Electronic Supplementary Information for 'Seeing' chemistry: investigating the contribution of mental imagery strength on students' thinking in relation to visuospatial problem solving in chemistry

Lauren Baade,[a] Effie Kartsonaki,[a] Hassan Khosravi[b] and Gwendolyn A. Lawrie*[a]

[a] *School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane City, QLD 4072, Australia. E-mail: g.lawrie@uq.edu.au*
[b] *Institute for Teaching and Learning Innovation, The University of Queensland, Brisbane City, QLD 4072, Australia*

S1: The Vividness of Object and Spatial Imagery (VOSI) questionnaire (with minor adjustments), used during the quantitative study

Try to imagine each of the items below to the best of your ability and then evaluate the vividness of your subjective mental imagery (pictures in your mind). For each of the items, please rate, on a scale from 1 to 5, the vividness of the evoked mental images. The scale is as follows:

5 = 'Perfectly clear and vivid as normal vision';
4 = 'Clear and reasonably vivid';
3 = 'Moderately clear and moderately vivid';
2 = 'Vague and dim';
1 = 'No image at all, you only "know" that you are thinking of the object'

Please rate the vividness of your subject mental imagery:

- Appearance of a candle fire
- Colour mixing in a watercolour painting of a floral bouquet
- Appearance of colours of the sun reflecting on the water
- Cutting-out and folding paper to create a 3D cube
- Fine details and shape of a jellyfish
- Rotation of 3D Tetris piece (3D shape) in order to fit a particular slot
- Shape and colour of an autumn leaf
- Trajectory of an object moved by a force (e.g., in "angry birds" or pool)
- 3D structure of a toilet flushing system
- Shape of cloud in the sky
- Mechanism of a mechanical wall clock
- Fine details of zebra's skin
- Schematic outline of a tractor from the 3 sides
- Construction plan (three-dimensional schema) of a roller coaster
- Location of your house on a map of your city
- Pictorial details of the best friend's face
- Plan of a multi-level road junction
- Colour pattern on a butterfly wing
- Mechanism of a door handle
- Splashes of colours in fireworks
- Technical instruction for assembling a kitchen appliance (e.g., blender, food processor)
- Appearance of colours in a bubble
- Motion of the planets on a model of the solar system
- Patterns on a peacock tail
- Shapes and colours of a bonfire
- Schema (plan) of a computer connection to a printer
- Texture of your favourite clothes
- Finding the way-out path in a paper maze

S2: The Plymouth Sensory Imagery (Psi-Q) questionnaire (with minor adjustments), used during the quantitative study

Please try to form the images described below and rate each mental image on the following scale:

0 (no imagery at all) to 10 (imagery as clear and vivid as real life)

Imagine the appearance of:
- A bonfire
- A sunset
- A cat climbing a tree
- A friend you know well
- The front door of your home

Imagine the sound of:
- The sound of a car horn
- Hands clapping in applause
- An ambulance siren
- The sound of children playing
- The bark of a dog

Imagine the smell of:
- Newly cut grass
- Burning wood
- A rose
- Fresh paint
- Fresh bread

Imagine the taste of:
- Black pepper
- Lemon
- Mustard
- Toothpaste
- Sea water

Imagine touching:
- Fur
- Warm sand
- A soft towel
- Icy water
- The point of a pin

Imagine the bodily sensation of:
- Relaxing in a warm shower
- Walking briskly in the cold
- Jumping into a swimming pool
- Having a sore throat
- Building with blocks (Lego)

Imagine feeling:
- Excited
- Relieved
- Scared
- Angry
- Frustrated

S3: Participants' scores from the VOSI and Psi-Q, collected during the quantitative study

	VOSI (out of 5)		Psi-Q (out of 10)						
Participant*	**Visual-Object**	**Visual-Spatial**	**Vision**	**Sound**	**Smell**	**Taste**	**Touch**	**Bodily Sensation**	**Emotional Feeling**
1	3.4	2.9	8.4	9.2	7	6.8	8.6	6.8	4.6
2	4.4	1.9	8.2	4.8	7.4	8.4	8.8	8.2	7
3	3.9	2.8	7.6	6	6.4	9.6	10	8.4	9.6
4	2.6	2.1	5.2	6.8	6.2	7.4	7.2	8.2	4.8
Pat	4.1	2.4	8	5.4	4.8	7.2	6.2	5.8	6.4
Taylor	1.2	1.3	1.8	5.2	0.4	6.4	6.6	7.4	4.8
7	4.1	2.9	7.2	6.4	6	7	6.6	7	7.2
8	4.7	3.9	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bailey	4.9	3.6	9.6	10	8.6	6.8	10	8.4	7
Jordan	3.9	3.8	8.4	8	4.8	8.8	7.6	8.6	9.6
11	3.2	2.9	9	6.6	7.2	6	7	8.2	7.8
12	3.9	2.9	8	4.2	3.4	7.8	7.4	6.4	6
Jamie	4.8	3.7	8	9.4	3.8	1.6	3.6	6	9.6
Alex	3.8	2.6	8	6.6	6.6	7.4	7.8	6.4	5.8
Sam	4.4	3.1	10	5.6	9	8.8	9.8	9.2	7
16	4.9	4.7	9.8	9.6	9.8	9.6	10	10	9
17	1.4	1.1	1.6	0.4	0	0	0.8	0	0.2
18	3.6	2.9	7.4	3.4	1.2	8.2	7.8	7	8.4

* Named participants partook in interviews during the qualitative study.

S4: Summary of model '1/time versus format*representation' with random effect of participant, for SUVI-LO and SUVI-FG tasks

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time) (correct selections only)
- **Fixed effects:** Format, representation, format*representation
- **Random effect:** Participant

Assumptions:

- **Normality of residuals:**
  Deviation at lower tail (see Fig. S4.1), Shapiro-Wilk test fails ($p < .0001$).
- **Normality of random effects:**
  Normal distribution confirmed (see Fig. S4.2), Shapiro-Wilk test passes ($p = .99$).
- **Homoscedasticity:**
  Minor deviations observed at positive extremes (see Fig. S4.3).

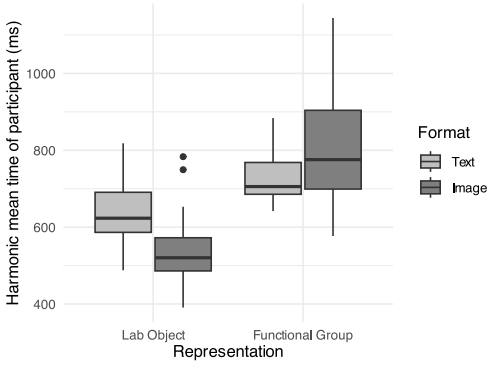


Fig. S4.0. Boxplot of the participants' harmonic mean times across the four representation / format combinations.

Results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	1.27×10^{-3}	$(1.17 \times 10^{-3}, 1.36 \times 10^{-3})$	4.71×10^{-5}	t(25.5) = 26.9	< .0001
Format (text)	1.20×10^{-4}	$(5.24 \times 10^{-5}, 1.85 \times 10^{-4})$	3.38×10^{-5}	t(906) = 3.51	4.73×10^{-4}
Representation (lab object)	6.30×10^{-4}	$(5.59 \times 10^{-4}, 6.99 \times 10^{-4})$	3.59×10^{-5}	t(906) = 17.5	< .0001
Format*representation (text, lab object)	-4.30×10^{-4}	$(-5.27 \times 10^{-4}, 3.27 \times 10^{-4})$	5.10×10^{-5}	t(906) = -8.37	< .0001

- **Model fit:**

Metric	Value
Marginal R^2	0.24
Conditional R^2	0.37
Log-likelihood	5915.74 (df=6)
AIC	-11819.47
BIC	-11790.48

Summary:

This model mostly meets statistical assumptions in that the random effects are normally distributed however the residuals show deviations from normality. Homoscedasticity is generally supported. The significant influence of the fixed effects and their interaction is evident. The fit indices suggest moderate explanatory power.

Figures:

Fig. S4.1 (Q-Q plot of residuals)

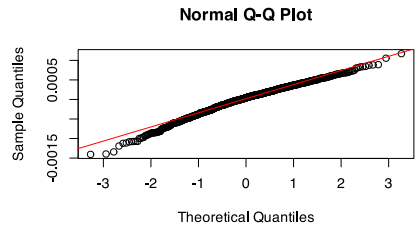


Fig. S4.2 (Q-Q plot of random effects)

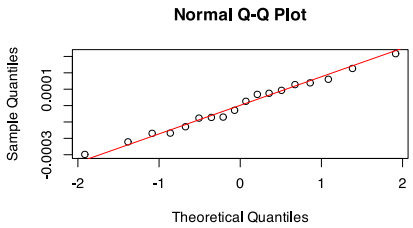
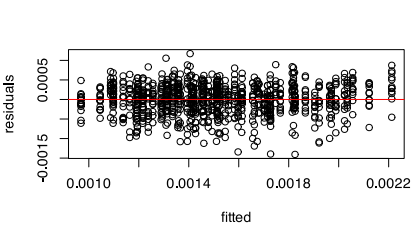


Fig. S4.3 (Plot of fitted against residuals)



S5: Pairwise comparisons showing that participants took statistically significantly different times to select the various representation/format combinations during the two SUVI tasks

Contrast	Estimate	SE	t-value	p-value
FG (images) – FG (text)	-1.19×10^{-4}	3.38×10^{-5}	t(906) = -3.51	.0027
FG (images) – LO (images)	-6.29×10^{-4}	3.58×10^{-5}	t(906) = -17.54	< .0001
FG (images) – LO (text)	-3.21×10^{-4}	3.61×10^{-5}	t(906) = -8.87	< .0001
FG (text) – LO (images)	-5.10×10^{-4}	3.59×10^{-5}	t(906) = -14.20	< .0001
FG (text) – LO (text)	-2.02×10^{-4}	3.62×10^{-5}	t(906) = -5.58	< .0001
LO (images) – LO (text)	3.08×10^{-4}	3.81×10^{-5}	t(906) = 8.08	< .0001

S6: Summary of estimated marginal means of time taken to correctly select each representation/format combination for the two SUVI tasks

Representation	Format	Estimated marginal mean (ms)	95% CI
Lab object	Text	630	(592, 671)
Lab object	Image	528	(503, 556)
Functional group	Text	722	(676, 775)
Functional group	Image	790	(735, 855)

S7: Summary of model '1/time versus VOSI Visual-Spatial score' with random effect of participant, for SUVI-FG (images only)

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time) (functional groups images, correct selections only)
- **Fixed effects:** VOSI Visual-Spatial score
- **Random effect:** Participant

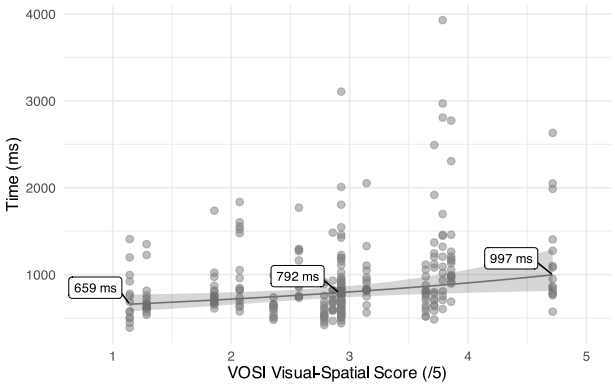


Fig. S7.0. Graph of the participants' times for SUVI-FG (images) against their VOSI Visual-Spatial scores, and the model.

Assumptions:

- **Normality of residuals:**
  Normal distribution confirmed (see Fig. S7.1), Shapiro-Wilk test passes (p = .76).
- **Normality of random effects:**
  Small deviations at extremes (see Fig. S7.2), Shapiro-Wilk test passes (p = .70).
- **Homoscedasticity:**
  Supports linearity and homoscedasticity (see Fig. S7.3).

Results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	1.68×10^{-3}	$(1.35 \times 10^{-3}, 2.02 \times 10^{-3})$	1.70×10^{-4}	t(16) = 9.85	< .0001
VOSI Visual-Spatial score	-1.44×10^{-4}	$(-2.55 \times 10^{-4}, -3.32 \times 10^{-5})$	5.69×10^{-5}	t(16) = -2.53	.022

- **Model fit:**

Metric	Value
Marginal R^2	0.08
Conditional R^2	0.27
Log-likelihood	1655.59 (df = 4)
AIC	-3303.19
BIC	-3288.93

Summary:

This model demonstrates a modest fit and meets statistical assumptions. The random effects and residuals are normally distributed and homoscedasticity is supported. The effect of VOSI Visual-Spatial score is statistically significant, however the marginal R^2 of 0.08 suggests weak explanatory power. The conditional R^2 (0.27) reflects an improvement in model fit due to the inclusion of random effects.

Figures:

Fig. S7.1 (Q-Q plot of residuals)

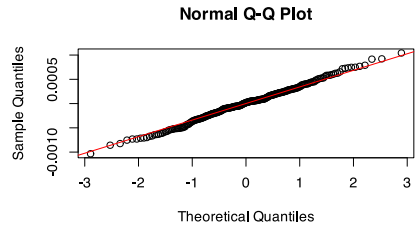


Fig S7.2 (Q-Q plot of random effects)

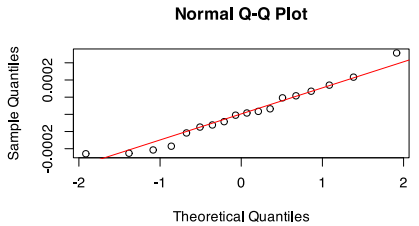
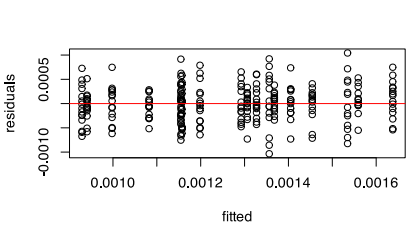


Fig. S7.3 (Plot of fitted against residuals)



S8: Summary of model 'Correctness versus VOSI Visual-Spatial score' with random effect of participant, for SUVI-FG (images only)

Overview:

- **Model type:** Generalised linear mixed (binomial distribution)
- **Dependent variable:** Correctness (binary outcome) (functional groups images)
- **Fixed effects:** VOSI Visual-Spatial score
- **Random effect:** Participant

Assumptions:

- **Linearity in logit:**
  Plot of the log of predicted probabilities versus VOSI Visual-Spatial score aligns reasonably with the fitted line (see Fig. S8).
- **Overdispersion:**
  Deviance residuals ratio of 0.25 suggests no evidence of overdispersion.

Results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	z-value	p-value
Intercept	0.97	(-2.13, 3.77)	1.19	0.82	.41
VOSI Visual-Spatial score	1.02	(0.11, 2.53)	0.51	2.01	.045

- **Model fit:**

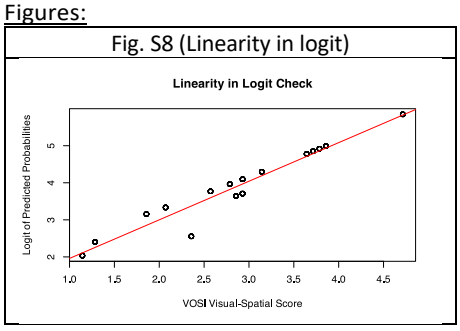
Metric	Value
Marginal R^2 (theoretical)	0.18
Conditional R^2 (theoretical)	0.28
Log-likelihood	-36.33 (df = 3)
AIC	78.66
BIC	89.45
AUC/ROC	0.83

Summary:

The model's statistical assumptions are well-met. The fixed effect of VOSI Visual-Spatial score on correctness is statistically significant, however its explanatory power is modest (marginal R^2 = 0.18). The AUC/ROC value of 0.83 indicates good model performance.

Figures:

Fig. S8 (Linearity in logit)



S9: Summary of model 'Correctness versus VOSI Visual-Spatial score' with random effect of participant, for SUVI-FG (text only)

Overview:

- **Model type:** Generalised linear mixed (binomial distribution)
- **Dependent variable:** Correctness (binary outcome) (functional groups text)
- **Fixed effects:** VOSI Visual-Spatial score
- **Random effect:** Participant

Assumptions checked:

- **Linearity in logit:**
  Plot of the log of predicted probabilities versus VOSI Visual-Spatial score aligns reasonably with the fitted line (see Fig. S9).
- **Overdispersion:**
  Deviance residuals ratio of 0.33 suggests no evidence of overdispersion.

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	z-value	p-value
Intercept	5.99	(3.43, 10.78)	1.60	3.73	.00019
VOSI Visual-Spatial score	-0.89	(-2.16, -0.08)	0.44	-2.03	.042

- **Model fit:**

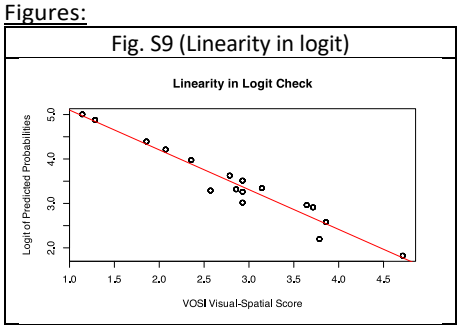
Metric	Value
Marginal R^2 (theoretical)	0.15
Conditional R^2 (theoretical)	0.22
Log-likelihood	-46.10 (df = 3)
AIC	98.19
BIC	108.99
AUC/ROC	0.79

Summary:

The model's statistical assumptions are well-met. The fixed effect of VOSI Visual-Spatial score on correctness is statistically significant, however its explanatory power is modest (marginal R^2 = 0.15). The AUC/ROC value of 0.79 indicates good model performance.

Figures:

Fig. S9 (Linearity in logit)



S10: Summary of model 'Correctness versus Psi-Q Bodily Sensation score' with random effect of participant, for SUVI-FG (images only)

Overview:

- **Model type:** Generalised linear mixed (binomial distribution)
- **Dependent variable:** Correctness (binary outcome) (functional groups images)
- **Fixed effects:** Psi-Q Bodily Sensation score
- **Random effect:** Participant

Assumptions checked:

- **Linearity in logit:**
  Plot of the log of predicted probabilities versus Psi-Q Bodily Sensation score aligns reasonably with the fitted line (see Fig. S10).
- **Overdispersion:**
  Deviance residuals ratio of 0.27 suggests no evidence of overdispersion.

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	z-value	p-value
Intercept	1.54	(-0.88, 3.86)	0.88	1.76	.078
Psi-Q Bodily Sensation score	0.30	(0.02, 0.75)	0.15	2.00	.046

- **Model fit:**

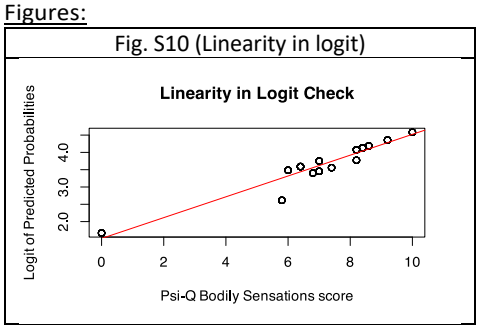
Metric	Value
Marginal R^2 (theoretical)	0.10
Conditional R^2 (theoretical)	0.18
Log-likelihood	-36.12 (df = 3)
AIC	78.23
BIC	88.86
AUC/ROC	0.84

Summary:

The model's statistical assumptions are well-met. The fixed effect of Psi-Q Bodily Sensation score on correctness is statistically significant, however its explanatory power is weak (marginal R^2 = 0.10). The AUC/ROC value of 0.84 indicates good model performance.

Figures:

Fig. S10 (Linearity in logit)



S11: Summary of model '1/time versus number of 90-degree rotations' with random effect of participant, for the PSVT:R_A

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time)
- **Fixed effects:** Number of 90-degree rotations
- **Random effect:** Participant

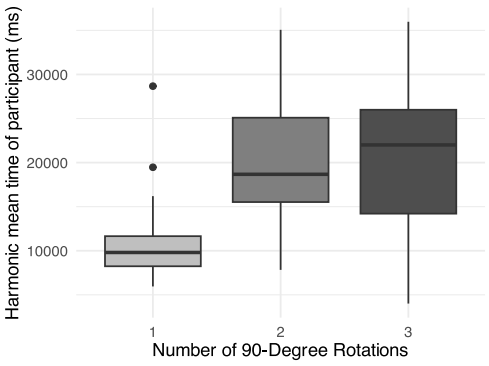


Fig. S11.0. Boxplots of the participants' harmonic mean times to solve PSVT:R_A questions.

Assumptions checked:

- **Normality of residuals:**
  Deviation at both extremes (see Fig. S11.1), and one outlier*, Shapiro-Wilk test fails ($p < .0001$).
- **Normality of random effects:**
  Normal distribution confirmed (see Fig. S11.2), Shapiro-Wilk test passes ($p = .86$).
- **Homoscedasticity:**
  Minor deviations observed at positive extremes and one outlier (see Fig. S11.3).

** The outlier was removed, and the model was refitted. The revised model showed only modest improvements in normality (Shapiro-Wilk test: $p = .010$) and fit (AIC: -2860.40, BIC: -2844.76), and the significance of the pairwise comparisons remained unaffected. Since the outlier was related to an incorrect response and subsequent analyses only use correct data, the impact is minimal. Hence, we present the results including the outlier for completeness. Fig. S11.0 excludes the outlier.*

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
(Intercept) 90-degree rotations = 1	1.03×10^{-4}	$(8.71 \times 10^{-5}, 1.19 \times 10^{-4})$	8.14×10^{-6}	t(24.4) = 12.7	< .0001
90-degree rotations = 2	-4.20×10^{-5}	$(-5.82 \times 10^{-5}, -2.59 \times 10^{-5})$	8.25×10^{-6}	t(151) = -5.09	< .0001
90-degree rotations = 3	-4.10×10^{-5}	$(5.95 \times 10^{-5}, 2.25 \times 10^{-5})$	9.46×10^{-6}	t t(151) = -4.33	< .0001

- **Model fit:**

Metric	Value
Marginal R^2	0.13
Conditional R^2	0.34
Log-likelihood	1411.30 (df = 5)
AIC	-2812.59
BIC	-2796.91

Summary:

The model fails to adhere to assumptions of normality of residuals, however random effects and homoscedasticity are satisfactory. Significant effects from the 90-degree rotations on the inverse of time are evident, with each step increase in rotations leading to a statistically significant increase in time. The marginal R^2 (0.13) suggests limited explanatory power by the fixed effect, however the conditional R^2 (0.27) shows an improvement in model fit due to the inclusion of random effects.

Figures:

Fig. S11.1 (Q-Q plot of residuals)

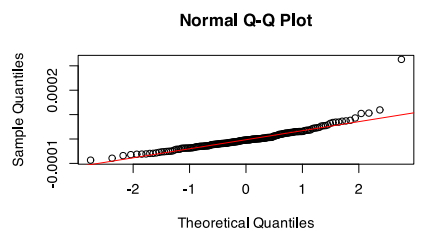


Fig. S11.2 (Q-Q plot of random effects)

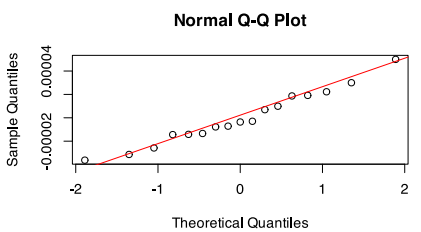
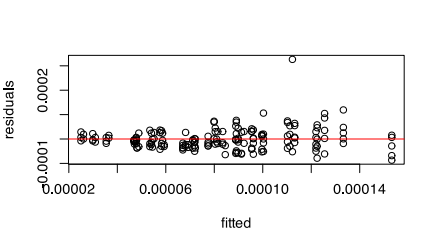


Fig. S11.3 (Plot of fitted against residuals)



S12: Pairwise comparisons showing participants took statistically different times to solve one and two, and one and three, 90-degree rotations questions, but not two and three 90-degree rotation questions, for the PSVT:R_A

Contrast	Estimate	SE	t-value	p-value
(90-degree rotations = 1) – (90-degree rotations = 2)	4.20×10^{-5}	8.25×10^{-6}	t(151) = 5.09	< . 0001
(90-degree rotations = 1) – (90-degree rotations = 3)	4.10×10^{-5}	9.45×10^{-6}	t(151) = 4.33	< .0001
(90-degree rotations = 2) – (90-degree rotations = 3)	-1.06×10^{-6}	1.03×10^{-5}	t(151) = -0.10	.99

S13: Summary of model '1/time versus VOSI Visual-Spatial score' with random effect of participant, for PSVT:R_A questions with two or three 90-degree rotations

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time) (two and three 90-degree rotation questions, correct selections only)
- **Fixed effects:** VOSI Visual-Spatial score
- **Random effect:** Participant

Assumptions:

- **Normality of residuals:**
  Generally follow a normal distribution with deviations at tails (see Fig. S13.1), Shapiro-Wilk test passes (p = .067).
- **Normality of random effects:**
  Normal distribution confirmed (see Fig. S13.2), Shapiro-Wilk test passes (p = .74).
- **Homoscedasticity:**
  Minor deviations at positive extreme (see Fig. S13.3).

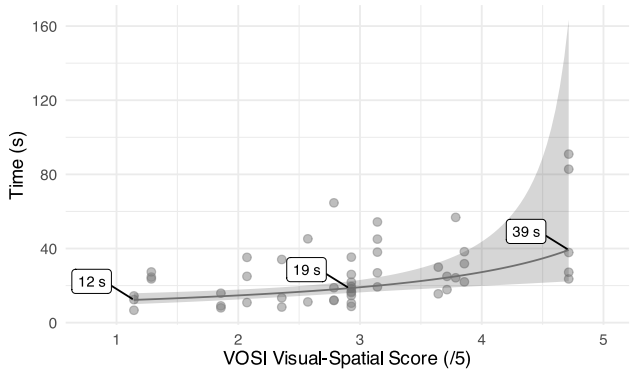


Fig. S13.0. Graph of the participants' times for PSVT:R_A questions (two or three 90-degree rotations, correct) against their VOSI Visual-Spatial scores, and the model.

Results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	9.93×10^{-6}	$(7.13 \times 10^{-5}, 1.27 \times 10^{-4})$	1.44×10^{-5}	t(13.7) = 6.90	< .0001
VOSI Visual-Spatial score	-1.57×10^{-5}	$(-2.49 \times 10^{-5}, -6.34 \times 10^{-6})$	4.76×10^{-6}	t(13.0) = -3.29	.0059

- **Model fit:**

Metric	Value
Marginal R^2	0.22
Conditional R^2	0.32
Log-likelihood	5915.74 (df = 6)
AIC	-872.72
BIC	-864.99

Summary:

This model adheres well to statistical assumptions and homoscedasticity is generally supported. Significant contribution from the VOSI Visual-Spatial score fixed effect is evident, and its explanatory power is moderate (marginal $R^2 = 0.22$).

Figures:

Fig. S13.1 (Q-Q plot of residuals)

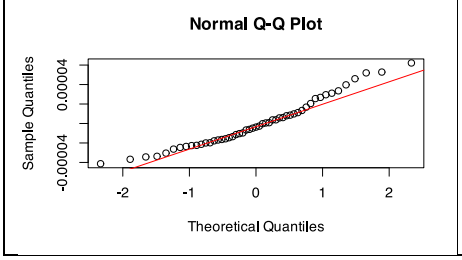


Fig. S13.2 (Q-Q plot of random effects)

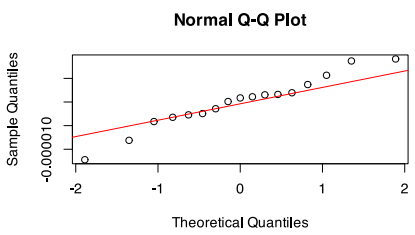
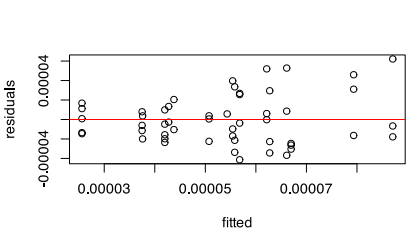


Fig. S13.3 (Plot of fitted against residuals)



S14: Summary of model '1/time versus Psi-Q Sound score' with random effect of participant, for PSVT:R_A questions with two or three 90-degree rotations

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time) (two and three 90-degree rotation questions, correct selections only)
- **Fixed effects:** Psi-Q Sound score
- **Random effect:** Participant

Assumptions checked:

- **Normality of residuals:**
  Generally follow a normal distribution with deviations at tails (see Fig. S14.1), Shapiro-Wilk test fails (p = .0079).
- **Normality of random effects:**
  Follow a normal distribution with deviations at tails (see Fig. S14.2), Shapiro-Wilk test passes (p = .53).
- **Homoscedasticity:**
  Supports linearity and homoscedasticity (see Fig. S14.3).

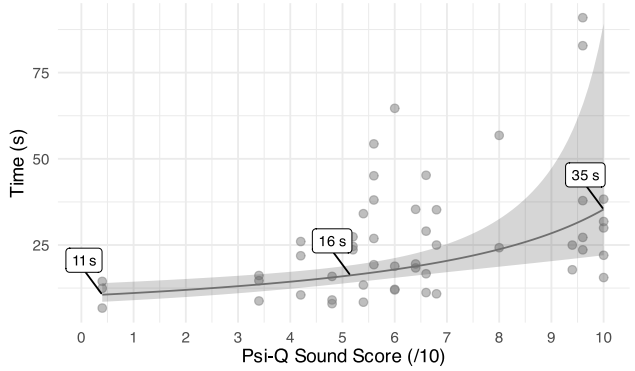


Fig. S14.0. Graph of the participants' times for PSVT:R_A questions (two or three 90-degree rotations, correct) against their Psi-Q Sound scores, and the model.

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	9.73×10^{-5}	$(7.32 \times 10^{-5}, 1.21 \times 10^{-4})$	1.23×10^{-5}	t(14.7) = 7.89	< .0001
Psi-Q Sound score	-6.90×10^{-6}	$(-1.06 \times 10^{-5}, -3.16 \times 10^{-6})$	1.90×10^{-6}	t(15.8) = -3.63	.0023

- **Model fit:**

Metric	Value
Marginal R^2	0.25
Conditional R^2	0.32
Log-likelihood	412.28 (df = 4)
AIC	-816.56
BIC	-809.07

Summary:

The model fails to adhere to assumptions of normality of residuals, however random effects and homoscedasticity are satisfactory. Significant contribution from the Psi-Q Sound score fixed effect is evident, and its explanatory power is moderate (marginal R^2 = 0.25).

Figures:

Fig. S14.1 (Q-Q plot of residuals)

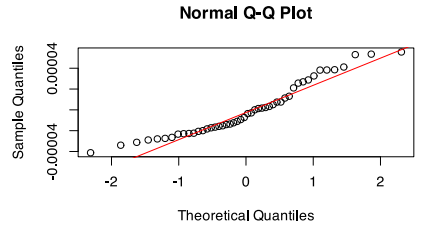


Fig. S14.2 (Q-Q plot of random effects)

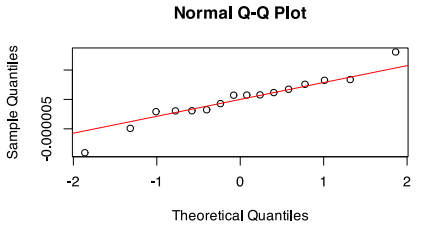
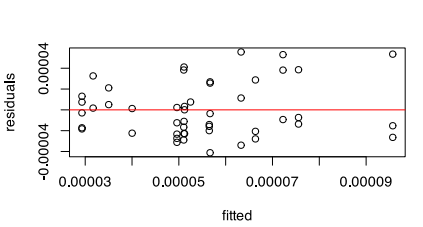


Fig. S14.3 (Plot of fitted against residuals)



S15: Summary of model '1/time versus Psi-Q Bodily Sensation score' with random effect of participant, for PSVT:R_A questions with two or three 90-degree rotations

Overview:

- **Model type:** Linear mixed
- **Dependent variable:** Inverse of time (1/time) (two and three 90-degree rotation questions, correct selections only)
- **Fixed effects:** Psi-Q Bodily Sensation score
- **Random effect:** Participant

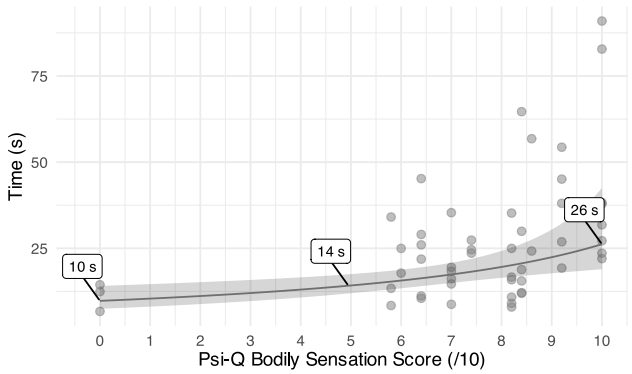


Fig. S15.0. Graph of the participants' times for PSVT:R_A questions (two or three 90-degree rotations, correct) against their Psi-Q Bodily Sensation scores, and the model.

Assumptions checked:

- **Normality of residuals:**
  Generally follow a normal distribution with deviations at tails (see Fig. S15.1), Shapiro-Wilk test fails (p = .050).
- **Normality of random effects:**
  Follow a normal distribution with deviations at tails (see Fig. S15.2), Shapiro-Wilk test fails (p = .011).
- **Homoscedasticity:**
  Supports linearity and homoscedasticity with minor deviations (see Fig. S15.3).

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	1.02×10^{-4}	$(7.09 \times 10^{-5}, 1.33 \times 10^{-4})$	1.61×10^{-5}	t(13.4) = 6.37	< .0001
Psi-Q Bodily Sensation score	6.43×10^{-6}	$(-1.05 \times 10^{-5}, -2.23 \times 10^{-6})$	2.11×10^{-6}	t(12.4) = -3.05	.0098

- **Model fit:**

Metric	Value
Marginal R^2	0.20
Conditional R^2	0.30
Log-likelihood	411.11 (df = 4)
AIC	-814.22
BIC	-806.74

Summary:

The model fails to adhere to assumptions of normality of residuals and random effects. Homoscedasticity is generally supported. Significant contributions from the Psi-Q Bodily Sensation score fixed effect is evident, and its explanatory power is moderate (marginal R^2 = 0.20).

Figures:

Fig. S15.1 (Q-Q plot of residuals)

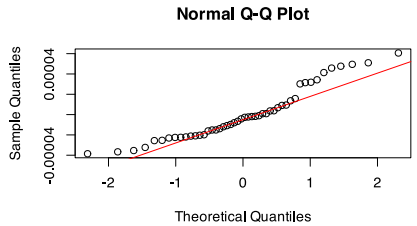


Fig. S15.2 (Q-Q plot of random effects)

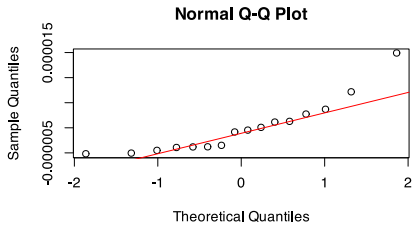
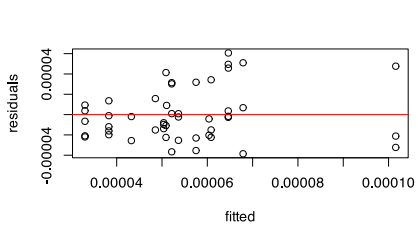


Fig. S15.3 (Plot of fitted against residuals)



S16: Excerpts of participants' descriptions of imagining a carboxylic acid, from interviews during the qualitative study

Participant	Excerpt Quote	Code
Bailey	"Essentially I saw the whole thing at once."	(1) All at once
Sam	"I saw the double bonded oxygen first." […] "The next step was I guess choosing whether or not I was going to use an oxygen with a negative sign or an OH."	(1) Carbonyl (oxygen) (2) Alcohol
Pat	"…started with the like functional group first and then it just kind of the chain." […] "I had the carbonyl and then the hydroxy popped out like that and then it went down."	(1) Carbonyl (2) Alcohol (3) Chain
Jamie	"I started with like sort of two carbons -cause, and some hydrogens. And then I thought, oh, carboxylic acid has a double bonded oxygen and I visualised that sort of coming off one of the carbons with a double bond and then next to it I said ohh, it's got a O with a H on it and I did that and I said, hey, I've made a carboxylic acid and then I said done."	(1) Chain (2) Carbonyl (oxygen) (3) Alcohol
Taylor	"There's a C, then double bond up to the O, and then out to the O-H."	(1) Carbonyl (carbon) (2) Carbonyl (oxygen) (3) Alcohol
Alex	"The first thing that popped into my head was the abbreviation C O O H, and then I kind of split it up into the like the alcohol and the carbonyl? Carbonyl group." […] "So the first thing is just like 'line C O O H', and then I think the line structure is probably how I first visualised it." […] *So the I noticed you drew like the line the two lines like that.* (Mhm). *And then the second line became the carbonyl, in a way.* (Mhm). *And, and so when, when you had the two lines like that, did they both appear together at once or did they appear one after the time, after the other.* "I think they'd appear how I draw them, like one, two."	(1) - COOH (2) Two lines (3) Alcohol (4) Carbonyl
Jordan	"I saw a line. And a C. A double bonded O at the top. And then an O-H group. That's how I saw it."	(1) Line (2) Carbonyl (carbon) (3) Carbonyl (oxygen) (4) Alcohol

S17: Summary of model '1/time versus number of steps to imagine a carboxylic acid'

Overview:

- **Model type:** Linear
- **Dependent variable:** Inverse of time (1/time)
- **Fixed effects:** Number of steps to imagine a carboxylic acid

Assumptions checked:

- **Normality of residuals:**
  Generally follow a normal distribution with deviations at tails (see Fig. S17.1), Shapiro-Wilk test passes (p = .15).
- **Homoscedasticity:**
  Supports linearity and homoscedasticity with minor deviations (see Fig. S17.2).

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	0.70	(0.38, 1.03)	0.13	t(5) = 5.54	.0026
Number of Steps	-0.12	(-0.23, -0.02)	0.04	t(5) = -2.93	.033

- **Model fit:**

Metric	Value
Multiple R^2	0.63
Adjusted R^2	0.56
Log-likelihood	6.70 (df = 3)
AIC	-7.40
BIC	-7.57

Summary:

The assumptions underlying the model are generally met. The fixed effect of Number of Steps has a significant impact on the inverse of time, suggesting that more steps slow the process. The model's fit is good, indicated by the R^2 values.

Figures:

Fig. S17.1 (Q-Q plot of residuals)

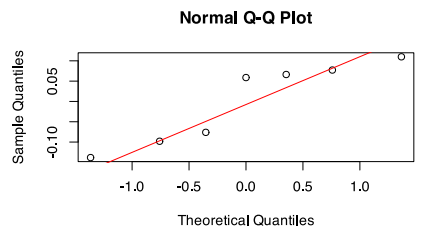
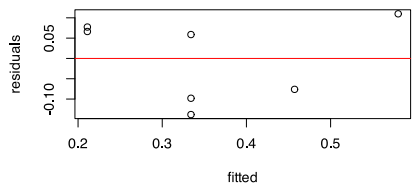


Fig. S17.2 (Plot of fitted against residuals)



S18: Summary of model 'Time to imagine a carboxylic acid versus Psi-Q Vision score'

Overview:

- **Model type:** Linear
- **Dependent variable:** Time
- **Fixed effects:** Psi-Q Vision score

Assumptions checked:

- **Normality of residuals:**
  Generally follow a normal distribution with deviations at tails (see Fig. S18.1), Shapiro-Wilk test passes (p = .99).
- **Homoscedasticity:**
  Supports linearity and homoscedasticity with minor deviations (see Fig. S18.2).

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
Intercept	5.95	(3.39, 8.51)	1.00	t(5) = 5.98	.0019
PsiQ Vision score	-0.35	(-0.66, - 0.03)	0.12	t(5) = -2.81	.037

- **Model fit:**

Metric	Value
Multiple R^2	0.61
Adjusted R^2	0.54
Log-likelihood	-7.37 (df = 3)
AIC	20.75
BIC	20.59

Summary:

The assumptions underlying the model are generally met. The fixed effect of Psi-Q Vision score has a significant impact on the time, suggesting that higher scores are associated with faster times. The model's fit is good, indicated by the R^2 values.

Figures:

Fig. S18.1 (Q-Q plot of residuals)

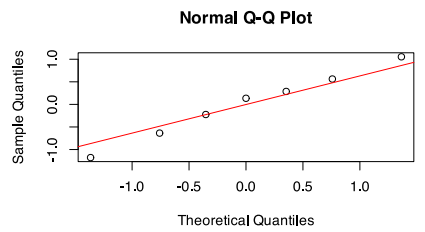
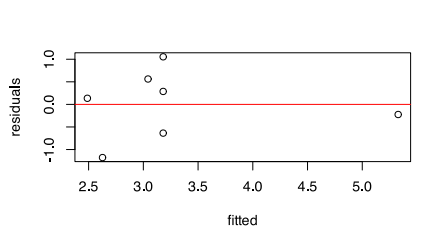


Fig. S18.2 (Plot of fitted against residuals)



S19: Summary of model 'VOSI Visual-Object score versus strategy to solve PSVT:R_A questions'

Overview:

- **Model type:** Linear
- **Dependent variable:** VOSI Visual-Object score
- **Fixed effects:** Strategy to solve PSVT:R_A questions

Statistical assumptions checked:

- **Normality of residuals:**
  Somewhat follow a normal distribution with deviations at the tails (see Fig. S19.1), Shapiro-Wilk test fails (p = .019).
- **Homoscedasticity:**
  Reasonably supports linearity and homoscedasticity (see Fig. S19.2).

Key model results:

- **Fixed effects:**

Variable	Coefficient	95% CI	SE	t-value	p-value
(Intercept) Mental Rotation	4.15	(3.95, 4.35)	0.094	t(17) = 44.1	< .0001
Multiple	0.69	(0.23, 1.14)	0.217	t(17) = 3.15	.0059
Projection	0.53	(-0.01, 1.07)	0.258	t(17) = 2.056)	.055
Verbal	-2.93	(-3.39, -2.48)	0.217	t(17) = -13.49	< .0001

- **F-statistic:**
  F(3,17) = 75.92, p < .0001

- **Model fit:**

Metric	Value
Multiple R^2	0.93
Adjusted R^2	0.92
Log-likelihood	-4.89 (df = 5)
AIC	19.79
BIC	25.01

Summary:

The model does not adhere to assumptions of normality of residuals, but homoscedasticity is reasonably supported. The Projection fixed effect does not meet conventions for statistical significance. The model explanatory power as indicated by the R^2 values is exceptionally high. Overall model is significant as indicated by the F-statistic (p < .0001). Given the small sample size and potential for overfitting, these results should be interpreted with caution.

Figures:

Fig. S19.1 (Q-Q plot of residuals)

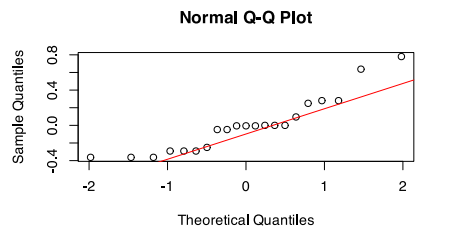


Fig. S19.2 (Plot of fitted against residuals)