will be discussed below.

Comparison of correlations of vividness ratings

The first point to note is that vividness ratings of the shapes rotating correlate significantly with performance on the spatial tests. This must be compared with the

Table 3. Correlations of ratings for the 2D shape ($N = 165$)

	CEVOKE	CDET	CCLAR	CMAINT	CCHDET	CCHCLAR	CPROP	CSIZE	CVIVID
CEVOKE	1.00								
CDET	.65**	1.00							
CCLAR	.73**	.67**	1.00						
CMAINT	.54**	.52**	.68**	1.00					
CCHDET	.04	.02	.11	.23**	1.00				
CCHCLAR	.00	.00	.13	.19*	.85**	1.00			
CPROP	.46**	.39**	.43**	.38**	.03	– .01	1.00		
CSIZE	.19*	.25**	.19*	.20*	.03	.05	.09	1.00	
CVIVID	.48**	.50**	.67**	.55**	.05	.08	.48**	.31**	1.00
CREASE	.30**	.28**	.32**	.36**	.04	– .03	.33**	.18*	.36**
CRCDIS	.09	.16*	.02	.05	.13	.12	.05	– .12	– .09
CRSIZCH	.05	– .03	– .01	.08	.14	.14	.18*	.09	.01
CRDETCH	.07	.13	.10	.15	.54**	.47**	– .04	– .08	.02
CRCLACH	– .01	.11	.03	.13	.40**	.47**	– .06	– .03	– .01
CRPROP	.32**	.39**	.28**	.33**	.10	.06	.53**	.18*	.36**
CRVIVID	.20**	.21**	.26**	.30**	.17*	.15	.34**	.21**	.40**

	CREASE	CRCDIS	CRSIZCH	CRDETCH	CRCLACH	CRPROP
CRCDIS	.11	1.00				
CRSIZCH	.04	– .14	1.00			
CRDETCH	.06	.23	.02	1.00		
CRCLACH	.07	.24**	.08	.75**	1.00	
CRPROP	.42**	.19*	.12	.07	.11	1.00
CRVIVID	.48**	.13	.10	.20*	.19*	.53**

*$p < .05$; **$p < .01$.

Table 4. Factor analysis of ratings for the 2D shape (oblimin rotation)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Cevoke	.86**	−.04	−.04	−.03
Cdet	.82**	.01	.04	−.16
Cclar	.95**	.05	−.08	−.03
Cmaint	.74**	.17	.04	.07
Cchdet	.07	.83**	−.09	.31
Cchclar	.08	.83**	−.14	.33
Cprop	.34**	−.14	.43**	.21
Csize	.22	−.08	.09	.40**
Cvivid	.65**	−.05	.21	.09
Crease	.07	−.05	.73**	−.07
CRsizch	−.17	.08	.09	.76**
CRdetch	.05	.84**	.12	−.29*
CRclach	−.05	.79**	.20	−.25
CRprop	.05	−.00	.77**	.09
CRvivid	−.09	.14	.84**	.05
Percentage variance	31.3	18.2	9.2	7.3

**Denotes highest loading variables.

Table 5. Factor correlation matrix for the 2D shape

	Factor 1	Factor 2	Factor 3
Factor 2	.03	1.00	
Factor 3	.43**	.07	1.00
Factor 4	.16	.02	.14

*$p < .05$; **$p < .01$.

failure of vividness ratings of 'real-life' objects from the VVIQ to correlate significantly with performance on the tests. VVIQ scores do, however, correlate significantly with vividness ratings for the shapes during imagined rotation. Significant correlations are obtained for ratings of vividness during rotation for the 2D shape, vividness of the 3D shape and vividness of the 3D shape rotating. The correlations obtained are negative because the rating scale on the VVIQ is the opposite direction from the rating scale on the new questionnaire.

Intercorrelation of questionnaire ratings

Examination of the correlations of the ratings for each shape (Tables 3 and 6) revealed that eight of the ratings for both the shapes strongly intercorrelate. These ratings are evoke, detail, clarity, maintain, proportion, vividness, ease of rotation, proportion during rotation and vividness during rotation. It was expected that there would be some relationship, simply because some properties or processes are prerequisites of others.

The obvious question to arise from so many ratings of the same shape correlating significantly with each other and performance on the test is whether all the ratings are

Table 6. Correlations of ratings for the 3D shape ($N = 165$)

	VEVOKE	VDET	VCLAR	VMAINT	VCHDET	VCHCLA	VPROP	VSIZE	VVIVID
VDET	.66**	1.00							
VCLAR	.63**	.71**	1.00						
VMAINT	.65**	.60**	.69**	1.00					
VCHDET	− .01	.08	.04	.17*	1.00				
VCHCLA	− .06	.04	.06	.14	.87**	1.00			
VPROP	.49**	.44**	.44**	.52**	.15*	.13	1.00		
VSIZE	.14	.10	.06	.19*	.00	− .01	.18*	1.00	
VVIVID	.46**	.48**	.63**	.61**	.16*	.17*	.49**	.28**	1.00
VREASE	.32**	.33**	.40**	.33**	.05	− .01	.33**	.13	.37**
VRCDIS	.08	.16*	.16*	.16*	.03	.04	.17*	.09	.18*
VRSIZCH	− .04	.04	.02	− .07	.02	.02	.08	− .09	.02
VRDETCH	− .11	.06	.01	.06	.48**	.45**	− .01	− .05	.08
VRCLACH	− .10	.04	.06	.02	.47**	.45**	− .05	− .00	.10
VRPROP	.30**	.25**	.31**	.27**	− .04	.01	.48**	.29**	.30**
VRVIVID	.25**	.30**	.33**	.19*	− .04	− .03	.30**	.17*	.40**

	VREASE	VRCDIS	VRSIZCH	VRDETCH	VRCLACH	VRPROP
VRCDIS	.44**	1.00				
VRSIZCH	.08	1.00				
VRDETCH	.32**	.09	.22**	1.00		
VRCLACH	.33**	.08	.26**	.90**	1.00	
VRPROP	.56**	.29**	.09	.09	.07	1.00
VRVIVID	.64**	.25**	.26**	.28**	.31**	.60**

* $p < .05$; ** $p < .01$.

Table 7. Factor analysis of ratings for the 3D shape (oblimin rotation)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Vevoke	.84**	−.14	−.01	−.03
Vdet	.86**	−.01	.01	−.19
Vclar	.89**	−.02	.06	−.16
Vmaint	.85**	.12	−.07	.07
Vdetch	.10	.92**	−.20	.11
Vclach	.07	.91**	−.20	.14
Vprop	.56**	.06	.17	.27
Vsize	−.07	.03	.27	.77**
Vvivid	.65**	.14	.16	.18
VRease	.23	.03	.73**	.02
VRsizch	−.04	.02	.39**	−.48**
VRdetch	−.12	.70**	.40**	−.27
VRclach	−.13	.70**	.43**	−.27
VRprop	.12	−.09	.71**	.39**
VRvivid	.12	−.05	.85**	.02
% variance	31.8	19.1	12.0	7.9

** Denotes highest loading variables.

Table 8. Factor correlation matrix for the 3D shape

	Factor 1	Factor 2	Factor 3
Factor 2	.04	1.00	
Factor 3	.25	.12	1.00
Factor 4	.20	−.07	−.03

$^{*}p < .05$; $^{**}p < .01$.

just measuring the same underlying factor or ability. There are two ways to examine this, grouping the correlations of the ratings and regressions of the ratings against performance on the spatial tests.

Factor analysis of the questionnaire ratings

To explore further the intercorrelations of the ratings (Tables 3 and 6), it was decided to factor-analyse the ratings for each shape (Tables 4, 5, 7 and 8).

Factor analysis with an oblique (oblimin) rotation (SPSS, 1988) obtained four factors for the ratings for both shapes (see Tables 4, 5, 7 and 8). The first two factors were the same for both shapes. The ratings making up the first factor were evoke, detail, clarity, maintain, proportion and vividness. These are all surface properties or processes associated with the static image. The second factor obtained loaded all four of the change ratings (detail and clarity change for both the static and rotated shape). Interestingly, maintain ratings did not load on this factor, so clearly there is a distinction that can be made between the rated ease of maintaining an image and the degree of pictorial stability of the image. For both shapes, the third factor loaded ratings for the rotated shape. For the 3D shape, all the ratings for the rotated shape loaded on this factor, whilst only proportion, ease of rotation, proportion rotated and vividness during rotation did so for the 2D shape. The fourth factor loaded the ratings for the relative size at which an image was imagined both static and during rotation.

From the factor analyses, it would seem that four main aspects of imagery are captured by the set of ratings: ease of forming an image, the pictorial stability of an image, the ease of rotating an image and the relative size at which the stimulus is imagined.

These four factors provide some support for the processes or properties identified in Kosslyn's (1980, 1994) model of the imagery system. Ease of forming an image corresponds to the generation processes; ease of rotation to the rotation sub-process operating on the visual buffer. The pictorial stability of an image could correspond to the maintenance processes operating on the visual buffer, and the relative size at which an item is imagined could reflect trade-offs between extent and detail in the visual buffer (i.e. structural properties of the visual buffer or attentional window). The factor structure observed indicates that the ratings of images are influenced in a significant way by the existence and operation of these processes and properties.

Regression of questionnaire ratings against spatial test performance

Examination of the correlations of the ratings with the tests (Table 2) reveals that ratings from all four factors correlate significantly. Considering that many of the ratings significantly correlate across factors, it is still uncertain whether it is variance unique to each factor or common to all factors (i.e. a single 'ability' to imagine) that accounts for significant proportions of the variance in the spatial test scores.

An initial exploration of the data using stepwise regression techniques was conducted to investigate whether performance on the spatial tests is accounted for by one block of variance common to all ratings or by unique variance associated with a particular group of ratings.

The stepwise regressions showed that removal of the variance associated with the rating that accounts for the largest proportion of the variance on the spatial tests does not remove all the variance that predicts performance on the tests from the other ratings.

For the regression against the scores on the CAB-S test, three variables were entered into the equation, proportion during rotation, evoke and detail change, resulting in a multiple R of .44 ($p < .001$). Interestingly, these three ratings come from the first three factors identified for the 2D shape ratings. That these factors (generation, pictorial stability and rotation) roughly correspond to classes of processes identified by the Kosslyn model could be interpreted as evidence to indicate that each group of processes plays a role in determining individual differences in spatial performance.

For the regression against the scores on the Vandenberg test, two variables were entered, ease of rotation and detail, resulting in a multiple R of .43 ($p < .001$). Again, these two ratings come from two of the factors (rotation and generation) identified for the ratings of the 3D shape. Again, a theoretically plausible explanation could be constructed to explain performance on a spatial test in terms of ease of rotation and ease of generating an image to be rotated.

Discussion of factor analysis and regressions

The results from the factor analyses and stepwise regressions suggest that the ratings are measuring at least four aspects of the imagery system that can be related to the classes of processes in Kosslyn's model, three of these factors appear to be related to performance on a spatial task.

Unfortunately, the story these analyses tell is not that simple, because stepwise regression fairly arbitrarily enters the variable with the highest variance. An alternative approach would be to force variables into the regression equation according to a theoretically plausible model of the quantitative role of the imagery factors involved in performing a spatial task.

The problem with a structural model such as that of Kosslyn (1980) is that although it allows a general prediction that each of the major classes of processes will be involved in performance of a spatial task, it does not predict the quantitative relative importance or even criticality of these processes when used on a rotation problem. Whilst the complete failure of, or severely impaired performance of, any of the hypothesized processes or structures would almost certainly affect manipulation of images (cf. Morton & Morris, 1995), this degree of failure probably does not occur with most normal participants. Quantitative differences in performance on a task involving imagery could be dependent on small differences of quality or speed in one critical process, or the interaction of these variations over many processes. At present, there exist no quantitative models for the use of imagery in spatial tasks; existing models simply predict that a process or property is used, not its relative importance or its degree of qualitative contribution.

Dean (1994) explored several forced regressions. These demonstrated that several combinations of ratings from the same factor were equally successful when placed in a regression equation. For example, for a forced regression against scores on the CAB-S test, three variables that load on the first factor (generation) all entered significant proportions of variance into the equation. These three variables represented pictorial resolution (or quality), formation and spatial extent. Counter to the explanation for the original stepwise regression, it could be argued, instead, that these three properties of imagery are crucial for performance on a spatial task. The extent to which these properties reflect the quality of the representation or the operation of underlying processes is unknown.

What the forced regressions showed is that although ratings on the same factor shared a large amount of variance, they also accounted for unique variance that had a predictive ability. However, since placement of these variables into the regression equation prevented any variables from other factors contributing significantly, it is probable that they were capturing variance shared with variables on other factors that then no longer contributed significantly to the regression equation. The most plausible explanation is that the ratings, as expected, were not pure measures of one particular property or process, probably because of their interdependence and the lack of precision inherent in introspective ratings. A tentative conclusion from the results obtained is that for 2D rotation, three separate classes of imagery processes all influence performance on the spatial test, rather than one process or property being critical to performance. For the 3D spatial test, two types of process seem to be critical with performance.

Correlations of questionnaire ratings across items

The correlations between the same ratings for each shape (see Table 9), reveal that all the ratings correlate across the shapes except for the maintain ratings. The significance of all the correlations is $p < .01$ except for that of the evoke ratings where $p < .05$. What is of interest is that this clearly demonstrates that a common ability is being measured by the ratings across the item type. This must be contrasted with the different predictive ability the ratings for each shape show against the spatial tests. Clearly, for some ratings, the specificity of the shape leads to the ratings accounting for variance that is not common to both shapes yet is shared with performance on the relevant spatial test. It adds to the evidence that different stimuli require the involvement or proportionally different involvement of sub-processes of the imagery system, and this is reflected in the surface properties of the image that can be assessed. Furthermore, the differences in the image that can be assessed seem to have some relationship with performance on the relevant spatial test.

Table 9. Correlations between ratings for the 2D and 3D shapes

Variable	Correlation
evoke	.20*
det	.28**
clar	.20**
maint	.15
chdet	.48**
chclar	.50**
prop	.32**
size	.37**
vivid	.28**
Rease	.25**
Rcdis	.43**
Rsizch	.38**
Rdetch	.60**
Rclach	.55**
Rprop	.41**
Rvivid	.30**

* $p < .05$; ** $p < .01$.

Differences in rating magnitude between items

It was hypothesized that the mean ratings for the 3D shape would be lower than the mean ratings for the 2D shape as the additional information contained in the 3D shape would place greater demands on the imagery system. The differences between mean ratings on each of the shapes are shown in Table 10. Significant differences between the mean scores for ratings on the two shapes exist for the following ratings. The 2D shape is imagined in more detail than the 3D shape. Surprisingly, the 3D shape is rated significantly easier to maintain than the 2D shape. This direction in the difference is also true for detail change ratings where again, surprisingly, there is significantly more change for the 2D shape. The mean rating for the 3D shape shows that the shape becomes more detailed as it is maintained, whilst the 2D shape becomes less detailed. For the proportion ratings, participants imagine significantly more of the 3D shape than the 2D shape. The 2D shape is rated significantly easier to imagine rotating than the 3D shape. During rotation, the detail of the 3D shape changes significantly more than that of the 2D shape. As with detail change during rotation, the 3D shape becomes significantly less clear during rotation than the 2D shape. Significantly less of the 3D shape is imagined during rotation than the 2D shape. The 3D shape during rotation is significantly less vivid than the 2D shape during rotation.

Table 10. Differences between mean ratings ($N = 165$)

Rating	Mean 2D shape	Mean 3D shape	*t*	*p*
evoke	6.30	5.99	1.47	.143
detail	6.25	5.79	2.52	.013*
clarity	5.95	5.84	0.69	.490
maintain	5.26	5.76	− 2.29	.023*
chdet	− 0.48	0.28	− 2.02	.045*
chcla	− 0.46	0.13	− 1.59	.114
proportion	6.80	7.36	− 2.20	.029*
size	5.31	5.22	0.82	.416
vivid	5.25	5.59	− 1.92	.056
Rease	5.88	3.59	11.81	<.001**
Rsizech	4.91	4.84	1.78	.424
Rdetch	− 0.09	− 1.32	3.36	.001**
Rclach	0.21	− 0.95	2.97	.003**
Rproportion	6.37	5.87	2.30	.023*
Rvivid	5.22	4.48	3.85	<.001**

*$p < .05$; **$p < .01$.

Rotation of an image is hypothesized to place greater demands on the imagery system, so it was expected that transformation of the shape would result in significantly lower mean ratings for those parameters that were rated for both the static and rotated shape. To examine this, differences between ratings for the stationary and rotated shape were calculated (see Table 11). For the ratings of the 2D shape, only the proportion ratings were significantly less during rotation than whilst static. No significant difference was found between the other pairs of ratings.

With the ratings for the 3D shape, all the ratings were significantly different in the expected direction. Proportion ratings were significantly less whilst rotating than static, and detail change was significantly more negative for the rotated shape than for the static shape. Clarity change was significantly more negative for the rotated shape than the static shape. Finally, vividness was significantly less for the rotated shape than the static shape. Clearly, then, for the 3D shape, rotation leads to less of the shape being imagined less vividly. Both the detail and the clarity of the shape change negatively during rotation, as opposed to a slight increase for both with the static shape. For the 2D shape, only proportion is affected by the action of rotation.

Analysis of sex differences

Details of the analysis of sex differences are given in Table 12. Significant differences were found between males and females on the CAB-S test and the Vandenberg test but not on the VVIQ. Significant differences in ratings for the CAB-S shape were found for ratings of proportion, vividness, ease of rotation, proportion during rotation and vividness during rotation. For the ratings on the Vandenberg shape, significant differences in ratings were found for evoke, detail, maintain, proportion, ease of rotation, size change during rotation, proportion during rotation and vividness during rotation.

Sex differences in questionnaire ratings of shapes are associated with ratings loading on factors of forming an image and rotating it, including vividness. The differences are greater for the ratings of the 3D shape. That the item imagined and operation performed

Table 11. Differences between mean ratings within the shape (N = 165)

	Mean static	Mean rotated	t	p
2D shape				
Proportion	6.80	6.37	2.22	.028*
Detail change	− 0.48	− 0.09	− 1.06	.291
Clarity change	− 0.46	0.21	− 1.71	.089
Vividness	5.25	5.22	0.23	.819
3D shape				
Proportion	7.36	5.87	6.65	.001**
Detail change	0.29	− 1.32	4.04	.001**
Clarity change	0.13	− 0.95	2.61	.010**
Vividness	5.59	4.49	6.24	.001*

*$p < .05$; **$p < .01$.

are responsible for the sex differences found is indicated by the lack of sex differences found on the VVIQ and differences between vividness during rotation ratings found on the new questionnaire.

The first question to ask is whether factors that cause variation between the two groups are the same as those that cause variation within the entire sample. The second question is whether there are any qualitative differences in the patterns of correlations between the groups.

Because significant differences exist between male and female scores, sex is obviously a significant predictor of performance. However, if sex is partialled out of the variance, the partial correlations of ratings with spatial tests still show that many of the ratings still correlate significantly with performance on the spatial tests (see Table 13). Therefore, the variation in performance on spatial tests related to self-reports of imagery is at least partially the result of different factors that cause the variation in performance between groups.

The variation in performance between males and females is far greater on the 3D mental rotation test than on the 2D test. A regression analysis forcing sex as a variable into the equation after the predictive ratings showed that for a regression against the CAB-S test, sex did not account for a significant amount of the variance after entering proportion during rotation, evoke and detail change ratings (partial $r = .14$). This is consistent with the findings of Mumaw *et al.* (1984), who found that that in a regression analysis, variance attributable to sex was negligible relative to other systematic individual difference variance within the entire sample. It should be noted that in this study, the other variables measured were rotation slope and intercept, which together are far more comprehensive measures of variance than test scores on their own.

For the regression against the Vandenberg test, sex was a significant predictor (partial $r = .24$, $p < .01$) after entering the ratings ease of rotation and detail change. That sex is still a significant predictor after entering the variation due to the imagery ratings indicates that there are multiple factors leading to the difference in performance between males and females on 3D spatial tests. Some of these factors are accounted for by variation in self-reported ratings of imagery, but evidently not all. It is probable that factors that underlie differences in performance on spatial tests are not fully

Table 12. Sex differences in test scores and ratings (*df* = 1, 163 for each test)

Variable	Female *M*	Male *M*	*F*	*p*
CABS	44.24	48.57	7.53	.01**
VAND	10.15	14.52	24.92	.01**
VVIQ	75.97	75.70	0.51	.48
Cevoke	6.15	6.74	3.05	.08
Cdet	6.19	6.48	0.74	.40
Cclar	5.85	6.30	1.60	.20
Cmaint	5.10	5.76	3.68	.05
Cchdet	− 0.49	− 0.26	0.07	.78
Cchclar	− 0.12	− 1.08	1.17	.28
Cprop	6.46	7.64	7.90	.01**
Csize	5.32	5.36	0.03	.83
Cvivid	5.09	5.74	4.96	.03*
CRease	5.65	6.38	4.29	.04*
CRsizch	4.87	4.98	0.51	.48
CRdetch	− 0.47	0.72	2.28	.13
CRclach	− 0.09	0.82	1.35	.25
CRprop	5.93	7.36	11.17	.01**
CRvivid	5.00	5.72	4.93	.03*
Vevoke	5.54	6.98	16.75	.01**
Vdet	5.50	6.34	6.69	.01**
Vclar	5.65	6.20	2.58	.11
Vmaint	5.36	6.64	13.51	.01**
Vdetch	0.24	0.30	0.01	.90
Vclach	0.04	0.30	0.10	.74
Vprop	6.84	8.44	12.31	.01**
Vsize	5.19	5.38	0.83	.37
Vvivid	5.41	5.96	3.12	.08
VRease	3.36	4.16	5.25	.02*
VRsizch	4.68	5.14	5.74	.02*
VRdetch	− 1.39	− 1.08	0.11	.73
VRclach	− 1.11	− 0.52	0.40	.53
VRprop	5.42	7.08	12.82	.01**
VRvivid	4.21	5.22	7.12	.01**

$*p < .05$; $**p < .01$.

captured by the ratings (unlike measurements of the slope and intercept on an experimental rotation task used by Mumaw *et al.*, 1984), thus leaving variance to be accounted for by factors not measured in this experiment that are partly responsible for sex differences.

Space does not permit full presentation of the tables of correlations for males and females separately to examine whether there are any qualitative differences in patterns of correlations and thus imagery use on spatial tests between the two groups. Interested readers are referred to Dean (1994), where a full presentation of these tables is available. In summary, in so far as the size of the sample for males will allow, there appear to be no qualitative differences in the use of imagery. The same types of imagery ratings predict performance on spatial tests for both the female and

Table 13. Correlations of ratings with VVIQ and spatial tests with variance attributable to sex differences partialled out

	VVIQ	CAB-S	Vandenberg
CABS	–	– 0.03	.03
VAND	– 0.03	–	.32**
VVIQ	.03	.32**	–
Cevoke	– 0.11	.31**	.05
Cdet	– 0.14	.25**	.13
Cclar	– 0.16	.15	.12
Cmaint	– 0.11	.11	.18*
Cchdet	– 0.08	– 0.12	.05
Cchclar	– 0.09	– 0.11	.02
Cprop	– 0.10	.27**	.09
Csize	– 0.05	.14	.06
Cvivid	– 0.10	.16*	.17*
CRease	– 0.18*	.23**	.08
CRsizch	.07	– 0.03	.02
CRdetch	– 0.24**	– 0.07	.05
CRclach	– 0.20*	– 0.05	.01
CRprop	– 0.20*	.32**	.11
CRvivid	– 0.21*	.19*	.13
Vevoke	– 0.06	.10	.14
Vdet	– 0.09	.18*	.27**
Vclar	– 0.14	.03	.19*
Vmaint	– 0.25**	.11	.20*
Vdetch	– 0.18*	– 0.12	– 0.01
Vclach	– 0.19*	– 0.12	– 0.04
Vprop	– 0.21*	.07	.06
Vsize	– 0.04	.05	.17*
Vvivid	– 0.26**	– 0.02	.12
VRease	– 0.31**	.00	.34**
VRsizch	.07	.02	.12
VRdetch	– 0.14	– 0.16*	.09
VRclach	– 0.12	– 0.21**	.12
VRprop	– 0.23*	.05	.20*
VRvivid	– 0.20*	– 0.04	.28**

*$p < .05$; **$p < .01$.

male groups, and the factor structure of the questionnaires is similar to that obtained for all participants.

Discussion

The starting point of this study was to apply theoretical findings about the imagery system to imagery questionnaires. It was hypothesized that the failure to find any relationship between self-reports of imagery and performance on spatial tests thought to require imagery was because existing self-reports only measured vividness. It was thought that an overall measure of imagery ability such as vividness would fail to capture or disguise the contribution of more precise properties identified by referring to Kosslyn's (1980, 1994) model. A second observation was of the startling difference in item type between imagery questionnaires and spatial tests. Existing questionnaires require imagining everyday items recalled or constructed from long-term memory, whereas spatial tests use line drawings of geometrical shapes that are perceived then imagined. Kosslyn's (1980, 1994) model does not predict that differences in item nature (especially its source) will affect the surface properties or conscious experience of an image. However, it seemed logical that if imagery is conceptualized as a collection of processes, these might be differentially involved according to the demands of the content of the stimulus (cf. Cooper, 1995).

The correlations of ratings with test scores obtained indicate that the different ratings employed measured aspects of imagery that play a role in performance on spatial tests and were not previously fully captured by vividness ratings. The pattern of correlations between the ratings and of the ratings with the spatial tests indicates that it is possible to make introspective ratings of imagery that capture at least three of the general classes of processes identified in Kosslyn's (1980) model: generation, rotation and maintenance. Not only can the ratings capture these processes, but in doing so, they can predict performance on spatial tests. The regression analysis shows that variables representing these factors capture significantly more variance in spatial test scores than do vividness ratings alone or any other single variable. That more than one variable predicts performance on the spatial tests is an indication that despite the strong intercorrelation of the ratings, they do capture more than one underlying variable. Furthermore, it is likely that quantitative differences in performance on spatial tests are affected by variations in quality or performance of all three factors rather than being a result of variations in a single critical variable. Whilst the total percentage of variance in spatial test scores accounted for is fairly small, this is probably to be expected given the inherently 'noisy' nature of introspective data.

Contrary to expectations, the results obtained seem to indicate that the major reason a relationship was not previously found between imagery questionnaires and spatial tests was because of the item type on the questionnaires. Vividness ratings for shapes, although not the best predictors of performance on the spatial tests, do correlate significantly. Compared with this, vividness ratings of real world objects (as represented by scores on the VVIQ) that are recalled or constructed from long-term memory fail to correlate significantly.

It must be noted that two major dimensions of the items have been changed: first, the item type (everyday scenes vs. line drawings of shapes), and second, the source of the image (recalled or constructed from long-term memory vs. perceived then imagined). The data from this study are not sufficient to enable us to conclude which of these factors is critical. Ratings for the 2D shape were the best predictors for performance on the CAB-S test, and ratings for the 3D shape were the best predictors for performance on the Vandenberg test. The small effects of item specificity displayed by the ratings against the 2D and 3D tests would seem to indicate that the type of item is more important than the source.

The present study yields results that seem to explain why a large number of previous studies found no relation between introspective reports of imagery and performance on spatial tests thought to require imagery. Examination of other individual differences studies (e.g. Kosslyn *et al.*, 1984; Kosslyn, Van Kleek & Kirby, 1990; Poltrock & Brown, 1984) where a relationship between self-reports of imagery and a variety of imagery measures and spatial tests was not observed reveals that items used on self-report measures were very different to those used on the other measures. It is clear from the results above that the effect of item type is a greater factor in obtaining predictive ratings than the nature of the ratings. Combinations of ratings are, however, better predictors of performance than vividness alone, as the analysis revealed that rating properties other than vividness accounted for more of the variance in the spatial tests than vividness alone. Thus, more specific ratings based on theory improved the correlation of ratings with spatial tests but had less of an effect than changing the items. Additionally, the new ratings seem to capture specific properties of the imagery system, providing evidence that casts doubt on the utility of measuring imagery using vividness scores as if imagery were a single ability.

The question that requires an answer is how we can relate our conscious experience of having an image to the structures identified in models of visual processing. Applying the results of this study to the models suggests several possibilities for revision or clarification. The first is that our experience of having an image is not just our experience of the processes of the visual buffer in operation but also our experience of other visual and spatial processes as well. This is a theoretical position at odds with Kosslyn's (1980, 1994) model but implicit within the working-memory model (Baddeley, 1990). Research has suggested that not only can the visuospatial sketchpad be decomposed into separable visual and spatial components, but it may also have separate rehearsal systems for each component (Logie, 1995; Morton & Morris, 1995). With regard to the issue of what processes our experience of imagining reflects, it is worth noting that because our phenomenological experience of imagining is unitary, it requires fairly subtle experimental manipulations to show separable subsystems by selective interference effects.

The second issue is, considering the predictive difference of ratings for 2D and 3D static shapes with performance on the tests, that even for perceived stimuli, visual processing takes place and is reflected in the reinforcement or maintenance of the image in the visual buffer. This corresponds to the suggestion by Logie (1995) that perceived stimuli are held in short-term store via the operation of long-term memory structures. Both these would explain the effect of item type on the ratings. However, this still leaves models needing to explain how the surface properties of the visual buffer are linked to, and affected by, the underlying processes. Furthermore, the role the information content or properties of the representation has in the solution of a spatial problem needs to be considered. Whilst a simple explanation of the extent that accuracy or detail of the information affecting manipulation might suffice, the precise way in which it does so has not yet been modelled. The key elements of a representation that lead to individual differences in performance on a spatial task need to be elucidated. To do this also requires a better understanding of what spatial tests and their equivalent experimental tasks actually measure. At present, precise quantitative models of the role of imagery processes in the performance of spatial tasks do not exist.

In conclusion, the data suggest that item origin and content have a measurable effect on both surface properties and underlying processes of the imagery system. Furthermore, the data indicate that it is possible to get participants to introspect successfully on several different properties of their images, rather than just use an overall measure such as vividness, and that these ratings capture more of the imagery process than vividness alone. These ratings also predict performance on spatial tests. Taken together, these conclusions suggest that investigations into the role of imagery in cognition using self-reports are possible but must pay careful attention to item content on the different measures and would more fruitfully employ a range of judgments about image properties.

References

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Hove, UK: Erlbaum.

Baddeley, A. D., & Andrade, J. (1998). Working memory and consciousness: An empirical approach. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds), *Theories of memory, volume II*. Hove, UK: Psychology Press.

Bat-Zion, N. (1986). The relation of imagery to other mental abilities. In D. G. Russell, D. F. Marks, & J. T. E. Richardson (Eds), *Imagery 2* (pp. 84–88). Dunedin, NZ: Human Performance Associates.

Betts, G. H. (1909). *The distribution and functions of mental imagery (Contributions to Education, No. 26)*. New York: Columbia University, Teachers College.

Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? *American Psychologist*, *40*, 786–799.

Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, volume 3*. Hillsdale, NJ: Erlbaum.

Cochran, K. F., & Wheatley, G. H. (1989). Ability and sex-related differences in cognitive strategies on spatial tasks. *The Journal of General Psychology*, *116*, 43–55.

Cocude, M., & Denis, M. (1988). Measuring the temporal characteristics of visual images. *Journal of Mental Imagery*, *12*, 89–101.

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience*, *111*, 845–849.

Cooper, L. A. (1991). Dissociable aspects of the mental representation of visual objects. In R. H. Logie & M. Denis (Eds), *Mental images in human cognition*. Oxford: Elsevier Science.

Cooper, L. A. (1995). Varieties of visual representation: How are we to analyze the concept of mental image? *Neuropsychologia*, *33*, 1575–1582.

Danaher, B. G., & Thoresen, C. E. (1972). Imagery assessment by self-report and behavioural measures. *Behaviour Research and Therapy*, *10*, 131–38.

Dean, G. M. (1994). *The function of imagery in cognition*. Ph.D. thesis, Lancaster University, UK.

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

DiVesta, F. J., Ingersoll, G., & Sunshine, P. (1971). A factor analysis of imagery tests. *Journal of Verbal Learning and Verbal Behaviour*, *10*, 471–479.

Durndell, A. J., & Wetherick, N. E. (1976a). The relation of reported imagery to cognitive performance. *British Journal of Psychology*, *67*, 501–506.

Durndell, A. J., & Wetherick, N. E. (1976b). Reported imagery and two spatial tests. *Perceptual and Motor Skills*, *43*, 1050.

Ernest, C. H. (1977). Imagery ability and cognition: A critical review. *Journal of Mental Imagery*, *2*, 181–216.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition*, *18*, 245–272.

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

Hakstian, A. R., & Cattell, R. B. (1976). *The comprehensive ability battery*. Champaign, IL: IPAT.

Hatakeyama, T. (1981). Individual differences in imagery ability and mental rotation. *Tohoku Psychologica Folia*, *40*, 6–23.

Hatakeyama, T. (1984). Individual differences in imagery ability and mental size comparison. *Tohoku Psychologica Folia*, *43*, 112–126.

Hinton, G. (1979). Some demonstrations of the effects of structural descriptions in mental imagery. *Cognitive Science*, *3*, 231–250.

Horan, P. F., & Rosser, R. A. (1984). A multivariable analysis of spatial abilities by sex. *Developmental Review*, *4*, 387–411.

Kimura, D. (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current Opinion in Neurobiology*, *6*, 259–263.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception and Psychophysics*, *14*, 90–94.

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

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

Kosslyn, S. M., Brunn, J. L., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, *18*, 195–243.

Kosslyn, S. M., Van Kleek, M. H., & Kirby, K. N. (1990). A neurologically plausible model of individual differences in visual mental imagery. In P. J. Hampson, D. F. Marks, & J. T. E. Richardson (Eds), *Imagery current developments.* London: Routledge.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds), *Human abilities.* Hillsdale, NJ: Erlbaum.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characteristics of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.

Logie, R. H. (1995). *Visuo-spatial working memory.* Hove, UK: Erlbaum.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds), *Human abilities.* Hillsdale, NJ: Erlbaum.

Lohman, D. F., Pellegrino, J. W., Alderton, D. L., & Regian, J. W. (1987). Dimensions and components of individual differences in spatial abilities. In S. H. Irvine & S. E. Newstead (Eds), *Intelligence and cognition: Contemporary frames of reference.* Dordrecht, The Netherlands: Martinus Nijhoff.

Lorenz, C., & Neisser, U. (1985). Factors of imagery and event recall. *Memory and Cognition*, *13*, 494–500.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*, 889–918.

McKelvie, S. J. (1995). 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*, *3 & 4*, 1–106.

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

Marks, D. F. (1989). Bibliography of research utilising the Vividness of Visual Imagery Questionnaire. *Perceptual and Motor Skills*, *69*, 707–718.

Morris, P. E., & Hampson, P. J. (1983). *Imagery and consciousness.* London: Academic Press.

Morris, P. E., & Reid, R. L. (1973). Recognition and recall: Latency and recurrence of images. *British Journal of Psychology*, *64*, 161–167.

Morton, N., & Morris, R. G. (1995). Image transformation dissociated from visuospatial working memory. *Cognitive Neuropsychology*, *12*, 767–791.

Mumaw, R. J., Pellegrino, J. W., Kail, R. J., & Carter, P. (1984). Different slopes for different folks: Process analysis of spatial aptitude. *Memory & Cognition*, *12*, 515–521.

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

Paivio, A. (1989). Basic puzzles in imagery research. In M. Denis, J. Engelkamp, & J. T. E. Richardson (Eds), *Cognitive and neuropsychological approaches to mental imagery* (pp. 3–16). Dordrecht, The Netherlands: Martinus Nijhoff.

Pellegrino, J. W., & Kail, R. Jr. (1982). Process analyses of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, volume 1.* London: Erlbaum.

Poltrock, S. E., & Agnoli, F. (1986). Are spatial visualization ability and visual imagery ability equivalent? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, volume 3.* Hillsdale, NJ: Erlbaum.

Poltrock, S. E., & Brown, P. (1984). Individual differences in visual imagery and spatial ability. *Intelligence*, *8*, 93–138.
Rehm, L. P. (1973). Relationships among measures of visual imagery. *Behaviour Research and Therapy*, *11*, 265–270.
Richardson, A. (1977). The meaning and measurement of memory imagery. *British Journal of Psychology*, *68*, 29–43.
Richardson, A. (1994). *Individual differences in imaging: Their measurement, origins and consequences*. New York: Baywood.
Richardson, J. T. E. (1991). Gender differences in imagery, cognition, and memory. In R. H. Logie & M. Denis (Eds), *Mental images in human cognition.* Oxford: Elsevier Science.
Richardson, J. T. E. (1999). *Imagery.* Hove, UK: Psychology Press.
Sheehan, P. W., & Neisser, U. (1969). Some variables affecting the vividness of imagery in recall. *British Journal of Psychology*, *60*, 71–80.
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.
SPSS (1988). *SPSS-X user's guide* (3rd ed.). Chicago: SPSS.
Stumpf, H. (1993). Performance factors and gender related differences in spatial ability: Another assessment. *Memory & Cognition*, *21*, 828–836.
Tapley, S. M., & Bryden, M. P. (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. *Canadian Journal of Psychology*, *31*, 122–130.
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualisation. *Perceptual and Motor Skills*, *47*, 599–604.
Wilson, J. R., & Vandenberg, S. G. (1978). Sex differences in cognition: Evidence from the Hawaii family study. In T. E. McGill, D. A. Dewsbury, & B. D. Sachs (Eds), *Sex and behavior: Stages and prospectus*. New York: Plenum.

Received 12 March 2001; revised version received 31 May 2002

Appendix

Shapes Questionnaire

The individual differences that have been found in peoples' ability to form mental images of items have long been of interest to psychologists. Some people have very lifelike 'mental pictures', whilst 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, and 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 questions underneath. (It will probably help if you imagine the shape each time you answer a question, rather than trying to answer all the questions from imagining the shape only once.) If you can imagine the shape more easily with your eyes open, do so, but DO NOT LOOK AT THE SHAPE WHILST IMAGINING.

The first question will always ask you how easily you can evoke an image of the shape in question and to rate this 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 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 very important, as you may find that your ratings of each property or quality are different. Concentrate on imagining 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 imagining to improve a particular property. Rather, we wish you to rate the level of that property that occurred 'automatically', when you imagined 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 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 losing concentration, take a rest!

REMEMBER, your accurate and honest answers are vital for the validity of this study.

You may begin when ready; if you have any doubts concerning 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 do not feel that the ratings capture something essential about your images. However, PLEASE TRY TO MAKE A RATING IN ALL CASES, then write down what is wrong. 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 your image?
NOT CLEAR AT ALL 1 2 3 4 5 6 7 8 9 VERY SHARP AND 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
More or less detailed?. . . . .

(b) The clarity
NOT AT ALL 1 2 3 4 5 6 7 8 9 A GREAT DEAL
More or less clear? . . . .

(6) How much of the shape can you form an image of, at any one time?
1/10th 1 2 3 4 5 6 7 8 9 10 ALL
PLEASE MARK ACCURATELY ON THE SHAPE ABOVE THE PARTS YOU IMAGINE (circle the shape if you can imagine all).

(7) How large is your image of the shape?
VERY SMALL 1 2 3 4 5 6 7 8 9 VERY LARGE
(AS IF SEEING IT FROM A GREAT DISTANCE) (AS IF SEEING IT FROM 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 imagine continuous or in discrete stages?

CONTINUOUS DISCRETE STAGES

(11) Does the size of the imagined shape (or the apparent distance at which it is imagined) change when you rotate it? If so, in what way does your image of the shape change?

(AS IF) MUCH SMALLER 1 2 3 4 5 6 7 8 9 (AS IF) MUCH LARGER

(12) Does rotating the shape affect (a) the detail and (b) the clarity of your image in any of the ways?

(a) The detail

NOT AT ALL 1 2 3 4 5 6 7 8 9 A GREAT DEAL

More or less detailed?. . . . .

(b) The clarity

NOT AT ALL 1 2 3 4 5 6 7 8 9 A GREAT DEAL

More or less clear? . . . .

(13) How much of the shape do you form an image of when rotating it.

1/10th 1 2 3 4 5 6 7 8 9 10 ALL

(14) If you only imagine parts of the shape whilst rotating it, can you mark these on the shape below?

(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, where 1 = VERY LITTLE?

(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