

# **INTEGRAL PERCEPTION IN AUGMENTED REALITY**

Michael K. McGee

Dissertation submitted to the Faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
in  
Industrial and Systems Engineering

Dr. Woodrow Barfield, Chairman

Dr. Jan Helge Bøhn

Dr. Thomas A. Dingus

Dr. Deborah Hix

Dr. Brian M. Kleiner

August 1999

Blacksburg, Virginia

Keywords: Augmented Reality, Augmented Environments, Integral Perception

Copyright 1999, Michael K. McGee

# Integral Perception in Augmented Reality

Michael K. McGee

(ABSTRACT)

Augmented reality, the superimposing of graphics or text onto an actual visual scene, is intended to enhance a user's understanding of the real world. This research examines the perceptual, cognitive, and human factors implications of combining integrally designed computer-generated imagery with real world scenes. Three experiments were conducted to test the theoretical and practical consequences of integral perception in augmented reality.

The first experiment was a psychophysical study that had participants subjectively assess the integrality of 32 scenes comprising four different augmented reality object environments (computer, brain, text, and liquid dynamic model), projected at two transparency levels (opaque, and semi-transparent), and presented with four different graphic textures (color, grayscale, white, and wireframe).

The second experiment expanded the psychophysical integrality assessment of augmented scenes to 32 different images composed of four new environments (housing development, computer lab, planetary photo, and trees in countryside), with multiple computer-generated graphics (two, four, six, and eight), at two levels of integrality as defined by experiment one (high, low).

The third experiment was an applied study that had two phases: 1) learning tasks using three augmented environments; and, 2) assembly tasks using eight augmented video instructions. The computer-generated graphics for each phase of experiment three were presented at two levels of integrality (high, low) as defined by experiment one.

The primary results of the three experiments show that augmented reality scenes with computer-generated imagery presented transparently and in color were perceived most integrally; increasing the number of graphics from two to eight decreased integral perception; and, high integral graphics aided performance in learning and real assembly tasks.

From the statistical results and experimenter observation of the three experiments, guidelines for designing integrally perceived graphics in augmented environments were compiled based on principles of human factors, perception, and graphic design. The key themes of the design guidelines were: 1) maintaining true shape information in the computer-generated graphics 2) using highly realistic graphics for naturalistic augmented settings; 3) considering the hardware limitations of the augmented system, particularly the display; and, 4) designing appropriately for the task (simple, complex, hands-on, cognitive, dynamic, static, etc.).

## ACKNOWLEDGMENTS

First, I would like to acknowledge ONR (NOO0149710388) for an equipment grant that has supported research at Virginia Tech in the area of augmented reality and wearable computers. Without the equipment, I would not have been able to conduct this research. Second, I would like to extend a global appreciation to all the fellow students I have encountered in the Virginia Tech Human Factors program in my nearly seven years of association with the university. They were my true educators during my long career here. Third, I would like to thank my committee for being genuinely helpful and encouraging throughout every step of this dissertation process. There were many times where I leaned on their expertise and good nature, and I was never let down. Next, I would like to thank Woodrow Barfield for constant mind-expanding inspiration, both with the dissertation and beyond. I look forward to years of engaging discussions on the role the human may or may not play in our developing high-tech world. And last in this list, but first in ultimate importance, I thank my parents for providing me every opportunity to succeed, in all things. There would have been nothing without them.

## TABLE OF CONTENTS

Acknowledgments	i
Table of Contents	ii
List of Figures	v
List of Tables	x
Chapter 1. Introduction	1
1.1. Problem Statement	3
1.2. Experimental Goals and Hypotheses	4
Chapter 2. Augmented Reality	6
2.1. Augmented Reality Systems	6
2.2. Augmented Reality and the User	16
2.3. Relation to Head-Up Displays	19
2.4. Relation to Virtual Reality	22
Chapter 3. Integral Perception	24
3.1. Integral/Separable Research	25
3.2. Related Perceptual Research	29
Chapter 4. Training	34
Chapter 5. Experiment One	37
5.1. Methodology	37
5.1.1. Experimental Design	37
5.1.2. Participants	38
5.1.3. Experimental Apparatus	38
5.1.3.1. Equipment Setup	38
5.1.3.2. Augmented Environments	41
5.1.3.3. Questionnaires	44
5.1.3.4. Vision Tests	45
5.1.4. Experimental Procedures	45
5.2. Expected Results	46
5.3. Results	47
5.3.1. Trial and Order ANOVAs	48
5.3.2. Augmented Object, Transparency, Texture ANOVAs	51
5.3.3. Nonparametric Tests	82
5.3.4. Subjective Rating Scales Factor Analyses	85
5.3.5. Correlations	96
5.4. Discussion	98
5.4.1. Order	98
5.4.2. Augmented Object, Transparency of Graphics, and Texture of Graphics	

.....	98
5.4.3. Subjective Rating Scales .....	103
5.4.4. Correlations .....	105
5.4.5. Design Guidelines .....	106
Chapter 6. Experiment Two .....	108
6.1. Methodology .....	108
6.1.1. Experimental Design .....	108
6.1.2. Participants .....	109
6.1.3. Experimental Apparatus .....	110
6.1.3.1. Equipment Setup .....	110
6.1.3.2. Augmented Environments .....	110
6.1.3.3. Questionnaires .....	112
6.1.3.4. Vision Tests .....	112
6.1.4. Experimental Procedures .....	112
6.2. Expected Results .....	113
6.3. Results .....	114
6.3.1. Trial and Order ANOVAs .....	114
6.3.2. Augmented Environment, Integralness, Quantity ANOVAs .....	118
6.3.3. Nonparametric Tests .....	141
6.3.4. Subjective Rating Scales Factor Analyses .....	146
6.3.5. Correlations .....	181
6.4. Discussion .....	183
6.4.1. Order .....	183
6.4.2. Augmented Environment, Integralness of Graphics, and Quantity of Graphics .....	183
6.4.3. Subjective Rating Scales .....	187
6.4.4. Correlations .....	190
6.4.5. Design Guidelines .....	190
Chapter 7. Experiment Three .....	192
7.1. Methodology .....	192
7.1.1. Experimental Design .....	192
7.1.2. Participants .....	194
7.1.3. Experimental Apparatus .....	194
7.1.3.1. Experimental Setup .....	194
7.1.3.2. Augmented Environments .....	196
7.1.3.3. Questionnaires .....	200
7.1.3.4. Vision Tests .....	200
7.1.4. Experimental Procedures .....	200
7.2. Expected Results .....	201
7.3. Results .....	203
7.3.1. Subjective Measure ANOVAs: Quiz Augmented Scenes .....	203
7.3.2. Quiz Answers and Subjective Measure Nonparametric Tests: Quiz Augmented Scenes .....	208
7.3.3. Wilcoxon-Mann-Whitney Test: Quiz Augmented Scene Drawings .....	213

7.3.4. Time and Subjective Measure ANOVAs: Assembly Tasks . . . . .	215
7.3.5. Error and Subjective Measure Nonparametric Tests: Assembly Tasks . . . . .	223
7.3.6. Correlations . . . . .	227
7.4. Discussion . . . . .	229
7.4.1. Quiz Augmented Scenes . . . . .	229
7.4.2. Assembly Tasks . . . . .	231
7.4.3. Correlations . . . . .	234
7.4.4. Design Guidelines . . . . .	236
Chapter 8. Overall Discussion and Conclusions . . . . .	238
8.1. Nature of Integral Graphics in Augmented Environments . . . . .	238
8.2. Design Guidelines . . . . .	239
8.3. Future Research . . . . .	244
8.4. Conclusions . . . . .	245
Chapter 9. References . . . . .	247
Appendix A: Experiment One Trial Questionnaire . . . . .	251
Appendix B: Universal General Questionnaire . . . . .	252
Appendix C: Universal Informed Consent Form . . . . .	253
Appendix D: Experiment Two Trial Questionnaire . . . . .	255
Appendix E: Experiment Three Quiz Questionnaires . . . . .	256
Appendix F: Experiment Three Assembly Questionnaire . . . . .	259
Vita . . . . .	260

## LIST OF FIGURES

Figure 1. Schematics for see-through (left) and video merging (right) techniques (adapted from Barfield et al., 1995). . . . .	8
Figure 1. A prototype head-mounted display for augmented reality applications ( <a href="http://www.wearcam.org/head-mounted-displays.html">http://www.wearcam.org/head-mounted-displays.html</a> ). . . . .	9
Figure 3. The i-glasses see-through head-mounted display from i-o Display Systems ( <a href="http://www.i-glasses.com/">http://www.i-glasses.com/</a> ). . . . .	9
Figure 4. The immersive head-mounted display used in the present studies. . . . .	10
Figure 5. The Milgram, Takemura, Utsumi, and Kishino (1994) reality-virtuality continuum. . . . .	13
Figure 6. Ellis' (1994) schematic of virtual reality. . . . .	22
Figure 7. Mental model of augmented environments. . . . .	24
Figure 8. Example of integral/separable dimensions. . . . .	26
Figure 9. Representation of transfer of training for traditional training methods and augmented reality as an on-the-job training tool. . . . .	35
Figure 10. The equipment setup for the three experiments. . . . .	40
Figure 11. The Virtual Research V8 head-mounted display. . . . .	40
Figure 12. Schematic showing equipment setup. . . . .	41
Figure 13. Computer-solid-white condition for experiment one. . . . .	43
Figure 14. Brain-solid-wireframe condition for experiment one. . . . .	43
Figure 15. Brain-transparent-color condition for experiment one. . . . .	43
Figure 16. Dynamic-transparent-color condition for experiment one. . . . .	43
Figure 17. Text-solid-gray condition for experiment one. . . . .	43
Figure 18. Dynamic-transparent-white condition for experiment one. . . . .	43
Figure 19. Example mental rotation test stimuli. . . . .	44
Figure 20. Example color vision test plate. . . . .	45
Figure 21. Graph of the main effect of augmented reality object, significant at the $p < 0.01$ level ( $p = 0.005$ ) as measured by goodness ratings. . . . .	53
Figure 22. Graph of the main effect of texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	53
Figure 23. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.005$ ) as measured by goodness ratings. . . . .	54
Figure 24. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	54
Figure 25. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	55
Figure 26. Graph of the main effect of transparency, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	57
Figure 27. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	57
Figure 28. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	58
Figure 29. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	58
Figure 30. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by integral ratings. . . . .	60
Figure 31. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by integral ratings. . . . .	60
Figure 32. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.004$ ) as measured by integral ratings. . . . .	61

Figure 33. Graph of the main effect of transparency, significant at the $p < 0.01$ level ( $p = 0.004$ ) as measured by distracting ratings. . . . .	63
Figure 34. Graph of the main effect of texture, significant at the $p < 0.01$ level ( $p = 0.004$ ) as measured by distracting ratings. . . . .	63
Figure 35. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distracting ratings. . . . .	64
Figure 36. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distracting ratings. . . . .	64
Figure 37. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distracting ratings. . . . .	65
Figure 38. Graph of the main effect of transparency, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by distinguishable ratings. . . . .	67
Figure 39. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distinguishable ratings. . . . .	67
Figure 40. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distinguishable ratings. . . . .	68
Figure 41. Graph of the interaction between transparency and texture, significant at the $p < 0.01$ level ( $p = 0.002$ ) as measured by distinguishable ratings. . . . .	68
Figure 42. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by distinguishable ratings. . . . .	69
Figure 43. Graph of the main effect of transparency, significant at the $p < 0.01$ level ( $p = 0.003$ ) as measured by blending ratings. . . . .	71
Figure 44. Graph of the main effect of texture, significant at the $p < 0.01$ level ( $p = 0.002$ ) as measured by blending ratings. . . . .	71
Figure 45. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by blending ratings. . . . .	72
Figure 46. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by blending ratings. . . . .	72
Figure 47. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by blending ratings. . . . .	73
Figure 48. Graph of the main effect of texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by usefulness ratings. . . . .	75
Figure 49. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.008$ ) as measured by usefulness ratings. . . . .	75
Figure 50. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by usefulness ratings. . . . .	76
Figure 51. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by usefulness ratings. . . . .	76
Figure 52. Graph of the main effect of transparency, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by unified scene ratings. . . . .	78
Figure 53. Graph of the main effect of texture, significant at the $p < 0.01$ level ( $p = 0.004$ ) as measured by unified scene ratings. . . . .	78
Figure 54. Graph of the interaction between augmented reality object and transparency, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by unified scene ratings. . . . .	79
Figure 55. Graph of the interaction between augmented reality object and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by unified scene ratings. . . . .	79
Figure 56. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by unified scene ratings. . . . .	80
Figure 57. Loadings plot for the solid-color factor analysis. . . . .	88

Figure 58. Loadings plot for the solid-gray factor analysis. . . . .	89
Figure 59. Loadings plot for the solid-white factor analysis. . . . .	90
Figure 60. Loadings plot for the solid-wire factor analysis. . . . .	91
Figure 61. Loadings plot for the transparent-color factor analysis. . . . .	92
Figure 62. Loadings plot for the transparent-gray factor analysis. . . . .	93
Figure 63. Loadings plot for the transparent-white factor analysis. . . . .	94
Figure 64. Loadings plot for the transparent-wire factor analysis. . . . .	95
Figure 65. Houses-low-six condition for experiment two. . . . .	111
Figure 66. Houses-high-four condition for experiment two. . . . .	111
Figure 67. Lab-high-four condition for experiment two. . . . .	111
Figure 68. Lab-low-eight condition for experiment two. . . . .	111
Figure 69. Planets-high-two condition for experiment two. . . . .	111
Figure 70. Trees-low-eight condition for experiment two. . . . .	111
Figure 71. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	119
Figure 72. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	119
Figure 73. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by goodness ratings. . . . .	120
Figure 74. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by integral ratings. . . . .	122
Figure 75. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by integral ratings. . . . .	122
Figure 76. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by integral ratings. . . . .	123
Figure 77. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by attend ratings. . . . .	125
Figure 78. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by attend ratings. . . . .	125
Figure 79. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.002$ ) as measured by attend ratings. . . . .	126
Figure 80. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by clutter ratings. . . . .	127
Figure 81. Graph of the interaction between augmented reality scene and quantity, significant at the $p < 0.01$ level ( $p = 0.002$ ) as measured by clutter ratings. . . . .	128
Figure 82. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	130
Figure 83. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.001$ ) as measured by separable ratings. . . . .	130
Figure 84. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by separable ratings. . . . .	131
Figure 85. Graph of the interaction between augmented reality scene and integralness, significant at the $p < 0.01$ level ( $p = 0.003$ ) as measured by separable ratings. . . . .	131
Figure 86. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by similar ratings. . . . .	133
Figure 87. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by similar ratings. . . . .	133
Figure 88. Graph of the interaction between augmented reality scene and integralness, significant at the $p < 0.01$ level ( $p = 0.003$ ) as measured by similar ratings. . . . .	134
Figure 89. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by recognition ratings. . . . .	135
Figure 90. Graph of the interaction between augmented reality scene and quantity, significant at the $p < 0.01$ level ( $p = 0.010$ ) as measured by recognition ratings. . . . .	136
Figure 91. Graph of the main effect of augmented reality scene, significant at the $p < 0.01$ level ( $p = 0.000$ ) as	

measured by usefulness ratings. . . . .	138
Figure 92. Graph of the main effect of quantity, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by usefulness ratings. . . . .	138
Figure 93. Graph of the main effect of integralness, significant at the $p < 0.01$ level ( $p = 0.000$ ) as measured by usefulness ratings. . . . .	139
Figure 94. Loadings plot for the houses-two-high factor analysis. . . . .	149
Figure 95. Loadings plot for the houses-two-low factor analysis. . . . .	150
Figure 96. Loadings plot for the houses-high-four factor analysis. . . . .	151
Figure 97. Loadings plot for the houses-four-low factor analysis. . . . .	152
Figure 98. Loadings plot for the houses-six-high factor analysis. . . . .	153
Figure 99. Loadings plot for the houses-six-low factor analysis. . . . .	154
Figure 100. Loadings plot for the houses-eight-high factor analysis. . . . .	155
Figure 101. Loadings plot for the houses-eight-low factor analysis. . . . .	156
Figure 102. Loadings plot for the lab-two-high factor analysis. . . . .	157
Figure 103. Loadings plot for the lab-two-low factor analysis. . . . .	158
Figure 104. Loadings plot for the lab-four-high factor analysis. . . . .	159
Figure 105. Loadings plot for the lab-four-low factor analysis. . . . .	160
Figure 106. Loadings plot for the lab-six-high factor analysis. . . . .	161
Figure 107. Loadings plot for the lab-six-low factor analysis. . . . .	162
Figure 108. Loadings plot for the lab-eight-high factor analysis. . . . .	163
Figure 109. Loadings plot for the lab-eight-low factor analysis. . . . .	164
Figure 110. Loadings plot for the planets-two-high factor analysis. . . . .	165
Figure 111. Loadings plot for the planets-two-low factor analysis. . . . .	166
Figure 112. Loadings plot for the planets-four-high factor analysis. . . . .	167
Figure 113. Loadings plot for the planets-four-low factor analysis. . . . .	168
Figure 114. Loadings plot for the planets-six-high factor analysis. . . . .	169
Figure 115. Loadings plot for the planets-six-low factor analysis. . . . .	170
Figure 116. Loadings plot for the planets-eight-high factor analysis. . . . .	171
Figure 117. Loadings plot for the planets-eight-low factor analysis. . . . .	172
Figure 118. Loadings plot for the trees-two-high factor analysis. . . . .	173
Figure 119. Loadings plot for the trees-two-low factor analysis. . . . .	174
Figure 120. Loadings plot for the trees-four-high factor analysis. . . . .	175
Figure 121. Loadings plot for the trees-four-low factor analysis. . . . .	176
Figure 122. Loadings plot for the trees-six-high factor analysis. . . . .	177
Figure 123. Loadings plot for the trees-six-low factor analysis. . . . .	178
Figure 124. Loadings plot for the trees-eight-high factor analysis. . . . .	179
Figure 125. Loadings plot for the trees-eight-low factor analysis. . . . .	180
Figure 126. Stereo components used in assembly task two of experiment three. . . . .	195
Figure 127. Stereo components used in assembly task five of experiment three. . . . .	195
Figure 129. Stereo components used in assembly task eight of experiment three. . . . .	195
Figure 128. Stereo components used in assembly task seven of experiment three. . . . .	195
Figure 132. Graph of the interaction between augmented reality scene and integralness, significant at the $p < 0.05$ level ( $p = 0.015$ ) as measured by integral ratings. . . . .	204
Figure 133. Graph of the main effect of augmented reality scene, significant at the $p < 0.05$ level ( $p = 0.017$ ) as measured by usefulness ratings. . . . .	206
Figure 134. Graph of the main effect of integralness, significant at the $p < 0.05$ level ( $p = 0.001$ ) as measured by usefulness ratings. . . . .	206
Figure 135. Graph of the interaction between augmented reality scene and integralness, significant at the $p < 0.05$ level ( $p = 0.007$ ) as measured by usefulness ratings. . . . .	207
Figure 136. Graph of the main effect of task, significant at the $p < 0.05$ level ( $p = 0.000$ ) as measured by time to complete task. . . . .	216
Figure 137. Graph of the interaction between task and integralness, significant at the $p < 0.05$ level ( $p = 0.001$ ) as measured by time to complete task. . . . .	216
Figure 138. Graph of the main effect of task, significant at the $p < 0.05$ level ( $p = 0.000$ ) as measured by ratings of	

the integral level of the graphics in the augmented instructions. ....	217
Figure 139. Graph of the main effect of integralness, significant at the $p < 0.05$ level ( $p = 0.033$ ) as measured by ratings of the integral level of the graphics in the augmented instructions. ....	218
Figure 140. Graph of the main effect of task, significant at the $p < 0.05$ level ( $p = 0.000$ ) as measured by ratings of the goodness of the graphics in the augmented instructions. ....	219
Figure 141. Graph of the main effect of task, significant at the $p < 0.05$ level ( $p = 0.000$ ) as measured by ratings of the usefulness of the graphics in the augmented instructions. ....	220
Figure 142. Graph of the interaction between task and integralness, significant at the $p < 0.05$ level ( $p = 0.001$ ) as measured by ratings of the usefulness of the graphics in the augmented instructions. ....	220
Figure 143. Graph of the main effect of task, significant at the $p < 0.05$ level ( $p = 0.000$ ) as measured by ratings of the overall effectiveness of the augmented instructions. ....	221
Figure 144. Graph of the interaction between task and integralness, significant at the $p < 0.05$ level ( $p = 0.008$ ) as measured by ratings of the overall effectiveness of the augmented instructions. ....	222

## LIST OF TABLES

Table 1. Augmentation available for the five main sensory systems. . . . .	7
Table 2. The impact of various technological augmented reality parameters. . . . .	12
Table 3. Classification of mixed and augmented reality. . . . .	15
Table 4. Subjective rating scale descriptions for experiment one. . . . .	38
Table 5. ANOVA summary table for overall order, measured by a goodness rating. . . . .	48
Table 6. ANOVA summary table for overall order, measured by a separable rating. . . . .	48
Table 7. ANOVA summary table for overall order, measured by an integral rating. . . . .	48
Table 8. ANOVA summary table for overall order, measured by a distracting rating. . . . .	48
Table 9. ANOVA summary table for overall order, measured by a distinguishable rating. . . . .	49
Table 10. ANOVA summary table for overall order, measured by a blending rating. . . . .	49
Table 11. ANOVA summary table for overall order, measured by a usefulness rating. . . . .	49
Table 12. ANOVA summary table for overall order, measured by a unified scene rating. . . . .	49
Table 13. ANOVA summary table for scene presentation order, measured by a goodness rating. . . . .	50
Table 14. ANOVA summary table for scene presentation order, measured by a separable rating. . . . .	50
Table 15. ANOVA summary table for scene presentation order, measured by an integral rating. . . . .	50
Table 16. ANOVA summary table for scene presentation order, measured by a distracting rating. . . . .	50
Table 17. ANOVA summary table for scene presentation order, measured by a distinguishable rating. . . . .	50
Table 18. ANOVA summary table for scene presentation order, measured by a blending rating. . . . .	50
Table 19. ANOVA summary table for scene presentation order, measured by a usefulness rating. . . . .	51
Table 20. ANOVA summary table for scene presentation order, measured by a unified scene rating. . . . .	51
Table 21. ANOVA summary table for the three independent variables, measured by a goodness rating. . . . .	51
Table 22. ANOVA summary table for the three independent variables, measured by a separable rating. . . . .	56
Table 23. ANOVA summary table for the three independent variables, measured by an integral rating. . . . .	59
Table 24. ANOVA summary table for the three independent variables, measured by a distracting rating. . . . .	62
Table 25. ANOVA summary table for the three independent variables, measured by a distinguishable rating. . . . .	66
Table 26. ANOVA summary table for the three independent variables, measured by a blending rating. . . . .	70
Table 27. ANOVA summary table for the three independent variables, measured by a usefulness rating. . . . .	74
Table 28. ANOVA summary table for the three independent variables, measured by a unified scene rating. . . . .	77
Table 29. Summary of significant results at the $p < 0.01$ level for the eight rating scale dependent measures for experiment one. . . . .	81
Table 30. Kruskal-Wallis test for augmented reality object as measured by goodness ratings. . . . .	82
Table 31. Kruskal-Wallis test for texture as measured by goodness ratings. . . . .	82
Table 32. Kruskal-Wallis test for transparency as measured by separable ratings. . . . .	82
Table 33. Kruskal-Wallis test for transparency as measured by distracting ratings. . . . .	83
Table 34. Kruskal-Wallis test for texture as measured by distracting ratings. . . . .	83
Table 35. Kruskal-Wallis test for transparency as measured by distinguishable ratings. . . . .	83
Table 36. Kruskal-Wallis test for transparency as measured by blending ratings. . . . .	83
Table 37. Kruskal-Wallis test for texture as measured by blending ratings. . . . .	84
Table 38. Kruskal-Wallis test for texture as measured by usefulness ratings. . . . .	84
Table 39. Kruskal-Wallis test for transparency as measured by unified scene ratings. . . . .	84
Table 40. Kruskal-Wallis test for texture as measured by unified scene ratings. . . . .	84
Table 41. Factor analysis results for the solid-gray condition extracting four factors using the principal components method without rotation. . . . .	86
Table 42. Factor analysis table for the data collapsed across solid-color conditions. . . . .	88
Table 43. Factor analysis table for the data collapsed across solid-gray conditions. . . . .	89
Table 44. Factor analysis table for the data collapsed across solid-white conditions. . . . .	90
Table 45. Factor analysis table for the data collapsed across solid-wire conditions. . . . .	91
Table 46. Factor analysis table for the data collapsed across transparent-color conditions. . . . .	92
Table 47. Factor analysis table for the data collapsed across transparent-gray conditions. . . . .	93
Table 48. Factor analysis table for the data collapsed across transparent-white conditions. . . . .	94

Table 49. Factor analysis table for the data collapsed across transparent-wire conditions. . . . .	95
Table 50. A list and description of the variables from the correlation tests for experiment one. . . . .	96
Table 51. Correlations and correlation p-values for tested variables from experiment one. . . . .	97
Table 52. Design guidelines from experiment one. . . . .	107
Table 53. Subjective rating scale descriptions for experiment two. . . . .	109
Table 54. ANOVA summary table for overall order, measured by a goodness rating. . . . .	115
Table 55. ANOVA summary table for overall order, measured by an integral rating. . . . .	115
Table 56. ANOVA summary table for overall order, measured by an attend rating. . . . .	115
Table 57. ANOVA summary table for overall order, measured by a clutter rating. . . . .	115
Table 58. ANOVA summary table for overall order, measured by a separable rating. . . . .	115
Table 59. ANOVA summary table for overall order, measured by a similar rating. . . . .	115
Table 60. ANOVA summary table for overall order, measured by a recognition rating. . . . .	116
Table 61. ANOVA summary table for overall order, measured by a usefulness rating. . . . .	116
Table 62. ANOVA summary table for scene presentation order, measured by a goodness rating. . . . .	116
Table 63. ANOVA summary table for scene presentation order, measured by an integral rating. . . . .	116
Table 64. ANOVA summary table for scene presentation order, measured by an attend rating. . . . .	116
Table 65. ANOVA summary table for scene presentation order, measured by a clutter rating. . . . .	117
Table 66. ANOVA summary table for scene presentation order, measured by a separable rating. . . . .	117
Table 67. ANOVA summary table for scene presentation order, measured by a similar rating. . . . .	117
Table 68. ANOVA summary table for scene presentation order, measured by a recognition rating. . . . .	117
Table 69. ANOVA summary table for scene presentation order, measured by a usefulness rating. . . . .	117
Table 70. ANOVA summary table for the three independent variables, measured by a goodness rating. . . . .	118
Table 71. ANOVA summary table for the three independent variables, measured by an integral rating. . . . .	121
Table 72. ANOVA summary table for the three independent variables, measured by an attend rating. . . . .	124
Table 73. ANOVA summary table for the three independent variables, measured by a clutter rating. . . . .	127
Table 74. ANOVA summary table for the three independent variables, measured by a separable rating. . . . .	129
Table 75. ANOVA summary table for the three independent variables, measured by a similar rating. . . . .	132
Table 76. ANOVA summary table for the three independent variables, measured by a recognition rating. . . . .	135
Table 77. ANOVA summary table for the three independent variables, measured by a usefulness rating. . . . .	137
Table 78. Summary of significant results at the $p < 0.01$ level for the eight rating scale dependent measures for experiment two . . . . .	140
Table 79. Kruskal-Wallis test for augmented reality scene as measured by goodness ratings. . . . .	141
Table 80. Kruskal-Wallis test for quantity as measured by goodness ratings. . . . .	141
Table 81. Kruskal-Wallis test for integralness as measured by goodness ratings. . . . .	141
Table 82. Kruskal-Wallis test for augmented reality scene as measured by integral ratings. . . . .	142
Table 83. Kruskal-Wallis test for quantity as measured by integral ratings. . . . .	142
Table 84. Kruskal-Wallis test for integralness as measured by integral ratings. . . . .	142
Table 85. Kruskal-Wallis test for augmented reality scene as measured by attend ratings. . . . .	142
Table 86. Kruskal-Wallis test for quantity as measured by attend ratings. . . . .	143
Table 87. Kruskal-Wallis test for integralness as measured by attend ratings. . . . .	143
Table 88. Kruskal-Wallis test for quantity as measured by clutter ratings. . . . .	143
Table 89. Kruskal-Wallis test for augmented reality scene as measured by separable ratings. . . . .	143
Table 90. Kruskal-Wallis test for quantity as measured by separable ratings. . . . .	144
Table 91. Kruskal-Wallis test for integralness as measured by separable ratings. . . . .	144
Table 92. Kruskal-Wallis test for augmented reality scene as measured by similar ratings. . . . .	144
Table 93. Kruskal-Wallis test for quantity as measured by recognition ratings. . . . .	144
Table 94. Kruskal-Wallis test for augmented reality scene as measured by usefulness ratings. . . . .	145
Table 95. Kruskal-Wallis test for quantity as measured by usefulness ratings. . . . .	145
Table 96. Kruskal-Wallis test for integralness as measured by usefulness ratings. . . . .	145
Table 97. Factor analysis results for the houses-two-low condition extracting four factors using the principal components method without rotation. . . . .	147
Table 98. Factor analysis table for the houses-two-high conditions. . . . .	149
Table 99. Factor analysis table for the houses-two-low conditions. . . . .	150
Table 100. Factor analysis table for the houses-four-high conditions. . . . .	151

Table 101. Factor analysis table for the houses-four-low conditions. . . . .	152
Table 102. Factor analysis table for the houses-six-high conditions. . . . .	153
Table 103. Factor analysis table for the houses-six-low conditions. . . . .	154
Table 104. Factor analysis table for the houses-eight-high conditions. . . . .	155
Table 105. Factor analysis table for the houses-eight-low conditions. . . . .	156
Table 106. Factor analysis table for the lab-two-high conditions. . . . .	157
Table 107. Factor analysis table for the lab-two-low conditions. . . . .	158
Table 108. Factor analysis table for the lab-four-high conditions. . . . .	159
Table 109. Factor analysis table for the lab-four-low conditions. . . . .	160
Table 110. Factor analysis table for the lab-six-high conditions. . . . .	161
Table 111. Factor analysis table for the lab-six-low conditions. . . . .	162
Table 112. Factor analysis table for the lab-eight-high conditions. . . . .	163
Table 113. Factor analysis table for the lab-eight-low conditions. . . . .	164
Table 114. Factor analysis table for the planets-two-high conditions. . . . .	165
Table 115. Factor analysis table for the planets-two-low conditions. . . . .	166
Table 116. Factor analysis table for the planets-four-high conditions. . . . .	167
Table 117. Factor analysis table for the planets-four-low conditions. . . . .	168
Table 118. Factor analysis table for the planets-six-high conditions. . . . .	169
Table 119. Factor analysis table for the planets-six-low conditions. . . . .	170
Table 120. Factor analysis table for the planets-eight-high conditions. . . . .	171
Table 121. Factor analysis table for the planets-eight-low conditions. . . . .	172
Table 122. Factor analysis table for the trees-two-high conditions. . . . .	173
Table 123. Factor analysis table for the trees-two-low conditions. . . . .	174
Table 124. Factor analysis table for the trees-four-high conditions. . . . .	175
Table 125. Factor analysis table for the trees-four-low conditions. . . . .	176
Table 126. Factor analysis table for the trees-six-high conditions. . . . .	177
Table 127. Factor analysis table for the trees-six-low conditions. . . . .	178
Table 128. Factor analysis table for the trees-eight-high conditions. . . . .	179
Table 129. Factor analysis table for the trees-eight-low conditions. . . . .	180
Table 130. A list and description of the variables from the correlation tests for experiment two. . . . .	181
Table 131. Correlations and correlation p-values for tested variables from experiment two. . . . .	182
Table 132. Design guidelines from experiment two. . . . .	191
Table 133. The eight verbal instructions for the second phase of experiment three. . . . .	197
Table 134. ANOVA summary table for augmented reality scene and integralness, as measured by integral ratings. . . . .	204
Table 135. ANOVA summary table for augmented reality scene and integralness, as measured by goodness ratings. . . . .	205
Table 136. ANOVA summary table for augmented reality scene and usefulness, as measured by usefulness ratings. . . . .	205
Table 137. Kruskal-Wallis test for integralness as measured by number of additional houses fitting on back hill in the houses augmented scene. . . . .	208
Table 138. Kruskal-Wallis test for integralness as measured by number of floors in the closest house in the houses augmented scene. . . . .	208
Table 139. Kruskal-Wallis test for integralness as measured by perceived average prices of the houses in the houses augmented scene. . . . .	209
Table 140. Kruskal-Wallis test for integralness as measured by number of graphically presented computers in the computer lab augmented scene. . . . .	209
Table 141. Kruskal-Wallis test for integralness as measured by the total number of computers shown in the computer lab augmented scene. . . . .	209
Table 142. Kruskal-Wallis test for integralness as measured by the monitor size of the computers shown in the computer lab augmented scene. . . . .	210
Table 143. Kruskal-Wallis test for integralness as measured by the row that did not have graphically presented computers in the computer lab augmented scene. . . . .	210
Table 144. Kruskal-Wallis test for integralness as measured by the number of brain structures in the brain	

augmented scene. . . . .	210
Table 145. Kruskal-Wallis test for integralness as measured by the number of layers of the brain shown in the brain augmented scene. . . . .	211
Table 146. Kruskal-Wallis test for integralness as measured by whether there were eyes represented in the brain augmented scene. . . . .	211
Table 147. Kruskal-Wallis test for integralness as measured by whether there was a brain indent represented in the brain augmented scene. . . . .	211
Table 148. Kruskal-Wallis test for augmented reality scene as measured by usefulness ratings. . . . .	212
Table 149. Kruskal-Wallis test for level of graphic integralness as measured by usefulness ratings. . . . .	212
Table 150. The rank values for the three quiz augmented scenes by the two expert judges. . . . .	213
Table 151. Rank-sum statistics and p-values for the three quiz augmented scene drawings. . . . .	214
Table 152. ANOVA summary table for integralness and task, as measured by time to complete task. . . . .	215
Table 153. ANOVA summary table for integralness and task, as measured by integral ratings of the graphics in the augmented instructions. . . . .	217
Table 154. ANOVA summary table for integralness and task, as measured by goodness ratings of the graphics in the augmented instructions. . . . .	218
Table 155. ANOVA summary table for integralness and task, as measured by usefulness ratings of the graphics in the augmented instructions. . . . .	219
Table 156. ANOVA summary table for integralness and task, as measured by the overall effectiveness ratings of the augmented instructions. . . . .	221
Table 157. Kruskal-Wallis test for task as measured by ratings of the integral level of the graphics in the augmented instructions. . . . .	223
Table 158. Kruskal-Wallis test for integralness as measured by ratings of the integral level of the graphics in the augmented instructions. . . . .	223
Table 159. Kruskal-Wallis test for task as measured by ratings of the goodness of the graphics in the augmented instructions. . . . .	224
Table 160. Kruskal-Wallis test for task as measured by ratings of the usefulness of the graphics in the augmented instructions. . . . .	224
Table 161. Kruskal-Wallis test for task as measured by ratings of the overall effectiveness of the augmented instructions. . . . .	224
Table 162. Chi-square analysis of integralness as measure by errors in task. . . . .	225
Table 163. Kruskal-Wallis test for integralness as measured by number of errors on task. . . . .	226
Table 164. Kruskal-Wallis test for task as measured by errors in task. . . . .	226
Table 165. A list and description of the variables from the correlation tests for experiment three. . . . .	227
Table 166. Correlations and correlation p-values for tested variables in experiment three. . . . .	228
Table 167. Design guidelines from experiment three. . . . .	237
Table 168. Overall design guidelines. . . . .	240

## CHAPTER 1. INTRODUCTION

Augmented reality is reported to complement human understanding and information processing by reducing the cognitive load associated with task performance (Neumann and Majoros, 1998). In practice, the superimposing of graphics onto real scenes through the use of display combining technology is supposed to aid our real world activities. However, the amount of aid, or enhanced understanding, may be impacted by the ability of users to perceive an augmented application as a single unified scene rather than separate parts. Performance and understanding may be facilitated by an augmented scene perceived as one integrated view and hindered by one perceived as multiple disparate views. The reported potential benefits of augmented reality could be severely compromised if users are unable to form appropriate single mental models of augmented scenes (Wickens and Baker, 1995). This dissertation explores these issues in a theoretical and applied investigation of integral perception in augmented environments.

Augmented reality, while not a new research area, has enjoyed a resurgence of interest in the last decade. However, the main focus of research efforts has been on developing the technology. To date there has been very little work studying the human as a component of augmented reality systems. While the technology is still maturing, it is important to establish a strong research emphasis on the human capabilities and limitations within augmented systems. One of the primary aims of this dissertation is to extend the philosophical ideas and theories of human integral perception into augmented reality.

The research described herein is an extension and expansion of the classic integral perception psychology research by Garner and Felfody (1970) that deals with the difficulty of processing two dimensions as a single entity. While Garner and Felfody's work serves as the primary influence for the dissertation, it is only one of several areas of research that contributes to the theoretical basis. Many other aspects of psychology, graphic design, perception, and display system design contribute to the knowledge base for this topic.

Garner and Felfody's (1970) classic psychology theory states that additional cognitive processing is necessary to view a scene as separable components, whereas integrally perceived scenes are

processed more easily. Separable dimensions are defined as dimensions that are perceived as unrelated, or analyzable, while integral dimensions are those that cannot be analyzed and are seen as one unified dimension (Garner and Felfody, 1970). An example of an integral stimulus would be a color. Humans are more likely to perceive an object as having a single color, for example red, rather than multiple specific dimensions of wavelength and brightness. An example separable stimulus would be a red square, where we would clearly perceive a color and a shape rather than one composite object. These two examples are typical of the simplistic research stimuli used in most integral perception research.

Augmented reality has two distinct complex visual components: computer-generated graphics and the real-world visual scene. The first research objective in this dissertation is to determine to what parameters of graphic design impact integral perception of a whole augmented scene. The next objective is to determine how to manipulate parameters of the augmented environment to make the graphic and real components more integral. Finally the dissertation addresses the question of whether integrally designed augmented reality applications aid performance in real world tasks.

These overall objectives were manifested into three experiments. The first experiment examined graphic texture and transparency within a variety of augmented object scenes. The second experiment investigated the level of perceived integralness as defined by experiment one across augmented environment scenes of multiple objects. The third experiment attempted to confirm the idea that integral perception aids performance in a real task.

The successful realization of the main goals of this dissertation was expected to provide substantial contributions to several individual research fields. Overall, this dissertation was expected to add considerably to the body of knowledge on the impact of integral perception in augmented reality systems. To these ends, this dissertation addresses the void of human-centered investigations in augmented reality by applying theories and principles from integral perception related research to augmented environments. It also answers the following two human factors questions: 1) Are augmented reality dimensions perceived as one unified integral view or two separate distinct dimensions that need to be processed individually?; and, 2) In what manner should augmented reality scenes be created to aid performance during real world tasks?

Furthermore, this dissertation expands the research methodology of integral perception related fields by demonstrating that augmented environments can serve as highly effective experimental testing stimuli. Lastly, practical information summarized in a framework of human factors guidelines is compiled from each of the three studies.

### ***1.1. Problem Statement***

The intent behind nearly all augmented reality applications is to provide a greater understanding of a real scene through information-adding graphics. The greater understanding afforded by augmented reality is then expected to aid real-task performance. However, thus far very little research has been conducted in the field of augmented reality that involves the human as an information processor where cognition and perception of an environment directly impact performance on a task. Furthermore, no augmented reality research has directly addressed the role graphics will play in the human information processor cycle. This dissertation addresses the void of human-centered investigations in augmented reality by applying theories and principles from integral perception related research to augmented environments.

As noted previously, augmented reality by definition consists of two distinct visual scenes; a computer-generated portion and a real world portion. In the context of augmented reality the principal human factors question is as follows: Are these dimensions perceived as one unified integral view or two separate distinct dimensions that need to be processed individually? Whether users perceive the two visual components as a single unified view can potentially affect the degree to which augmented reality enhances a user's understanding of a task environment.

In addition to the absence of human-centered research in augmented environments, there has been a lack of realistic tasks in the body of work related to integral perception. Historically, simplistic tasks and stimuli have been used to test perceptual issues. This dissertation will expand the research methodology of these fields by demonstrating that augmented environments can serve as a highly effective experimental testing resource for integral perception research.

## *1.2. Experimental Goals and Hypotheses*

The main goals of this dissertation are to describe parameters of computer-generated graphics and augmented environments that enhance integral perception, and to examine the effect of integral perception on task performance in augmented reality. Secondary goals include defining relationships among various subjective measures of integral perception, and describing any interrelationships between experimental variables.

The main hypothesis of this dissertation is that the perception of certain augmented scenes as unified views will enhance performance, and the perception of certain augmented scenes as two distinct entities will decrease performance. Most augmented applications will fit this hypothesis. However, these experiments will also investigate the possibility that integral perception can be detrimental in specific augmented situations. This dissertation tests the main hypothesis in three progressive experiments addressing subjective and practical implications of integral perception of computer-generated graphics in augmented reality.

Following the tradition of integral/separable research, the first experiment starts the examination of integral perception with relatively simple augmented scenes of single objects and graphics. The graphics were presented at four levels of realism through the manipulation of the visible texture for each object. To mimic both see-through and immersive augmented systems, graphics were superimposed onto real images in either an opaque or semi-transparent format. Integral perception was assessed with a series of subjective ratings based on integral and related perceptual research. The goal for experiment one was to establish which textures and transparency levels add to a more integral perception of the respective augmented scenes. The results were expected to directly address issues discussed in the literature review on realistic graphics and integral perception.

The second experiment extended the concept of integral perception to recognizing a scene of multiple similar graphic and real objects as an overall unified view. To address issues of clutter and selective attention, from two to eight additional graphics were added to the real scenes. Based on previous results, two sets of graphics for experiment two were developed for each scene

following the procedures that led to the two most extreme results of integral and non-integral perception in experiment one. As before, integral perception was assessed with a series of subjective ratings based on integral and related perceptual research. The main goal for experiment two was to extend the integral/separable methodology into visual scenes of multiple similar objects and to determine any negative effects of multiple graphics on subjective perceptions of integralness.

The third experiment investigated the effects of integrally developed augmented scenes on learning and task performance. Three learning tasks and eight assembly tasks guided by augmented reality instructional tapes were constructed using either high or low integral methods as indicated by the extreme results for integral perception from experiment one. Experiment three's augmented reality scenes showed varying numbers of computer-generated graphics in both the learning and assembly phase. The main goal of experiment three was to establish that the theoretical constructs formed in experiments one and two made a practical difference in representative real-world tasks.

## CHAPTER 2. AUGMENTED REALITY

Augmented reality is commonly characterized as computer-generated graphics superimposed onto a natural scene (Barfield, Rosenberg, and Lotens, 1995). Each term in the phrase ‘augmented reality’ provides insight into what exactly is meant when discussing this technology. The term ‘augmented’ by definition is an addition meant to increase or sharpen understanding. The term ‘reality’ by definition is the state or quality of being real or of existing in fact. Thus, augmented reality deals with enhancing the understanding of events that are real and happening in the present time. An augmented environment, as opposed to reality, is also possible when the scene itself is a recording, a model, or some other computer generated environment. Augmented environments are not limited by what is real or existing in fact; that is, any visual scene could be augmented, real or otherwise.

In practice, given the state-of-the-art, augmentation typically takes the form of simple graphics, schematic overlays, complex graphics, or text that annotates a real environment. In particular augmented reality technologies are effective for tasks that benefit from being presented in three dimensions. Feiner, MacIntyre, and Sellgmann (1993) describe the future of augmented reality as an automated presentation design system customized for each user, task, and situation that dynamically responds to the real world by presenting appropriate knowledge-based graphic overlays. Already augmented systems have been applied to a large variety of research areas: maintenance, repair, entertainment, military, medicine, personal portable computers, architecture, manufacturing, tele-commuting, annotation, visualization, collaborative work, robot path planning, consumer design, and training (Azuma, 1997; Barfield, Rosenberg, and Lotens, 1995; Vallino, 1997). The actual amount of research conducted and type of augmented reality system used varies considerably per application domain.

### *2.1. Augmented Reality Systems*

In theory there are no restrictions as to which sensory system of the human body can benefit from enhanced input signals. Indeed, there are many such devices used every day that are not typically called augmented reality. Glasses improve poor vision, hearing aids enhance audio signals, and

artificial noses and tongues have been developed to recover sensory information. However, augmentation in the current context generally applies to additional information projected into the overall environment rather than improving the capabilities of an individual sensory system. Table 1, adapted from Barfield et al., 1995, shows example augmentation for each of the five main senses.

Table 1. Augmentation available for the five main sensory systems.

<b>Sense</b>	<b>Type of Augmentation Possible</b>	<b>Examples</b>
Vision	text, schematics, simple graphics, complex graphics, pictures, animation, video	Annotated text, wiring diagrams, navigation arrows, etc. This is the most common augmentation.
Hearing	simple beeps, signals, music, spoken words, dialogue	The sounds of a factory could be simulated while a worker trains in a secluded work area.
Touch	simple objects, environmental, complex objects, tools	Temperatures could be adjusted to reinforce actual environmental conditions.
Smell	simple, complex	Some poisonous gases are mixed with identifying odors. A sample smell could augment a reading on that poison gas.
Taste	sweet, sour, bitter	Artificial tastes could be provided for novice cooks as they view various foods.

In actuality the typical augmented reality system limits itself to the visual sense, presenting computer graphics superimposed onto the real world using a display. The two main methods that are used to create these systems are see-through displays and video-merging techniques. Figure 1 shows schematic diagrams highlighting the architecture for the see-through and video-merging techniques of augmented reality.

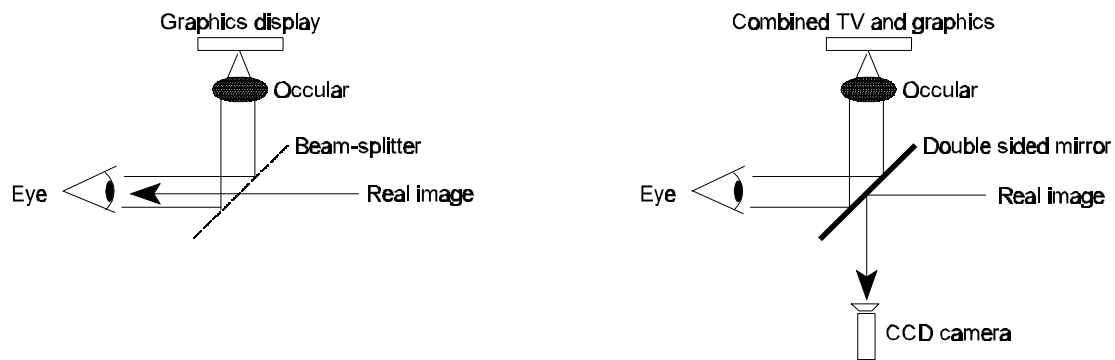


Figure 1. Schematics for see-through (left) and video merging (right) techniques (adapted from Barfield et al., 1995).

See-through displays use semi-transparent glasses or goggles to show computer-generated graphics that overlay whatever is in the real world scene. The semi-transparent display is usually manifested in a head-mounted device although other devices have been demonstrated. Rekimoto (1997) showed how a see-through hand held display could be used to view the world through a portable augmented window. Rekimoto envisions the device would be used as a palmtop computer, without the cumbersome head-mounted devices that are traditionally used. Figure 2 shows an example of a large prototype see-through head-mounted system.



Figure 1. A prototype head-mounted display for augmented reality applications (<http://www.wearcam.org/head-mounted-displays.html>).

While large unwieldy head-mounted displays are used in some research settings, the more common devices use far less hardware to accomplish similar see-through abilities. A typical semi-transparent head-mounted device used in augmented reality applications is the i-o Display Systems i-glasses shown in Figure 3.



Figure 3. The i-glasses see-through head-mounted display from i-o Display Systems (<http://www.i-glasses.com/>).

See-through displays show the real world directly rather than with a video signal, albeit with a slightly grayed out appearance due to the semi-transparent glasses. Adversely, the graphics in semi-transparent displays tend not to be as vivid and compelling as in the video-merging technique.

Video-merging methods combine a video feed of the real world and the computer graphics into one scene. As with the see-through method, the display is typically a head-mounted device, though, any display device could be used. Lion, Rosenberg, and Barfield (1993) developed a stereoscopic real-time augmented interface displayed on a regular computer monitor. The authors pointed out that the computer monitor provided crisper images than other display devices, such as head-mounted displays, which is still true today.

Figure 4 shows the immersive head-mounted display used in the current studies. It is shown with an attached video camera used to film the view directly in front of the user. This view mimics what the user would see if the opaque helmet was not there.



Figure 4. The immersive head-mounted display used in the present studies.

The video-merging technique for augmented reality allows graphics to be presented as solid objects rather than translucent images. This feature could potentially lead to more vividly and realistically perceived graphics in the video-merging technique over the semi-transparent method. However, unlike the see-through technique that allows users to view the real world directly, the

video-merging method requires that all visual information be converted to video signals. The users view of the real world would be limited to the resolution, brightness, field of view, and other display parameters of the viewing device. Many subtleties of the visually rich real world may be lost in such a display, which could alter a users perception of the augmented system.

In both augmented reality methods described above there are many limitations associated with the newness of the technology. Resolution, brightness, and field of view are only a few of the technological parameters that impact augmented reality research (Azuma, 1997; Barfield et al., 1995; Mckenna and Zeltzer, 1992). Table 2 lists several concerns and general implications for augmented reality along with possible affects they may have on integral perception.

Table 2. The impact of various technological augmented reality parameters.

<b>Technological Parameter</b>	<b>General Implications</b>	<b>Effect on Integral Perception</b>
Pixel resolution	In video-merged systems, the real view is limited to the resolution of the display. See-through systems maintain real-world resolution, limited only by the human eye. In both methods, the graphics are limited to the display resolution capabilities.	An “equal” resolution between real and artificially created objects may actually increase integral perception. However, less resolution overall may hurt other perceptions about the augmented system as a whole.
Brightness	As with resolution, the video-merged method limits the brightness of the real scene to the capabilities of the display. The semi-transparent visor used in see through systems causes two problems. First, because the visor is not opaque, graphics are translucent. Second, the real view is dimmed.	As with resolution, equal levels of brightness may aid integral perception, while overall perceptions of the system could be negatively affected. Translucent graphics may decrease integral perception.
Field of view	It is more difficult to correctly display images in wide field of view augmented reality systems, particularly see-through displays due to complex optics (Azuma, 1997).	A correctly imaged wider field of view would likely increase integral perception.
Registration	One of the largest technical challenges in current and near future applications for augmented reality is the correct registration of graphics with the real scene.	Unregistered graphics will greatly decrease integral perception.
Graphic realism	Magnificent computer graphics can be created; however, developing graphics that are visually and measurably indistinguishable from real-world images is still a great challenge (Greenberg, 1999).	More realistic graphics are expected to yield higher integral perception.
Depth resolution	Along with registration, the display of graphics at correct depth is one of the hardest technical challenges to be overcome. Currently, nearly all systems overlay graphics onto one plane in a real scene.	Graphics objects projected at correct depth should increase the integral perception of augmented systems.
Color	As with other parameters, color in video signals is limited by the display system. While most systems strive to include color, it is still an expensive addition.	Color graphics should enhance the integral perception of augmented reality systems over non-colored images.
Motion	Some applications require motion in the real scene and corresponding graphics. Motion itself requires animation with graphics, however, the largest impact is the complications to other technological parameters, specifically registration.	Augmented scenes with realistic motion should greatly enhance integral perception. However, incorrectly represented motion may have large adverse impact on integral perception.

Display technology is changing rapidly in the field of virtual and augmented environments, and Table 2 represents several main areas of interest; however, there are additional display factors that developers may need to consider. McKenna and Zeltzer (1992) provide a comprehensive list of display factors that affect virtual and augmented environments beyond the factors listed in Table 2 (e.g., refresh rate, jitter, and shading). Generally, like resolution and brightness, these factors are related to the fidelity of the display in representing the real and virtual world. Thus each of these factors, as dictated by the display used, are likely to influence the perception of the integrality of augmented reality systems. The choice of displays when designing an augmented reality system is not straightforward as numerous trade-offs are involved. The strengths and limitations of the various display systems need to be considered in conjunction with the needs of the task.

That augmented reality commonly utilizes head-mounted devices is not an artefact leftover from virtual reality research. In actuality what is considered the first 'virtual reality' system was a see-through augmented system (Sutherland, 1965). Whether used in virtual or augmented reality applications, head-mounted displays allow the participant to view the world by turning their head to items of interest rather than passively watching what is shown. Furthermore, for augmented systems the hands are free and visible along with the real world task. Virtual reality systems usually require some hand-operated input device. That is one of the key advantages of augmented reality applications: hands-free operation.

With reasonable portability and hands-free operation, augmented reality could serve a useful role in almost every facet of human interaction (conceivably any human interaction, but particularly with the applications listed earlier). Indeed some researchers are eager to show that a portable augmented reality system, a wearable computer, can be used at all times to help with everyday activities (Barfield and Baird, 1998; Starner, Mann, Rhodes, Levine, Healey, Kirsch, Picard, and Pentland, 1997). The interactiveness allowed by head movement and hands-free operation is what allows for an interesting discussion on using augmented reality in training. A worker using an augmented reality system has constant exposure to information that may be used for enhancing training.

Milgram, Takemura, Utsumi, and Kishino (1994) discuss augmented reality as it falls on the reality-virtuality continuum. Augmented reality is described as one of many mixed reality display systems. Mixed reality systems are those that have real and virtual objects juxtaposed to one extent or another. A system with more real items would be labeled augmented reality and systems with more virtual imagery would be labeled augmented virtuality. The ends of the continuum are completely real or graphic. Figure 5 shows the Milgram et al. reality-virtuality continuum.

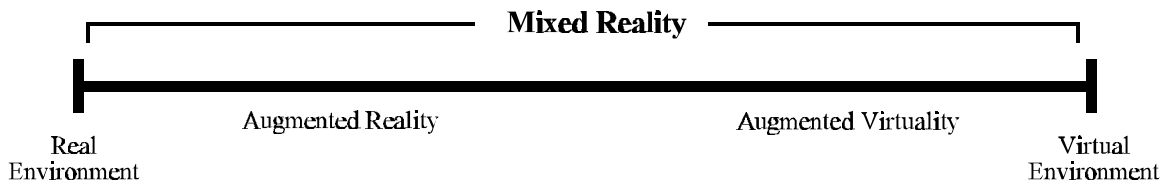


Figure 5. The Milgram, Takemura, Utsumi, and Kishino (1994) reality-virtuality continuum.

In this context, the term augmented reality is defined as “augmenting natural feedback to the operator with simulated cues.” Conversely, the term augmented virtuality refers to “augmenting simulated feedback to the operator with real cues.” Based on the reality-virtuality continuum, Milgram et al. (1994) further classified mixed and augmented reality systems, as shown in the adapted Table 3. The table first characterizes a mixed reality system by describing the source of the real and graphic signals in column one. The second column indicates whether the predominant visual information would be real or synthetic. The next column denotes whether the real view is direct line-of-sight or a scanned video signal. The last column signifies whether the user would be outside the real environment (exocentric) or be inside the real environment (egocentric).

Table 3. Classification of mixed and augmented reality.

<b>Class of Mixed Reality System</b>	<b>Real or Computer Graphics World</b>	<b>Direct or Scanned View</b>	<b>Exocentric or Egocentric Frame of Reference</b>
Monitor-based video, with computer graphic overlays	Real	Scanned	Exocentric
Head-mounted display-based video, with computer graphic overlays	Real	Scanned	Egocentric
Head-mounted display-based optical see-through, with computer graphic overlays	Real	Direct	Egocentric
Head-mounted display-based video see-through, with computer graphic overlays	Real	Scanned	Egocentric
Monitor/computer graphic world, with video overlays	Computer Graphics	Scanned	Exocentric
Head-mounted display/computer graphic world, with video overlays	Computer Graphics	Scanned	Egocentric
Computer graphic based world, with real object intervention	Computer Graphics	Direct or Scanned	Egocentric

The system used in the current work is a head-mounted display-based video, with computer graphic overlays. The visual scene is predominantly real, via a scanned video signal, and users are part of the environment rather than watching it from a distance.

A goal for the future, according to Milgram et al. (1994), is to eliminate users awareness of what kind of system they are using. As computer imaging techniques advance, a time will come where it will not be discernible whether the primary world is real or simulated, or whether the view is direct or scanned. Essentially, the ultimate in integral perception in augmented reality systems would be achieved. But, as Milgram et al., Greenberg (1994), and others point out, we are a long way from resolving the complex issues of perfectly integral augmented environments.

## *2.2. Augmented Reality and the User*

There has been minimal rigorous research in the field of augmented reality that focuses on how this emerging technology will interact with those who are using it. The role of the human is often only as a test subject to demonstrate the next new augmented reality system. The current research hopes in part to extend human factors experimental principles into augmented reality research so that the user in augmented reality is not overlooked. There are only a few efforts that have begun to consider the capabilities and limitations of the human in augmented reality research.

Feiner and various colleagues have presented a variety of demonstration augmented reality systems with some experimentation on user task performance. Their operational definition of augmented reality is the enrichment of the real world with a complementary virtual world. This enrichment is achieved without the replacement of the real world, rather supplementing by superimposing information into it.

The Touring Machine presented by Feiner, MacIntyre, Höllerer, and Webster (1997) examines a prototype mobile augmented reality system with a limited registration of graphics to real objects. In what was rather informal data collection, the authors themselves went through prescribed scenarios where they walked around the Columbia University campus using the augmented system with projected graphical information to learn more about the surrounding buildings. They wore a see-through display to view the real and synthetic visual information, used a handheld computer the size of a personal data assistant to interact with the system, and carried supplementary equipment in a backpack. The augmented information consisted of textual labels on buildings in addition to a semi-transparent image of a windows desktop operating system.

From the walkthrough scenarios the authors came up with several implications for the user of their system. The most critical one was that even a very small number of augmented graphics can create a good deal of confusion depending on the task. The non-integral graphics made it very difficult to multitask walking and using the augmented reality system. Even what was originally believed to be a relatively sparsely filled view of graphics interfered with the real view and

navigation. In proposed future work they intend to look at algorithms that move graphics around to avoid blocking real objects. However, before such actions are taken, an effort should be made to determine how many graphics are appropriate for one visual scene. This dissertation in part examines the issue of the appropriate number of graphics for one visual scene.

In more recent work, Webster, Feiner, MacIntyre, Massie, and Krueger (1999) studied subjects assembling a space frame of cylindrical struts with the aid of an augmented reality system. The augmented environment allowed clicking on certain pieces of the space frame to provide users with textual, verbal, and graphical information. Its purpose was to demonstrate the ability of augmented reality to be used as an assembly aid. Only preliminary information has been made available, and no informal or formal data collection has been reported.

A major contribution to augmented reality that has incorporated human factors principles was presented by Ellis (1998). Ellis conducted rigorous experimental research on the depth perception of virtual and real objects in an augmented reality setting. Ellis' work concentrated on augmented environments where near-focus was a primary concern rather than objects focused at large distances ( $> 2$  m). Ellis states that the original experimental concept was to examine calibration and alignment of displays, but the importance of perceptual phenomena was deemed more important considering the eventual development of near perfect calibrated systems.

In Ellis' (1998) work, subjects were tasked with moving a real pointer to match the depth of a virtual object in a variety of viewing conditions and augmented environments. Ellis reports a host of interesting significant results involving depth perception and augmented environments. The most interesting result to the current work may be an induced transparency that occurs in users wearing a see-through display when a virtual object is placed behind a real normally opaque object. The placement of objects causes a perceptual transparency in the real object that creates a heightened sense of integral perception. This effect may enhance integral perception, however, the objects perceptually lose their correct depth.

Ellis and others investigating depth perception in augmented reality (Drascic and Milgram, 1996) conclude that much more work is needed to correctly display virtual and real objects in correct

relational depth. The perceptual and physical challenges to positioning synthetic items in correct depth in augmented systems is beyond current technology in all but the most controlled conditions. There are certainly many interactions between integral perception and depth of objects in augmented environments, but the difficulties in producing correct depth precludes a meaningful investigation at this time.

Recent work by Baird (1999) and Barfield and Baird (1998) showed the benefits of using a wearable computing augmented reality system in a manufacturing assembly. Baird compared single-eye opaque and see-through augmented reality systems against traditional computer-display and paper-based instructional methods. Subjects assembled computer boards using the four instructional techniques. Baird found that both augmented instructional methods were superior to either the computer-display or paper-based methods. Subjects were much faster at assembly with the augmented systems and made less errors. Participants did report some usability problems with the augmented systems which included some of the technological parameters discussed earlier in Table 2 in addition to certain comfort issues. While no effort was made to make the augmented instructions integral with any object in the real scene, Baird did show experimentally that augmented reality can aid in manufacturing assembly.

The one well known success story in industry involving augmented reality and real users is the use of augmented environments to aid the manufacturing and assembly process of some Boeing airplanes (Caudell and Mizell, 1992). Boeing researchers describe their problem as an issue of figuring out how to successfully make 5 million parts fly in close formation. In an effort to aid this massive assembly process, Boeing began over a decade ago to experiment with the use of augmented reality.

Boeing planes are designed on CAD systems. In the past, the resulting CAD drawings were transferred to the assembly floor in a complicated set of assembly guides, templates, mylar drawings, wiring lists, and location markings on sheet metal (Caudell, 1994). What resulted was a large collection of mismatched media all relaying similar information that was difficult to decipher as the plane design adapted to its final form. Errors in design were extremely difficult to detect and changes propagated slowly through the engineering process. Computers became the natural

application for storing design information, and a “see-thru” virtual reality system was devised to display visual images to the workers as they stood at their assembly stations.

Informal sampling by Boeing has shown that the augmented reality system has brought substantial benefits to the assembly line, finding 20-50% speed improvement in divisions that incorporated augmented reality into the assembly work process (Nash, 1997). The success of the Boeing system was unknown for several years after the initial research endeavor until a recent surge of publicity showed that the system indeed was still being used successfully (Kaplan, 1997; Nash, 1997; and Nellis, 1997).

### ***2.3. Relation to Head-Up Displays***

Head-up displays are closely related in concept to augmented reality; indeed the definitions can be nearly identical. McCann and Foyle (1994) characterize head-up displays as superimposing symbology in the forward field of view. However, their full definition highlights the specificity of head-up displays in comparison to augmented reality: superimposing *instrument* symbology in the *pilot's* forward field of view. Head-up displays typically augment or replace the displays available in either a car or an airplane cockpit. The graphic objects are projected onto a plane near or in front of the windshield.

Differences between head-up displays and augmented reality lie more in the approach and application area rather than technical definition. Head-up displays are often meant to be used in place of regular physical displays to avoid the pilot or driver from removing their line-of-sight from the forward field of view. Human factors research in the field of head-up displays tends to deal with the display symbology and proving that head-up displays are more effective than traditional head-down instrument panels. Augmented reality prefers to enhance and add to a real scene rather than replace information. Despite the differences in approach, both use visual symbols and/or text to enhance a user's understanding of his or her situation.

A review of human factors research in head-up displays by Weintraub and Ensing (1992)

presented limited related work to the current research effort. Head-up displays are designed to be clearly separate from the real world and not integrally perceived. However, this issue of dividing attentional resources between synthetic symbology and real-world objects is relevant to both integral and purposely non-integral perception of synthetic objects and the real world.

Cognitive effort to discretely attend to one source and then the other may decrement the performance on the real world task whether using a head-up display in an aircraft or an augmented system for assembly. Preliminary and anecdotal head-up display research confirms this hypothesis. For instance, head-up displays are often turned off by pilots in critical missions because the non-integral head-up displays interfere with the task. A similar result could be expected in augmented reality applications. Graphics most likely would not be turned off in an assembly task, but a surgeon relying on integral graphics to aid an operation would likely abandon an augmented system if it was perceived to be a hinderance.

A major issue in head-up displays is the attentional demand that projected information may capture away from the main task. McCann and Foyle (1994) point out that superimposed graphics form a perceptual group based on their similar design compared to the rest of the real-world scene. It is theorized that the perceptual grouping of graphics captures attention from the more important real visual information. Furthermore, if superimposed symbology is organized into subgroups, then multiple perceptual groups could be formed causing serially processing of all the augmented information. McCann and Foyle state that head-up displays are supposed to reduce the amount of time needed to process instrument information and if attention is captured when instrument information is not needed, then head-up displays lose their effectiveness. The situation described is akin to non-integral graphics in augmented environments. It is hypothesized that integrally designed graphics would avoid this excess cognitive processing due to attention capturing graphic perceptual groups.

McCann and Foyle (1994) conducted a study where perceptual groups were purposely created in superimposed symbology. The clear results showed that when attention was needed to switch between perceptual groups, time to decide on target selection increased. One of the authors' proposed design solutions was to present virtual objects, instead of head-up displays, that appear

to be physically part of the real world. This concept is exactly what is meant by integral perception.

A secondary issue related to attentional demand in head-up displays is clutter (Stokes, Wickens, and Kite, 1990). Rather than stealing attention from the pilot, cluttered graphics or symbology can lose their effectiveness for enhancing understanding. Too much added information can overwhelm the real task, particularly when flying an aircraft. Stokes et al. emphasize the importance of using a few well designed synthetic displays over a broad collection.

Too many cluttered graphics can lead to obscuring of critical information. Head-up displays are generally designed to cover a minimal portion of the forward view; however, violations of occlusion are a common problem in augmented environments. Graphics cannot simultaneously be placed both in front and behind objects in the real scene using standard equipment. A single video mixing console allows only two levels of layering. Multiple video mixing consoles could be linked to effectively provide more layers, however, video quality would decrease at each stage without film studio caliber equipment. Alternate approaches involving graphic design techniques could be used to give the appearance that graphics are at multiple depths, but they too are difficult to implement and have their own limitations.

In other work, Wickens and Baker (1995) discuss how in some head-up display navigation tasks schematic figures surpassed photographic images in aiding navigation task performance. As was discussed in other head-up display research areas, the mission critical tasks of pilots and drivers demand synthetic additions to the real scene to be very appropriate for the task. This need to focus on appropriate graphics for the task is likely to occur in augmented applications as well. Some applications may not benefit from integral graphic stimuli, or benefit little over purposely non-integral projected graphics. The current work addresses this issue by testing a variety of augmented reality settings.

## 2.4. Relation to Virtual Reality

The term augmented reality is often referred to as a variation of virtual reality; technically, this is not the case. Referring to the reality-virtuality continuum of Milgram et al. (1994), virtual reality is a completely graphical world that users interact with. As has already been discussed, augmented reality is selected graphics superimposed onto the real world. In both cases, similar input and output devices are used to interact with either augmented or virtual environments.

Disagreeing with the distinction shown on the reality-virtuality continuum, Ellis (1994) views augmented reality as a form of virtual environments at a conceptual level. Ellis stated that virtual environments are a new communication medium for human-machine interaction that is concerned with the conveying of information. Ellis further described virtual environments as sensors, effectors, and special purpose hardware and software that links the human operator and some environment. Furthermore, the virtual worlds and their interactions with human operators are defined by their content, geometry, and dynamics. Figure 6 shows a schematic breakdown of a general virtual reality setup.

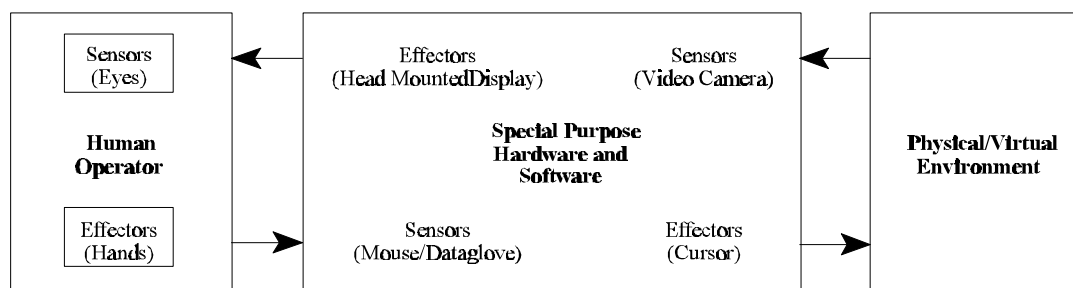


Figure 6. Ellis' (1994) schematic of virtual reality.

In the context described by Ellis (1994), augmented and virtual environments are of the same vein. Virtual, and to a lesser extent augmented environments, allow visual, auditory, and haptic

interaction modes totally unrealizable in purely physical environments. In actual practice, there are many differences between the two research areas.

The main difference, as the reality-virtuality continuum denotes, is that augmented reality always involves the real world in some fashion, while virtual reality utilizes a fully computer-generated environment. Virtual reality has received the bulk of the public interest, and therefore, has been more heavily researched. Consequently, there have been a variety of psychophysical research from areas such as sickness in virtual environments (McGee, 1998) to the illusion of being present in a virtual environment (Snow, 1996). Other than similar technological parameter studies, the second research area listed, presence, has the most intriguing connection to integral perception in augmented reality.

Consider again the reality-virtuality continuum of Milgram et al. (1994) in Figure 5. Integral perception is not limited to graphics being perceived as part of the real world (augmented reality); or vice versa, perceiving real items as belonging to a mostly graphic world (augmented virtuality). The concept of presence conceptually extends integral perception to include perceiving one's real self as being part of a 100 % artificial world. In essence, there is a psychological extension to the reality-virtuality continuum that includes psychological perception. The counterpart on the reality-virtuality continuum to the sense of presence, would be perceiving non-real objects to be part of a 100 % real world. Another psychological term, not associated with either augmented or virtual reality research that could fit this description would be hallucination.

Beyond the interesting thought extension to the reality-virtuality continuum, there is no reason to discuss virtual environments further. There are equipment overlaps common in the research, however, the applications for the respective endeavors are very different. Furthermore, a totally graphical environment precludes any real-time task performance in the actual world that requires hand-eye coordination which is a primary interest of the current work.

### CHAPTER 3. INTEGRAL PERCEPTION

Perception is an inferential process that is continually updated with new evidence provided by the senses. There are several levels to perception that depend on the characteristics of the task and user ranging from detection to identification and recognition (Sanders and McCormick, 1987). Many different cues contribute to the perceptual hypothesis of any given situation beyond simple sensory stimuli, including prior experiences, physiology, learned associations, and current context. Even a simple perceptual act causes a complex cognitive decision making process. More intricate perceptual activity, such as integral perception in augmented environments, leads to rich, diverse, and difficult to understand mental processing mechanisms (Prinz and Bridgeman, 1996).

Reducing cognitive processing is a commonly studied issue in human factors and psychology research. As stated earlier, one intention of augmented reality is to reduce the cognitive overhead associated with completing tasks. One classic area of psychology research that examines an issue of perception and cognitive processing intuitively related to augmented reality is the integral/separable work pioneered by Garner (1974) and Garner and Felfody (1970).

Garner and Felfody (1970) originally studied integral/separable dimensions to discover whether redundant information would aid tasks and orthogonal information would interfere with tasks. Redundant information is when two or more dimensions vary consistently together (e.g., the color and position of lights in a traffic signal), and whether orthogonal information is when two or more dimensions vary independently of one another (e.g., the number of signals at a traffic stop and the number of Ford vehicles present at any given time). From their work, Garner and Felfody discovered a much more intricate perceptual phenomenon than redundancy in user's reactions to separate and integral dimensions. Separable dimensions are defined as components or properties of a stimulus perceived as multiple discernable entities. Integral dimensions are defined as components or properties of a stimulus that are perceived as a single unified entity.

Barfield et al. (1995) present a simple schematic on how the perception of augmented reality as a single mental model may aid cognitive processing. Figure 7 shows a simple progression of perception from two separably perceived entities, the real and virtual world, to one integrally

perceived entity, which is augmented reality.

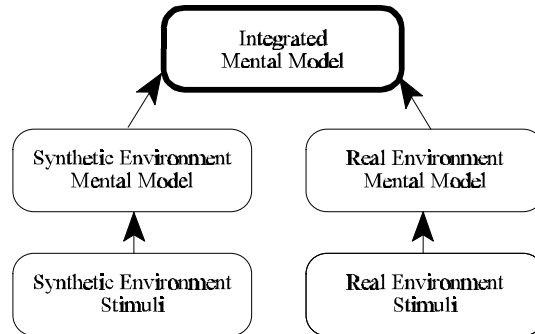


Figure 7. Mental model of augmented environments.

Whether users actually form one integrated mental model of augmented reality and the subsequent effect on performance is a primary research question of the current work. Furthermore, the degree of integral perception attained is sure to affect task performance in certain augmented environments more than others. Learning how to answer these and other questions in augmented reality will in part be accomplished by a review of Garner and Felfody's (1970) classic work and related research.

### ***3.1. Integral/Separable Research***

The theory behind Garner and Felfody's (1970) integral and separable dimensions work is that extra attentional and cognitive processing is needed to perceive separable dimensions. To test this premise, simplistic stimuli were used to try and track down the exact difference in processing. Figure 8 shows a example of a stimulus from Garner and Felfody's work.

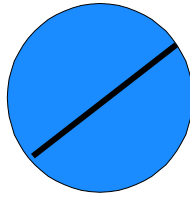


Figure 8. Example of integral/separable dimensions.

In Figure 8 there is a blue circle with a diameter line shown at a certain angle. The perceptual question asks whether the two dimensions are perceived as one unanalyzable object or two. Clearly in this case there are two distinct components; the line and the colored circle. However, the color of the circle is likely perceived as just blue and not with a dimension of hue and a dimension of brightness. Consequently, according to the theory, tasks involving some level of perception should take longer to do with the separable dimension objects like Figure 8 than non-separable objects (consider just the blue color). Indeed, Garner and Felfody's (1970) work showed just that.

Unfortunately the stimuli in Garner and Felfody's (1970) work and most subsequent research has been limited to extremely simplistic target object dimensions such as color, shape, and line angles on paper note cards. This research goal was to establish the theory rather than to test the ideas in a more complex environment such as augmented reality. This bodes well for pure psychology research, but areas dealing with more complicated objects, like augmented reality, are left with very applicable research ideas and very un-applicable research methodology.

The common experimental task for integral/separable research has been speeded judgements of similarity or simple classifications based on characteristics of the stimuli presented. The judged similarity of stimuli and method of classification is supposed to determine whether the dimensions of the test stimuli are perceived integrally or separably. Various subjects may judge targets differently on the continuum that exists between separate and integral, but whatever level is used by each individual subject does not change.

Based on their results, Garner and Felfody (1970) concluded that integral dimensions lead to a

speed and accuracy gain and non-integral dimensions lead to interference, in some cases. However, not all of Garner and Felfody's work produced clear results. Some dimensions expected to be perceived as integral were not, and the perception of other dimensions produced inconclusive results. They concluded that it is difficult a priori to determine even with simple dimensions which will be perceived separate and which integral. This suggests that complex stimuli will be even more difficult to predict beforehand whether they will be integrally perceived.

The body of research that followed Garner and Felfody's (1970) classic integral/separable work investigated a wide range of stimuli and conditions. Shepard (1991) reviewed over 25 years of integral/separable research with the goal of converging various experimental evidence into one theoretical basis.

Shepard (1991) detailed how integral/separable dimensions became a highlight in the ongoing psychological debate on how perceptual processing occurs in humans. Integral/separable dimensions proved to be a phenomena that could demonstrate two different types of perceptual processing metrics. Various psychophysical research showed that most perceptual processing occurred on a continuous scale (Gescheider, 1985). Integral processing appears to follow the general schema for perceptual scaling. However, separable processing shows discrete steps of processing, necessitated by the distinct dimensions, rather than a continuous scale. Shepard explains how much of the integral/separable work over several decades focused on the metrics of integral dimensions rather than advancing the research methodology of integral perception (Foard and Nelson, 1984; Garner, 1976; Gottwald and Garner, 1975; Grau and Nelson, 1988; Handel and Imai, 1972; Nosofsky, 1987; Shepard, 1986; Shepard, 1964) .

After the long emphasis on perceptual scaling, a few studies began to apply the integral/separable theories to modern-day applications. As was mentioned, most integral/separable studies used simple tasks and object dimensions, however, Jacob and Sibert (1992) devised a more complex experimental stimuli that involved interactive graphical manipulation.

Jacob and Sibert (1992) attempted to match input devices to tasks based on their respective control and perceptual structure. The phrase perceptual structure refers to the aforementioned

perception of stimuli as having either integral or separable dimensions (or structure). The phrase control structure in Jacob and Sibert's work referred to whether an input device acts integrally or separably in Euclidean space. Specifically, input devices that can move from any point in space to any other in a straight line, are integral (e.g., a three-dimensional input device, a three-dimensional tracker), while devices that require a stair-step pattern to travel from one point to another are separable (e.g., a two-dimensional input device, a mouse).

Jacob and Sibert (1992) found that task performance was best when perceptual and control structures matched. The most salient result was a condition where only two dimensional point to point movement was needed. This condition avoided the confound of designing a method for the mouse to move a cursor in three-dimensional space.

In the two dimensional movement tasks, the mouse performed better than the three-dimensional tracker. The extra movement dimension of the tracker apparently interfered with the movement of the other two dimensions. While these results were interesting, the larger contribution was the extension of integral/separable methodology beyond simplistic stimuli and tasks.

Barfield, Sanford, and Foley (1988) extended the simplistic visual stimuli of earlier integral/separable research into realistic computer-generated graphics similar to those used in augmented reality. Realistic graphics are created by adding light sources, surface shading, texture, shadows, color, etc., to basic shapes. One of the research questions was whether realistic graphic objects were perceived as having perceptually distinct components. Integral perception of graphic objects would increase their perceptual realism as well. Barfield et al. aimed to determine which realism cues would be perceived integrally in a whole graphic image.

Barfield et al. (1988) used mental rotation tasks based on the Shepard and Metzler (1971) mental rotation tests, and subjective realism judgements to assess whether more realistic graphics aided task performance. The results showed that realism does aid task performance, however, there appears to be diminishing returns for added realism beyond a certain point. There was not any apparent interference by separably perceived realism cues as task times did not appreciably decrease for any additional realism cue. The results supported the hypothesis that realism cues

are perceived integrally in graphics.

While the integral/separable methodology has begun to extend beyond simple tasks (Jacob and Sibert, 1992) and stimuli (Barfield et al., 1988), no study has attempted to apply the integral/separable rationale to as complex an environment as augmented reality or as an applied task as assembly.

### ***3.2. Related Perceptual Research***

While integral and separable dimension work, primarily from Garner and Felfody (1970), provided the inspirational research interest for integral perception in augmented reality, a broader examination of related perceptual issues is necessary. The concept of augmented reality fits the philosophy of integral/separable research very well, however, other perceptual notions are also relevant to the perception of augmented reality as one unified visual scene.

The first related perceptual concepts to be discussed are far more ‘classic’ than the separable/integral research discussed earlier. The Gestalt laws of perceptual organization are century-old psychological constructs that consider the “separation of nonhomogeneous fields” (Kehrer and Meinecke, 1996). Fundamentally the laws examine whether visual configurations are perceived as holistic entities from the specific visual properties of the related parts. Like integral/separable research, the demonstrations of Gestalt laws are generally conducted with simplistic representations involving dots or line drawings. Of the various Gestalt laws, two are related to the current investigation of integral perception in augmented reality.

The first is the law of similarity which states that similar stimuli are perceived as belonging more to each other than dissimilar ones. There are two facets to this law that are specifically related to augmented realities composed of multiple objects. If an augmented reality scene contains multiple real and similar computer-generated objects, then they may form a perceptual grouping that should aid integral perception. On the other hand, if a visual scene contains multiple real and graphic objects where there are physical characteristics or anomalies of the graphic objects that draw attention, then these objects are likely to form a perceptual grouping that would hinder

integral perception.

The second law considered in this dissertation is the law of good Gestalt or Prägnanz. This law of goodness states that the more interrelated the elements of a whole are to each other and the more weakly they are related to outside elements, the more the whole is seen as good (Hoffmann, 1996). That somewhat vague description is often utilized in research by not defining for subjects what exactly goodness is supposed to mean. Subjects are left to develop their own construct of goodness when they view a visual scene. The current dissertation utilizes the goodness law in this way by not specifically biasing the subject on how to rate goodness of augmented scenes.

The next related research area to integral perception is recognition-by-component parts. The basic premise is that in any scene there are component parts that lead to an overall recognition of the whole (Biederman, 1987). This idea accepts that there is no visual scene of any complexity that will be integrally perceived as one unitary object as defined in classic integral/separable research. However, the overall scene can be recognized as one whole, that happens to be made up of individual parts. In essence, integral perception becomes a hierarchy of parts and wholes in so-called 'partonomies' (Hoffmann, 1996).

Biederman (1987) tests the concepts of recognition-by-components with intuitively meaningless do-it-yourself objects and partially drawn objects. Do-it-yourself objects are contrived pictures of unrelated geometric components such as blocks, cylinders, wedges, and cones. The assumption of the theory is that detectable edge properties in the do-it-yourself objects lead to pattern recognition, where meaning is attached to the whole of the senseless connection of parts.

With partial objects, Biederman (1987) discusses how perception can readily fill in missing information, or even recover from certain incorrect information in a visual scene. Again the emphasis is on detectable edge properties such as curvature, collinearity, symmetry, parallelism, and co-termination. Images, especially those generally included in augmented realities, are not usually explicit line drawings with obvious edge properties. Biederman responds that any image with recognizable shapes and objects has some edge properties. Only a completely uniform colored scene would be devoid of all edge properties. In addition, Biederman states that texture

properties such as color, brightness, and repeating patterns are sure to also contribute to the overall recognition of parts as a whole.

The ability to overcome deficiencies in a visual scene is a powerful element of the recognition-by-components theory. Do-it-yourself objects are purposely designed with senseless parts to see how much meaning is attached to the whole. It is assumed in augmented applications that the computer-generated graphics add meaningful information and content to the overall visual scene. Perception of the visual scene as whole or integral would likely be affected by the extent that subjects perceive the superimposed graphics as adding useful and appropriate information to the overall visual scene.

All conceivable augmented reality applications would fit the various tenets of the recognition-by-components theory. No combined computer-generated graphic and real image would truly be viewed as integral, however, through the process of integral perception, the scene could be perceived as a unified whole. With this logic, it would be expected that no augmented scene would be rated as perfectly integral, though, certain augmented scenes would be perceived as more integral than others. This dissertation examines this supposition.

Perception is closely related to several other human processes, among them are sensation and attention. In both of these areas consideration is given to fields of study related to integral perception. Parallel processing is a research topic in the area of sensation that seems to have related impact to integral perception. However, parallel processing is generally concerned with the ability of the visual system to, in parallel, transmit information to the brain (where subsequent perception would occur) (van der Heijden, 1996). A number of physical stimulus manipulation experiments have been conducted to investigate this phenomena, but they have all stopped short of addressing perception.

While sensation is a vital precursor to perception, attention is much more important for what is actually perceived. All theories of attention and perception have at some level related attention to conscious perception (Neumann, 1996). Even perceptual theories that do not claim to be conscious acts, like integral/separable perception, include an element of selective attention.

Rosch and Lloyd (1978) discussed the role of selective attention in integral perception. They pointed out that in order for dimensions to be perceived separately, they must first be selectively attended to, consciously or otherwise. That is, separable perception cannot occur without selective attention. However, integral perception can occur whether dimensions can be selectively attended to or not. Rosch and Lloyd allude to the underlying continuum of integral perception that would in part be dictated by the degree of selective attention possible within individual stimuli.

Related to the idea of selective attention is context of a scene. The visual makeup of a complex scene, like those typically used in augmented environments, is sure to influence what dimensions or objects a subject would attend to. The degree to which subjects focus on computer-generated graphics or real objects could influence integral perception.

Hoffmann (1996) discusses several aspects of context that can influence object or scene recognition. An object is rarely presented in absence of all other objects in our everyday viewing of the real world. In most augmented reality applications, that would certainly be the case. Hoffmann's primary assumption is that global features of a scene influence perception. This is in contrast to the earlier discussion that parts are hierarchically perceived until a whole is understood. Hoffmann states that the two styles of perception will be influenced by each other in the overall perception of a visual scene. Furthermore, global recognition of a visual scene can aid the uniform perception of individual objects within the scene. Therefore, the degree that subjects recognize an augmented scene should influence whether computer-generated graphics are viewed as integral to the whole scene.

The last related perceptual research area to be discussed concerns the realism of computer-generated graphics. The previous topics in this section focus on the psychological phenomena associated with a particular perceptual action. The stimuli typically studied in experiments in these areas are used only to propagate the premises of each specific theory. Unlike these other topics, the perspective considered by Greenberg (1999) on realistic synthetic images does not focus on intricacies of the perceptual phenomenon, but rather on methods to create the most realistic stimulus.

Greenberg (1999) states the ultimate goal is to create synthetic images that are measurably indistinguishable from real-world images. One of the consequences of reaching such a goal is the use of simulations to predict user performance in reality. While not specifically mentioning augmented reality, Greenberg alludes to such a simulation as being the perfect testing ground for realistic image creation and integral perception.

Three main research areas are discussed by Greenberg (1999) that need to advance to reach the stated goal. Being a specialist in computer graphics, Greenberg first mentions the difficulties of modeling correct lighting conditions in computer graphic design. Then, in a shift of focus, Greenberg details how a greater understanding of the human visual system is needed. Lastly, Greenberg discusses the last “underappreciated yet important” aspect of image synthesis, the conditions under which the image will be viewed and the perceptual characteristics of the observer. Greenberg’s analysis shows that the importance of integral perception is not overlooked in such a partisan focus as the generation of computer graphics.

Prinz and Bridgeman (1996), and Sanders and McCormick (1987) characterize perception as being a rich, diverse, and complex field of study. This description certainly fits the multifaceted phenomenon of integral perception as discussed here. Augmented environments appear to be an excellent task environment to continue the rich investigation of integral perception.

## CHAPTER 4. TRAINING

Training in augmented environments can be useful for physical tasks where it is desirable to avoid the dangers of the real world (e.g., hitting a baseball or performing a surgery) (Barfield et al., 1995). However, one aspect of training, on-the-job training, seems particularly applicable for potential manufacturing uses of augmented reality training. Assembly task training is the context of the task environment for the third experiment presented in this dissertation. While training is not specifically manipulated or tested, a brief review of some the main concepts of training will be presented for context.

Training is defined as the systematic acquisition of skills, rules, concepts, or attitudes that result in improved performance in another environment (Goldstein, 1993). Generally, the first stage in training is learning the lower level basic tasks, the second stage is putting together the basic tasks into meaningful patterns, and the third stage is understanding when and how to use the meaningful patterns of basic tasks (Swezey and Llaneras, 1997). Presumably each stage would need to be carried out to have a trained worker that successfully learns the whole real task. To achieve this, a training program would need to select appropriate training tools to match each stage of learning. Many methods and technologies can be used to implement a training program; lectures, books, videos, and simulators are common examples of training media. Augmented reality could theoretically satisfy each of the stages of learning.

Augmented reality has shown to be effective for medical training (Kancherla, Rolland, Wright, and Burdea, 1995), but the use of most systems implemented thus far does not meet the definition for training. Augmented reality systems are generally not intended to alter performance in another environment but the currently existing one. However, augmented reality can be used as an on-the-job training tool.

Traditional training follows a linear progression from training, to transfer of training, to actual task performance as depicted in Case 1 of Figure 9. In actuality, training does not end with simulator practice or lectures in the classroom. Goldstein (1993) states that absolute task learning cannot be completed in training alone; on-the-job experience must also be obtained to fully learn

the intricacies of the job and task.

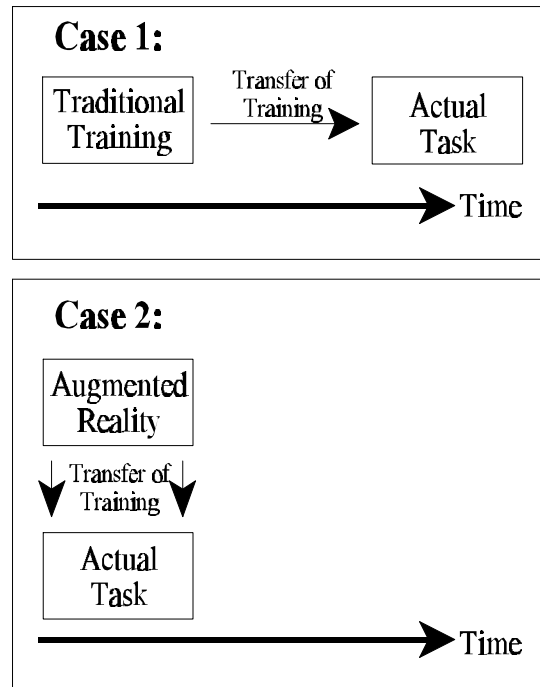


Figure 9. Representation of transfer of training for traditional training methods and augmented reality as an on-the-job training tool.

On-the-job training is the name given to the supplementary training that inherently must continue once the job starts. On-the-job training is usually an informal process meant to finish off the learning that did not occur during formal training. This informal approach is the main argument against using on-the-job training as a primary training source (Goldstein, 1993).

Goldstein continues by stating there are several advantages to the untested potential of on-the-job training. High on the list is the ultimate in high fidelity, the workplace itself, rather than a representation of the workplace. In addition to the actual workplace, the exact behaviors and procedures that will be used on the job can be used during training. Finally, actual work would be taking place. The parallel relationship of augmented reality training and actual task performance is depicted in Case 2 of Figure 9.

In fact, Goldstein (1993) foresaw that technologies like augmented reality would be used for just

such a purpose. Goldstein gives a brief description of embedded training describing technologies that are integrally involved with the task equipment itself, perhaps even aware of augmented technologies when writing it: “An example of embedded training is an office machine with graphic displays that show where a problem has occurred and then lists the steps (thus training the worker) needed to remedy it” (Goldstein, 1993, p. 230).

On-the-job augmented reality training is not the right choice for all work situations, however. None of us would want to take a flight with a pilot that has never flown before, even if he or she is equipped with an augmented reality training system. However, on-the-job training using augmented reality has the potential to greatly alter the way training is conducted in many other situations. The greater issue is therefore figuring out what types of tasks are appropriate for augmented reality training, and whether integrally or separably perceived augmented worlds will affect training performance. Intuitively it would be desirable to have an augmented reality system that is perceived as integrally as possible for an effective training paradigm.

## CHAPTER 5. EXPERIMENT ONE

### 5.1. Methodology

#### 5.1.1. Experimental Design

The first experiment was an investigation into the effects of different manipulations in graphic design on the perceived integration of graphics and reality in single object augmented environments. Three independent variables and eight dependent measures comprised the main components for an all within-factor design. Supplementary analyses were made into the relationships among the eight subjective ratings and correlations among the various experimental parameters.

The three independent variables consisted of texture (color, grayscale, solid white, wireframe), transparency (solid, transparent), and augmented object (computer, brain, text, fluid). They were factorially combined to form a 4x2x4 design (32 unique stimuli). The 32 unique scenes were shown twice to each subject for a total of 64 trials, each presented in random order.

The texture levels were chosen to provide a wide range of realism given the capabilities of the augmented reality system. The augmented object stimuli represented augmented reality scenes that might be encountered in real world usage of augmented reality and theoretical combinations of graphics and real objects. The transparency level was set to approximately imitate how a see-through system would operate.

Eight subjective dependent measures were collected from participants for each trial. The questionnaire (shown in Appendix A) used to gather dependent measure data for the main experiment was based on eight 7-point rating scales. These rating scales encompassed a variety of questions about the augmented scene related to integral/separable research and associated perceptual areas. In summary, the ratings were on goodness, separable, integral, distracting, distinguishable, blending, usefulness, and unified scene. Specifically, the goodness question stems from the Gestalt law of goodness; the separable, integral, distinguishable, and unified questions

are directly from integral/separable research; the distracting and usefulness questions are based on the recognition-by-components theory; and, the blending question is a combination of ideas from realism in graphics and integral/separable research. Table 4 summarizes the eight dependent measures as they were defined for experiment one. Appendix A shows the exact rating scale questionnaire used by the subjects to assess each trial scene in experiment one.

Table 4. Subjective rating scale descriptions for experiment one.

<b>Variable Description</b>
What is the goodness rating of the augmented scene (1 high, 7 low).
How separable were graphics from the real scene rating of the augmented scene (1 very, 7 not at all).
How integral were graphics with the real scene rating of the augmented scene (1 very, 7 not at all).
Were graphics distracting rating of augmented scene (1 very, 7 not at all).
How distinguishable were graphics from real scene rating of the augmented scene (1 very, 7 not at all).
How well did graphics and real scene blend rating of the augmented scene (1 very, 7 not at all).
Were the graphics useful to the overall scene rating of the augmented scene (1 very, 7 not at all).
Unified or distinct perception of whole scene rating of the augmented scene (1 unified, 7 distinct).

### *5.1.2. Participants*

Ten participants, aged 24-37, came from the general Virginia Tech university population. To be eligible for the study, participants needed to correctly perceive at least 7 out of 8 items on the 20/40 line of an eye chart and correctly identify 7 out of 8 color vision trials. No participant failed to meet these requirements.

### *5.1.3. Experimental Apparatus*

5.1.3.1. Equipment Setup. The equipment setup to run the augmented reality portion of the experiment was comprised of a VCR, video mixer, head-mounted display, video capture card, and two monitors. The video cassette recorders were Panasonic models AG-7500 and 7500A.

The video mixer was a Videonics MX-1 digital video mixer. The head-mounted display was a Virtual Research Systems, Inc. V8. The video capturing was achieved through a Hauppauge Win/TV-Cinema Pro digital video board that allows “live video” to be displayed on a PC VGA monitor. The Hauppauge card was installed in a Dell Dimensions XPS M200s PC. The two monitors used to view the augmented environments by the experimenter were a Sony Trinitron color TV and Sony color video monitor. Figure 10 shows two digital pictures of the equipment setup. Figure 11 is a closeup digital picture of the V8 head-mounted display. Figure 12 shows a schematic for how the equipment was connected.



Figure 10. The equipment setup for the three experiments.



Figure 11. The Virtual Research V8 head-mounted display.

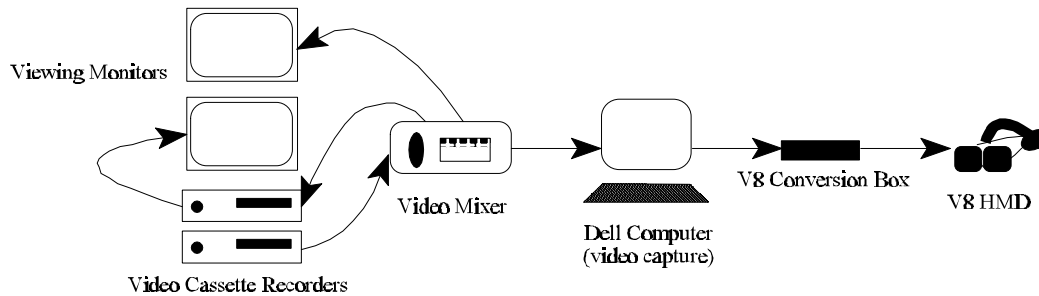


Figure 12. Schematic showing equipment setup.

The arrows in Figure 12 represent video signals being transferred from one part of the system to another. Converting the video signal to correct input formats was one of the larger challenges of setting up this augmented reality system. In summary, the V8 head-mounted display accepted only a standard monitor 15-pin video input. The video mixer had input and output in only RCA or S/VHS video formats. The rest of the system was used to convert or view signals as appropriate. Beyond the different formats required at various stages, all video signals needed to be limited to a 640x480 resolution and 60 Hz to be viewed correctly in the V8 head-mounted display.

5.1.3.2. Augmented Environments. The equipment used to create the augmented environments consisted of all the equipment used to run the experiment plus a Hyperconverter scan converter, an Intergraph PC, a model No. R-1084 Ricoh video camera recorder, PaintShop Pro Version 5.01 Image Editor software, and Sense8 WorldUp 3D Content Authoring Tool software.

The additional complexities that are not shown in Figure 10 were an extra computer used to develop the graphics (Intergraph) and a scan converter. The computer output standard monitor 15-pin video that was converted into S/VHS format by the scan converter before being transferred to the video mixer. Real scenes were fed directly into the video mixer from the video camera with RCA format.

The stimuli for experiment one represented augmented reality scenes that might be encountered in real-world usage of augmented reality and theoretical combinations of graphics and real objects.

Graphics were created using a 3D modeling program or images were found on the World-Wide Web (WWW) and modified for appropriate use with image editing software.

The real objects in the visual scene were a partially disassembled computer, a seated person shown from an overhead view, a partially printed sentence on a plain piece of paper, and a dynamic fluid motion model. The graphic objects were intended to be meaningful additions to the real visual scene. The partially disassembled computer had a graphic outline over most of its components, representing a maintenance augmented reality task. The seated person had an image of a brain superimposed over the skull, representing a medical augmented reality task. The partially printed sentence was completed with graphic letters and words, demonstrating text augmented capabilities. Lastly, one graphic globule was added to the dynamic fluid motion model, demonstrating dynamic action in augmented reality.

The real objects were video-taped and the graphics were then superimposed with the use of a chroma key feature in the video mixer. Chroma keying allows the removal of a color or set of colors in one image to be replaced by visual information from a second image (commonly used in weather forecasts). The final tapes used in experiment one were composed of 64 twenty second clips presented in random order. Figures 13 through 18 show six example images from the first study.

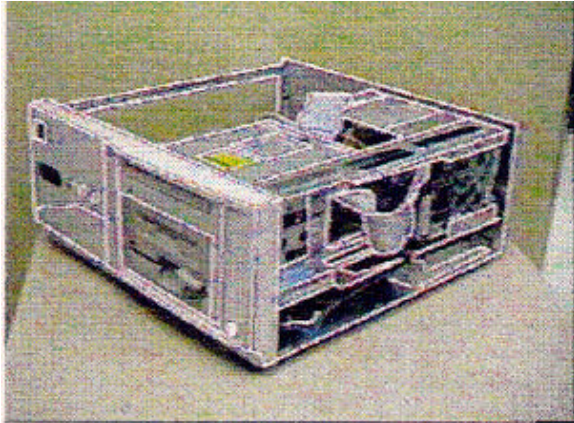


Figure 13. Computer-solid-white condition for experiment one.



Figure 14. Brain-solid-wireframe condition for experiment one.

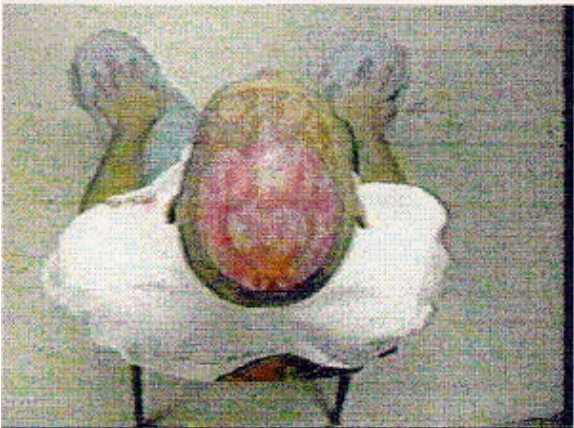


Figure 15. Brain-transparent-color condition for experiment one.

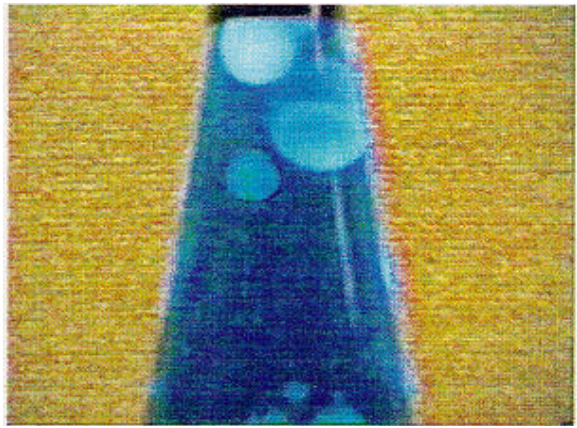


Figure 16. Dynamic-transparent-color condition for experiment one.

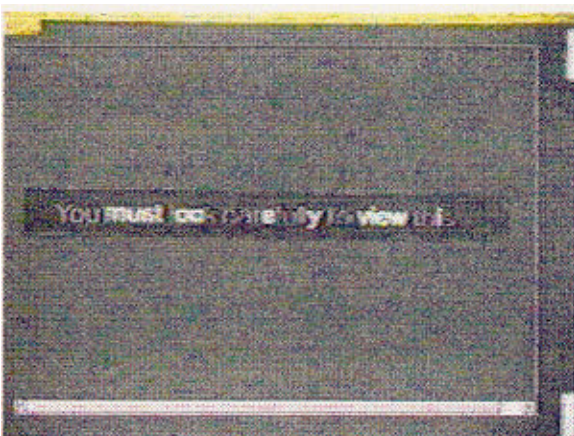


Figure 17. Text-solid-gray condition for experiment one.

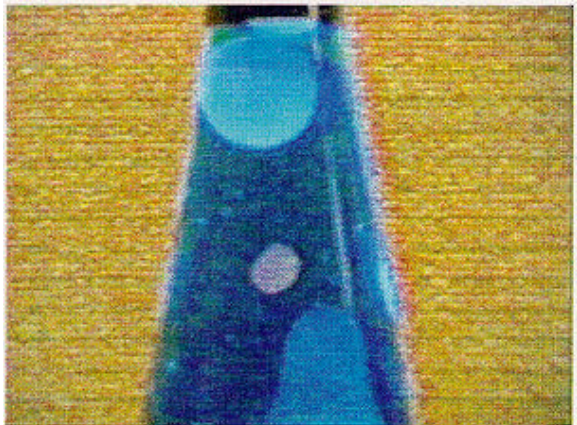


Figure 18. Dynamic-transparent-white condition for experiment one.

5.1.3.3. Questionnaires. There were two questionnaires used in the first experiment: a general information questionnaire and a mental rotation test. The general information questionnaire asked questions on the participants age, gender, handedness, computer usage, computer game playing, and a self-assessment of mental visualization and rotation ability. This questionnaire was administered to determine if any individual differences in subjects would consistently relate to the dependent measures. The subjects in all three experiments for this dissertation completed the same universal general questionnaire, shown in Appendix B.

A mental rotation test was administered to determine if a paper-and-pencil spatial ability test could be used to predict tendencies for integral perception. The mental rotation test used was based on the work of Shepard and Metzler (1971). Shepard and Metzler (1971) developed three-dimensional line drawings that were used in discrimination comparison tests. An example test stimuli is shown in Figure 19. The stimuli were adapted to a computerized test form by Vandenberg and Kuse (1978), and later revised by Crawford and Christensen (1995) to a written test. The Crawford and Christensen (1995) version of the test was used for this experiment.

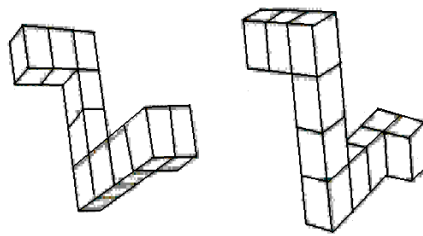


Figure 19. Example mental rotation test stimuli.

The Crawford and Christensen (1995) test has 20 trials that assess the ability of participants to mentally rotate three-dimensional line drawings. Participants are shown one primary object and asked to identify the same object in a set of four other objects. In the set of four secondary objects, two are the same as the primary (rotated slightly in one or the other direction) and two are different drawings entirely. The test is scored based on the number of correct and incorrect answers.

5.1.3.4. Vision Tests. Two vision tests were used in all three experiments: a visual acuity test, and a color blindness test. The visual acuity eye test was a Good-Lite Co. Illiterate “E” 20 foot equivalent eye chart. This test is a series of horizontal rows of variously oriented capital letter “E’s” that are viewed from ten feet. The E’s are presented in either the up, right, down, or left orientation. There are 10 lines of progressively smaller E’s representing visual acuity from 20/200 to 20/16.

The color blindness test was a pseudo isochromatic color vision test based on 16 plates in a 13 ring binder. Figure 20 shows an example of one of the test stimuli. The plates within the binder itself are precisely made to have the exact color and brightness levels to correctly perform the test.

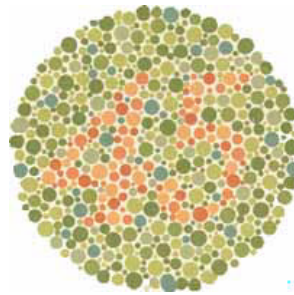


Figure 20. Example color vision test plate.

#### *5.1.4. Experimental Procedures*

Participants were first greeted and asked to read and sign the consent form (universal to all three studies, shown in Appendix C). Upon agreeing to participate, subjects completed the general questionnaire, the mental rotation test, and the two visual tests. The main portion of the study, the subjective assessments of augmented scenes, was then explained.

First, the nature of the experiment was described verbally by the experimenter. Participants were told that they were going to make eight subjective ratings for 64 video clips of pre-recorded augmented object scenes. The participants then read the subjective ratings sheet (see Appendix A) that was to be used to make the assessments. Acknowledging that they understood the ratings, the participants were then shown a demonstration video clip by the experimenter of four example scenes from the study. The participants were asked to practice making judgements

“within their head” and ask any questions. Lastly, the use of the head-mounted display was explained and fitted to the participants head. The subjects then began viewing the 64 augmented video clip scenes and making subjective assessments. The experimenter paused the videotape after each 20 second clip to record the subjective assessments. At this time, the participant verbally expressed subjective ratings to the experimenter after lifting the front of the head-mounted display to view the subjective ratings sheet. Any questions participants had during this or any portion of the experiment were answered by the experimenter. Upon completing the subjective assessments of the augmented objects, the participants were informed the experiment was over and thanked for their time.

## ***5.2. Expected Results***

Experiment one was designed to delineate between differently designed augmented object scenes with a variety of integral perception related measures. It was hypothesized that the solid-color graphics would be judged ‘best’ and the transparent-wireframe augmented scenes the ‘worst’ across the dependent measures. However, several interactions involving the augmented scene independent variable were expected to show tendencies for different best and worst case perceptual ratings. For example, solid-white graphics were expected to be judged better than other textures for the computer outline and text augmented conditions. In other univariate results, practice and order effects were not expected to reach significance as the presentation of the 64 augmented scenes was randomly ordered for each subject.

The multivariate factor analyses of the relationships between the dependent measures were expected to show specific clustering of dependent measures that are perceived by subjects as being similar. Furthermore, the factor analyses were expected to show that select pairs of subjective assessments are rated in an opposite fashion. For example, the integral rating value was expected to be directly opposite in scale of the separable rating for the same stimuli. It was also expected that certain dependent measures would dominate the calculated analysis factors that explain the variance in the data. It was hypothesized that subjective ratings that come closest to a true measure of integral perception would dominate the factor analysis summary graphs. Lastly, the calculated analysis factors were expected to result in easily interpretable combinations of

factors that may characterize the data better than individual measures.

The last analyses to be presented in the results section, the correlations, were expected to show several significant relationships. The dependent measures were purposely designed to ask similar questions, therefore, several significant correlations between them were expected to result. In addition, the scores on the mental rotation test were expected to closely relate to subjects self-assessment of their ability to mentally image objects. Other significant results were likely; however, no other specific relationships were purposely designed for in this experiment.

Beyond the statistical information, it was expected that the results for experiment one would lead to several implications for design. These implications were expected to deal with human factors, perception, and graphic design guidelines for integral perception in augmented reality.

### **5.3. Results**

Arguments can be made for considering the dependent measure rating scales as either ordinal or interval scale data, thus affecting whether non-parametric or parametric ANOVA tests should be used for evaluation (Keppel, 1991; Meister, 1985; Siegel and Castellan, 1988). The intent here is not to claim the use of one test is more appropriate over another, rather, the purpose is to adequately and accurately describe the summary data. Therefore, parametric ANOVA tests are presented in conjunction with related non-parametric tests. To further limit the discussion to robust effects, statistical results in the two sets of three-factor within subjects analyses in experiments one and two only were considered significant with p-values < 0.01.

All graphs and figures presented throughout all the results sections show significant effects with trend lines. The trend line graphs were chosen to highlight the rate of change between various levels of variables in significant results. Any significant effects with a p-value < 0.05/0.01 as indicated are shown with **bold** highlighting in the summary tables.

### 5.3.1. Trial and Order ANOVAs

Two different analyses were conducted to investigate any order effects that may have been present in experiment one. One-way ANOVAs were conducted for the overall order (trials 1 through 64) and the scene presentation order of each individual augmented scene (first and second). Each of the eight dependent rating scale measures were tested. Tables 5 through 12 show the ANOVA summary tables for overall order as measured by the eight questionnaire ratings (goodness, separable, integral, distracting, distinguishable, blending, usefulness, unified scene).

Table 5. ANOVA summary table for overall order, measured by a goodness rating.

Source	DF	SS	MS	F	P
order	63	180.311	2.862	0.90	0.701
Error	576	1840.300	3.195		
Total	639	2020.611			

Table 6. ANOVA summary table for overall order, measured by a separable rating.

Source	DF	SS	MS	F	P
order	63	219.823	3.489	0.88	0.722
Error	576	2271.300	3.943		
Total	639	2491.123			

Table 7. ANOVA summary table for overall order, measured by an integral rating.

Source	DF	SS	MS	F	P
order	63	150.100	2.383	0.62	0.990
Error	576	2213.000	3.842		
Total	639	2363.100			

Table 8. ANOVA summary table for overall order, measured by a distracting rating.

Source	DF	SS	MS	F	P
order	63	185.911	2.951	0.77	0.900
Error	576	2201.700	3.822		
Total	639	2387.611			

Table 9. ANOVA summary table for overall order, measured by a distinguishable rating.

Source	DF	SS	MS	F	P
order	63	206.586	3.279	0.87	0.745
Error	576	2163.100	3.755		
Total	639	2369.686			

Table 10. ANOVA summary table for overall order, measured by a blending rating.

Source	DF	SS	MS	F	P
order	63	195.975	3.111	0.80	0.870
Error	576	2251.400	3.909		
Total	639	2447.375			

Table 11. ANOVA summary table for overall order, measured by a usefulness rating.

Source	DF	SS	MS	F	P
order	63	189.061	3.001	0.76	0.911
Error	576	2269.500	3.940		
Total	639	2458.561			

Table 12. ANOVA summary table for overall order, measured by a unified scene rating.

Source	DF	SS	MS	F	P
order	63	238.436	3.785	0.94	0.600
Error	576	2308.500	4.008		
Total	639	2546.936			

Tables 13 through 20 show the ANOVA summary tables for scene presentation order (spo) as measured by the eight trial questionnaire ratings (goodness, separable, integral, distracting, distinguishable, blending, usefulness, unified scene).

Table 13. ANOVA summary table for scene presentation order, measured by a goodness rating.

Source	DF	SS	MS	F	P
spo	1	7.439	7.439	2.36	0.125
Error	638	2013.172	3.155		
Total	639	2020.611			

Table 14. ANOVA summary table for scene presentation order, measured by a separable rating.

Source	DF	SS	MS	F	P
spo	1	0.002	0.002	0.00	0.984
Error	638	2491.122	3.905		
Total	639	2491.123			

Table 15. ANOVA summary table for scene presentation order, measured by an integral rating.

Source	DF	SS	MS	F	P
spo	1	11.556	11.556	3.14	0.077
Error	638	2351.544	3.686		
Total	639	2363.100			

Table 16. ANOVA summary table for scene presentation order, measured by a distracting rating.

Source	DF	SS	MS	F	P
spo	1	0.352	0.352	0.09	0.759
Error	638	2387.259	3.742		
Total	639	2387.611			

Table 17. ANOVA summary table for scene presentation order, measured by a distinguishable rating.

Source	DF	SS	MS	F	P
spo	1	2.377	2.377	0.64	0.424
Error	638	2367.309	3.711		
Total	639	2369.686			

Table 18. ANOVA summary table for scene presentation order, measured by a blending rating.

Source	DF	SS	MS	F	P
spo	1	3.025	3.025	0.79	0.375
Error	638	2444.350	3.831		
Total	639	2447.375			

Table 19. ANOVA summary table for scene presentation order, measured by a usefulness rating.

Source	DF	SS	MS	F	P
spo	1	0.352	0.352	0.09	0.763
Error	638	2458.209	3.853		
Total	639	2458.561			

Table 20. ANOVA summary table for scene presentation order, measured by a unified scene rating.

Source	DF	SS	MS	F	P
spo	1	1.502	1.502	0.38	0.540
Error	638	2545.434	3.990		
Total	639	2546.936			

### 5.3.2. Augmented Object, Transparency, Texture ANOVAs

Three-factor within subjects ANOVAs were performed for the independent variables augmented object, transparency, and texture, as measured by the dependent measures from the trial questionnaire rating scales (goodness, separable, integral, distracting, distinguishable, blending, usefulness, and unified scene). Tables 21 through 28 show the ANOVA summary tables for the three independent variables as measured by the eight trial questionnaire ratings. Corresponding figures for all significant effects are shown.

Table 21. ANOVA summary table for the three independent variables, measured by a goodness rating.

Source	DF	SS	MS	F	P
subj	9	256.189	28.465	2.57	
<b>ar</b>	<b>3</b>	<b>161.842</b>	<b>53.947</b>	<b>5.39</b>	<b>0.005</b>
trans	1	47.852	47.852	8.84	0.016
<b>texture</b>	<b>3</b>	<b>86.342</b>	<b>28.781</b>	<b>15.54</b>	<b>0.000</b>
subj*ar	27	270.267	10.010	2.11	
subj*trans	9	48.727	5.414	1.74	
subj*texture	27	50.017	1.852	0.52	
<b>ar*trans</b>	<b>3</b>	<b>42.080</b>	<b>14.027</b>	<b>5.37</b>	<b>0.005</b>
<b>ar*texture</b>	<b>9</b>	<b>255.702</b>	<b>28.411</b>	<b>9.19</b>	<b>0.000</b>
trans*texture	3	13.930	4.643	3.16	0.041
subj*ar*trans	27	70.530	2.612	2.69	
subj*ar*texture	81	250.502	3.093	3.19	
subj*trans*texture	27	39.680	1.470	1.51	
<b>ar*trans*texture</b>	<b>9</b>	<b>37.714</b>	<b>4.190</b>	<b>4.32</b>	<b>0.000</b>
Error	401	389.239	0.971		
Total	639	2020.611			

Figures 21 through 25 show the graphs for the significant results of the goodness measure. Figure 21 highlights higher average goodness ratings (indicating worse goodness perception) for text augmented conditions. Figure 22 shows an average higher rating for goodness for the wireframe texture. Figure 23 indicates a decrease in the goodness ratings for the text-transparent condition. Figure 24 shows a shift in goodness trend for several augmented scenes in the white texture condition. Figure 25 presents the complex three-way interaction.

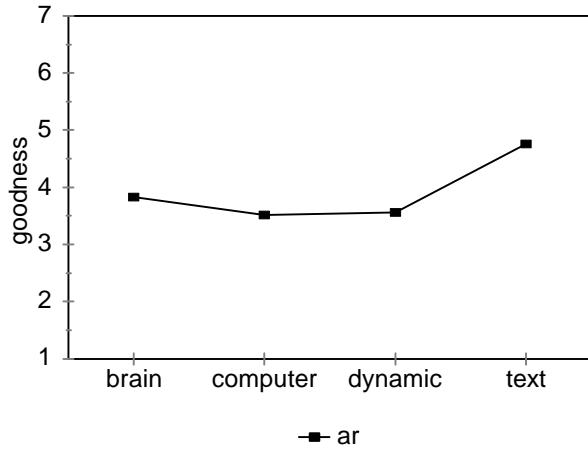


Figure 21. Graph of the main effect of augmented reality object, significant at the  $p < 0.01$  level ( $p = 0.005$ ) as measured by goodness ratings.

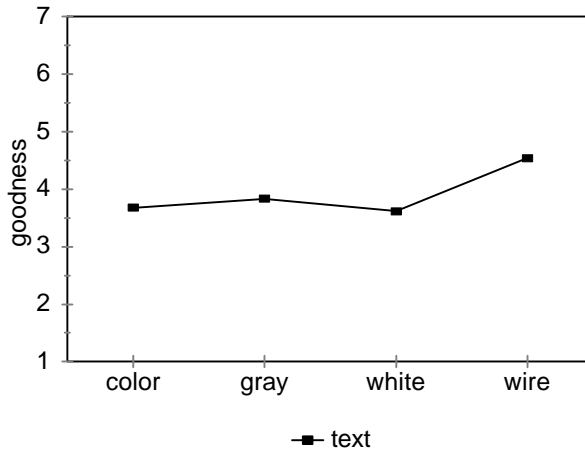


Figure 22. Graph of the main effect of texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

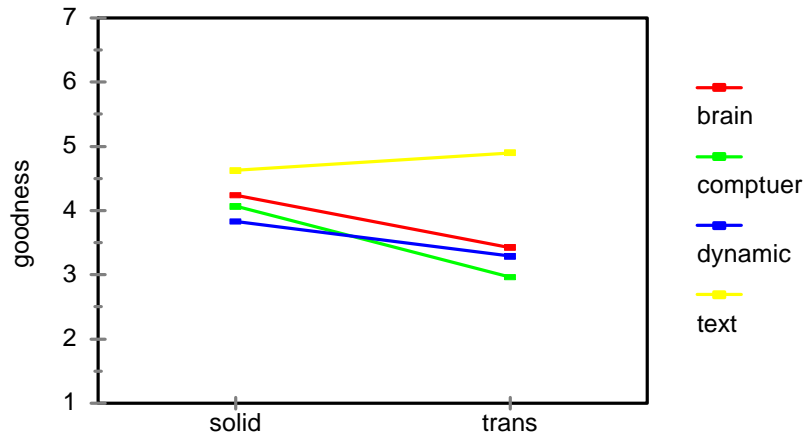


Figure 23. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.005$ ) as measured by goodness ratings.

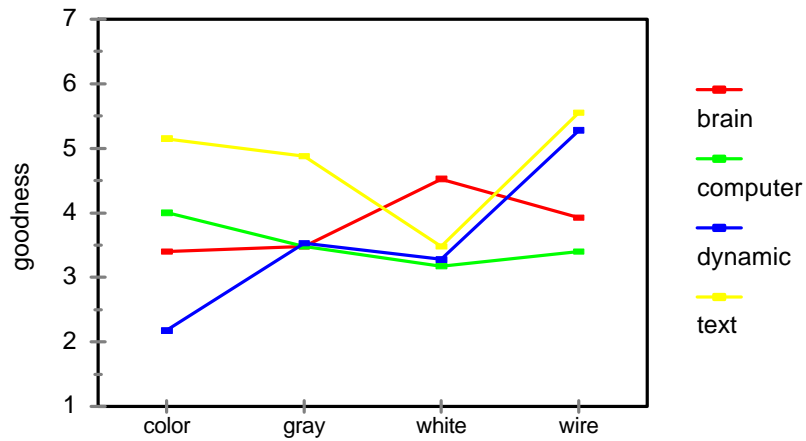


Figure 24. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

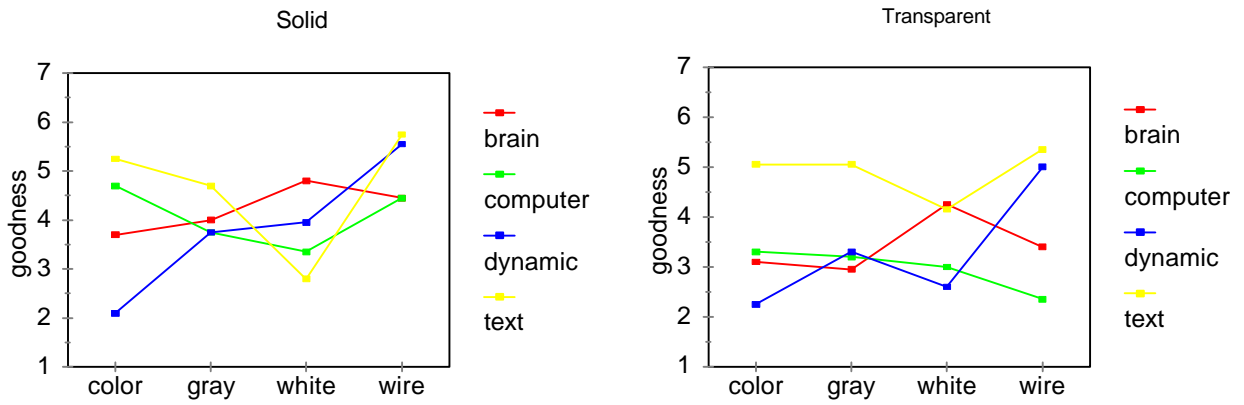


Figure 25. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

Table 22. ANOVA summary table for the three independent variables, measured by a separable rating.

Source	DF	SS	MS	F	P
subj	9	281.608	31.290	2.82	
ar	3	60.955	20.318	1.52	0.231
<b>trans</b>	<b>1</b>	<b>126.914</b>	<b>126.914</b>	<b>122.55</b>	<b>0.000</b>
texture	3	18.167	6.056	1.66	0.198
subj*ar	27	359.873	13.329	2.54	
subj*trans	9	9.320	1.036	0.30	
subj*texture	27	98.286	3.640	1.00	
<b>ar*trans</b>	<b>3</b>	<b>134.342</b>	<b>44.781</b>	<b>13.66</b>	<b>0.000</b>
<b>ar*texture</b>	<b>9</b>	<b>324.327</b>	<b>36.036</b>	<b>10.42</b>	<b>0.000</b>
trans*texture	3	7.255	2.418	1.44	0.252
subj*ar*trans	27	88.486	3.277	2.20	
subj*ar*texture	81	280.158	3.459	2.32	
subj*trans*texture	27	45.198	1.674	1.12	
<b>ar*trans*texture</b>	<b>9</b>	<b>58.014</b>	<b>6.446</b>	<b>4.32</b>	<b>0.000</b>
Error	401	598.220	1.492		
Total	639	2491.123			

Figures 26 through 29 show the graphs for the significant results of the separable measure.

Figure 26 shows lower average separable ratings (indicating more separable perception) for the solid graphics condition. Figure 27 highlights an increase in separable ratings for the transparent-computer augmented reality condition. Figure 28 indicates a shift in separable trends for several augmented objects at the white texture condition. Figure 29 presents the complex three-way interaction.

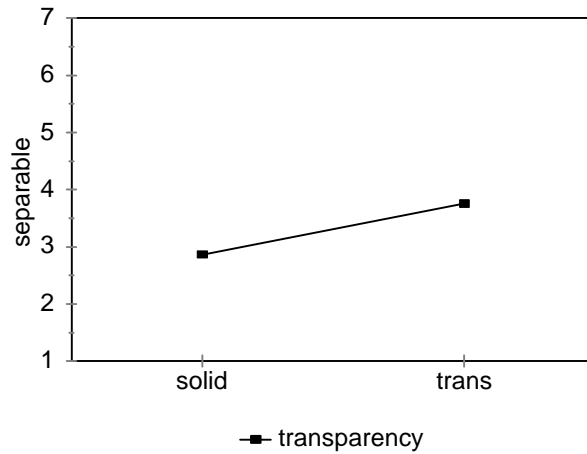


Figure 26. Graph of the main effect of transparency, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

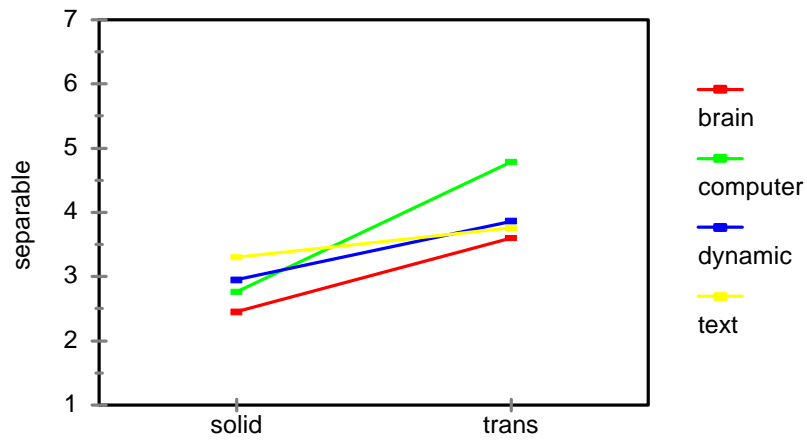


Figure 27. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

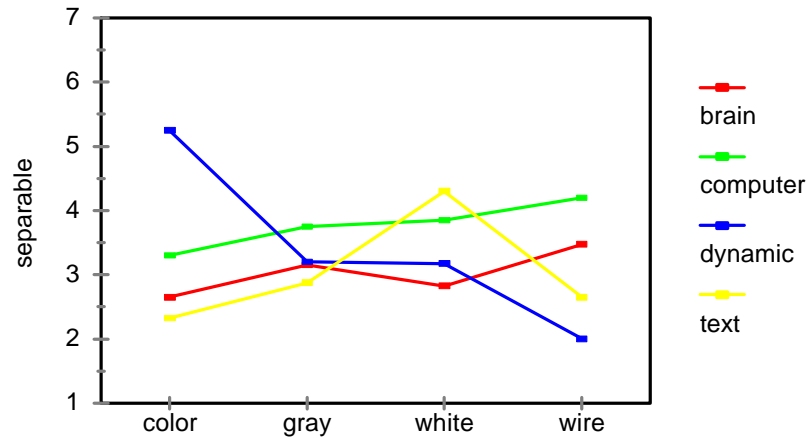


Figure 28. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

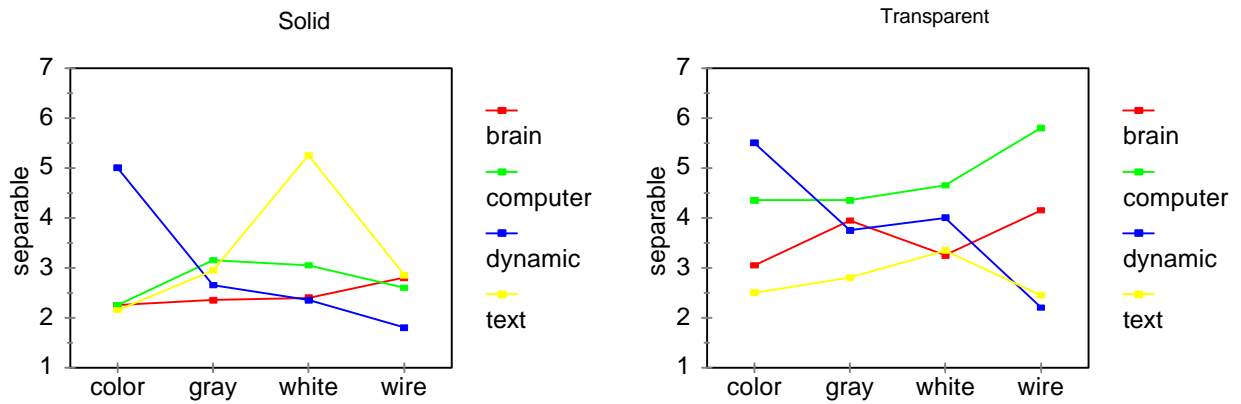


Figure 29. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

Table 23. ANOVA summary table for the three independent variables, measured by an integral rating.

Source	DF	SS	MS	F	P
subj	9	245.537	27.282	1.64	
ar	3	3.725	1.242	0.08	0.968
trans	1	50.625	50.625	8.44	0.017
texture	3	40.488	13.496	3.90	0.019
subj*ar	27	396.963	14.702	2.70	
subj*trans	9	54.000	6.000	1.46	
subj*texture	27	93.325	3.456	0.88	
<b>ar*trans</b>	<b>3</b>	<b>85.950</b>	<b>28.650</b>	<b>7.85</b>	<b>0.001</b>
<b>ar*texture</b>	<b>9</b>	<b>241.587</b>	<b>26.843</b>	<b>7.75</b>	<b>0.000</b>
trans*texture	3	4.037	1.346	0.63	0.600
subj*ar*trans	27	98.550	3.650	2.19	
subj*ar*texture	81	280.475	3.463	2.07	
subj*trans*texture	27	57.338	2.124	1.27	
<b>ar*trans*texture</b>	<b>9</b>	<b>41.188</b>	<b>4.576</b>	<b>2.74</b>	<b>0.004</b>
Error	401	669.313	1.669		
Total	639	2363.100			

Figures 30 through 32 show the graphs for the significant results of the integral measure. Figure 30 shows increased integral ratings (indicating less integral perception) for the transparent-text condition. Figure 31 shows three standouts from the general trend for augmented object and texture: the dynamic-color condition (lower integral ratings), the dynamic-wireframe condition (higher integral ratings), and the text-white condition (lower integral ratings). Figure 32 presents the complex three-way interaction.

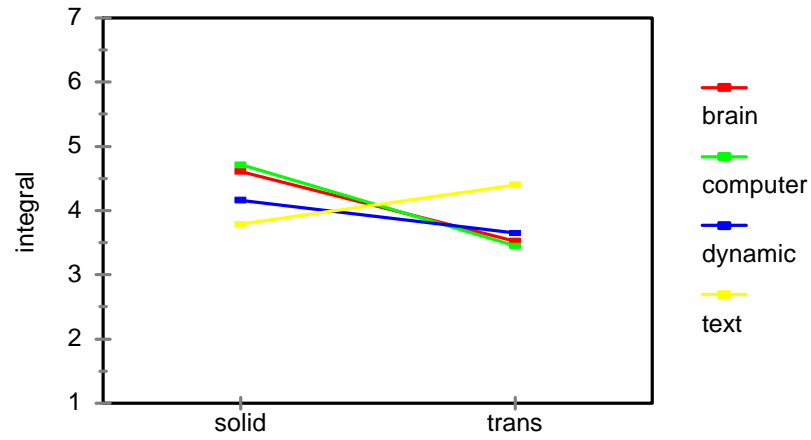


Figure 30. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by integral ratings.

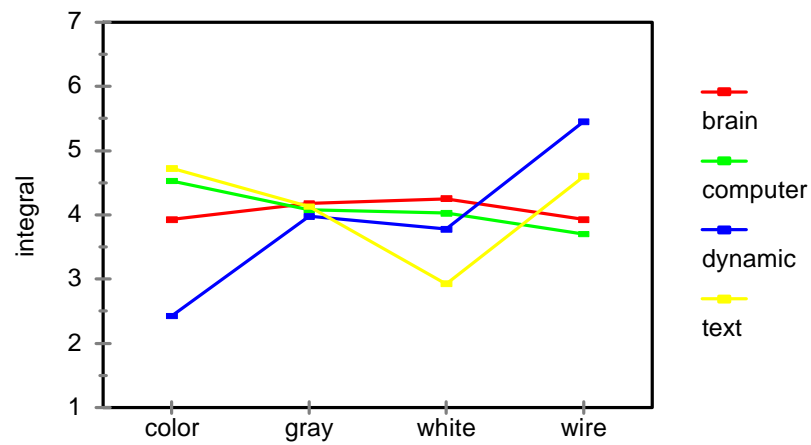


Figure 31. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by integral ratings.

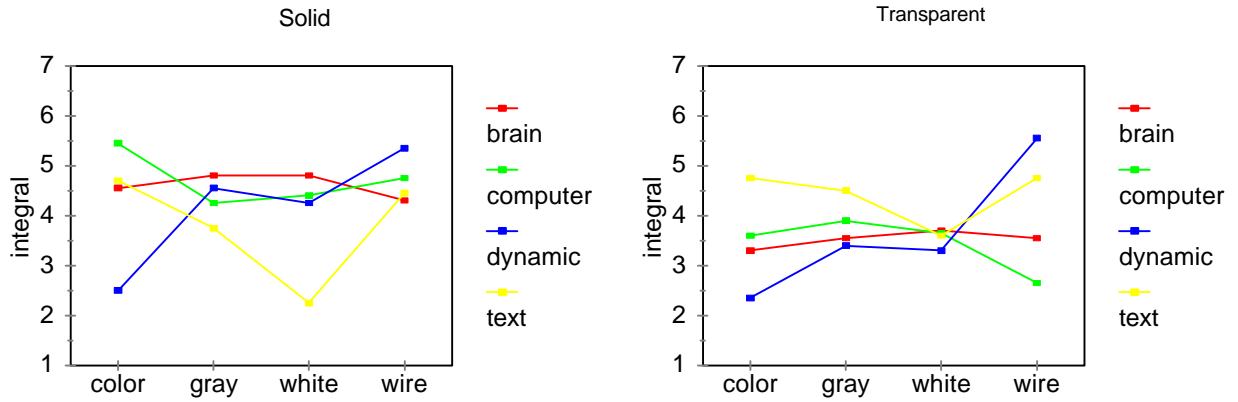


Figure 32. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.004$ ) as measured by integral ratings.

Table 24. ANOVA summary table for the three independent variables, measured by a distracting rating.

Source	DF	SS	MS	F	P
subj	9	157.845	17.538	0.75	
ar	3	52.855	17.618	0.96	0.426
<b>trans</b>	<b>1</b>	<b>101.602</b>	<b>101.602</b>	<b>14.69</b>	<b>0.004</b>
<b>texture</b>	<b>3</b>	<b>120.605</b>	<b>40.202</b>	<b>11.11</b>	<b>0.000</b>
subj*ar	27	496.098	18.374	4.44	
subj*trans	9	62.258	6.918	2.58	
subj*texture	27	97.723	3.619	1.28	
<b>ar*trans</b>	<b>3</b>	<b>78.717</b>	<b>26.239</b>	<b>10.00</b>	<b>0.000</b>
<b>ar*texture</b>	<b>9</b>	<b>309.827</b>	<b>34.425</b>	<b>12.48</b>	<b>0.000</b>
trans*texture	3	17.392	5.797	4.43	0.012
subj*ar*trans	27	70.861	2.624	2.10	
subj*ar*texture	81	223.408	2.758	2.21	
subj*trans*texture	27	35.311	1.308	1.05	
<b>ar*trans*texture</b>	<b>9</b>	<b>63.064</b>	<b>7.007</b>	<b>5.62</b>	<b>0.000</b>
Error	401	500.045	1.247		
Total	639	2387.611			

Figures 33 through 37 show the graphs for the significant results of the distracting measure. Figure 33 shows lower average distracting ratings (indicating more distracting perception) for the solid graphics condition. Figure 34 presents a lower distracting rating for the wireframe texture condition. Figure 35 indicates a difference in distracting rating trends for the transparent-text condition. Figure 36 shows a shift in distracting ratings for several augmented objects across several texture conditions. Figure 37 presents the complex three-way interaction.

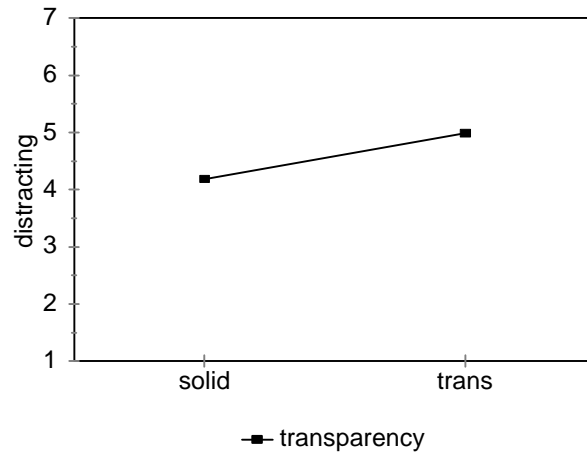


Figure 33. Graph of the main effect of transparency, significant at the  $p < 0.01$  level ( $p = 0.004$ ) as measured by distracting ratings.

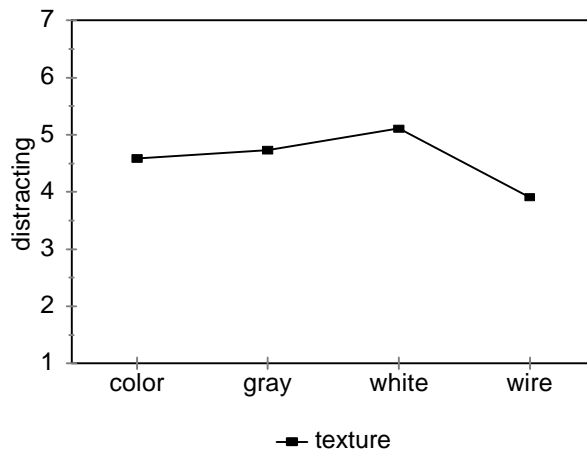


Figure 34. Graph of the main effect of texture, significant at the  $p < 0.01$  level ( $p = 0.004$ ) as measured by distracting ratings.

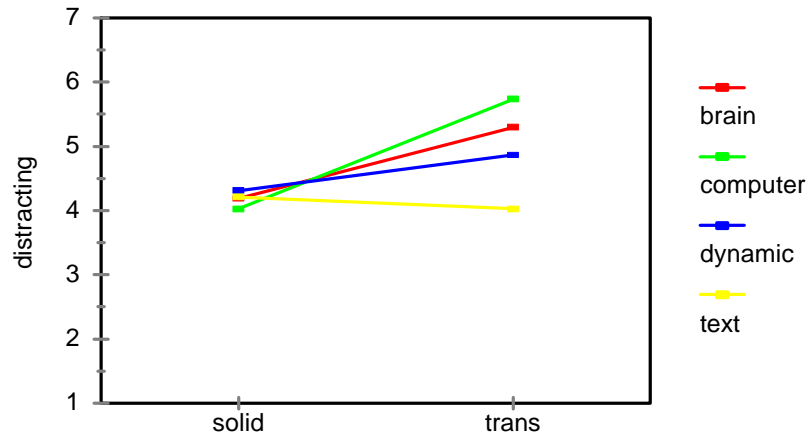


Figure 35. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distracting ratings.

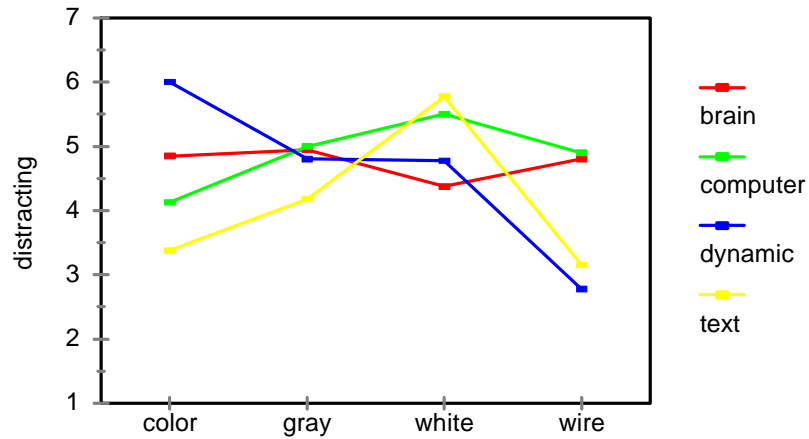


Figure 36. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distracting ratings.

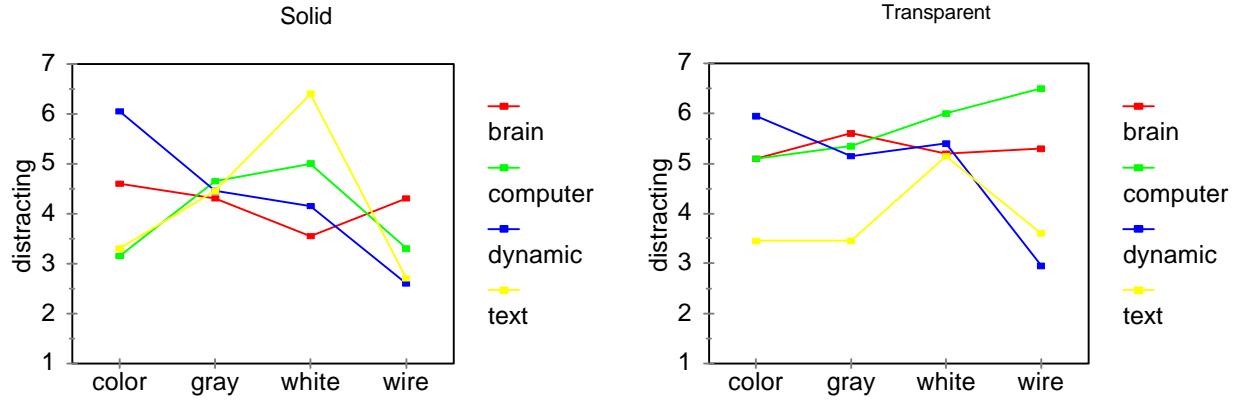


Figure 37. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distracting ratings.

Table 25. ANOVA summary table for the three independent variables, measured by a distinguishable rating.

Source	DF	SS	MS	F	P
subj	9	330.358	36.706	2.72	
ar	3	77.555	25.852	1.88	0.156
<b>trans</b>	<b>1</b>	<b>83.377</b>	<b>83.377</b>	<b>25.78</b>	<b>0.001</b>
texture	3	8.930	2.977	1.09	0.371
subj*ar	27	370.461	13.721	3.05	
subj*trans	9	29.108	3.234	0.97	
subj*texture	27	73.836	2.735	0.83	
<b>ar*trans</b>	<b>3</b>	<b>124.167</b>	<b>41.389</b>	<b>14.09</b>	<b>0.000</b>
<b>ar*texture</b>	<b>9</b>	<b>283.527</b>	<b>31.503</b>	<b>10.92</b>	<b>0.000</b>
<b>trans*texture</b>	<b>3</b>	<b>34.217</b>	<b>11.406</b>	<b>6.66</b>	<b>0.002</b>
subj*ar*trans	27	79.286	2.937	2.23	
subj*ar*texture	81	233.770	2.886	2.19	
subj*trans*texture	27	46.236	1.712	1.30	
<b>ar*trans*texture</b>	<b>9</b>	<b>65.664</b>	<b>7.296</b>	<b>5.53</b>	<b>0.000</b>
Error	401	529.195	1.320		
Total	639	2369.686			

Figures 38 through 42 show the graphs for the significant results of the distinguishable measure. Figure 38 shows lower average distinguishable ratings (indicating more distinguishable perception) for the solid graphics condition. Figure 39 highlights an increase in distinguishable ratings for the transparent-text condition. Figure 40 shows several shifts in distinguishable ratings trends for the augmented object and texture interaction. Figure 41 indicates a difference in distinguishable ratings from the white texture condition across the two levels of transparency. Figure 42 presents the complex three-way interaction.

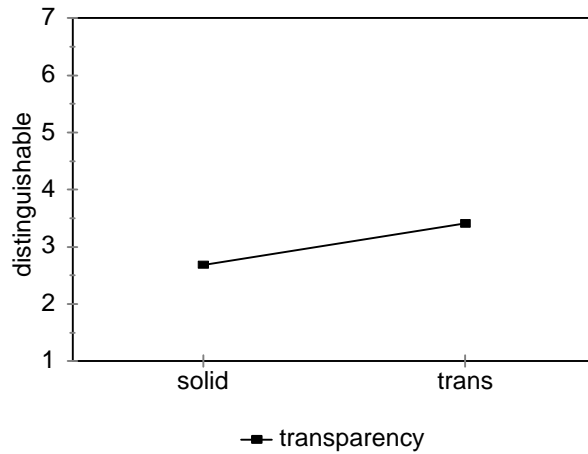


Figure 38. Graph of the main effect of transparency, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by distinguishable ratings.

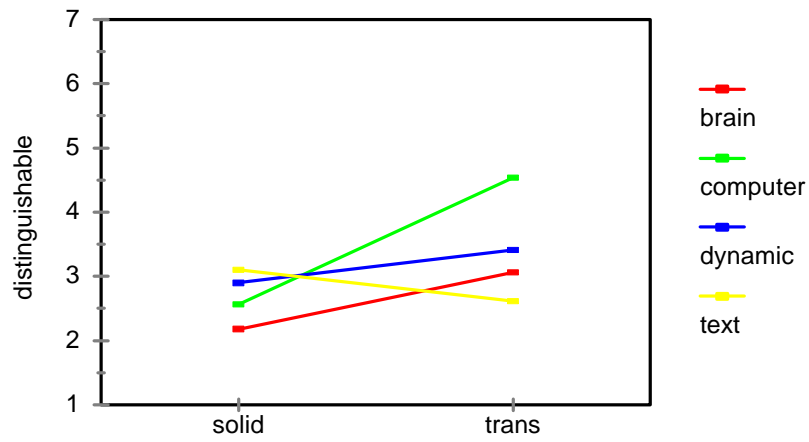


Figure 39. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distinguishable ratings.

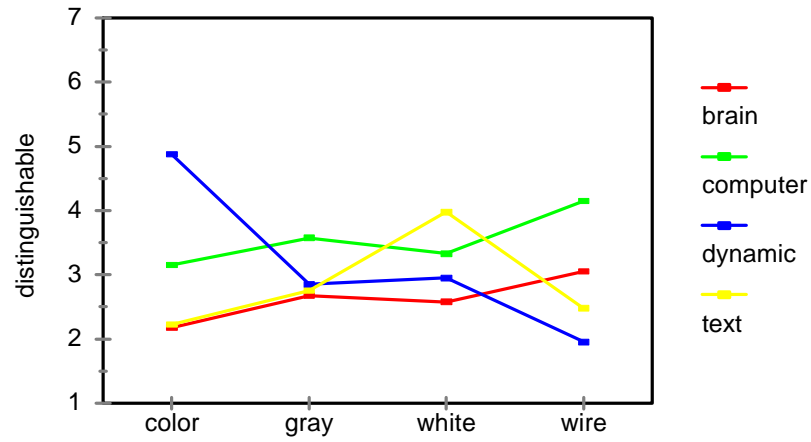


Figure 40. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distinguishable ratings.

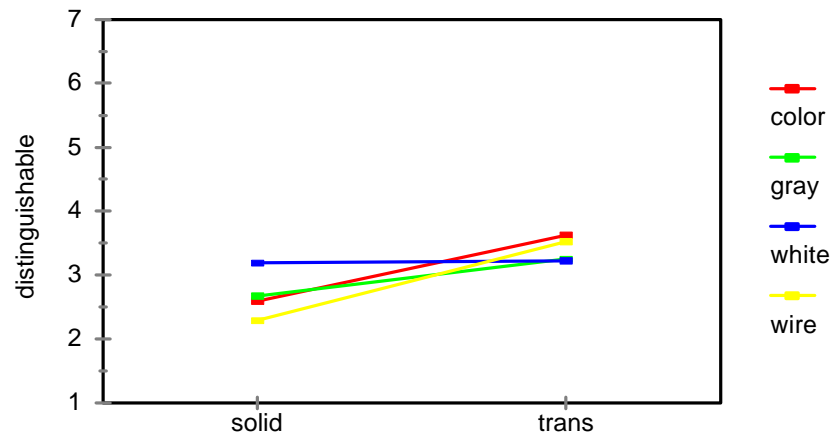


Figure 41. Graph of the interaction between transparency and texture, significant at the  $p < 0.01$  level ( $p = 0.002$ ) as measured by distinguishable ratings.

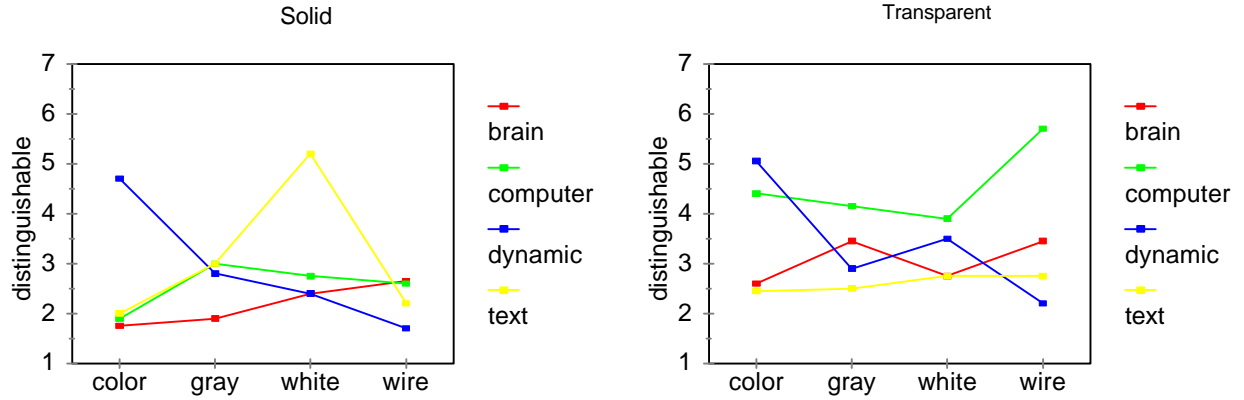


Figure 42. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by distinguishable ratings.

Table 26. ANOVA summary table for the three independent variables, measured by a blending rating.

Source	DF	SS	MS	F	P
subj	9	328.250	36.472	2.28	
ar	3	44.900	14.967	0.95	0.432
<b>trans</b>	<b>1</b>	<b>66.306</b>	<b>66.306</b>	<b>15.52</b>	<b>0.003</b>
<b>texture</b>	<b>3</b>	<b>75.388</b>	<b>25.129</b>	<b>6.66</b>	<b>0.002</b>
subj*ar	27	426.975	15.814	2.73	
subj*trans	9	38.444	4.272	0.91	
subj*texture	27	101.862	3.773	0.94	
<b>ar*trans</b>	<b>3</b>	<b>100.469</b>	<b>33.490</b>	<b>8.73</b>	<b>0.000</b>
<b>ar*texture</b>	<b>9</b>	<b>286.838</b>	<b>31.871</b>	<b>10.12</b>	<b>0.000</b>
trans*texture	3	14.881	4.960	2.41	0.089
subj*ar*trans	27	103.531	3.834	3.20	
subj*ar*texture	81	255.162	3.150	2.63	
subj*trans*texture	27	55.494	2.055	1.72	
<b>ar*trans*texture</b>	<b>9</b>	<b>68.694</b>	<b>7.633</b>	<b>6.37</b>	<b>0.000</b>
Error	401	480.181	1.197		
Total	639	2447.375			

Figures 43 through 47 show the graphs for the significant results of the blending measure. Figure 43 shows higher average blending ratings (indicating less blending perception) for the solid graphics condition. Figure 44 presents higher blending ratings for the wireframe texture condition. Figure 45 indicates higher blending ratings for the transparent-text condition. Figure 46 shows several differences in blending trends for the augmented object and texture interaction. Figure 47 presents the complex three-way interaction.

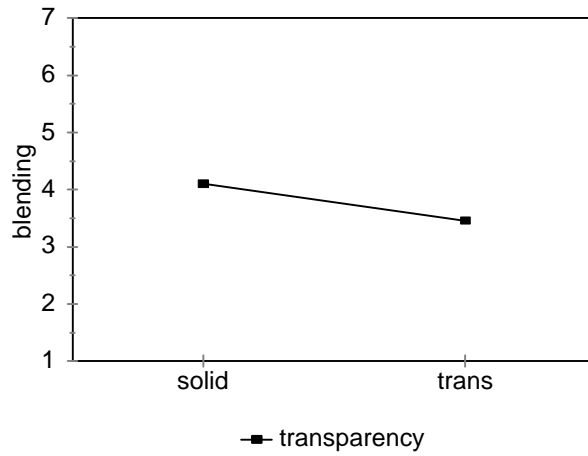


Figure 43. Graph of the main effect of transparency, significant at the  $p < 0.01$  level ( $p = 0.003$ ) as measured by blending ratings.

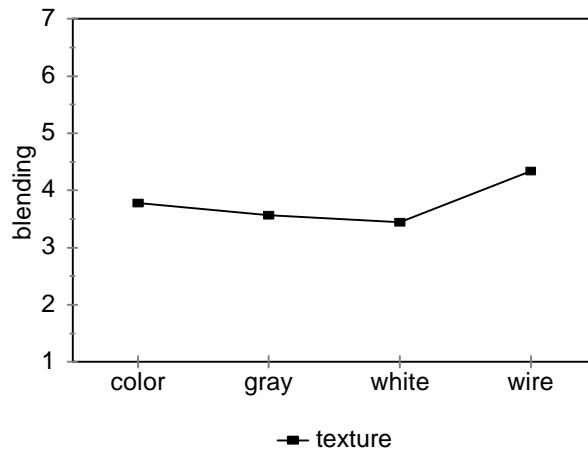


Figure 44. Graph of the main effect of texture, significant at the  $p < 0.01$  level ( $p = 0.002$ ) as measured by blending ratings.

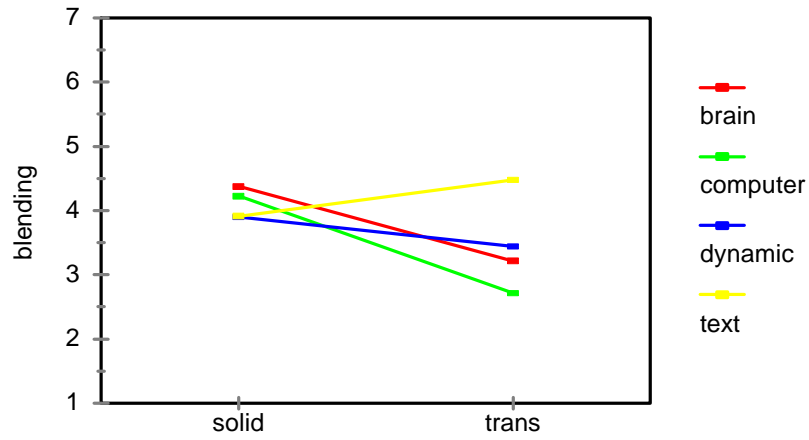


Figure 45. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by blending ratings.

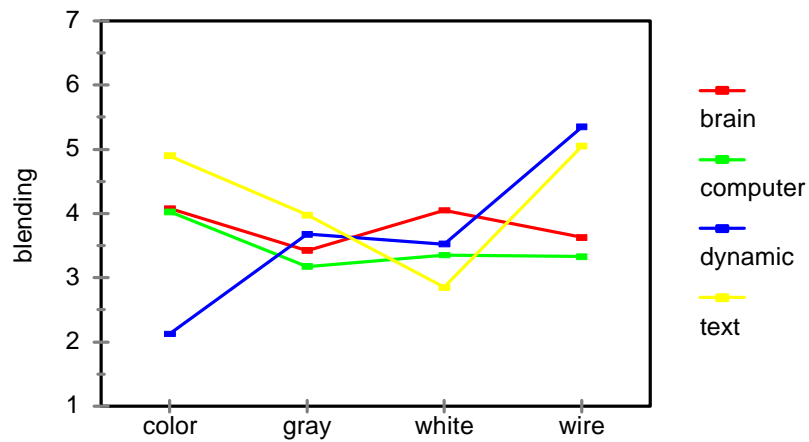


Figure 46. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by blending ratings.

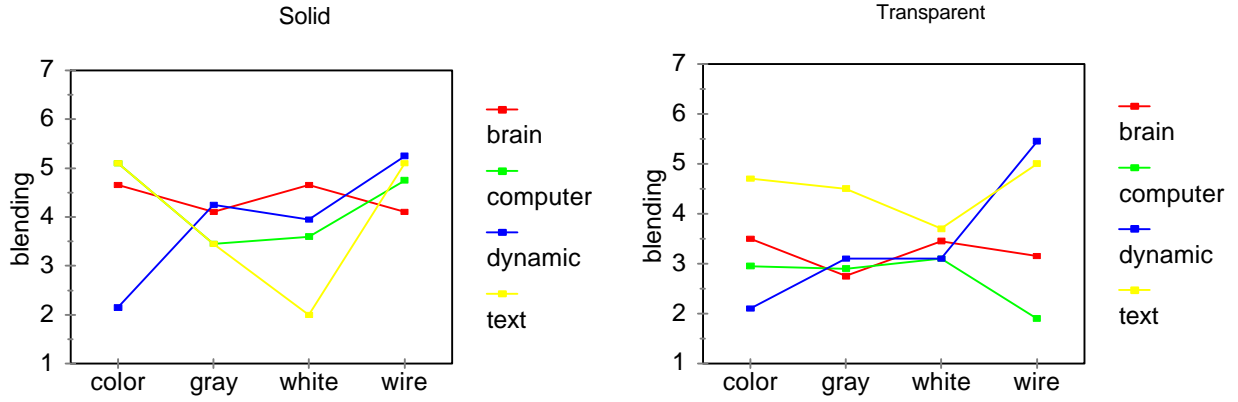


Figure 47. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by blending ratings.

Table 27. ANOVA summary table for the three independent variables, measured by a usefulness rating.

Source	DF	SS	MS	F	P
subj	9	258.389	28.710	1.05	
ar	3	40.880	13.627	0.56	0.645
trans	1	30.189	30.189	3.71	0.086
<b>texture</b>	<b>3</b>	<b>159.392</b>	<b>53.131</b>	<b>21.22</b>	<b>0.000</b>
subj*ar	27	655.855	24.291	4.30	
subj*trans	9	73.264	8.140	2.26	
subj*texture	27	67.592	2.503	0.55	
<b>ar*trans</b>	<b>3</b>	<b>42.030</b>	<b>14.010</b>	<b>4.81</b>	<b>0.008</b>
<b>ar*texture</b>	<b>9</b>	<b>179.364</b>	<b>19.929</b>	<b>5.14</b>	<b>0.000</b>
trans*texture	3	10.767	3.589	1.96	0.143
subj*ar*trans	27	78.580	2.910	2.55	
subj*ar*texture	81	314.339	3.881	3.40	
subj*trans*texture	27	49.342	1.827	1.60	
<b>ar*trans*texture</b>	<b>9</b>	<b>40.389</b>	<b>4.488</b>	<b>3.93</b>	<b>0.000</b>
Error	401	458.189	1.143		
Total	639	2458.561			

Figures 48 through 51 show the graphs for the significant results of the usefulness measure.

Figure 48 shows higher average usefulness ratings (indicating less usefulness perception) for the wireframe texture condition. Figure 49 highlights higher usefulness ratings for the transparent-text condition. Figure 50 shows several differences in usefulness ratings trends for the augmented object and texture interaction. Figure 51 presents the complex three-way interaction.

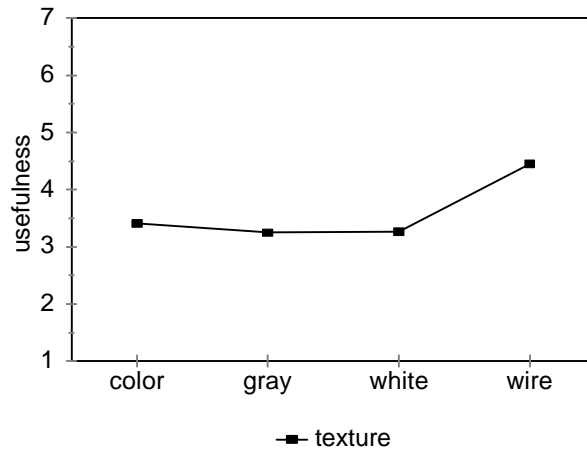


Figure 48. Graph of the main effect of texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

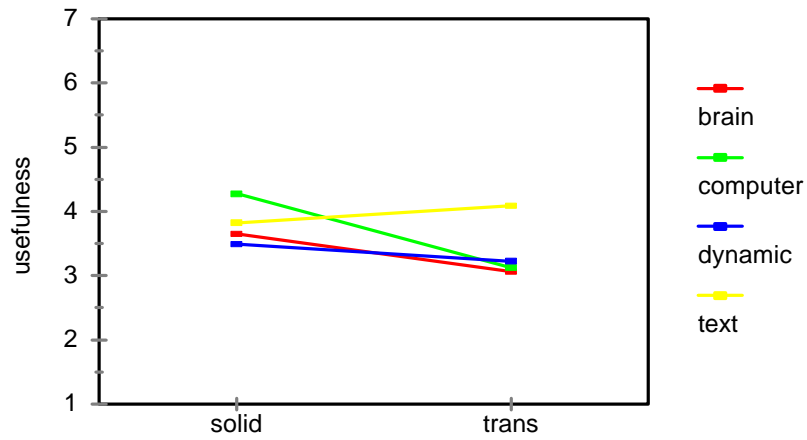


Figure 49. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.008$ ) as measured by usefulness ratings.

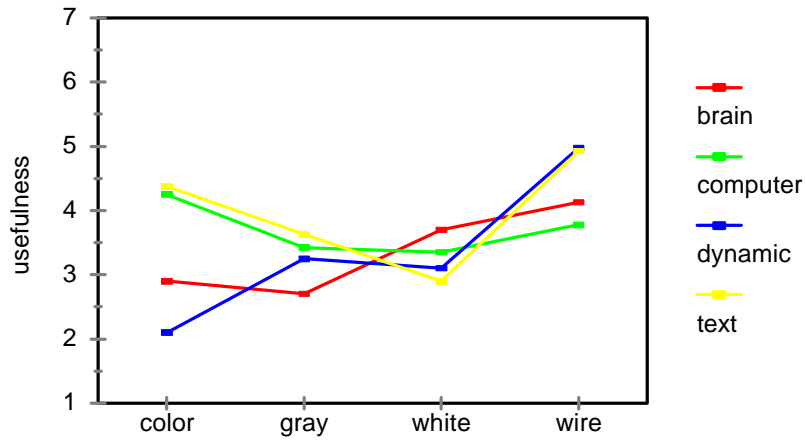


Figure 50. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

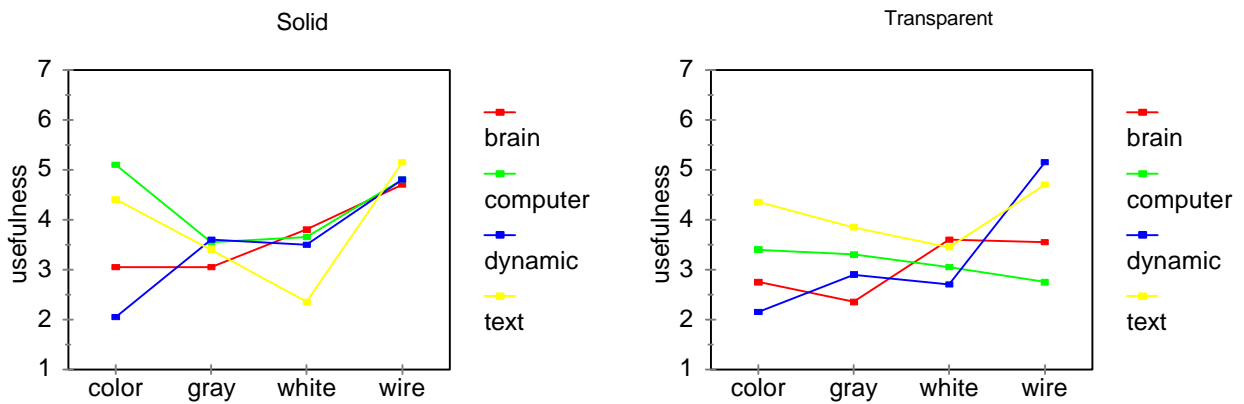


Figure 51. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

Table 28. ANOVA summary table for the three independent variables, measured by a unified scene rating.

Source	DF	SS	MS	F	P
subj	9	388.389	43.154	1.90	
ar	3	33.355	11.118	0.49	0.690
<b>trans</b>	<b>1</b>	<b>87.764</b>	<b>87.764</b>	<b>21.75</b>	<b>0.001</b>
<b>texture</b>	<b>3</b>	<b>50.092</b>	<b>16.697</b>	<b>5.71</b>	<b>0.004</b>
subj*ar	27	608.255	22.528	4.50	
subj*trans	9	36.314	4.035	0.99	
subj*texture	27	79.017	2.927	0.84	
<b>ar*trans</b>	<b>3</b>	<b>77.680</b>	<b>25.893</b>	<b>7.70</b>	<b>0.001</b>
<b>ar*texture</b>	<b>9</b>	<b>305.464</b>	<b>33.940</b>	<b>12.32</b>	<b>0.000</b>
trans*texture	3	17.567	5.856	3.20	0.039
subj*ar*trans	27	90.805	3.363	3.02	
subj*ar*texture	81	223.114	2.754	2.47	
subj*trans*texture	27	49.417	1.830	1.64	
<b>ar*trans*texture</b>	<b>9</b>	<b>53.264</b>	<b>5.918</b>	<b>5.32</b>	<b>0.000</b>
Error	401	446.439	1.113		
Total	639	2546.936			

Figures 52 through 56 show the graphs for the significant results of the unified measure. Figure 52 shows higher average unified ratings (indicating less unified perception) for the solid graphics condition. Figure 53 presents higher unified ratings for the wireframe texture condition and lower unified ratings for the white texture condition. Figure 54 highlights higher unified ratings for the transparent-text condition. Figure 55 shows several differences in unified ratings trends for the augmented object and texture interaction. Figure 56 presents the complex three-way interaction.

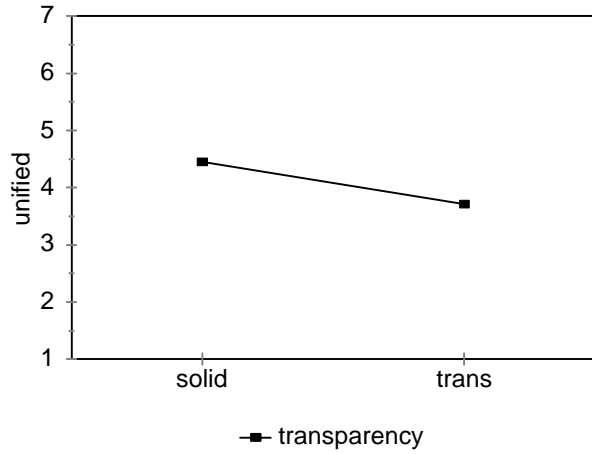


Figure 52. Graph of the main effect of transparency, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by unified scene ratings.

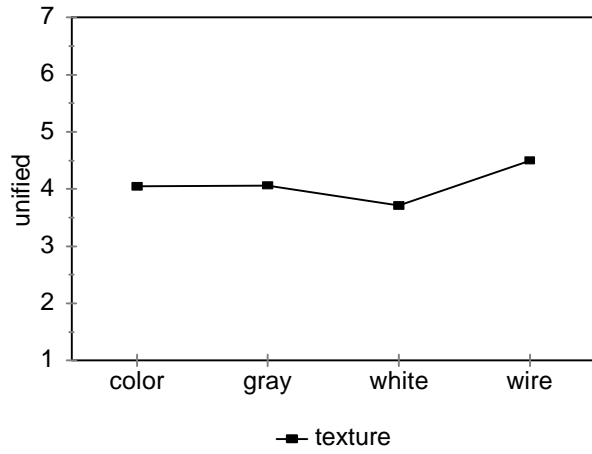


Figure 53. Graph of the main effect of texture, significant at the  $p < 0.01$  level ( $p = 0.004$ ) as measured by unified scene ratings.

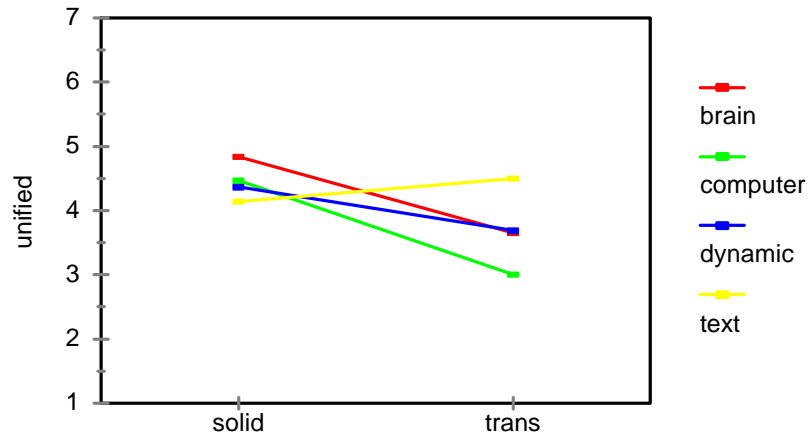


Figure 54. Graph of the interaction between augmented reality object and transparency, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by unified scene ratings.

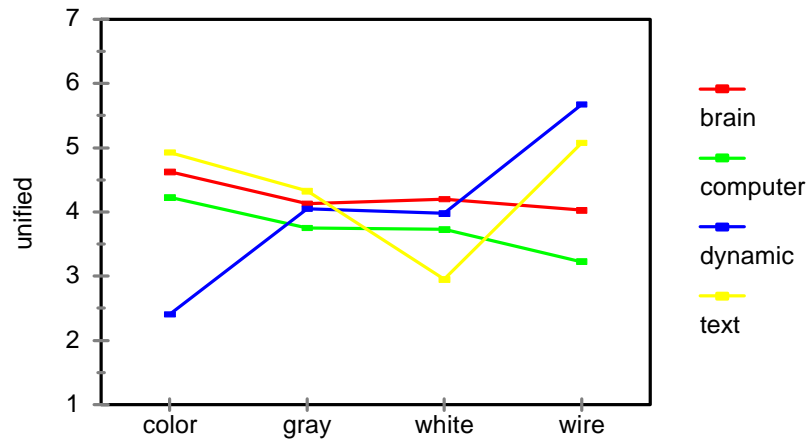


Figure 55. Graph of the interaction between augmented reality object and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by unified scene ratings.

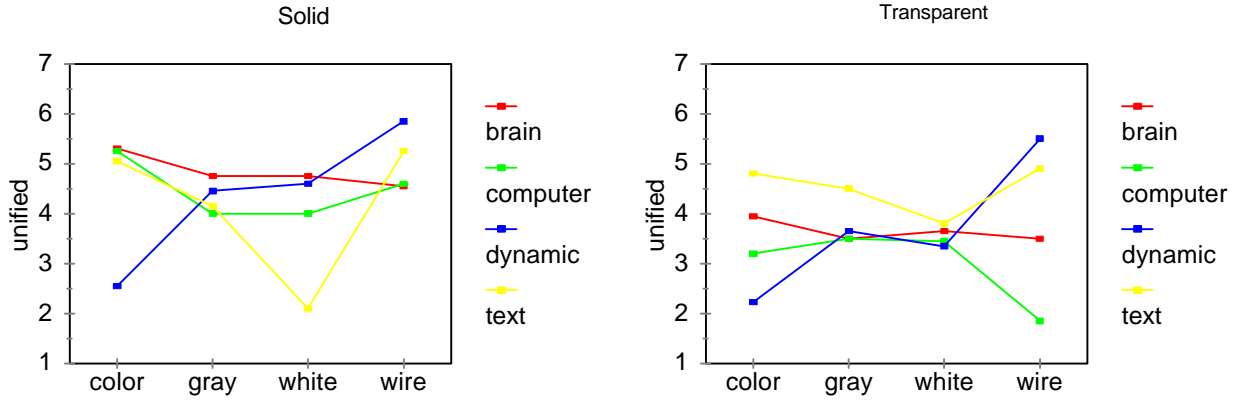


Figure 56. Graphs of the interaction between augmented reality object and texture for both the solid and transparent levels, shown to represent the 3-way interaction between augmented reality, transparency, and texture, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by unified scene ratings.

Table 29 summarizes all the significant effects for the three independent variables of experiment one for the eight subjective assessment dependent measures.

Table 29. Summary of significant results at the  $p < 0.01$  level for the eight rating scale dependent measures for experiment one.

	<b>goodness</b>	<b>separable</b>	<b>integral</b>	<b>distract.</b>	<b>disting.</b>	<b>blend</b>	<b>useful</b>	<b>unified</b>
<b>ar</b>	x							
<b>trans</b>		x		x	x	x		x
<b>text</b>	x			x		x	x	x
<b>ar*trans</b>	x	x	x	x	x	x		
<b>ar*text</b>	x	x	x	x	x	x	x	x
<b>tans*text</b>					x			
<b>ar*trans*text</b>	x	x	x	x	x	x	x	x

### 5.3.3. Nonparametric Tests

Where main effect significant results were indicated in the above parametric ANOVAs, non-parametric tests were performed to confirm those significant results. Tables 30 through 40 show summary table results of Kruskal-Wallis Tests for the significant main effects from the parametric ANOVAs.

Table 30. Kruskal-Wallis test for augmented reality object as measured by goodness ratings.

ar	N	Median	Ave Rank	Z
brain	160	4.000	313.3	-0.57
computer	160	3.000	277.9	-3.37
dynamic	160	3.000	284.5	-2.85
text	160	5.000	406.4	6.78
Overall	640		320.5	

H = 49.33 DF = 3 P = 0.000  
H = 50.91 DF = 3 P = 0.000 (adjusted for ties)

Table 31. Kruskal-Wallis test for texture as measured by goodness ratings.

texture	N	Median	Ave Rank	Z
color	160	3.000	294.9	-2.02
gray	160	3.000	313.8	-0.53
white	160	3.000	290.6	-2.36
wire	160	5.000	382.6	4.91
Overall	640		320.5	

H = 25.53 DF = 3 P = 0.000  
H = 26.35 DF = 3 P = 0.000 (adjusted for ties)

Table 32. Kruskal-Wallis test for transparency as measured by separable ratings.

trans	N	Median	Ave Rank	Z
solid	320	2.000	279.2	-5.65
trans	320	3.000	361.8	5.65
Overall	640		320.5	

H = 31.94 DF = 1 P = 0.000  
H = 33.14 DF = 1 P = 0.000 (adjusted for ties)

Table 33. Kruskal-Wallis test for transparency as measured by distracting ratings.

<u>trans</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
solid	320	5.000	282.6	-5.19
trans	320	6.000	358.4	5.19
Overall	640		320.5	

H = 26.96 DF = 1 P = 0.000  
H = 27.94 DF = 1 P = 0.000 (adjusted for ties)

Table 34. Kruskal-Wallis test for texture as measured by distracting ratings.

<u>texture</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
color	160	5.000	321.9	0.11
gray	160	5.000	327.9	0.59
white	160	6.000	366.3	3.62
wire	160	3.000	265.8	-4.32
Overall	640		320.5	

H = 24.09 DF = 3 P = 0.000  
H = 24.96 DF = 3 P = 0.000 (adjusted for ties)

Table 35. Kruskal-Wallis test for transparency as measured by distinguishable ratings.

<u>trans</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
solid	320	2.000	285.0	-4.86
trans	320	3.000	356.0	4.86
Overall	640		320.5	

H = 23.63 DF = 1 P = 0.000  
H = 24.80 DF = 1 P = 0.000 (adjusted for ties)

Table 36. Kruskal-Wallis test for transparency as measured by blending ratings.

<u>trans</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
solid	320	4.000	350.6	4.12
trans	320	3.000	290.4	-4.12
Overall	640		320.5	

H = 16.97 DF = 1 P = 0.000  
H = 17.47 DF = 1 P = 0.000 (adjusted for ties)

Table 37. Kruskal-Wallis test for texture as measured by blending ratings.

texture	N	Median	Ave Rank	Z
color	160	3.000	317.4	-0.25
gray	160	3.000	302.9	-1.39
white	160	3.000	291.1	-2.32
wire	160	5.000	370.6	3.96
Overall	640		320.5	

H = 17.31 DF = 3 P = 0.001  
H = 17.81 DF = 3 P = 0.000 (adjusted for ties)

Table 38. Kruskal-Wallis test for texture as measured by usefulness ratings.

texture	N	Median	Ave Rank	Z
color	160	3.000	297.1	-1.85
gray	160	3.000	293.9	-2.10
white	160	3.000	293.5	-2.13
wire	160	5.000	397.5	6.08
Overall	640		320.5	

H = 37.04 DF = 3 P = 0.000  
H = 38.31 DF = 3 P = 0.000 (adjusted for ties)

Table 39. Kruskal-Wallis test for transparency as measured by unified scene ratings.

trans	N	Median	Ave Rank	Z
solid	320	5.000	353.7	4.54
trans	320	3.000	287.3	-4.54
Overall	640		320.5	

H = 20.63 DF = 1 P = 0.000  
H = 21.17 DF = 1 P = 0.000 (adjusted for ties)

Table 40. Kruskal-Wallis test for texture as measured by unified scene ratings.

texture	N	Median	Ave Rank	Z
color	160	4.000	317.3	-0.26
gray	160	4.000	319.2	-0.10
white	160	3.000	286.4	-2.69
wire	160	5.000	359.1	3.05
Overall	640		320.5	

H = 12.48 DF = 3 P = 0.006  
H = 12.80 DF = 3 P = 0.005 (adjusted for ties)

#### 5.3.4. *Subjective Rating Scales Factor Analyses*

Multivariate factor analyses were conducted on the eight rating scale measures (goodness, separable, integral, distracting, distinguishable, blending, usefulness, unified scene) for eight combined conditions of experiment one. Factor analyses were conducted to examine underlying relationships between the eight rating scale measures, and investigate the possibility of reducing the dimensionality of the data set (i.e., determine the extent that the dependent measures explained variance in respective factor loadings matrices).

Factor analysis provides the most understandable results when there are a large number of observations per condition. In data sets of few observations, interpretation is more speculative. There are no specific agreed upon rules, however, two rules of thumb are at least 5 observations per measure and/or at least 100 total observations per condition (Smith, 1998). With the 32 unique conditions in experiment one there are 20 total observations for each condition. To produce factor analyses that are more interpretable, data set combinations need to be considered. Further examination of the summary of experiment one results in Table 29 shows a strong pattern of significance for the transparency and texture independent variables, however, only one significant result for augmented reality object. This limited impact of the augmented object independent variable on the overall separation of the data provides an excellent option for developing a stronger data set for factor analysis. Collapsing the data across the augmented object conditions yields eight sets of data with 10 observations per measure and 80 total observations per condition.

With the combined data sets, preliminary factor analyses were first conducted by extracting four factors using the principal components method without rotation. Table 41 shows the results from this analysis on the solid-gray condition.

Table 41. Factor analysis results for the solid-gray condition extracting four factors using the principal components method without rotation.

Variable	Factor1	Factor2	Factor3	Factor4	Communality
goodness	-0.555	-0.666	-0.374	-0.293	0.977
separabl	0.807	-0.393	0.240	0.164	0.891
integral	-0.786	0.022	0.529	-0.215	0.945
distract	0.889	0.060	0.000	-0.061	0.797
disting.	0.659	-0.622	0.244	-0.182	0.914
blend	-0.885	-0.094	0.109	0.016	0.804
useful	-0.649	-0.609	0.018	0.391	0.946
unified	-0.888	0.176	0.043	0.025	0.823
<b>Variance</b>	<b>4.7952</b>	<b>1.4002</b>	<b>0.5510</b>	<b>0.3492</b>	<b>7.0955</b>
% Var	0.599	0.175	0.069	0.044	0.887

The variance accounted for by each factor, indicated by the eigenvalue line highlighted by **bold** in Table 41, shows only two factors that meet Kaiser's  $> 1$  selection criteria of variance accounted for. This pattern was universal in preliminary tests of various conditions; therefore, only two factors were extracted in subsequent factor analyses.

Table 41 also shows loadings for the first two factors (shown in *italics*) in relatively narrow ranges of high loadings (near  $\pm 1$ ), particularly in factor one. Therefore, varimax rotation of the loadings matrices were performed on all subsequent factor analyses to minimize the number of high loadings on each factor which yielded reduced dimensionality and simplified interpretation of the factors.

The communality column in Table 41 approximately represents the variance explained by the variables within the factors. Thus, it can loosely be compared to an  $R^2$  measure. From the varimax rotation technique, these values will all be artificially high and, therefore, will not be considered in the overall interpretation of the factor analyses.

Tables 42 through 49 show summary tables of the two calculated factors for the eight combined conditions of experiment one using the principal components factor analysis method and varimax rotation. Loadings for factors one and two are indicated in **bold** if the loadings plots and experimental evidence indicate they are important variables for interpreting the respective factor. In terms of the loadings values, scores closer to -1 or +1 are expected to have greater impact on

the respective factor. In addition, the relative size of each variable loading in comparison with other loadings in the respective factors is also indicative of interpretation importance.

The loadings plots for factor analyses are often more informative than the summary tables in interpreting the factors and dependent variable relationships. Of particular relevance are the spatial clusters of variables, the length of loading lines in the plots, and directly polar variables. These events are apparent in many of the loadings plot figures presented here with the summary tables.

Table 42. Factor analysis table for the data collapsed across solid-color conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.849</b>	0.338	0.836
separable	-0.292	<b>-0.908</b>	0.910
integral	<b>0.746</b>	0.502	0.808
distract	<b>-0.741</b>	-0.485	0.784
disting.	-0.320	<b>-0.898</b>	0.909
blend	<b>0.716</b>	0.614	0.891
useful	<b>0.940</b>	0.217	0.930
unified	0.501	<b>0.771</b>	0.845
Variance	3.6613	3.2515	6.9128
% Var	0.458	0.406	0.864

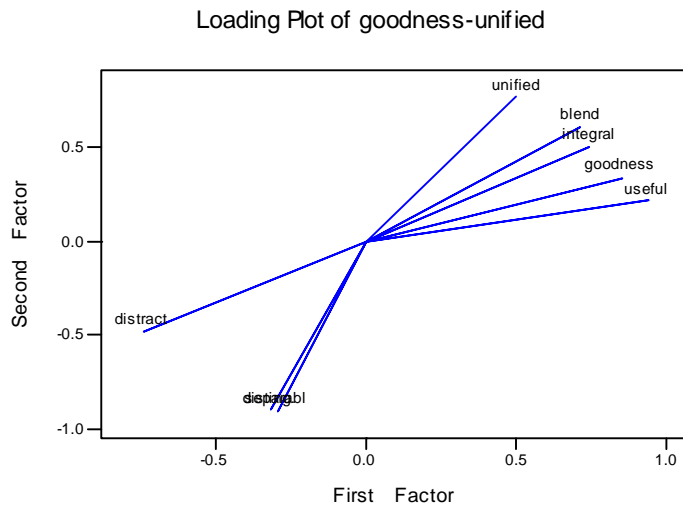


Figure 57. Loadings plot for the solid-color factor analysis.

Table 43. Factor analysis table for the data collapsed across solid-gray conditions.

Variable	Factor1	Factor2	Communality
goodness	0.063	<b>0.865</b>	0.751
separable	<b>-0.885</b>	-0.151	0.806
integral	<b>0.651</b>	0.440	0.618
distract	-0.687	<b>-0.567</b>	0.793
disting.	<b>-0.898</b>	0.121	0.822
blend	<b>0.664</b>	0.592	0.792
useful	0.172	<b>0.874</b>	0.793
unified	<b>0.824</b>	0.375	0.820
Variance	3.6404	2.5550	6.1954
% Var	0.455	0.319	0.774

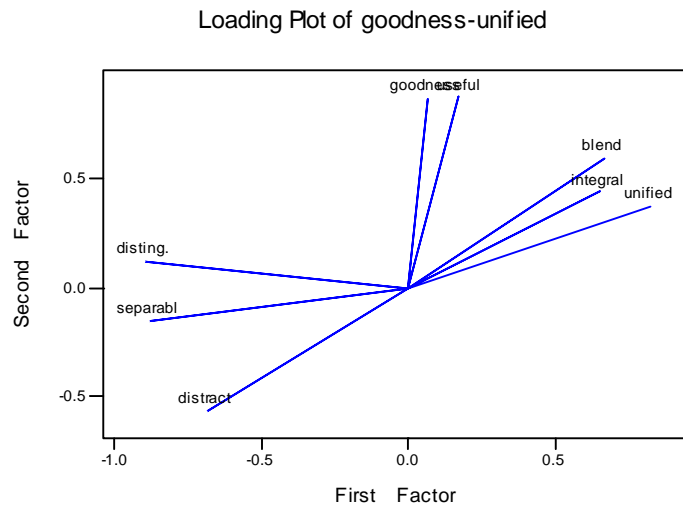


Figure 58. Loadings plot for the solid-gray factor analysis.

Table 44. Factor analysis table for the data collapsed across solid-white conditions.

Variable	Factor1	Factor2	Communality
goodness	0.387	<b>0.732</b>	0.686
separable	<b>-0.917</b>	-0.165	0.867
integral	<b>0.790</b>	0.452	0.828
distract	-0.686	<b>-0.547</b>	0.770
disting.	<b>-0.904</b>	-0.237	0.874
blend	<b>0.689</b>	0.569	0.798
useful	0.161	<b>0.911</b>	0.856
unified	<b>0.820</b>	0.400	0.833
Variance	4.0745	2.4369	6.5114
% Var	0.509	0.305	0.814

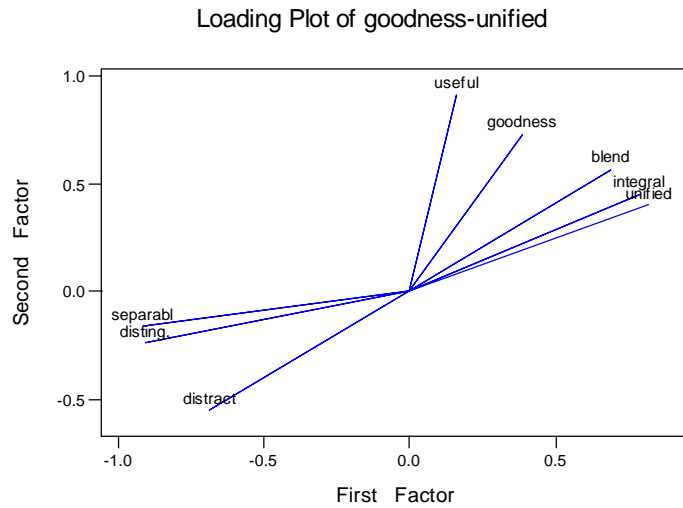


Figure 59. Loadings plot for the solid-white factor analysis.

Table 45. Factor analysis table for the data collapsed across solid-wire conditions.

Variable	Factor1	Factor2	Communality
goodness	0.066	<b>0.864</b>	0.750
separable	<b>-0.905</b>	0.036	0.820
integral	<b>0.746</b>	0.306	0.650
distract	-0.501	<b>-0.587</b>	0.596
disting.	<b>-0.860</b>	-0.194	0.777
blend	0.626	0.626	0.783
useful	0.164	<b>0.817</b>	0.695
unified	<b>0.745</b>	0.426	0.737
Variance	3.3439	2.4642	5.8081
% Var	0.418	0.308	0.726

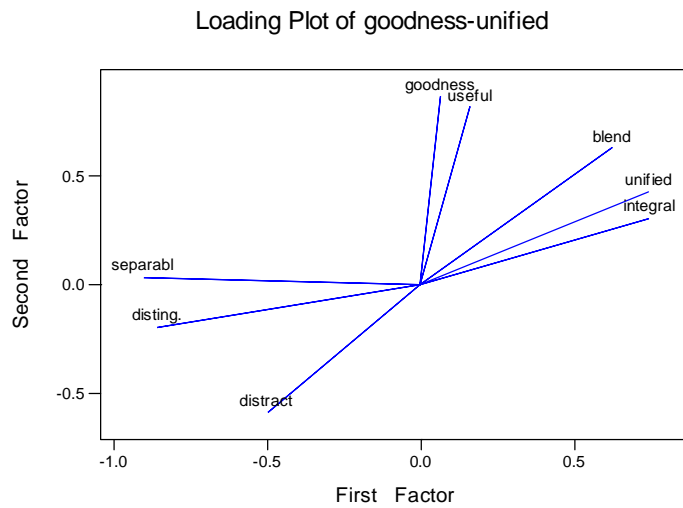


Figure 60. Loadings plot for the solid-wire factor analysis.

Table 46. Factor analysis table for the data collapsed across transparent-color conditions.

Variable	Factor1	Factor2	Communality
goodness	0.197	<b>0.886</b>	0.823
separable	<b>-0.893</b>	-0.221	0.847
integral	<b>0.588</b>	0.575	0.675
distract	-0.592	<b>-0.657</b>	0.782
disting.	<b>-0.921</b>	-0.132	0.866
blend	<b>0.764</b>	0.502	0.836
useful	0.229	<b>0.899</b>	0.861
unified	<b>0.796</b>	0.465	0.851
Variance	3.6510	2.8901	6.5411
% Var	0.456	0.361	0.818

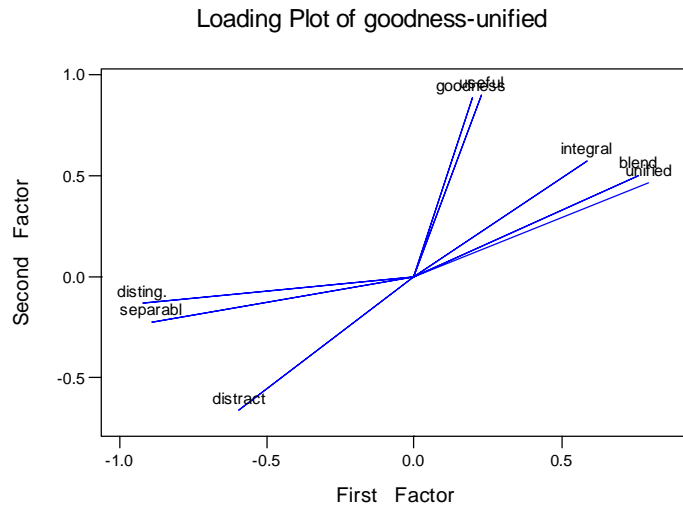


Figure 61. Loadings plot for the transparent-color factor analysis.

Table 47. Factor analysis table for the data collapsed across transparent-gray conditions.

Variable	Factor1	Factor2	Communality
goodness	0.260	<b>-0.804</b>	0.714
separable	<b>-0.917</b>	0.115	0.853
integral	0.646	-0.477	0.644
distract	-0.689	<b>0.576</b>	0.806
disting.	<b>-0.847</b>	0.100	0.727
blend	<b>0.762</b>	-0.485	0.816
useful	0.126	<b>-0.900</b>	0.825
unified	<b>0.818</b>	-0.375	0.809
Variance	3.7816	2.4139	6.1954
% Var	0.473	0.302	0.774

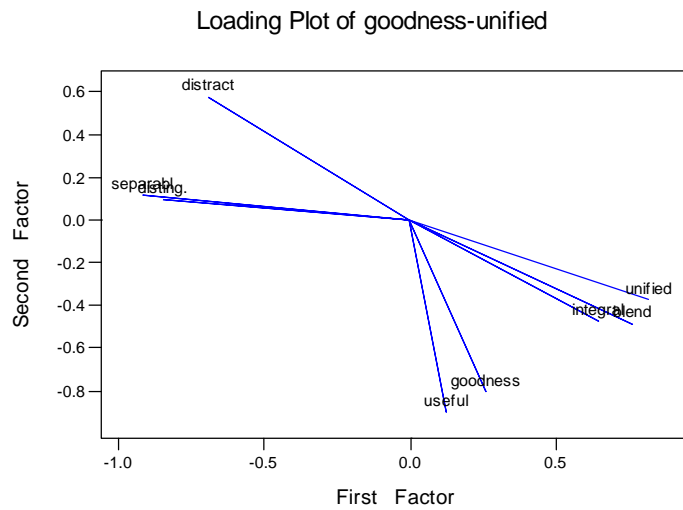


Figure 62. Loadings plot for the transparent-gray factor analysis.

Table 48. Factor analysis table for the data collapsed across transparent-white conditions.

Variable	Factor1	Factor2	Communality
goodness	0.123	-0.845	0.729
separable	-0.878	0.162	0.797
integral	0.717	-0.404	0.678
distract	-0.730	0.423	0.712
disting.	-0.862	-0.027	0.744
blend	0.773	-0.423	0.776
useful	0.234	-0.869	0.809
unified	0.890	-0.225	0.842
Variance	4.0199	2.0676	6.0875
% Var	0.502	0.258	0.761

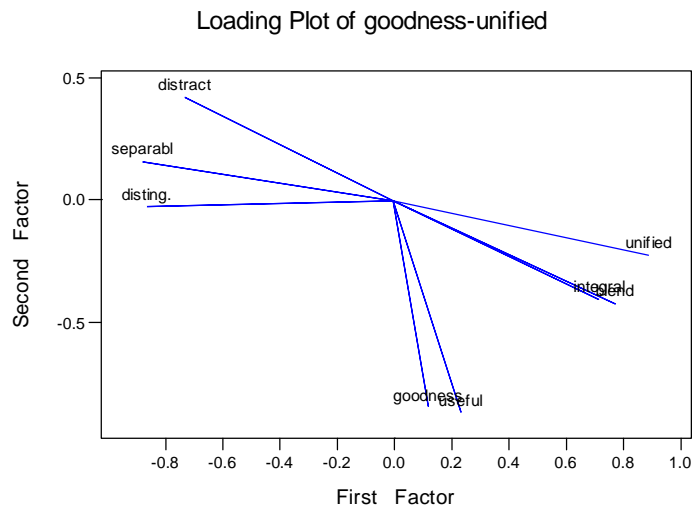


Figure 63. Loadings plot for the transparent-white factor analysis.

Table 49. Factor analysis table for the data collapsed across transparent-wire conditions.

Variable	Factor1	Factor2	Communality
goodness	0.430	<b>-0.793</b>	0.813
separable	<b>-0.883</b>	0.296	0.868
integral	<b>0.690</b>	-0.511	0.737
distract	-0.701	<b>0.508</b>	0.749
disting.	<b>-0.870</b>	0.233	0.811
blend	<b>0.742</b>	-0.518	0.819
useful	0.263	<b>-0.906</b>	0.890
unified	<b>0.832</b>	-0.455	0.900
Variance	4.0012	2.5857	6.5869
% Var	0.500	0.323	0.823

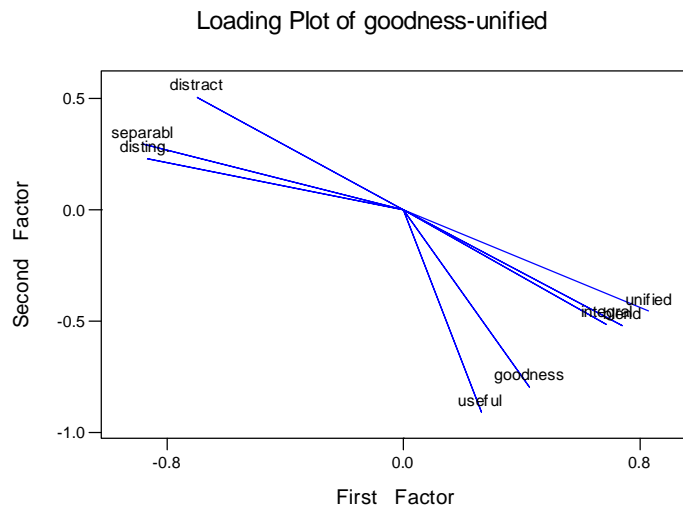


Figure 64. Loadings plot for the transparent-wire factor analysis.

### 5.3.5. Correlations

Correlations and corresponding significance p-values were calculated to find the degree of relationship between the tested variables from experiment one. Table 50 lists and describes the various variables that were included in the correlation computations. Table 51 shows the correlations and corresponding p-values for the tested variables. Significant correlations and their p-values are highlighted in **bold** in Table 51.

Table 50. A list and description of the variables from the correlation tests for experiment one.

Variable	Variable Description
q1	What is the goodness rating of the augmented scene (1 high, 7 low).
q2	How separable were graphics from the real scene rating of the augmented scene (1 very, 7 not at all).
q3	How integral were graphics with the real scene rating of the augmented scene (1 very, 7 not at all).
q4	Were graphics distracting rating of augmented scene (1 very, 7 not at all).
q5	How distinguishable were graphics from real scene rating of the augmented scene (1 very, 7 not at all).
q6	How well did graphics and real scene blend rating of the augmented scene (1 very, 7 not at all).
q7	Were the graphics useful to the overall scene rating of the augmented scene (1 very, 7 not at all).
q8	Unified or distinct perception of whole scene rating of the augmented scene (1 unified, 7 distinct).
rotate	Mental rotation test score (higher is better).
age	Age of the subjects (years).
gender	Gender of subjects (1 male, 2 female).
comp	Self-rating for time spent working on computers each week (1 low, 6 high).
games	Self-rating of time spent playing computer games (1 low, 7 high).
mental	Self-assessment of the ability to mentally manipulate objects within the mind (1 high, 7 low).

Table 51. Correlations and correlation p-values for tested variables from experiment one.

	q1	q2	q3	q4	q5	q6	q7	q8	rotate	age	gender	comp	games
<b>q2</b>	0.116												
	0.750												
<b>q3</b>	0.063	<b>-0.737</b>											
	0.863	<b>0.015</b>											
<b>q4</b>	-0.152	<b>0.718</b>	<b>-0.733</b>										
	0.674	<b>0.019</b>	<b>0.016</b>										
<b>q5</b>	0.012	<b>0.902</b>	<b>-0.683</b>	<b>0.690</b>									
	0.973	<b>0.000</b>	<b>0.029</b>	<b>0.027</b>									
<b>q6</b>	0.629	-0.532	0.562	<b>-0.822</b>	-0.539								
	0.052	0.113	0.091	<b>0.004</b>	0.108								
<b>q7</b>	0.588	-0.132	0.067	-0.508	-0.201	<b>0.651</b>							
	0.074	0.715	0.854	0.134	0.578	<b>0.041</b>							
<b>q8</b>	0.307	<b>-0.843</b>	<b>0.730</b>	<b>-0.677</b>	<b>-0.818</b>	<b>0.661</b>	0.216						
	0.389	<b>0.002</b>	<b>0.016</b>	<b>0.031</b>	<b>0.004</b>	<b>0.037</b>	0.549						
<b>rotate</b>	-0.200	-0.447	0.166	-0.270	-0.439	0.090	-0.217	0.484					
	0.579	0.196	0.648	0.451	0.205	0.804	0.547	0.156					
<b>age</b>	0.154	<b>0.735</b>	<b>-0.651</b>	0.512	<b>0.769</b>	-0.343	-0.209	-0.513	-0.220				
	0.671	<b>0.015</b>	<b>0.041</b>	0.130	<b>0.009</b>	0.333	0.561	0.129	0.541				
<b>gender</b>	-0.108	-0.008	0.034	-0.083	-0.341	0.044	-0.054	-0.046	-0.138	-0.231			
	0.767	0.982	0.925	0.819	0.334	0.905	0.882	0.899	0.703	0.521			
<b>comp</b>	-0.485	-0.221	0.441	-0.571	-0.253	0.119	0.004	-0.012	0.144	-0.352	0.327		
	0.156	0.539	0.202	0.085	0.481	0.743	0.992	0.974	0.692	0.318	0.356		
<b>games</b>	-0.209	0.218	0.133	0.397	0.213	-0.371	-0.346	-0.380	-0.387	-0.317	0.063	0.055	
	0.562	0.546	0.715	0.257	0.554	0.291	0.327	0.279	0.270	0.372	0.864	0.881	
<b>mental</b>	-0.084	0.039	0.201	0.132	0.015	-0.148	0.130	-0.262	<b>-0.670</b>	-0.375	0.060	0.053	<b>0.737</b>
	0.817	0.914	0.579	0.716	0.967	0.682	0.721	0.465	<b>0.034</b>	0.286	0.869	0.885	<b>0.015</b>

## **5.4. Discussion**

### *5.4.1. Order*

Each participant in experiment one made eight subjective assessments for 64 randomly ordered augmented scenes (comprising 32 unique scenes presented twice). Subjects were given demonstrations and asked to practice making assessments, however, some subtle practice effects were likely to occur during the experiment itself. Between the large number of conditions and the random order, these effects were hoped to be mitigated. The ANOVAs for both overall order (1-64) and individual scene presentation order (1,2) showed no significant effects for order as measured by any of the eight subjective assessments.

### *5.4.2. Augmented Object, Transparency of Graphics, and Texture of Graphics*

Experiment one began the investigation into the theoretical background of integral graphics in augmented reality and the development of design guidelines. The main goal was to establish which texture and transparency levels would influence integral perception in a variety of augmented scenes. The four augmented scenes tested were a brain, a computer outline, text, and a liquid dynamic model. There were a variety of interesting results concerning these four augmented object applications across the other two independent variables and the eight subjective assessment measures. Patterns in the results rather than specific individual effects were used to convey the most information from the statistical analyses. The discussion will first cover the main effect results, followed by the two-way interactions. The three-way interactions are considered in the same context as the two-way interactions where they are most applicable.

The augmented reality object (computer, brain, text, dynamic) only showed a significant main effect in the ANOVA performed on the goodness dependent measure. Figure 21, showing this effect, indicated that the text condition received goodness ratings much worse than any of the other three conditions. The text condition was the simplest visual scene among the four augmented objects tested in experiment one. This simple presentation of real white text was

perhaps most affected by any superimposed graphics that deviated from white letters. In the text augmented scene, any texture condition other than white may have led to poorer overall goodness ratings of the graphics. Examining the many significant interactions involving the independent variable augmented object will provide more information on this topic.

The transparency independent variable (solid, transparent) showed significant main effects when measured by separable, distracting, distinguishable, and unified scene ratings. For each of the transparency effects, the corresponding figures showed that the transparent condition was always rated more favorably than the solid graphic conditions. This result was a mild surprise as the more vivid and realistic graphics that the head-mounted display is capable of is considered a strength of that display medium over see-through systems. If the design goal is integrally perceived graphics, it would seem that the use of see-through display systems would be entirely appropriate and that head-mounted displays would need to use some device to make their graphics semi-transparent. As with the independent variable augmented object, the significant interactions provide more specified information on the transparency effect.

The texture independent variable (color, gray, white, wireframe) showed significant main effects when measured by goodness, distracting, blending, usefulness, and unified scene ratings. Examining the trend lines for this main effect show that the wireframe texture condition was universally viewed as the poorest condition, receiving the worst ratings for each of the measures listed above. This result is not surprising as the wireframe condition visually is the most distorted of the four texture levels.

In addition to the poor wireframe ratings, in all the graphs there is an indication for slightly more positive ratings for the white texture condition. Previously the text augmented condition was mentioned as achieving the highest integral perception when presented with the white textured graphics. We will later see that the computer augmented scene also took greatest advantage of the white texture conditions. These two conditions likely caused the positive influence on the increased white texture integral perceptions across the different subjective measures. Again, the interaction discussion will more clearly illuminate the impact of this independent variable.

While the main effects begin the discussion on the independent variables for experiment one, it is the interactions that provide the most insight on the true significance of the independent variables in this study. The first interaction to be examined will be between the independent variables augmented object and transparency, which were significant in six of the eight subjective measures. In the goodness measure, this interaction highlights what was alluded to by the significant main effect of augmented object (which was the only measure that showed augmented object to be significant). Figure 23 shows that the text condition receives slightly worse ratings than each of the other augmented scenes when the objects were presented transparently. The integral ratings also show this effect in its augmented object and transparency interaction, as well as showing that the text conditions received higher ratings than the other conditions when the graphics were presented solidly. This further confirms that the text condition needed graphics as vivid and white as possible to get the best ratings.

In another result in the augmented object and transparency interaction, the separable and distinguishable ratings, and to a lesser extent the blending ratings, show the computer scene as having better overall integral perception than the other augmented scenes for the transparency condition. This result is likely an artefact of the computer-transparent scenes, especially when presented with the wireframe texture. These augmented scenes were difficult for some subjects to see in the head-mounted display. When this occurred, instead of thinking the graphics were poor, subjects assumed that the graphics were so well blended and not separable or distinguishable that they needed to give them high ratings. This highlights a limitation in the head-mounted display rather than a true result of the computer and transparency conditions.

Even more poignant than the previous interaction results is the information found in the interaction between the independent variables augmented object and texture, significant across all eight subjective assessments. It is perhaps the most consistent, clear, and meaningful result in the full body of this dissertation. Each of the four augmented scenes show a distinct pattern across the texture conditions that is consistent throughout all eight measures.

Earlier in this discussion, the text scene was presented as the worst case condition of the augmented object independent variable when measured by goodness ratings. An explanation was

put forth that the text condition would be most impacted by any texture other than matching white graphics with its white real letters. This is clearly shown in the interaction of augmented object and texture. The text condition in each of the interaction graphics shows a large dip in the ratings (or an increase if that leads to better values) for the white texture condition. While this 'better' rating for text lies in the middle of the ratings for augmented object and texture assessments, the other three texture conditions for text are the worst in the entire set of graphs (other than one liquid dynamic condition to be discussed next). This shows that the text condition had a much narrower range of acceptance for less than perfect integral graphics.

Other than the text condition, the liquid dynamic model has the most obvious extremes in ratings. The color texture condition for the dynamic scene clearly has the best ratings, not only within the liquid dynamic scene, but across all augmented scenes and textures. This was true for each of the eight subjective measures for this interaction. In contrast to the best ratings for the dynamic-color condition, the dynamic-wireframe condition was consistently rated by subjects as among the worst scenes overall. The dynamic-wireframe conditions received the worst, or very nearly the worst ratings for all the dependent measures when compared to all other independent variable combinations. The dynamic nature of this augmented scene appeared to overly highlight the poor quality of the wireframe graphics for subjects.

The computer condition in the interaction between augmented object and texture was the one augmented scene that maintained constant ratings across the four texture levels. In nearly all the graphs there is little variation as the computer trend lines transition from color, gray, white, and wireframe. Since the computer augmented object was an outline rather than a traditional graphic (or a unique situation like the text condition), it seemed more impervious to changing texture conditions than the other augmented objects.

The image of the brain condition in the augmented object and texture interaction did not have as pronounced a pattern as either the text or liquid dynamic conditions. In a few interaction graphs, the brain image did show poor ratings when combined with white texture, however, this effect did not occur in all the dependent measures. In fact, in several cases the brain augmented condition trend lines were nearly as flat as the computer condition across all textures. The exact impact of

the brain image on integral perception ratings will be further explained by the three-way interactions discussed later, which were also significant for all eight subjective measures.

The next two-way interaction to discuss is the interaction between the independent variables transparency and texture, significant only for the distinguishable ratings. At first glance, the corresponding graph for this effect does not appear to offer much in terms of practical significance as the ratings values are all similar. However, the white texture condition does stand out as it remains relatively constant across both transparency conditions while the other three textures all have increased distinguishable ratings (meaning the scenes were less distinguishable) in the transparency condition. This effect is likely due to the strength of the text-white and computer-white conditions ratings as was discussed previously.

The three-way interactions were presented as two two-way interactions, displaying the augmented object by texture graphs for both the solid and transparent conditions. Generally, the three-way interactions confirm other findings, though more magnified in certain cases. The most obvious pattern from the three-way interaction trend lines is the clear difference between the text-solid-white and text-transparent-white conditions. In all the three-way interaction graphs the text-solid-white conditions are displayed with much higher ratings than the text-transparent-white conditions. This again highlights the need for exact graphic matches with the text augmented object condition.

The next most obvious characteristic of the three-way interaction is the difference between the brain image condition at the white texture level. Specifically notable in the solid graphics figure, the brain-white condition receives much poorer ratings throughout the dependent measures. While this is interesting in itself, the various effects that were discussed as showing more positive ratings in the white condition can be considered even stronger. The brain-white condition is counteracting the effects of the text-white and computer-white conditions which yield more positive white texture ratings overall.

Another pattern that can be seen when considering all the three-way graphs is the generally worse ratings for all the solid conditions, except for the text and dynamic-color conditions. The text

effect has been discussed several times already. The dynamic-color conditions were rated much more integrally in both the solid and transparent graphs, however, the other texture levels for the dynamic-solid conditions fared much worse. This effect alludes to the need for dynamic augmented realities to have more integrally perceived graphics than other applications, much like the text augmented scenes with white textures.

Another effect that was often demonstrated in the three-way interaction graphs was the much worse integral perception for the computer-solid-color condition. This was not surprising as the use of color in the computer outline does not add any useful information to the computer.

The corresponding nonparametric tests of the significant main effects from the ANOVA analyses added little to the overall discussion on the independent variables. It was interesting to confirm that the one single significant result for the independent variable augmented object as measured by goodness was supported by a significant nonparametric result. All the other main effects results were confirmed as well by the nonparametric tests.

#### *5.4.3. Subjective Rating Scales*

Factor analyses were completed for the eight subjective rating scale measures to examine underlying relationships between the dependent variables. The main expectations from these analyses were clusters of dependent measures, meaningful and interpretable factors, and possibly dominant factors apparent in the graphs of the factor analyses loadings plots.

The first goal, to see if similar subjective measures would cluster, proved to be very successful. The loadings plots broke down into four easily identifiable clusters across all the factor analyses performed. The first cluster at one extreme of factor one (either positive or negative depending on the exact data included in the factor analysis) contained the integral, blending, and unified scene measures. The second cluster at the other extreme of the first factor had the separable and distinguishable measures. There was one cluster on the second factor containing the goodness and usefulness measures. The distracting measure was the only remaining measure, which formed its own cluster generally opposite both the goodness and useful measures, although there is

obvious influence from both factors one and two on the distracting measure. All the graphs show this general distribution of factors, however, some analyses are clearer than others. Figure 61 showing the loadings plot for the transparent-color conditions amply demonstrates the distribution of clusters.

The relationships highlighted in the clusters are not that surprising considering the generally accepted meaning of the words used in the eight dependent measures. It is interesting to note that the distracting measure did not more closely couple with the separable and distinguishable ratings. There was also a mild expectation that subjects would more closely associate the psychological term goodness with integralness. This turned out not to be the case as the goodness measure was clustered with usefulness on factor two while the integral measure was on factor one. The usefulness measure was intended to capture a different segment of information than the other integrally defined rating scales, and it appears that this was achieved in conjunction with the goodness construct.

Cluster results, depending on how tightly associated the measures are, indicate a very close relationship between dependent measures. Thus the integral, blending, and unified scene rating scales appear to be based on the same underlying meaning. As in turn are the separable and distinguishable questions, and lastly the usefulness and goodness measures. The distracting measure appears to have no underlying commonalities with the other measures. Another pattern to look for in factor analyses loadings plots is measures that are directly opposite each other, forming a near straight line between two points.

Examining the loadings plots for directly opposite measures reveals that it is not the classically used term separable that appears opposite the integral measure, but the solo distracting measure. This is a very interesting result considering the long established tradition in the literature for considering the integral and separable terms as polar opposites. This would indicate that it is more appropriate to discuss non-integral features of augmented environments, not as separate from the real scene, but distracting. That is a more powerful implication than mere separation. One possible confound to this conclusion is that factor one is possibly tainted by the weighting given to the separable and distinguishable questions rather than more clearly being associated with

the distracting measure.

Further interpretation of the factor analyses can be garnered from dominant measures in the loadings plots. Dominant factor measures would be identified by obviously long dependent measure lines emanating from the center over other shorter measure lines. There does not appear to be any consistently dominant measure in these factor analyses results. That is not surprising considering how many of them were related. This finding indicates that there are many appropriate ways of subjectively capturing integral perception information.

Aside from the very interesting loadings plots, there are also the calculated factors themselves to consider. The factors are the axes lines in the loadings plots. However, the factors take on additional meaning when considering the factor values for the respective measures shown in the factor analyses tables. The apparent pattern is that for factor one the integral, blending, and unified scene measures are at one extreme and the separable and distinguishable measures are at the other. For factor two the goodness and usefulness measures are one end, with the distracting measure at least nominally representing the other extreme. The underlying meaning behind the two factors seems to represent a global integral perception for the first factor and a general overall quality quotient for the second factor.

#### *5.4.4. Correlations*

There were almost 20 significant correlations, with all but five between pairs of various subjective rating scales. These subjective measures were purposely designed to be somewhat related, as was already covered in detail by the factor analysis discussion, so the significant correlations are no surprise.

There were also three significant correlations that involved gender and one or another of the subjective assessment rating scale measures. While these correlations are significant, tempered conclusions should be made since there are only two values for gender (1,2) and there were only two female participants in the first experiment. More data would be needed before serious discussion of these results could be undertaken.

The last two significant correlations involved participant self-assessment of mental manipulation ability. The first was a significant correlation with scores on the mental rotation test. While this was an expected result, it was good to see that people are decently able to assess their own spatial ability. The last significant correlation was between self-assessment of mental manipulation ability and time spent playing computer games. This is not an unusual result as frequent play in 3D computer games is likely to increase spatial ability. At least, as the significant correlations indicates, the perception that spatial ability increases.

#### *5.4.5. Design Guidelines*

There was an abundance of useful information in the data analysis and discussion for the first experiment that can be summarized in design guidelines. Table 52 states design guidelines as they pertain to either human factors, perception, and graphic design in augmented reality.

Table 52. Design guidelines from experiment one.

<b>Guideline</b>	<b>Implication</b>
1. For more simplistic augmented applications, greater attention must be paid to creating more integral graphics.	Graphic Design
2. Use integral graphics to avoid the perception that the graphics are distracting to the overall scene.	Perception
3. All things being equal, transparent graphics should be used for greater integral perception in augmented reality applications.	Perception, Graphic Design
4. Three-dimensional augmented scenes should be designed with transparent graphics to aid the illusion of depth.	Graphic design
5. Dynamic augmented reality applications need very integrally designed graphics to maintain acceptable levels of integral perception.	Perception, Graphic Design
6. To achieve higher levels of integral perception with outline graphics, they should be displayed clearly, although high quality graphics are not necessary.	Graphic Design
7. Use color textures for increased integral perception only if color adds meaningful information to the augmented scene.	Perception, Graphic Design
8. See-through display systems should be avoided when high integral perception is desired in augmented applications.	Human Factors
9. The quality of the display system used to view the augmented scenes plays a key role in the ultimate integral perception of augmented graphics.	Human Factors
10. For assessing integral perception, use subjective measures defined as integral, blending, or unified scene.	Human Factors

These guidelines are based on the premise that integrally perceived graphics are scenes composed of both real and computer-generated imagery that is perceived as a unified whole. The guidelines here are lessons learned from experiment one. They comprise guidelines for the general use of integral graphics in augmented reality, creating integral graphics for augmented applications, using integral graphics in appropriate augmented systems and applications, and assessing integral perception augmented reality experimentally.

## CHAPTER 6. EXPERIMENT TWO

### 6.1. Methodology

#### 6.1.1. Experimental Design

The second experiment continued the efforts of the first study in investigating the influence of various parameters of augmented environments on integral perception. Unlike experiment one, in this study the specific focus was on augmented environments of multiple objects rather than single images. As with the first experiment, three independent variables and eight subjective dependent measures comprised an all within-factor experimental design.

The independent variables tested were level of integrality (high/low), as defined by experiment one, number of graphic objects included in the augmented scene (2,4,6,8), and augmented environment (houses, computer lab, planets, trees). The three variables were factorially combined to form a 4x2x4 design (32 unique stimuli). The 32 unique scenes were shown twice to each subject for a total of 64 trials, each presented in random order.

The integrality variable was included to specifically compare high and low integrally designed augmented environments and to confirm results from experiment one. The varying number of graphics were chosen to determine the effects of increasing number of graphics on integral perception in augmented environments. The augmented environments independent variable represented possible applications where a scene of real objects would be filled out with similar graphics objects.

Again following the methodology of experiment one, eight subjective 7-point rating scale dependent measures were collected for each trial presentation of an augmented scene. The second study used a similar set of subjective ratings to experiment one, however, there were some additions and subtractions to the eight ratings. The questions for experiment two involved both the level of integration of the graphics into the overall augmented scene and the number of graphics that were present in the scene. In summary, the ratings were on goodness, integral,

attend, clutter, separable, similar, recognition, and usefulness. Specifically, the goodness, integral, separable, and usefulness questions were included based on the same reasons discussed in the experiment one section; the attend question stems from the selective attention premise; the clutter question is a direct extension of the issues presented earlier on clutter; the similarity question is based on the Gestalt law of similarity; and, the recognition question stems from the recognition-by-components theory. Table 53 summarizes the eight dependent measures as they were defined for experiment two. Appendix D shows the exact rating scale questionnaire used by the subjects to assess each trial scene in experiment two.

Table 53. Subjective rating scale descriptions for experiment two.

<b>Variable Description</b>
What is the goodness rating of the augmented scene (1 high, 7 low).
How separable were graphics from the real scene rating of the augmented scene (1 very, 7 not at all).
How integral were graphics with the real scene rating of the augmented scene (1 very, 7 not at all).
Tendency to focus on graphics or real scene rating of augmented scene (1 graphics, 7 real).
How cluttered did graphics make the whole scene rating of the augmented scene (1 very, 7 not at all).
How similar were graphics and real objects rating of the augmented scene (1 very, 7 not at all).
Recognition of scene affected by graphics rating of the augmented scene (1 very, 7 not at all).
Were the graphics useful to the overall scene rating of the augmented scene (1 very, 7 not at all).

### *6.1.2. Participants*

Ten participants, aged 22-35, came from the general Virginia Tech university population. To be eligible for the study, participants needed to correctly perceive at least 7 out of 8 items on the 20/40 line of the eye chart and correctly identify 7 out of 8 color vision trials. No participant failed to meet these requirements.

### *6.1.3. Experimental Apparatus*

6.1.3.1. Equipment Setup. The equipment to run the augmented reality portion of experiment two was the exact same as used in experiment one, previously described in the equipment setup section of experiment one and shown in Figures 10, 11, and 12.

6.1.3.2. Augmented Environments. The equipment used to create the augmented environments in experiment two was the same as that used in experiment one, as was described in the augmented environments section of experiment one. The stimuli themselves, however, were different from those in experiment one.

As with experiment one, an attempt was made to create augmented environments that matched up with either actual or potential real-world uses of augmented reality. Unlike experiment one, the intent was to examine augmented environments rather than augmented objects. In experiment two, the real scenes were physical areas that contained multiple objects, with room for additional similar objects. Graphics of objects similar to those contained in the real scene were then used to fill in the spaces with a specified number of additional objects.

The real scenes were a partially completed housing development, a half empty computer lab, a planetary scene of outer space, and a view of a countryside sparsely filled with trees. As in experiment one, the graphics were superimposed onto the real scenes with the use of a chroma key feature of a video mixer. The final augmented scenes were then transferred to tape in 64 twenty second clips presented in random order. Figures 65 through 70 show six example images from the second study.



Figure 65. Houses-low-six condition for experiment two.



Figure 66. Houses-high-four condition for experiment two.



Figure 67. Lab-high-four condition for experiment two.



Figure 68. Lab-low-eight condition for experiment two.



Figure 69. Planets-high-two condition for experiment two.



Figure 70. Trees-low-eight condition for experiment two.

6.1.3.3. Questionnaires. There were two questionnaires used in the second experiment: a general information questionnaire (shown in Appendix B) and a mental rotation test. Both of these questionnaires are the same ones used and described in the questionnaires section of experiment one.

6.1.3.4. Vision Tests. The vision tests and criteria for participation were the same as used in experiment one. The two tests, acuity and color vision, were described in the vision tests section of experiment one.

#### *6.1.4. Experimental Procedures*

The procedures for the second experiment are nearly identical to the first. As in experiment one, participants were first greeted and asked to read and sign the consent form, see Appendix C. Upon agreeing to participate, subjects completed the general questionnaire, the mental rotation test, and the two visual tests. The main portion of the study, the subjective assessments of augmented scenes, was then explained.

First, the nature of the experiment was described verbally by the experimenter. Participants were then told that they were going to make eight subjective ratings for 64 video clips of pre-recorded augmented environment scenes. Different from the first study is that the augmented environments were explained as scenes of objects shown with additional graphic objects. The participants then read the subjective ratings sheet (see Appendix D) that was to be used to make the assessments. Acknowledging that they understood the ratings, the participants were shown a demonstration video clip by the experimenter of four scenes from the study. The participants were asked to practice making judgements “within their head” and ask any questions. Lastly, the use of the head-mounted display was explained and fitted to the participants head. The subjects then began viewing the 64 augmented video clip scenes and making subjective assessments. The experimenter paused the videotape after each 20 second clip to record subjective assessments. At this time, the participant verbally expressed subjective ratings to the experimenter after lifting the front of the head-mounted display to view the subjective ratings sheet. Any questions participants had during this or any portion of the experiment were answered by the experimenter.

Upon completing the subjective assessments of the augmented environments, the participants were informed the experiment was over and were thanked for their time.

## **6.2. *Expected Results***

Experiment two was designed to elicit differences in subjective assessment of multiple object augmented environments. The main hypothesis was that the multi-object augmented environments can be perceived integrally, however, integral perception would decrease with increasing numbers of graphic objects. Furthermore, it was expected that the environments judged 'best' would be the two-object integrally designed augmented scenes. The environments judged 'worst' were expected to be the eight-object non-integrally designed augmented scenes.

Other than the main hypotheses, the results were expected to confirm that the high and low integrally designed graphics would be perceived with congruent integral/separable assessments. As for the interactions, it was expected that augmented scenes with low-integral designed graphics would be more negatively affected by increasing numbers of graphics than augmented scenes with high-integral designed graphics. Furthermore, augmented scenes with high-integral graphics were expected to show evidence for a plateau in negative assessments of augmented scenes with a higher number of graphics. In contrast, negative assessments of augmented scenes with lower number of graphics were expected to continue to worsen across the four levels of graphics quantity. As with experiment one, practice and order effects were not expected as the presentation of the 64 augmented scenes is randomly ordered differently for each subject.

The multivariate factor analyses for experiment two were expected to show the same types of effects as experiment one: clustering of variables, polar pairs of variables, and dominant variables in the two calculated factors. As with experiment one, it was expected that the integral related assessments would show clusters and dominate the calculated factors.

The correlations for experiment two were expected to show similar relationships among dependent measures as in experiment one. A significant relationship between the mental rotation test scores and spatial ability self-assessment ratings was expected. Given the expansive

correlation investigation, other significant results were likely, however, no specific results were designed for.

### **6.3. Results**

As the experimental design and methods were the same for the first two experiments, the same statistical analyses were completed for experiment two as experiment one. Trial and order effects are first presented, then main effects and interactions for the independent variables are shown, followed by corresponding nonparametric tests, next multivariate factor analyses of the eight dependent measures, and lastly correlations among variables.

As in experiment one, parametric ANOVA tests are presented in conjunction with related non-parametric tests. In addition, statistical results in the three-factor within subjects analyses only were considered significant with p-values < 0.01.

All graphs and figures show significant effects using trend lines. The trend line graphs were chosen to highlight the rate of change between various levels of variables in significant results. Any significant effects with a p-value < 0.05/0.01 as indicated are shown with **bold** highlighting in the summary tables.

#### *6.3.1. Trial and Order ANOVAs*

Two different analyses were conducted to investigate any order effects that may have been present in experiment two. One-way ANOVAs were conducted for the overall order (trials 1 through 64) and the scene presentation order of each individual augmented scene (first and second). Each of the eight dependent rating scale measures was tested. Tables 54 through 61 show the ANOVA summary tables for overall order as measured by the eight questionnaire ratings (goodness, integral, attend, clutter, separable, similar, recognition, usefulness).

Table 54. ANOVA summary table for overall order, measured by a goodness rating.

Source	DF	SS	MS	F	P
order	63	121.14	1.92	0.80	0.860
Error	576	1378.30	2.39		
Total	639	1499.44			

Table 55. ANOVA summary table for overall order, measured by an integral rating.

Source	DF	SS	MS	F	P
order	63	159.29	2.53	0.90	0.692
Error	576	1617.70	2.81		
Total	639	1776.99			

Table 56. ANOVA summary table for overall order, measured by an attend rating.

Source	DF	SS	MS	F	P
order	63	88.24	1.40	0.57	0.997
Error	576	1413.20	2.45		
Total	639	1501.44			

Table 57. ANOVA summary table for overall order, measured by a clutter rating.

Source	DF	SS	MS	F	P
order	63	184.70	2.93	0.81	0.856
Error	576	2094.40	3.64		
Total	639	2279.10			

Table 58. ANOVA summary table for overall order, measured by a separable rating.

Source	DF	SS	MS	F	P
order	63	127.38	2.02	0.64	0.986
Error	576	1827.40	3.17		
Total	639	1954.78			

Table 59. ANOVA summary table for overall order, measured by a similar rating.

Source	DF	SS	MS	F	P
order	63	146.55	2.33	0.71	0.955
Error	576	1890.50	3.28		
Total	639	2037.05			

Table 60. ANOVA summary table for overall order, measured by a recognition rating.

Source	DF	SS	MS	F	P
order	63	87.02	1.38	0.39	1.000
Error	576	2051.90	3.56		
Total	639	2138.92			

Table 61. ANOVA summary table for overall order, measured by a usefulness rating.

Source	DF	SS	MS	F	P
order	63	141.15	2.24	0.81	0.845
Error	576	1585.10	2.75		
Total	639	1726.25			

Tables 62 through 69 show the ANOVA summary tables for scene presentation order (spo) as measured by the eight trial questionnaire ratings (goodness, integral, attend, clutter, separable, similar, recognition, usefulness).

Table 62. ANOVA summary table for scene presentation order, measured by a goodness rating.

Source	DF	SS	MS	F	P
spo	1	0.00	0.00	0.00	0.979
Error	638	1499.43	2.35		
Total	639	1499.44			

Table 63. ANOVA summary table for scene presentation order, measured by an integral rating.

Source	DF	SS	MS	F	P
spo	1	2.14	2.14	0.77	0.381
Error	638	1774.85	2.78		
Total	639	1776.99			

Table 64. ANOVA summary table for scene presentation order, measured by an attend rating.

Source	DF	SS	MS	F	P
spo	1	1.06	1.06	0.45	0.503
Error	638	1500.39	2.35		
Total	639	1501.44			

Table 65. ANOVA summary table for scene presentation order, measured by a clutter rating.

Source	DF	SS	MS	F	P
spo	1	0.90	0.90	0.25	0.616
Error	638	2278.20	3.57		
Total	639	2279.10			

Table 66. ANOVA summary table for scene presentation order, measured by a separable rating.

Source	DF	SS	MS	F	P
spo	1	0.62	0.62	0.20	0.652
Error	638	1954.15	3.06		
Total	639	1954.77			

Table 67. ANOVA summary table for scene presentation order, measured by a similar rating.

Source	DF	SS	MS	F	P
spo	1	0.08	0.08	0.02	0.877
Error	638	2036.97	3.19		
Total	639	2037.05			

Table 68. ANOVA summary table for scene presentation order, measured by a recognition rating.

Source	DF	SS	MS	F	P
spo	1	0.08	0.08	0.02	0.880
Error	638	2138.85	3.35		
Total	639	2138.92			

Table 69. ANOVA summary table for scene presentation order, measured by a usefulness rating.

Source	DF	SS	MS	F	P
spo	1	0.04	0.04	0.01	0.904
Error	638	1726.21	2.71		
Total	639	1726.25			

### 6.3.2. Augmented Environment, Integralness, Quantity ANOVAs

Three-factor within subjects ANOVAs were performed for the independent variables augmented environment scene, quantity, and integralness, as measured by the dependent measures from the trial questionnaire rating scales (goodness, integral, attend, clutter, separable, similar, recognition, usefulness). Tables 70 through 74 show the ANOVA summary tables for the three independent variables as measured by the eight trial questionnaire ratings. Corresponding figures for all significant effects are shown.

Table 70. ANOVA summary table for the three independent variables, measured by a goodness rating.

Source	DF	SS	MS	F	P
subj	9	119.452	13.272	1.18	
<b>ar</b>	<b>3</b>	<b>178.517</b>	<b>59.506</b>	<b>9.01</b>	<b>0.000</b>
<b>quantity</b>	<b>3</b>	<b>101.230</b>	<b>33.743</b>	<b>15.29</b>	<b>0.000</b>
<b>integralness</b>	<b>1</b>	<b>248.752</b>	<b>248.752</b>	<b>35.88</b>	<b>0.000</b>
subj*ar	27	178.405	6.608	1.78	
subj*quantity	27	59.567	2.206	1.99	
subj*integralness	9	62.389	6.932	2.08	
ar*quantity	9	12.914	1.435	1.28	0.259
ar*integralness	3	22.205	7.402	2.21	0.110
quantity*integralness	3	0.817	0.272	0.37	0.775
subj*ar*quantity	81	90.602	1.119	1.50	
subj*ar*integralness	27	90.342	3.346	4.48	
subj*quantity*integralness	27	19.855	0.735	0.98	
ar*quantity*integralness	9	14.952	1.661	2.22	0.020
Error	401	299.439	0.747		
Total	639	1499.436			

Figures 71, 72, and 73 show the graphs for the significant results of the goodness measure. Figure 71 shows lower average goodness ratings (indicating worse goodness perception) for the planets augmented condition. Figure 72 displays lower goodness ratings for the two graphics condition. Figure 73 indicates lower goodness ratings for high integral augmented environments.

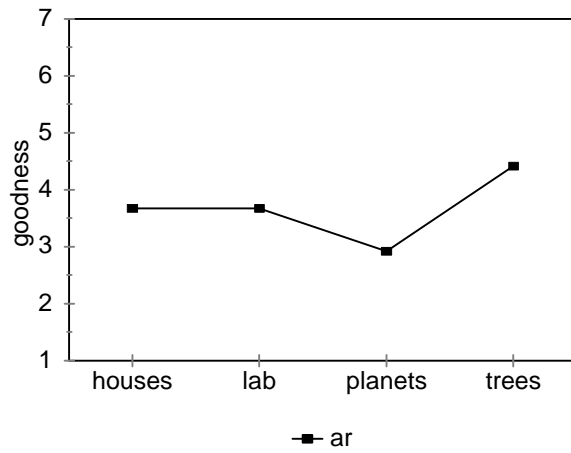


Figure 71. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

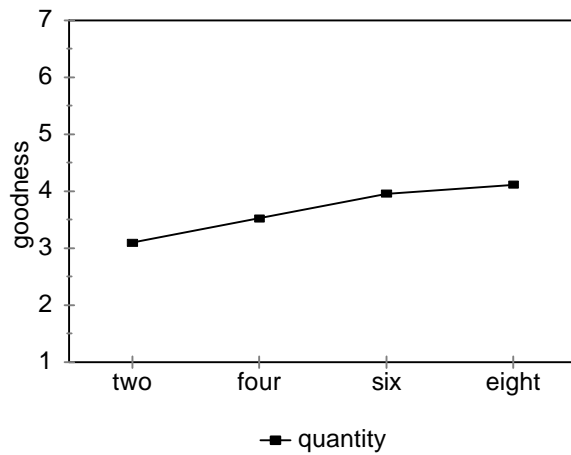


Figure 72. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

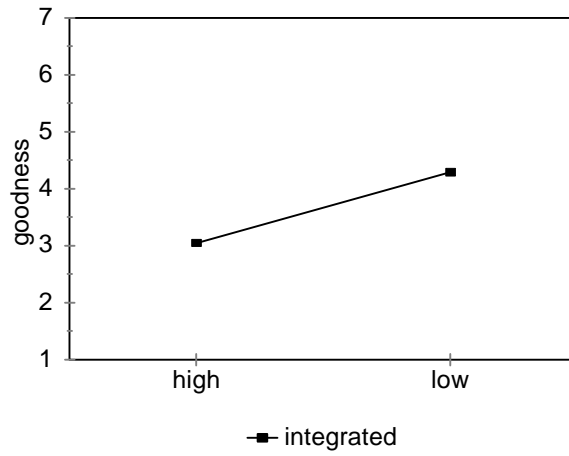


Figure 73. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by goodness ratings.

Table 71. ANOVA summary table for the three independent variables, measured by an integral rating.

Source	DF	SS	MS	F	P
subj	9	187.939	20.882	2.01	
<b>ar</b>	<b>3</b>	<b>233.705</b>	<b>77.902</b>	<b>11.53</b>	<b>0.000</b>
<b>quantity</b>	<b>3</b>	<b>73.705</b>	<b>24.568</b>	<b>14.07</b>	<b>0.000</b>
<b>integralness</b>	<b>1</b>	<b>392.189</b>	<b>392.189</b>	<b>58.75</b>	<b>0.000</b>
subj*ar	27	182.405	6.756	2.00	
subj*quantity	27	47.155	1.746	1.01	
subj*integralness	9	60.077	6.675	1.83	
ar*quantity	9	16.652	1.850	1.63	0.119
ar*integralness	3	32.230	10.743	3.52	0.028
quantity*integralness	3	1.980	0.660	0.47	0.707
subj*ar*quantity	81	91.677	1.132	1.40	
subj*ar*integralness	27	82.317	3.049	3.76	
subj*quantity*integralness	27	38.067	1.410	1.74	
ar*quantity*integralness	9	11.777	1.309	1.61	0.109
Error	401	325.114	0.811		
Total	639	1776.986			

Figures 74, 75, and 76 show the graphs for the significant results of the integral measure. Figure 74 shows lower average integral ratings (indicating more integral perception) for the planets condition and higher average integral ratings (indicating less integral perception) for the trees augmented condition. Figure 75 displays increasing integral ratings across the four levels of the quantity variable. Figure 76 indicates lower integral ratings for high integral augmented environments.

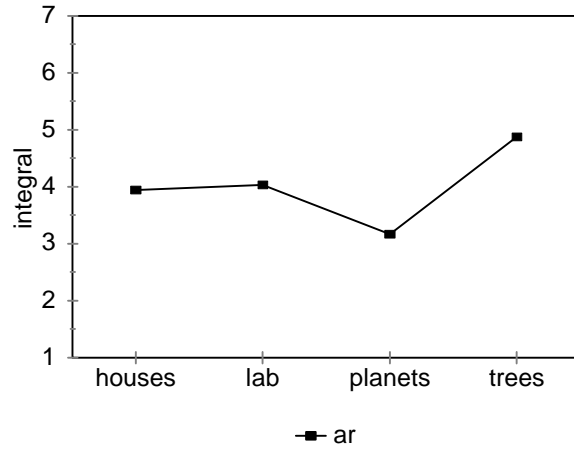


Figure 74. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by integral ratings.

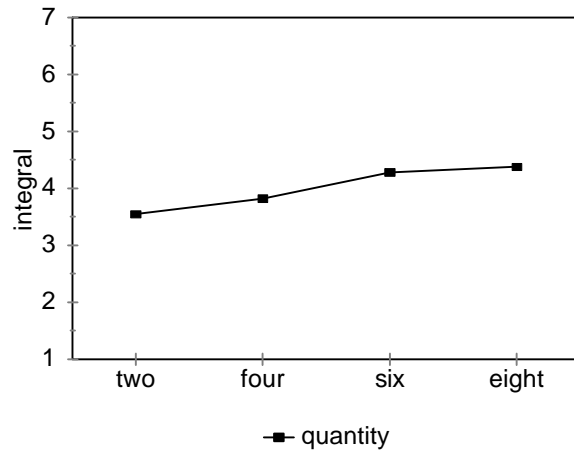


Figure 75. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by integral ratings.

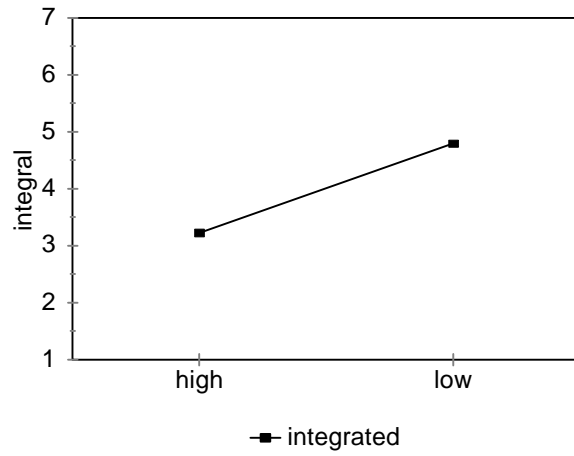


Figure 76. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by integral ratings.

Table 72. ANOVA summary table for the three independent variables, measured by an attend rating.

Source	DF	SS	MS	F	P
subj	9	248.287	27.587	1.86	
<b>ar</b>	<b>3</b>	<b>128.506</b>	<b>42.835</b>	<b>6.85</b>	<b>0.001</b>
<b>quantity</b>	<b>3</b>	<b>98.969</b>	<b>32.990</b>	<b>30.65</b>	<b>0.000</b>
<b>integralness</b>	<b>1</b>	<b>218.556</b>	<b>218.556</b>	<b>19.84</b>	<b>0.002</b>
subj*ar	27	168.900	6.256	2.19	
subj*quantity	27	29.063	1.076	1.15	
subj*integralness	9	99.163	11.018	4.23	
ar*quantity	9	18.819	2.091	2.20	0.030
ar*integralness	3	28.706	9.569	3.65	0.025
quantity*integralness	3	1.569	0.523	0.75	0.532
subj*ar*quantity	81	76.900	0.949	1.33	
subj*ar*integralness	27	70.825	2.623	3.68	
subj*quantity*integralness	27	18.837	0.698	0.98	
ar*quantity*integralness	9	8.319	0.924	1.30	0.237
Error	401	286.025	0.713		
Total	639	1501.444			

Figures 77, 78, and 79 show the graphs for the significant results of the attend measure. Figure 77 shows higher average attend ratings (indicating more real scene attending) for the planets condition and lower average attend ratings (indicating more graphics attending) for the trees augmented condition. Figure 78 displays decreasing attend ratings across the four levels of the quantity variable. Figure 79 indicates higher attend ratings for high integral augmented environments.

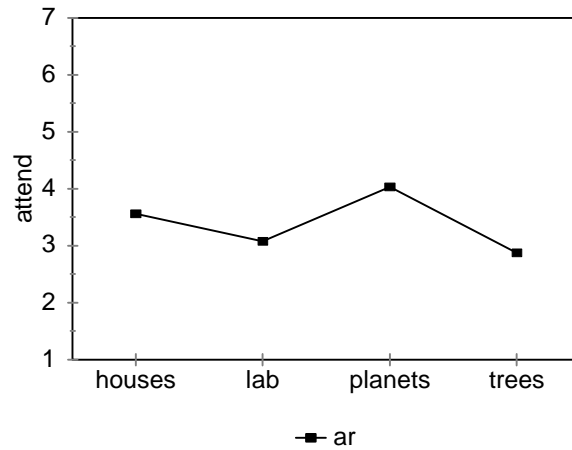


Figure 77. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by attend ratings.

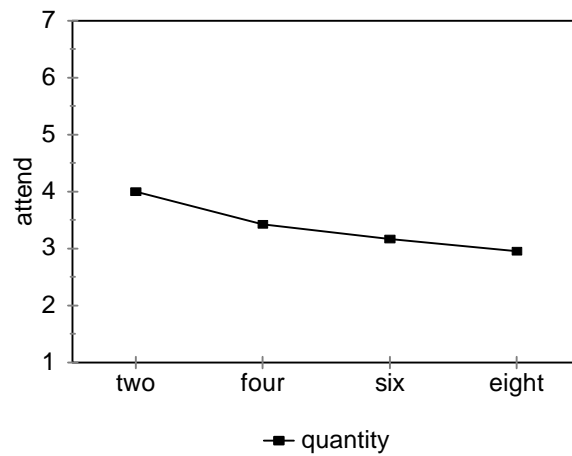


Figure 78. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by attend ratings.

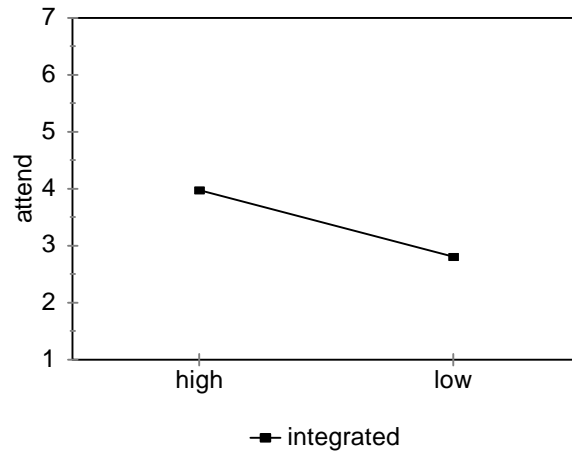


Figure 79. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.002$ ) as measured by attend ratings.

Table 73. ANOVA summary table for the three independent variables, measured by a clutter rating.

Source	DF	SS	MS	F	P
subj	9	294.413	32.713	2.38	
ar	3	15.350	5.117	1.10	0.364
<b>quantity</b>	<b>3</b>	<b>844.162</b>	<b>281.387</b>	<b>59.11</b>	<b>0.000</b>
integralness	1	47.306	47.306	6.32	0.033
subj*ar	27	125.088	4.633	2.57	
subj*quantity	27	128.525	4.760	3.21	
subj*integralness	9	67.381	7.487	4.28	
<b>ar*quantity</b>	<b>9</b>	<b>40.838</b>	<b>4.538</b>	<b>3.29</b>	<b>0.002</b>
ar*integralness	3	11.319	3.773	2.29	0.101
quantity*integralness	3	8.806	2.935	2.22	0.109
subj*ar*quantity	81	111.725	1.379	1.13	
subj*ar*integralness	27	44.494	1.648	1.35	
subj*quantity*integralness	27	35.756	1.324	1.08	
ar*quantity*integralness	9	13.719	1.524	1.25	0.265
Error	401	490.219	1.222		
Total	639	2279.100			

Figures 80, and 81 show the graphs for the significant results of the clutter measure. Figure 80 displays decreasing clutter ratings (indicating increasing clutter perception) across the four levels of the quantity variable. Figure 81 shows a difference in the houses-six condition of the interaction between augmented environment and graphics quantity.

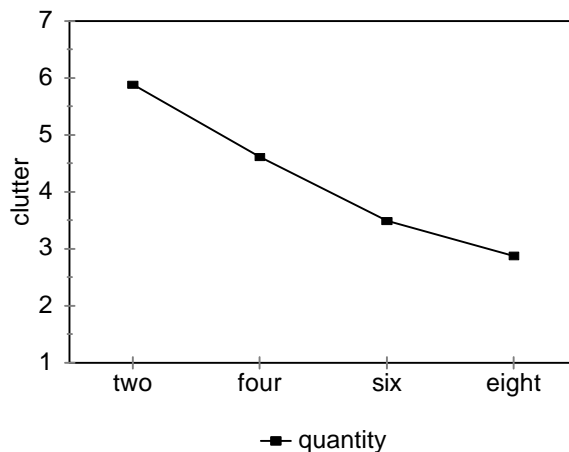


Figure 80. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by clutter ratings.

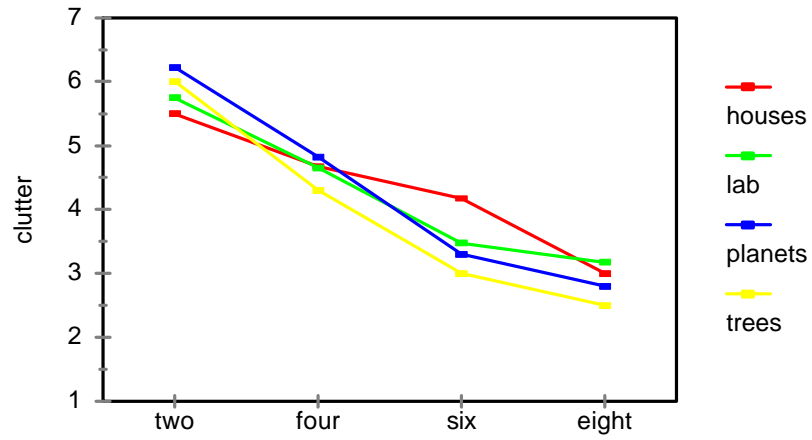


Figure 81. Graph of the interaction between augmented reality scene and quantity, significant at the  $p < 0.01$  level ( $p = 0.002$ ) as measured by clutter ratings.

Table 74. ANOVA summary table for the three independent variables, measured by a separable rating.

Source	DF	SS	MS	F	P
subj	9	117.369	13.041	1.65	
<b>ar</b>	<b>3</b>	<b>416.888</b>	<b>138.963</b>	<b>21.20</b>	<b>0.000</b>
<b>quantity</b>	<b>3</b>	<b>28.912</b>	<b>9.637</b>	<b>7.18</b>	<b>0.001</b>
<b>integralness</b>	<b>1</b>	<b>452.256</b>	<b>452.256</b>	<b>82.29</b>	<b>0.000</b>
subj*ar	27	177.019	6.556	1.52	
subj*quantity	27	36.244	1.342	0.84	
subj*integralness	9	49.462	5.496	1.32	
ar*quantity	9	10.125	1.125	0.86	0.566
<b>ar*integralness</b>	<b>3</b>	<b>67.781</b>	<b>22.594</b>	<b>5.82</b>	<b>0.003</b>
quantity*integralness	3	0.356	0.119	0.10	0.958
subj*ar*quantity	81	106.219	1.311	1.50	
subj*ar*integralness	27	104.750	3.880	4.45	
subj*quantity*integralness	27	31.175	1.155	1.32	
ar*quantity*integralness	9	6.756	0.751	0.86	0.560
Error	401	349.463	0.871		
Total	639	1954.775			

Figures 82 through 85 show the graphs for the significant results of the separable measure. Figure 82 shows higher average separable ratings (indicating less separable perception) for the planets condition and lower average separable ratings (indicating more separable perception) for the trees augmented condition. Figure 83 displays decreasing separable ratings across the four levels of the quantity variable. Figure 84 indicates higher separable ratings for high integral augmented environments. Figure 85 indicates a converging of separable ratings across the low integral augmented environment conditions.

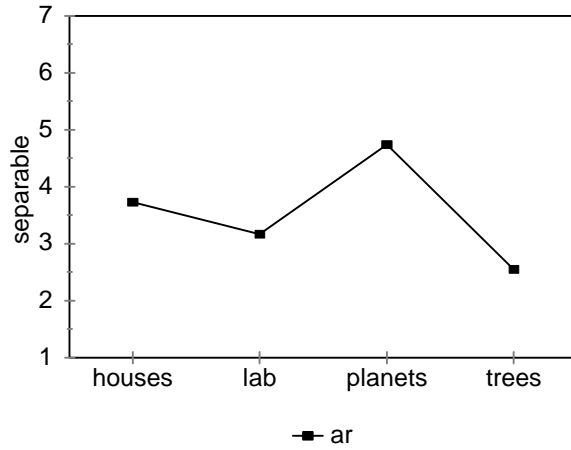


Figure 82. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

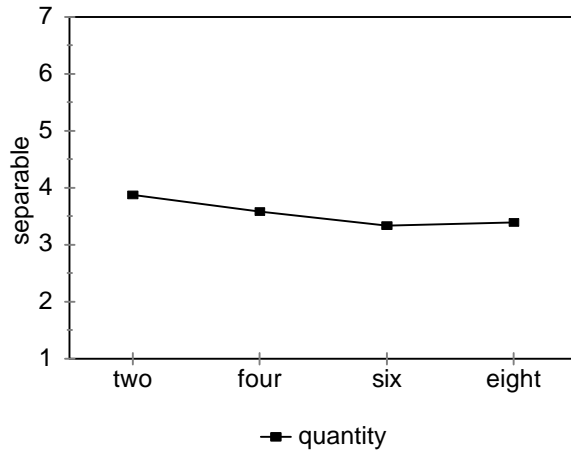


Figure 83. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.001$ ) as measured by separable ratings.

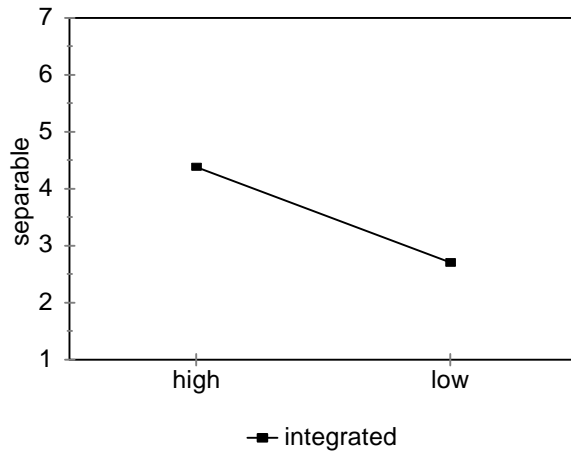


Figure 84. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by separable ratings.

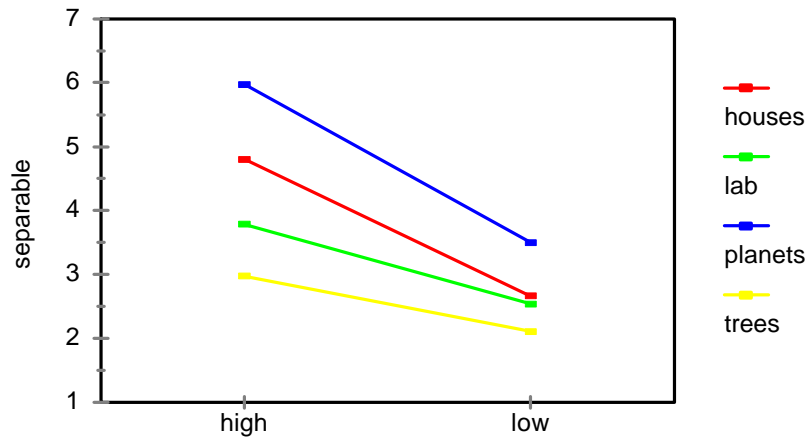


Figure 85. Graph of the interaction between augmented reality scene and integralness, significant at the  $p < 0.01$  level ( $p = 0.003$ ) as measured by separable ratings.

Table 75. ANOVA summary table for the three independent variables, measured by a similar rating.

Source	DF	SS	MS	F	P
subj	9	123.220	13.691	1.39	
<b>ar</b>	<b>3</b>	<b>359.517</b>	<b>119.839</b>	<b>20.30</b>	<b>0.000</b>
quantity	3	5.455	1.818	2.25	0.105
<b>integralness</b>	<b>1</b>	<b>658.127</b>	<b>658.127</b>	<b>79.66</b>	<b>0.000</b>
subj*ar	27	159.373	5.903	1.49	
subj*quantity	27	21.811	0.808	0.62	
subj*integralness	9	74.358	8.262	2.00	
ar*quantity	9	9.602	1.067	1.09	0.380
<b>ar*integralness</b>	<b>3</b>	<b>70.017</b>	<b>23.339</b>	<b>6.13</b>	<b>0.003</b>
quantity*integralness	3	0.905	0.302	0.26	0.850
subj*ar*quantity	81	79.320	0.979	1.19	
subj*ar*integralness	27	102.811	3.808	4.61	
subj*quantity*integralness	27	30.798	1.141	1.38	
ar*quantity*integralness	9	10.777	1.197	1.45	0.164
Error	401	330.958	0.825		
Total	639	2037.048			

Figures 86, 87, and 88 show the graphs for the significant results of the similar measure. Figure 86 shows lower average similar ratings (indicating more similar perception) for the planets condition and higher average similar ratings (indicating less similar perception) for the trees augmented condition. Figure 87 indicates lower similar ratings for high integral augmented environments. Figure 88 indicates a converging of similar ratings across the low integral augmented environment conditions.

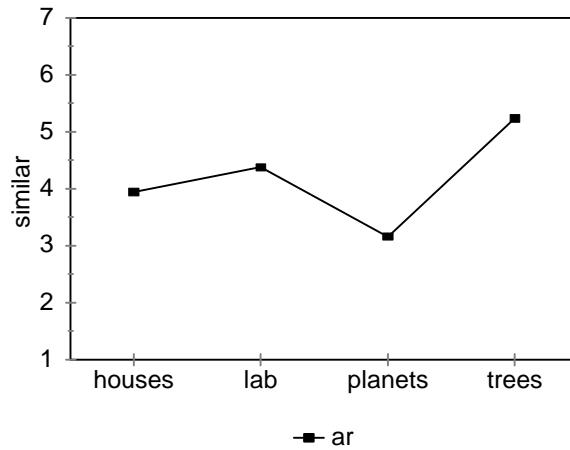


Figure 86. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by similar ratings.

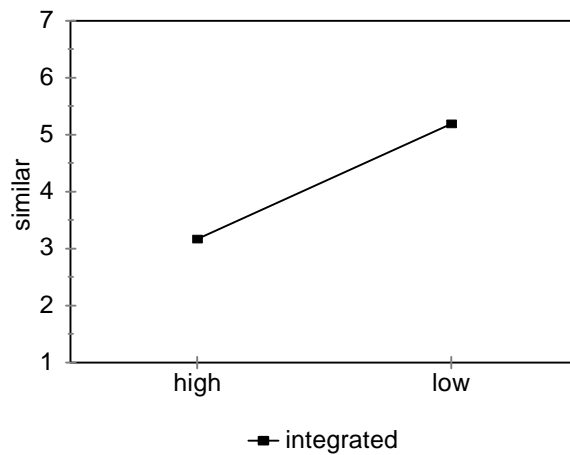


Figure 87. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by similar ratings.

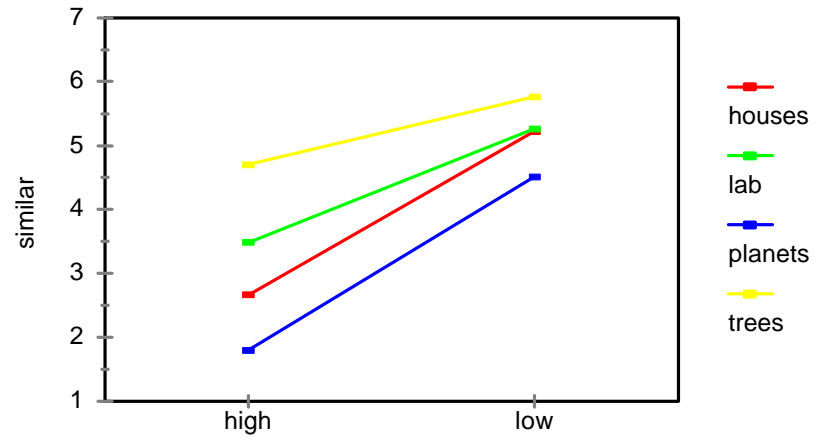


Figure 88. Graph of the interaction between augmented reality scene and integralness, significant at the  $p < 0.01$  level ( $p = 0.003$ ) as measured by similar ratings.

Table 76. ANOVA summary table for the three independent variables, measured by a recognition rating.

Source	DF	SS	MS	F	P
subj	9	982.595	109.177	4.94	
ar	3	3.642	1.214	0.14	0.937
<b>quantity</b>	<b>3</b>	<b>52.580</b>	<b>17.527</b>	<b>22.34</b>	<b>0.000</b>
integralness	1	13.514	13.514	0.79	0.396
subj*ar	27	239.748	8.880	3.49	
subj*quantity	27	21.186	0.785	0.35	
subj*integralness	9	153.033	17.004	5.01	
<b>ar*quantity</b>	<b>9</b>	<b>28.027</b>	<b>3.114</b>	<b>2.62</b>	<b>0.010</b>
ar*integralness	3	8.980	2.993	1.27	0.305
quantity*integralness	3	8.992	2.997	1.47	0.246
subj*ar*quantity	81	96.395	1.190	1.18	
subj*ar*integralness	27	63.786	2.362	2.34	
subj*quantity*integralness	27	55.148	2.043	2.02	
ar*quantity*integralness	9	5.839	0.649	0.64	0.761
Error	401	405.458	1.011		
Total	639	2138.923			

Figures 89 and 90 show the graphs for the significant results of the recognition measure. Figure 89 displays decreasing recognition ratings (indicating more graphics recognition influence) across the four levels of the quantity variable. Figure 90 shows several differences among conditions in the augmented environment and quantity interaction.

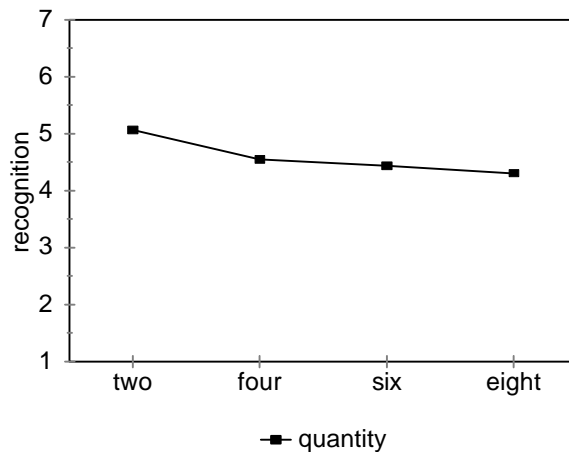


Figure 89. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by recognition ratings.

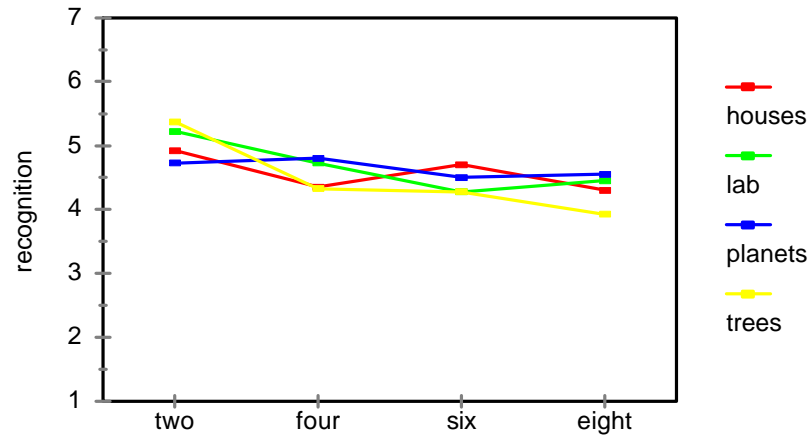


Figure 90. Graph of the interaction between augmented reality scene and quantity, significant at the  $p < 0.01$  level ( $p = 0.010$ ) as measured by recognition ratings.

Table 77. ANOVA summary table for the three independent variables, measured by a usefulness rating.

Source	DF	SS	MS	F	P
subj	9	205.889	22.877	2.11	
<b>ar</b>	<b>3</b>	<b>242.542</b>	<b>80.847</b>	<b>10.10</b>	<b>0.000</b>
<b>quantity</b>	<b>3</b>	<b>33.817</b>	<b>11.272</b>	<b>10.98</b>	<b>0.000</b>
<b>integralness</b>	<b>1</b>	<b>323.477</b>	<b>323.477</b>	<b>43.71</b>	<b>0.000</b>
subj*ar	27	216.130	8.005	1.75	
subj*quantity	27	27.730	1.027	0.74	
subj*integralness	9	66.602	7.400	1.66	
ar*quantity	9	12.764	1.418	1.24	0.283
ar*integralness	3	36.992	12.331	2.93	0.051
quantity*integralness	3	1.417	0.472	0.46	0.715
subj*ar*quantity	81	92.627	1.144	1.46	
subj*ar*integralness	27	113.492	4.203	5.36	
subj*quantity*integralness	27	27.942	1.035	1.32	
ar*quantity*integralness	9	10.489	1.165	1.49	0.150
Error	401	314.339	0.784		
Total	639	1726.248			

Figures 91, 92, and 93 show the graphs for the significant results of the usefulness measure.

Figure 91 shows lower average usefulness ratings (indicating more usefulness perception) for the planets condition and higher average usefulness ratings (indicating less usefulness perception) for the trees augmented condition. Figure 92 displays increasing usefulness ratings across the four levels of the quantity variable. Figure 93 indicates lower usefulness ratings for high integral augmented environments.

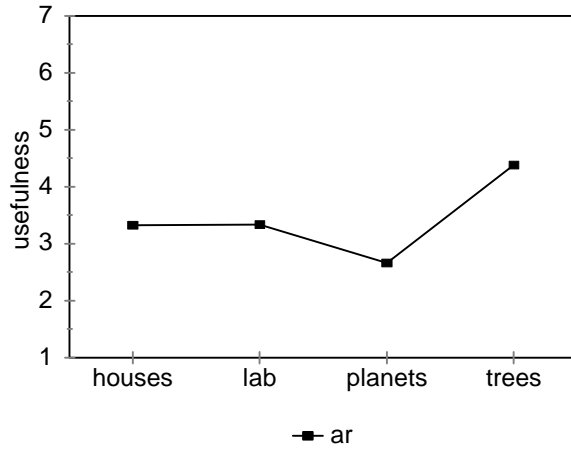


Figure 91. Graph of the main effect of augmented reality scene, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

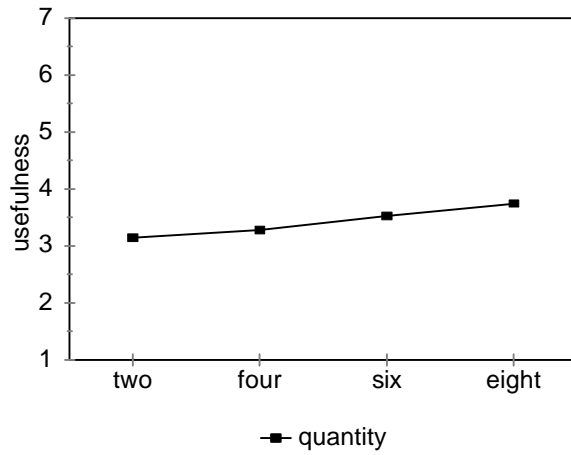


Figure 92. Graph of the main effect of quantity, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

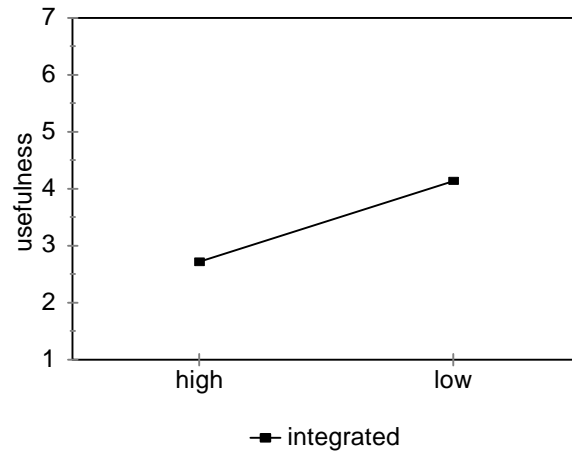


Figure 93. Graph of the main effect of integralness, significant at the  $p < 0.01$  level ( $p = 0.000$ ) as measured by usefulness ratings.

Table 78 summarizes all the significant effects for the three independent variables of experiment two for the eight subjective assessment dependent measures.

Table 78. Summary of significant results at the  $p < 0.01$  level for the eight rating scale dependent measures for experiment two

	<b>goodness</b>	<b>integral</b>	<b>attend</b>	<b>clutter</b>	<b>separable</b>	<b>similar</b>	<b>recognition</b>	<b>useful</b>
<b>ar</b>	x	x	x		x	x		x
<b>quantity</b>	x	x	x	x	x		x	x
<b>integralness</b>	x	x	x		x	x		x
<b>ar*quantity</b>				x			x	
<b>ar*integralness</b>					x	x		

### 6.3.3. Nonparametric Tests

Where main effect significant results were indicated in the above parametric ANOVAs, non-parametric tests were performed to confirm those significant results. Tables 79 through 96 show summary table results of Kruskal-Wallis Tests for the significant main effects from the parametric ANOVAs.

Table 79. Kruskal-Wallis test for augmented reality scene as measured by goodness ratings.

<u>ar</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
houses	160	3.000	321.6	0.09
lab	160	3.000	321.6	0.09
planets	160	3.000	231.5	-7.03
trees	160	5.000	407.3	6.85
Overall	640		320.5	

H = 72.33 DF = 3 P = 0.000  
H = 75.20 DF = 3 P = 0.000 (adjusted for ties)

Table 80. Kruskal-Wallis test for quantity as measured by goodness ratings.

<u>quantity</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
two	160	3.000	253.3	-5.31
four	160	3.000	304.2	-1.29
six	160	4.000	352.9	2.56
eight	160	4.000	371.6	4.04
Overall	640		320.5	

H = 39.50 DF = 3 P = 0.000  
H = 41.07 DF = 3 P = 0.000 (adjusted for ties)

Table 81. Kruskal-Wallis test for integralness as measured by goodness ratings.

<u>integralness</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
high	320	3.000	244.9	-10.34
low	320	4.000	396.1	10.34
Overall	640		320.5	

H = 107.01 DF = 1 P = 0.000  
H = 111.25 DF = 1 P = 0.000 (adjusted for ties)

Table 82. Kruskal-Wallis test for augmented reality scene as measured by integral ratings.

ar	N	Median	Ave Rank	Z
houses	160	4.000	312.7	-0.62
lab	160	4.000	323.3	0.22
planets	160	3.000	230.8	-7.09
trees	160	5.000	415.3	7.49
Overall	640		320.5	

H = 80.03 DF = 3 P = 0.000  
H = 82.61 DF = 3 P = 0.000 (adjusted for ties)

Table 83. Kruskal-Wallis test for quantity as measured by integral ratings.

quantity	N	Median	Ave Rank	Z
two	160	3.500	271.3	-3.89
four	160	4.000	301.4	-1.51
six	160	5.000	350.1	2.34
eight	160	5.000	359.3	3.06
Overall	640		320.5	

H = 24.16 DF = 3 P = 0.000  
H = 24.94 DF = 3 P = 0.000 (adjusted for ties)

Table 84. Kruskal-Wallis test for integralness as measured by integral ratings.

integralness	N	Median	Ave Rank	Z
high	320	3.000	235.2	-11.67
low	320	5.000	405.8	11.67
Overall	640		320.5	

H = 136.29 DF = 1 P = 0.000  
H = 140.69 DF = 1 P = 0.000 (adjusted for ties)

Table 85. Kruskal-Wallis test for augmented reality scene as measured by attend ratings.

ar	N	Median	Ave Rank	Z
houses	160	3.000	340.3	1.57
lab	160	3.000	288.4	-2.53
planets	160	4.000	395.7	5.94
trees	160	3.000	257.6	-4.97
Overall	640		320.5	

H = 51.61 DF = 3 P = 0.000  
H = 53.69 DF = 3 P = 0.000 (adjusted for ties)

Table 86. Kruskal-Wallis test for quantity as measured by attend ratings.

<u>quantity</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
two	160	4.000	393.7	5.78
four	160	3.000	324.8	0.34
six	160	3.000	296.2	-1.92
eight	160	3.000	267.3	-4.21
Overall	640		320.5	

H = 41.22 DF = 3 P = 0.000  
H = 42.88 DF = 3 P = 0.000 (adjusted for ties)

Table 87. Kruskal-Wallis test for integralness as measured by attend ratings.

<u>integralness</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
high	320	4.000	389.4	9.43
low	320	3.000	251.6	-9.43
Overall	640		320.5	

H = 88.86 DF = 1 P = 0.000  
H = 92.43 DF = 1 P = 0.000 (adjusted for ties)

Table 88. Kruskal-Wallis test for quantity as measured by clutter ratings.

<u>quantity</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
two	160	6.000	484.0	12.92
four	160	5.000	357.0	2.88
six	160	3.000	250.8	-5.51
eight	160	2.000	190.3	-10.29
Overall	640		320.5	

H = 233.46 DF = 3 P = 0.000  
H = 240.94 DF = 3 P = 0.000 (adjusted for ties)

Table 89. Kruskal-Wallis test for augmented reality scene as measured by separable ratings.

<u>ar</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
houses	160	3.000	341.0	1.62
lab	160	3.000	286.3	-2.70
planets	160	5.000	436.9	9.20
trees	160	2.000	217.8	-8.11
Overall	640		320.5	

H = 120.22 DF = 3 P = 0.000  
H = 125.14 DF = 3 P = 0.000 (adjusted for ties)

Table 90. Kruskal-Wallis test for quantity as measured by separable ratings.

quantity	N	Median	Ave Rank	Z
two	160	3.000	355.1	2.73
four	160	3.000	324.1	0.29
six	160	3.000	299.6	-1.65
eight	160	3.000	303.2	-1.37
Overall	640		320.5	

H = 9.12 DF = 3 P = 0.028  
H = 9.49 DF = 3 P = 0.023 (adjusted for ties)

Table 91. Kruskal-Wallis test for integralness as measured by separable ratings.

integralness	N	Median	Ave Rank	Z
high	320	5.000	406.1	11.71
low	320	2.000	234.9	-11.71
Overall	640		320.5	

H = 137.16 DF = 1 P = 0.000  
H = 142.77 DF = 1 P = 0.000 (adjusted for ties)

Table 92. Kruskal-Wallis test for augmented reality scene as measured by similar ratings.

integralness	N	Median	Ave Rank	Z
high	320	5.000	406.1	11.71
low	320	2.000	234.9	-11.71
Overall	640		320.5	

H = 137.16 DF = 1 P = 0.000  
H = 142.77 DF = 1 P = 0.000 (adjusted for ties)

Table 93. Kruskal-Wallis test for quantity as measured by recognition ratings.

quantity	N	Median	Ave Rank	Z
two	160	6.000	368.5	3.79
four	160	5.000	317.2	-0.26
six	160	4.000	305.8	-1.16
wight	160	4.000	290.5	-2.37
Overall	640		320.5	

H = 16.06 DF = 3 P = 0.001  
H = 16.65 DF = 3 P = 0.001 (adjusted for ties)

Table 94. Kruskal-Wallis test for augmented reality scene as measured by usefulness ratings.

ar	N	Median	Ave Rank	Z
houses	160	3.000	307.9	-0.99
lab	160	3.000	313.3	-0.57
planets	160	2.000	235.4	-6.72
trees	160	4.500	425.4	8.28
Overall	640		320.5	

H = 86.32 DF = 3 P = 0.000  
H = 89.10 DF = 3 P = 0.000 (adjusted for ties)

Table 95. Kruskal-Wallis test for quantity as measured by usefulness ratings.

quantity	N	Median	Ave Rank	Z
two	160	3.000	290.0	-2.41
four	160	3.000	306.9	-1.08
six	160	3.000	331.0	0.83
eight	160	4.000	354.1	2.66
Overall	640		320.5	

H = 11.03 DF = 3 P = 0.012  
H = 11.38 DF = 3 P = 0.010 (adjusted for ties)

Table 96. Kruskal-Wallis test for integralness as measured by usefulness ratings.

integralness	N	Median	Ave Rank	Z
high	320	2.000	240.4	-10.96
low	320	4.000	400.6	10.96
Overall	640		320.5	

H = 120.17 DF = 1 P = 0.000  
H = 124.04 DF = 1 P = 0.000 (adjusted for ties)

#### *6.3.4. Subjective Rating Scales Factor Analyses*

Multivariate factor analyses for experiment two were conducted on the eight rating scale measures (goodness, integral, attend, clutter, separable, similar, recognition, usefulness) for all 32 unique combinations of the independent variables. Factor analyses were conducted to examine underlying relationships between the eight rating scale measures, and investigate the possibility of reducing the dimensionality of the data set (i.e., determine the extent that the dependent measures explained variance in respective factor loadings matrices).

As was discussed in the factor analysis section of experiment one, the most understandable results are obtained from large data sets exceeding two rules of thumb: at least 5 observations per measure and/or at least 100 total observations per condition (Smith, 1998). In experiment one there was a variable available to collapse the data that yielded 10 observations per measure. Upon examination of the summary results for experiment two (Table 78), there was not an equivalent candidate variable for collapsing the data set. Therefore, it was necessary to carry out the factor analyses despite not meeting the informal rule-of-thumb criteria. This type of technical violation can make interpretation of the factor analyses results more difficult. Fortunately in the analyses for experiment two, there were 32 factor analyses to compare. The many similar factor analyses allow for meaningful interpretation that would otherwise be very difficult with limited data sets.

From the 32 data sets, preliminary factor analyses were first conducted by extracting four factors using the principal components method without rotation. Table 97 shows the results from this analysis on the houses-two-low condition.

Table 97. Factor analysis results for the houses-two-low condition extracting four factors using the principal components method without rotation.

Variable	Factor1	Factor2	Factor3	Factor4	Communality
goodness	-0.870	0.219	-0.134	-0.094	0.832
integral	-0.862	0.012	-0.254	-0.173	0.838
attend	0.874	-0.112	-0.335	-0.056	0.892
clutter	0.449	-0.675	0.119	-0.557	0.981
separabl	0.821	0.343	0.088	0.045	0.802
similar	-0.719	-0.434	0.498	0.140	0.973
recog.	-0.048	-0.849	-0.315	0.363	0.954
useful	-0.905	0.036	-0.192	-0.165	0.885
<b>Variance</b>	<b>4.4801</b>	<b>1.5449</b>	<b>0.6005</b>	<b>0.5327</b>	<b>7.1582</b>
% Var	0.560	0.193	0.075	0.067	0.895

The variance accounted for by each factor, indicated by the eigenvalue line highlighted by **bold** in Table 97, shows only two factors that meet Kaiser's > 1 selection criteria of variance accounted for. This pattern was prevalent in tests of most conditions; therefore, for simplicity and consistency, only two factors were extracted in subsequent factor analyses.

Table 97 also shows loadings for the first two factors (shown in *italics*) in generally narrow ranges of high loadings (near +/- 1), particularly in factor one. Therefore, varimax rotation of the loadings matrices were performed on all subsequent factor analyses to minimize the number of high loadings on each factor which yielded reduced dimensionality and simplified interpretation of the factors.

The communality column in Table 97 approximately represents the variance explained by the variables within the factors. Thus, it can loosely be compared to an  $R^2$  measure. A consequence of using the varimax rotation causes the communality values to be artificially high and uninformative for interpreting the factors.

Tables 98 through 129 show summary tables of the two calculated factors for the 32 combinations of the independent variables of experiment two using the principal components factor analysis method and varimax rotation. Loadings for factors one and two are indicated in **bold** if the loadings plots and experimental evidence indicate they are important variables for interpreting the respective factor. With the loadings values, scores closer to -1 or +1 are expected

to have greater impact on the respective factor. In addition, the relative size of each variable loading in comparison with other loadings in the respective factor is also indicative of interpretation importance.

The loadings plots for factor analyses are often more informative than the summary tables in interpreting the factors and dependent variable relationships. Of particular relevance are the spatial clusters of variables, the length of loading lines in the plots, and directly polar variables. These events are apparent in many of the loadings plot figures presented here.

Table 98. Factor analysis table for the houses-two-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.778</b>	0.554	0.911
integral	<b>0.752</b>	0.502	0.818
attend	<b>-0.648</b>	-0.377	0.562
clutter	<b>-0.844</b>	-0.129	0.729
separable	-0.267	<b>-0.857</b>	0.806
similar	0.233	<b>0.831</b>	0.744
recog.	<b>-0.798</b>	-0.209	0.680
useful	<b>0.674</b>	0.557	0.765
Variance	3.5189	2.4964	6.0153
% Var	0.440	0.312	0.752

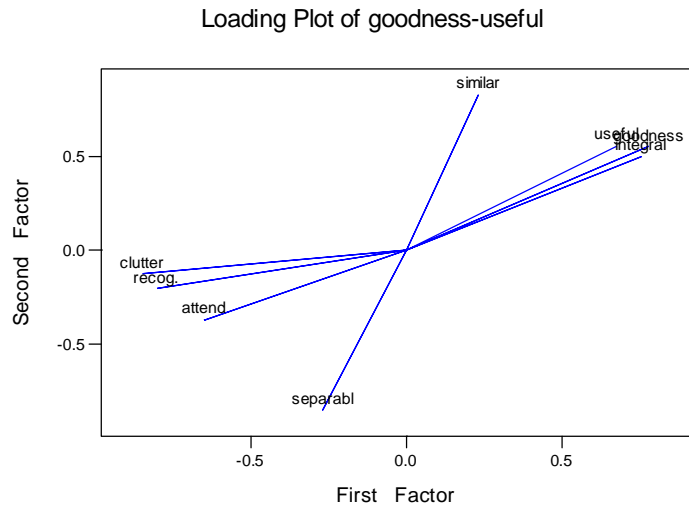


Figure 94. Loadings plot for the houses-two-high factor analysis.

Table 99. Factor analysis table for the houses-two-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.835</b>	-0.329	0.806
integral	<b>0.854</b>	-0.122	0.744
attend	<b>-0.853</b>	0.223	0.777
clutter	-0.359	<b>0.727</b>	0.657
separable	<b>-0.858</b>	-0.236	0.792
similar	<b>0.769</b>	0.338	0.706
recog.	0.156	<b>0.836</b>	0.723
useful	<b>0.893</b>	-0.152	0.821
Variance	4.4322	1.5928	6.0250
% Var	0.554	0.199	0.753

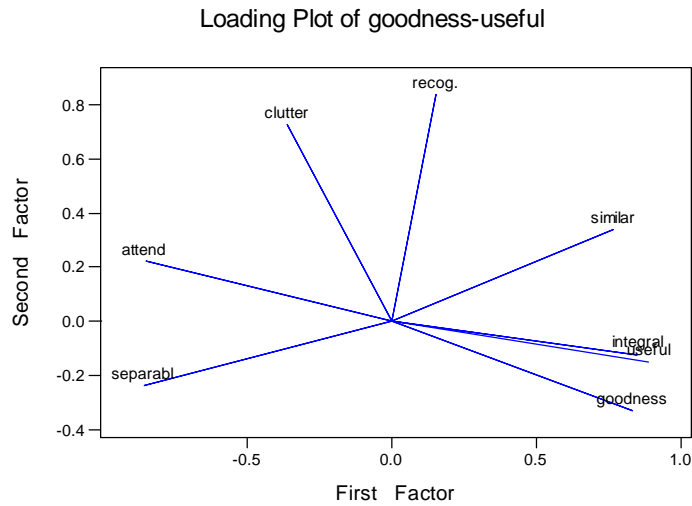


Figure 95. Loadings plot for the houses-two-low factor analysis.

Table 100. Factor analysis table for the houses-four-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.875</b>	0.341	0.881
integral	<b>0.856</b>	0.400	0.894
attend	<b>-0.907</b>	-0.133	0.841
clutter	<b>-0.827</b>	-0.303	0.776
separable	<b>-0.856</b>	-0.329	0.841
similar	0.506	<b>0.523</b>	0.529
recog.	-0.128	<b>-0.917</b>	0.857
useful	0.553	<b>0.729</b>	0.837
Variance	4.3161	2.1391	6.4552
% Var	0.540	0.267	0.807

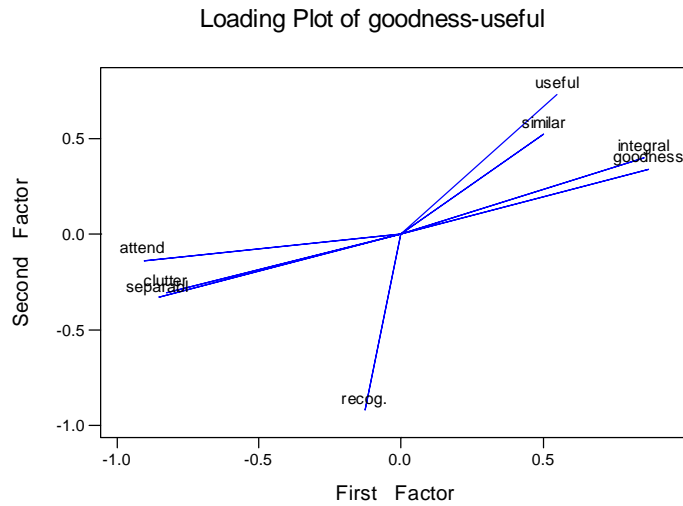


Figure 96. Loadings plot for the houses-high-four factor analysis.

Table 101. Factor analysis table for the houses-four-low conditions.

Variable	Factor1	Factor2	Communality
goodness	0.461	<b>0.723</b>	0.735
integral	<b>0.888</b>	0.204	0.829
attend	<b>-0.829</b>	-0.093	0.696
clutter	-0.279	<b>-0.867</b>	0.830
separable	<b>-0.864</b>	-0.198	0.786
similar	<b>0.916</b>	0.054	0.842
recog.	0.470	<b>-0.750</b>	0.783
useful	0.598	<b>0.529</b>	0.638
Variance	3.9303	2.2093	6.1395
% Var	0.491	0.276	0.767

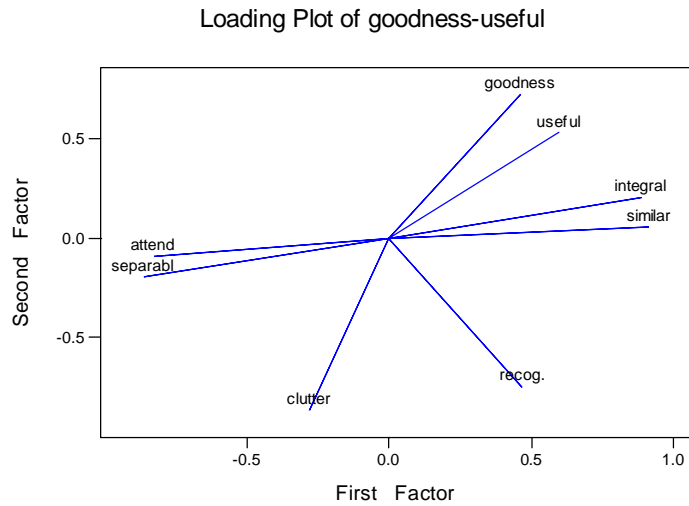


Figure 97. Loadings plot for the houses-four-low factor analysis.

Table 102. Factor analysis table for the houses-six-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.745</b>	-0.007	0.555
integral	0.692	0.339	0.594
attend	<b>-0.917</b>	0.089	0.850
clutter	-0.409	-0.509	0.426
separable	<b>-0.826</b>	-0.343	0.799
similar	0.111	<b>0.532</b>	0.295
recog.	0.175	<b>-0.852</b>	0.757
useful	0.556	0.610	0.681
Variance	3.0757	1.8799	4.9556
% Var	0.384	0.235	0.619

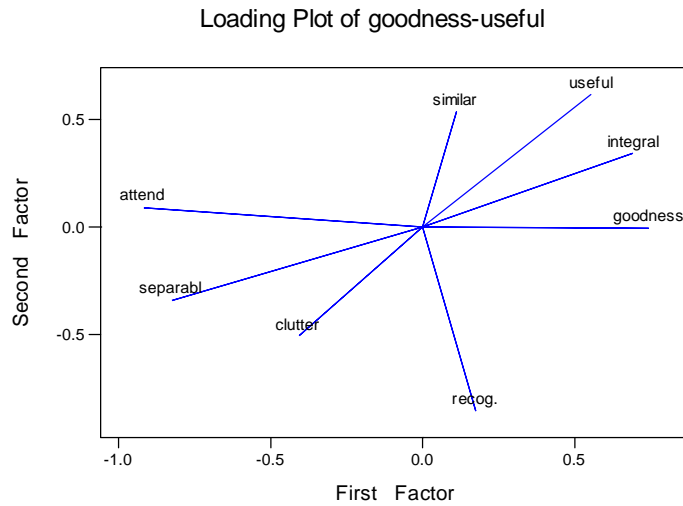


Figure 98. Loadings plot for the houses-six-high factor analysis.

Table 103. Factor analysis table for the houses-six-low conditions.

Variable	Factor1	Factor2	Communality
goodness	0.613	0.675	0.830
integral	<b>0.854</b>	0.401	0.890
attend	<b>-0.786</b>	-0.383	0.765
clutter	-0.355	-0.772	0.722
separable	<b>-0.932</b>	-0.115	0.882
similar	<b>0.923</b>	0.072	0.857
recog.	-0.002	<b>-0.889</b>	0.791
useful	<b>0.712</b>	0.208	0.550
Variance	4.0758	2.2120	6.2878
% Var	0.509	0.276	0.786

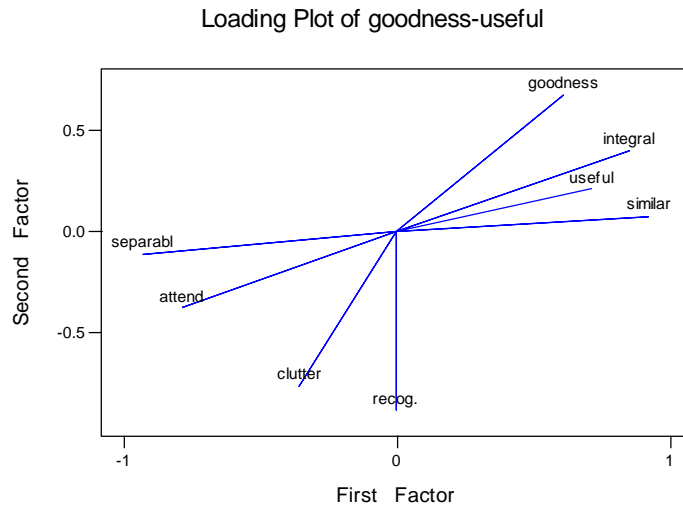


Figure 99. Loadings plot for the houses-six-low factor analysis.

Table 104. Factor analysis table for the houses-eight-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.938</b>	-0.157	0.905
integral	<b>0.954</b>	-0.165	0.938
attend	<b>-0.869</b>	0.014	0.756
clutter	-0.680	0.234	0.517
separable	-0.642	0.621	0.798
similar	0.168	<b>-0.816</b>	0.694
recog.	-0.005	<b>0.663</b>	0.440
useful	0.584	-0.543	0.636
Variance	3.7893	1.8929	5.6822
% Var	0.474	0.237	0.710

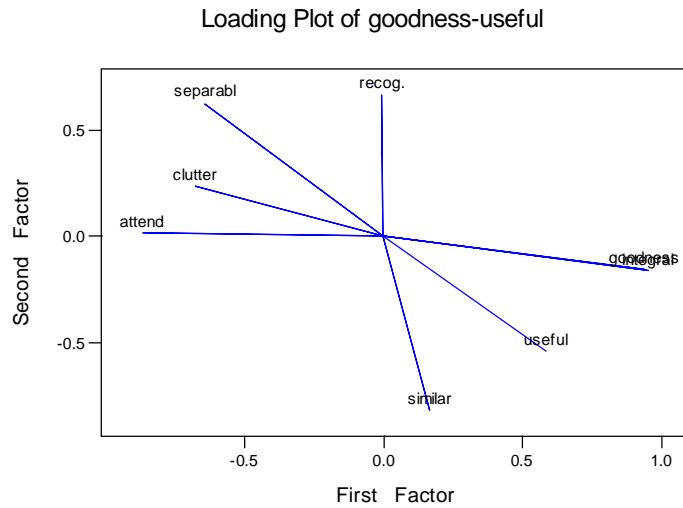


Figure 100. Loadings plot for the houses-eight-high factor analysis.

Table 105. Factor analysis table for the houses-eight-low conditions.

Variable	Factor1	Factor2	Communality
goodness	0.714	0.517	0.777
integral	<b>0.890</b>	0.268	0.863
attend	<b>-0.665</b>	-0.444	0.640
clutter	<b>-0.647</b>	-0.509	0.678
separable	<b>-0.727</b>	-0.525	0.804
similar	<b>0.793</b>	-0.011	0.629
recog.	-0.078	<b>-0.913</b>	0.841
useful	<b>0.880</b>	0.209	0.818
Variance	4.1008	1.9499	6.0507
% Var	0.513	0.244	0.756

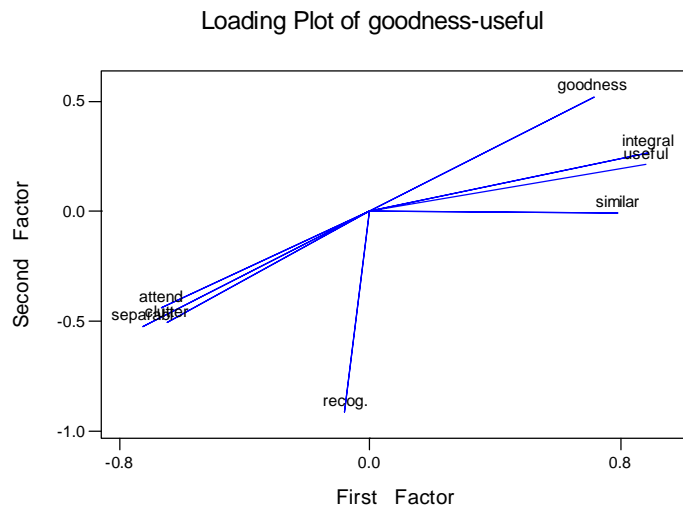


Figure 101. Loadings plot for the houses-eight-low factor analysis.

Table 106. Factor analysis table for the lab-two-high conditions.

Variable	Factor1	Factor2	Communality
goodness	0.369	<b>0.779</b>	0.744
integral	<b>0.813</b>	0.414	0.832
attend	0.065	<b>-0.910</b>	0.832
clutter	-0.310	-0.692	0.575
separable	<b>-0.833</b>	-0.307	0.788
similar	<b>0.830</b>	0.056	0.692
recog.	<b>-0.644</b>	-0.208	0.458
useful	<b>0.816</b>	0.050	0.669
Variance	3.3613	2.2287	5.5900
% Var	0.420	0.279	0.699

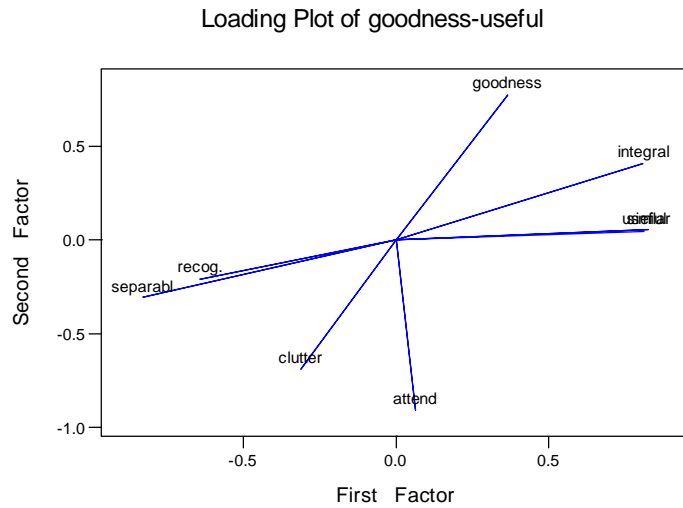


Figure 102. Loadings plot for the lab-two-high factor analysis.

Table 107. Factor analysis table for the lab-two-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>-0.719</b>	0.035	0.518
integral	-0.393	<b>0.640</b>	0.565
attend	<b>0.804</b>	-0.176	0.677
clutter	<b>0.794</b>	-0.074	0.637
separable	0.069	<b>-0.845</b>	0.718
similar	0.320	<b>0.569</b>	0.427
recog.	<b>0.919</b>	-0.094	0.854
useful	-0.292	<b>0.815</b>	0.750
Variance	2.9866	2.1593	5.1459
% Var	0.373	0.270	0.643

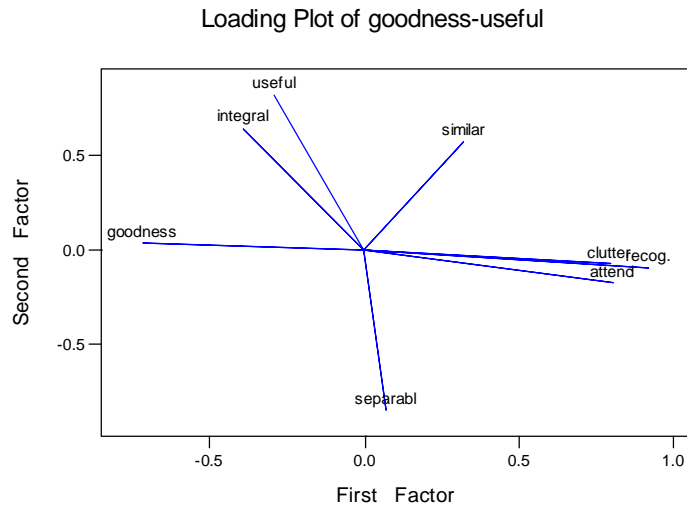


Figure 103. Loadings plot for the lab-two-low factor analysis.

Table 108. Factor analysis table for the lab-four-high conditions.

Variable	Factor1	Factor2	Communality
goodness	0.765	-0.375	0.726
integral	<b>0.855</b>	-0.165	0.758
attend	0.009	<b>0.796</b>	0.634
clutter	-0.012	<b>0.765</b>	0.585
separable	-0.377	0.629	0.537
similar	<b>0.849</b>	0.113	0.733
recog.	-0.460	<b>0.751</b>	0.776
useful	0.699	-0.624	0.878
Variance	2.8794	2.7481	5.6275
% Var	0.360	0.344	0.703

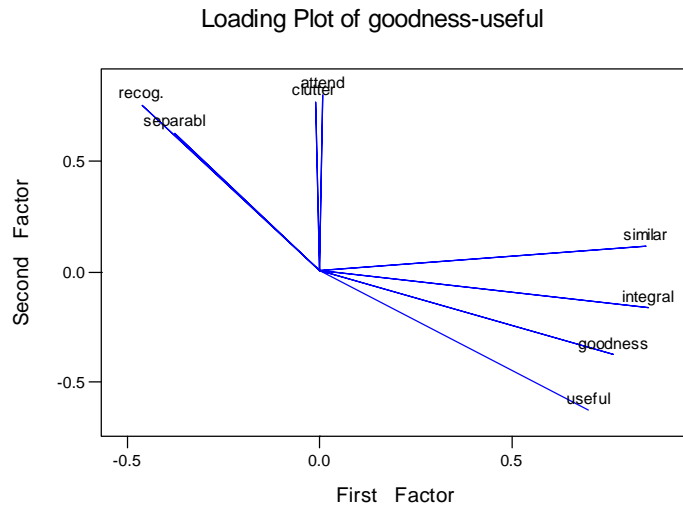


Figure 104. Loadings plot for the lab-four-high factor analysis.

Table 109. Factor analysis table for the lab-four-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.753</b>	-0.223	0.617
integral	<b>0.869</b>	-0.024	0.756
attend	<b>-0.836</b>	0.279	0.777
clutter	-0.415	0.680	0.634
separable	<b>-0.839</b>	0.129	0.720
similar	0.546	-0.121	0.313
recog.	0.003	<b>0.927</b>	0.860
useful	<b>0.807</b>	-0.126	0.666
Variance	3.8464	1.4972	5.3436
% Var	0.481	0.187	0.668

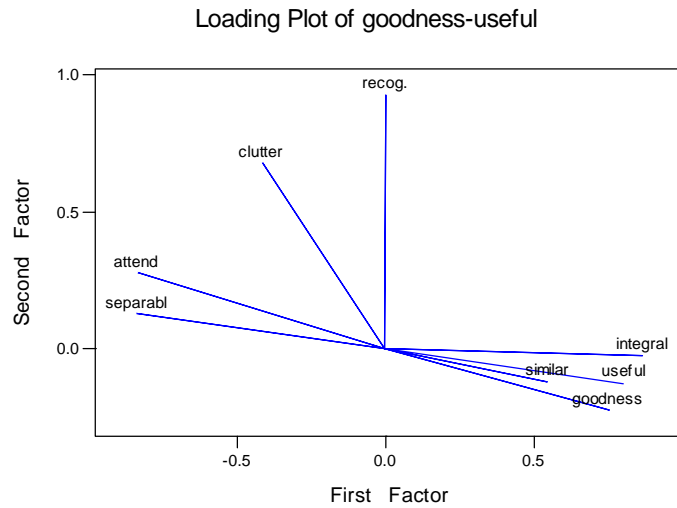


Figure 105. Loadings plot for the lab-four-low factor analysis.

Table 110. Factor analysis table for the lab-six-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.788</b>	-0.315	0.720
integral	<b>0.846</b>	0.156	0.739
attend	-0.562	-0.192	0.352
clutter	-0.369	-0.373	0.275
separable	<b>-0.854</b>	-0.286	0.811
similar	0.585	<b>0.602</b>	0.705
recog.	0.016	<b>-0.917</b>	0.840
useful	<b>0.756</b>	0.346	0.691
Variance	3.4314	1.7030	5.1345
% Var	0.429	0.213	0.642

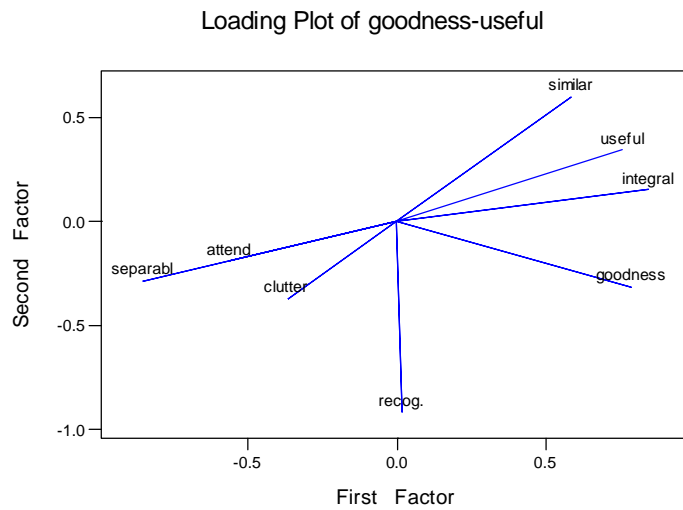


Figure 106. Loadings plot for the lab-six-high factor analysis.

Table 111. Factor analysis table for the lab-six-low conditions.

Variable	Factor1	Factor2	Communality
goodness	0.509	<b>0.714</b>	0.770
integral	<b>0.797</b>	0.376	0.777
attend	-0.082	<b>-0.911</b>	0.837
clutter	-0.234	-0.703	0.548
separable	<b>-0.800</b>	-0.390	0.793
similar	<b>0.919</b>	0.044	0.847
recog.	-0.536	-0.283	0.367
useful	<b>0.850</b>	0.173	0.752
Variance	3.4512	2.2390	5.6902
% Var	0.431	0.280	0.711

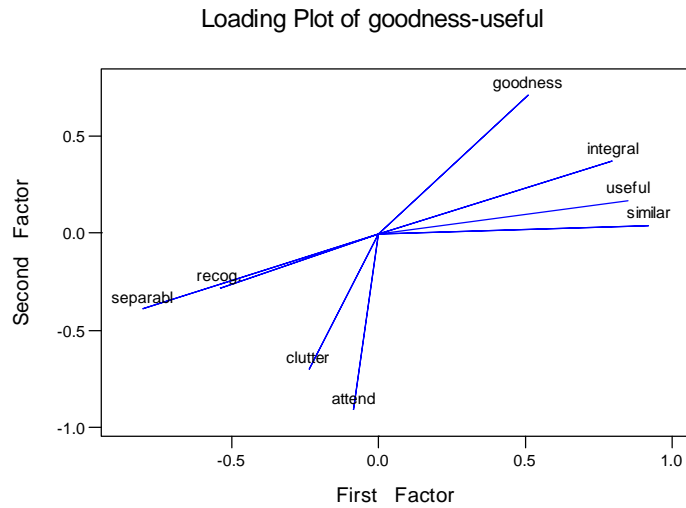


Figure 107. Loadings plot for the lab-six-low factor analysis.

Table 112. Factor analysis table for the lab-eight-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.737</b>	-0.432	0.730
integral	0.562	-0.674	0.770
attend	<b>-0.858</b>	0.092	0.744
clutter	<b>-0.909</b>	-0.070	0.831
separable	-0.672	0.446	0.651
similar	0.003	<b>-0.841</b>	0.708
recog.	-0.100	<b>0.737</b>	0.553
useful	0.250	<b>-0.830</b>	0.752
Variance	2.9457	2.7943	5.7400
% Var	0.368	0.349	0.718

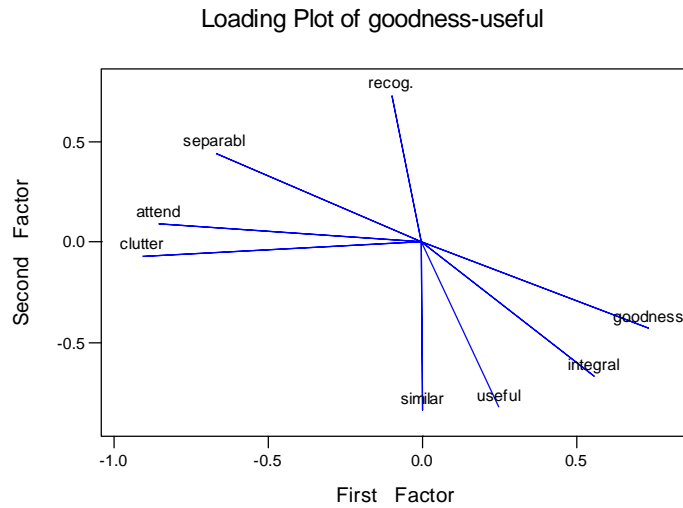


Figure 108. Loadings plot for the lab-eight-high factor analysis.

Table 113. Factor analysis table for the lab-eight-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>-0.883</b>	-0.301	0.871
integral	<b>-0.830</b>	-0.458	0.898
attend	0.727	0.365	0.662
clutter	0.710	-0.067	0.509
separable	0.615	<b>0.653</b>	0.805
similar	0.068	<b>-0.883</b>	0.785
recog.	0.802	-0.291	0.727
useful	<b>-0.875</b>	-0.169	0.794
Variance	4.2930	1.7577	6.0508
% Var	0.537	0.220	0.756

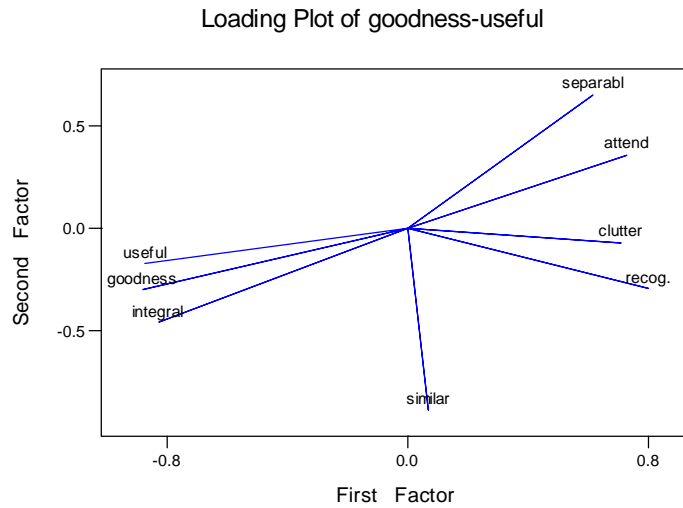


Figure 109. Loadings plot for the lab-eight-low factor analysis.

Table 114. Factor analysis table for the planets-two-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.697</b>	-0.182	0.518
integral	<b>0.742</b>	-0.608	0.919
attend	-0.499	-0.299	0.338
clutter	<b>-0.899</b>	0.075	0.813
separable	<b>-0.832</b>	0.490	0.933
similar	0.101	<b>-0.765</b>	0.595
recog.	0.023	<b>0.556</b>	0.310
useful	0.329	<b>-0.697</b>	0.595
Variance	2.9028	2.1178	5.0206
% Var	0.363	0.265	0.628

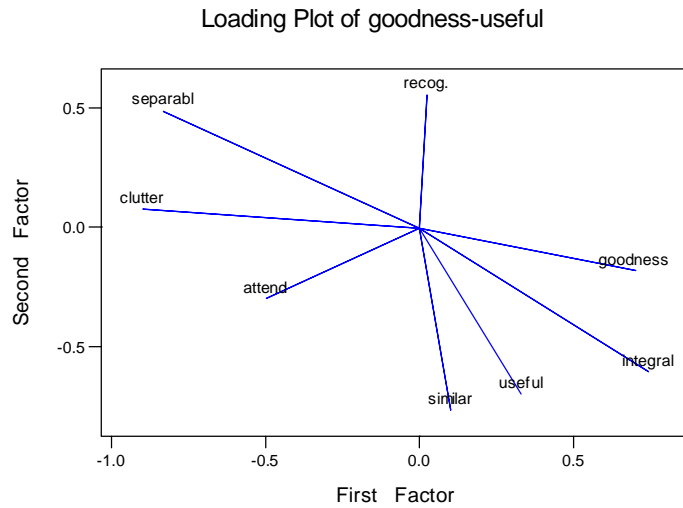


Figure 110. Loadings plot for the planets-two-high factor analysis.

Table 115. Factor analysis table for the planets-two-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.862</b>	0.189	0.779
integral	<b>0.939</b>	0.042	0.883
attend	-0.381	0.698	0.632
clutter	<b>-0.465</b>	0.266	0.287
separable	<b>-0.560</b>	0.576	0.646
similar	0.029	<b>-0.777</b>	0.604
recog.	0.138	<b>0.799</b>	0.658
useful	0.779	-0.278	0.684
Variance	2.9279	2.2467	5.1746
% Var	0.366	0.281	0.647

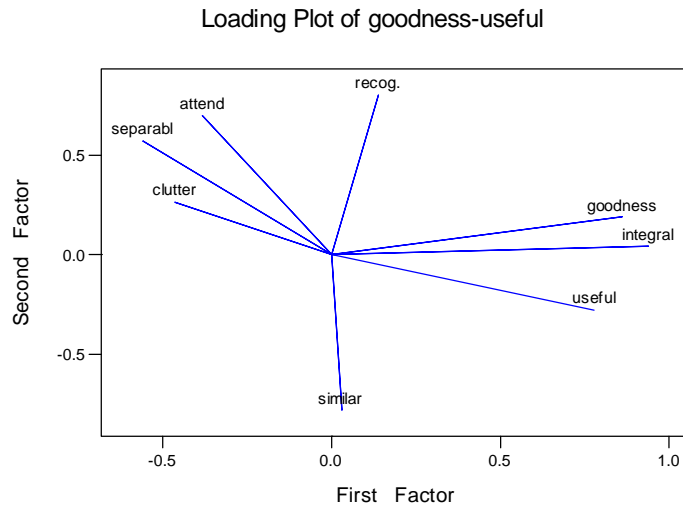


Figure 111. Loadings plot for the planets-two-low factor analysis.

Table 116. Factor analysis table for the planets-four-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.853</b>	-0.115	0.740
integral	<b>0.937</b>	-0.024	0.879
attend	-0.597	<b>0.624</b>	0.745
clutter	-0.521	0.050	0.273
separable	<b>-0.877</b>	-0.009	0.769
similar	<b>0.899</b>	0.140	0.827
recog.	-0.186	<b>-0.942</b>	0.922
useful	<b>0.942</b>	-0.002	0.887
Variance	4.7304	1.3116	6.0420
% Var	0.591	0.164	0.755

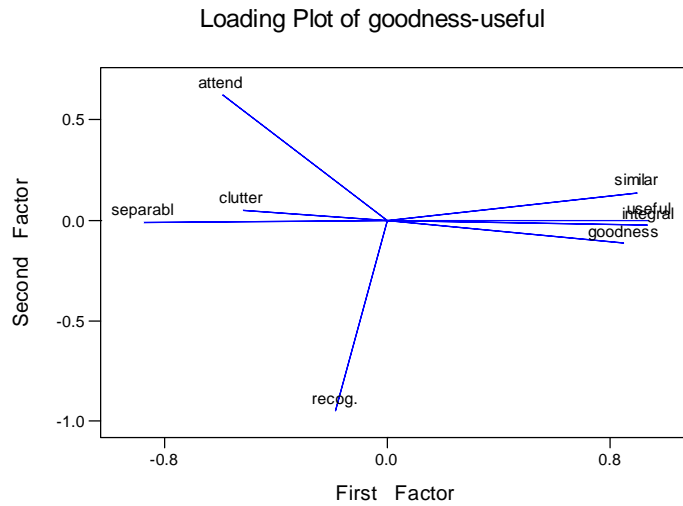


Figure 112. Loadings plot for the planets-four-high factor analysis.

Table 117. Factor analysis table for the planets-four-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.803</b>	0.271	0.719
integral	<b>0.722</b>	0.548	0.822
attend	<b>-0.901</b>	-0.040	0.814
clutter	<b>-0.832</b>	-0.024	0.693
separable	<b>-0.742</b>	0.008	0.551
similar	0.000	<b>0.762</b>	0.581
recog.	-0.136	<b>-0.660</b>	0.454
useful	<b>0.762</b>	0.355	0.708
Variance	3.8220	1.5190	5.3410
% Var	0.478	0.190	0.668

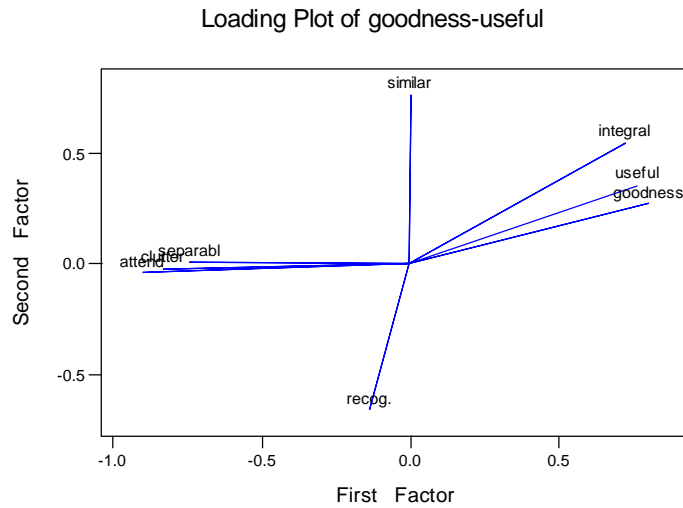


Figure 113. Loadings plot for the planets-four-low factor analysis.

Table 118. Factor analysis table for the planets-six-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>-0.858</b>	-0.217	0.783
integral	<b>-0.916</b>	-0.068	0.844
attend	0.489	0.461	0.452
clutter	<b>0.793</b>	-0.082	0.636
separable	0.216	<b>0.833</b>	0.740
similar	0.118	<b>-0.845</b>	0.728
recog.	0.508	0.085	0.265
useful	<b>-0.693</b>	-0.094	0.490
Variance	3.2431	1.6950	4.9381
% Var	0.405	0.212	0.617

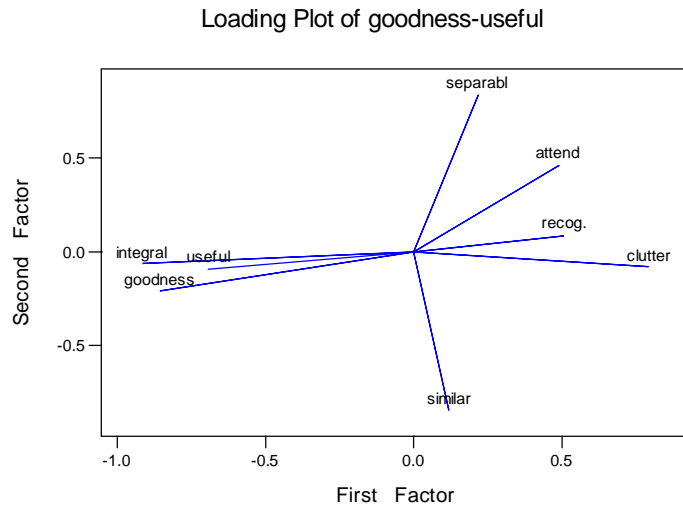


Figure 114. Loadings plot for the planets-six-high factor analysis.

Table 119. Factor analysis table for the planets-six-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.859</b>	0.139	0.757
integral	<b>0.877</b>	0.317	0.870
attend	<b>-0.884</b>	0.098	0.790
clutter	<b>-0.801</b>	-0.086	0.650
separable	<b>-0.884</b>	-0.236	0.837
similar	0.107	<b>0.900</b>	0.822
recog.	-0.139	<b>-0.737</b>	0.562
useful	<b>0.891</b>	0.325	0.900
Variance	4.5352	1.6519	6.1871
% Var	0.567	0.206	0.773

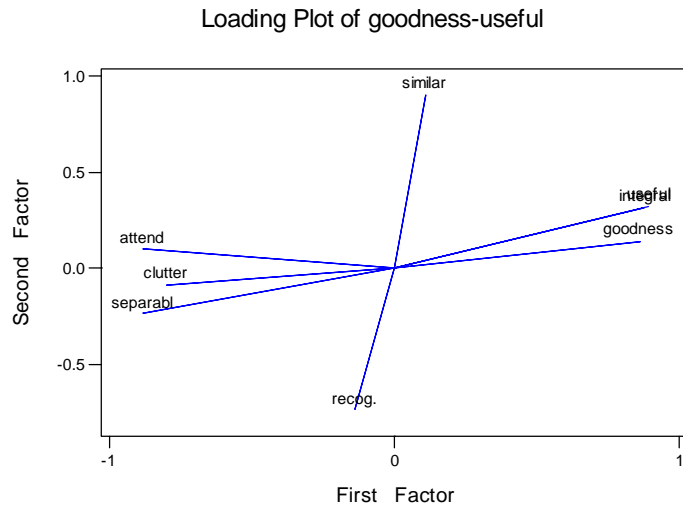


Figure 115. Loadings plot for the planets-six-low factor analysis.

Table 120. Factor analysis table for the planets-eight-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.855</b>	0.076	0.736
integral	<b>0.885</b>	-0.274	0.859
attend	-0.484	-0.589	0.580
clutter	<b>-0.732</b>	-0.243	0.594
separable	-0.163	<b>-0.904</b>	0.843
similar	0.088	<b>0.933</b>	0.878
recog.	-0.030	0.093	0.010
useful	<b>0.945</b>	0.071	0.898
Variance	3.2109	2.1864	5.3972
% Var	0.401	0.273	0.675

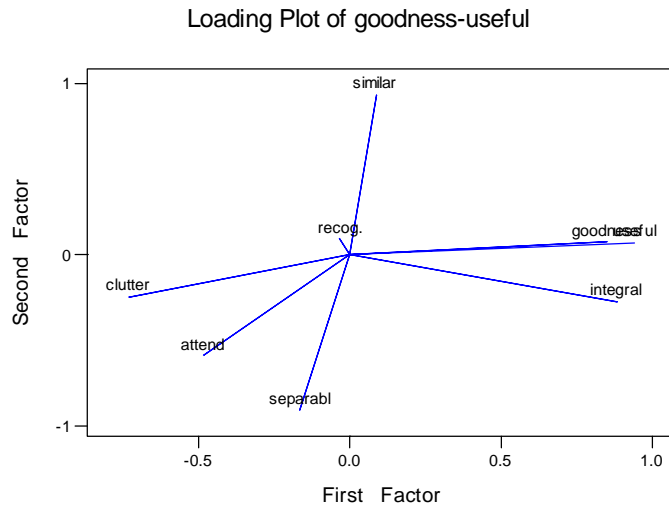


Figure 116. Loadings plot for the planets-eight-high factor analysis.

Table 121. Factor analysis table for the planets-eight-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.880</b>	-0.036	0.776
integral	<b>0.919</b>	-0.087	0.852
attend	<b>-0.923</b>	0.053	0.856
clutter	<b>-0.865</b>	0.047	0.750
separable	-0.076	<b>0.716</b>	0.518
similar	0.337	0.380	0.258
recog.	-0.331	<b>0.796</b>	0.743
useful	<b>0.858</b>	-0.319	0.838
Variance	4.1857	1.4057	5.5913
% Var	0.523	0.176	0.699

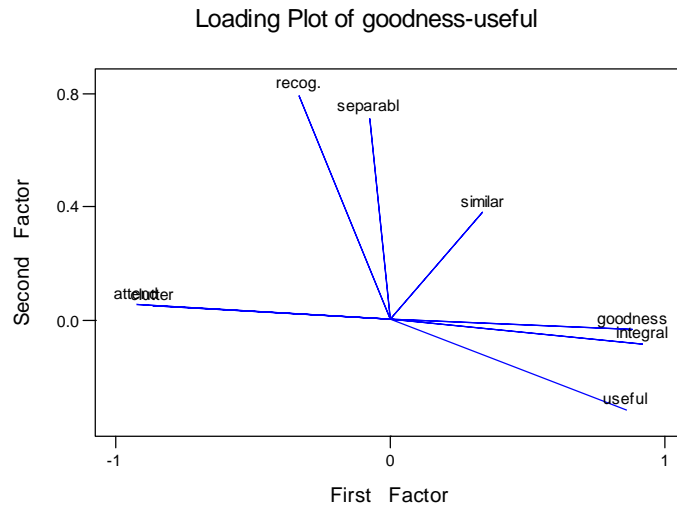


Figure 117. Loadings plot for the planets-eight-low factor analysis.

Table 122. Factor analysis table for the trees-two-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.873</b>	0.102	0.773
integral	<b>0.836</b>	-0.217	0.746
attend	<b>-0.740</b>	0.013	0.548
clutter	0.037	<b>-0.825</b>	0.681
separable	<b>-0.765</b>	-0.108	0.596
similar	0.536	-0.172	0.317
recog.	0.045	<b>-0.862</b>	0.746
useful	<b>0.866</b>	-0.086	0.757
Variance	3.6343	1.5300	5.1643
% Var	0.454	0.191	0.646

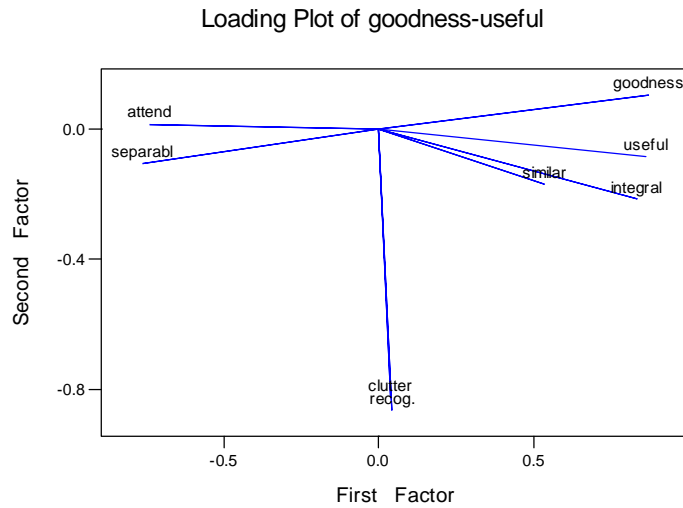


Figure 118. Loadings plot for the trees-two-high factor analysis.

Table 123. Factor analysis table for the trees-two-low conditions.

Variable	Factor1	Factor2	Communality
goodness	0.126	<b>-0.699</b>	0.504
integral	0.667	-0.572	0.772
attend	-0.172	<b>0.837</b>	0.730
clutter	0.479	0.561	0.544
separable	<b>-0.835</b>	0.026	0.699
similar	0.630	0.160	0.422
recog.	0.465	0.627	0.610
useful	<b>0.896</b>	-0.266	0.873
Variance	2.8326	2.3210	5.1536
% Var	0.354	0.290	0.644

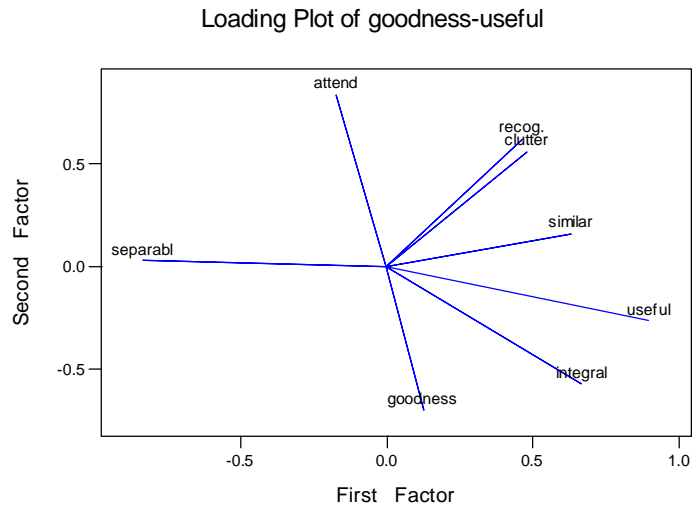


Figure 119. Loadings plot for the trees-two-low factor analysis.

Table 124. Factor analysis table for the trees-four-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.816</b>	-0.218	0.714
integral	<b>0.878</b>	-0.001	0.771
attend	<b>-0.869</b>	0.097	0.765
clutter	-0.444	<b>0.767</b>	0.785
separable	<b>-0.819</b>	0.355	0.798
similar	<b>0.834</b>	0.198	0.735
recog.	-0.184	<b>-0.819</b>	0.705
useful	<b>0.812</b>	-0.018	0.659
Variance	4.4491	1.4821	5.9312
% Var	0.556	0.185	0.74

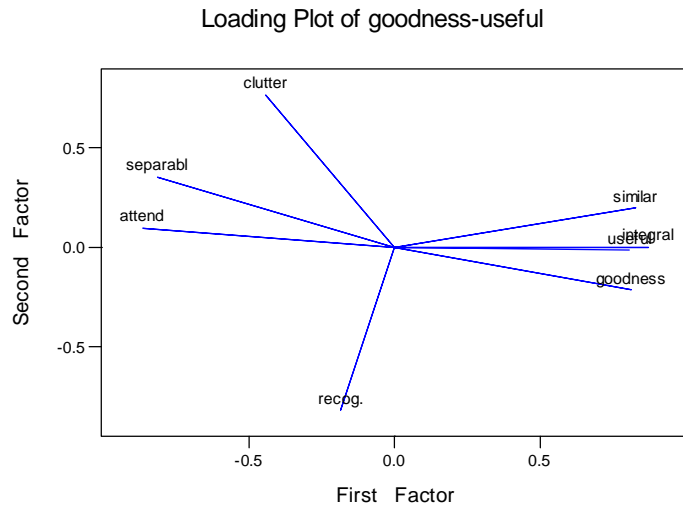


Figure 120. Loadings plot for the trees-four-high factor analysis.

Table 125. Factor analysis table for the trees-four-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>0.874</b>	0.028	0.765
integral	<b>0.893</b>	0.126	0.814
attend	<b>-0.872</b>	-0.054	0.763
clutter	-0.165	-0.672	0.479
separable	<b>-0.774</b>	-0.098	0.609
similar	0.342	-0.654	0.545
recog.	-0.161	<b>-0.877</b>	0.796
useful	<b>0.888</b>	-0.142	0.808
Variance	3.8802	1.6975	5.5777
% Var	0.485	0.212	0.697

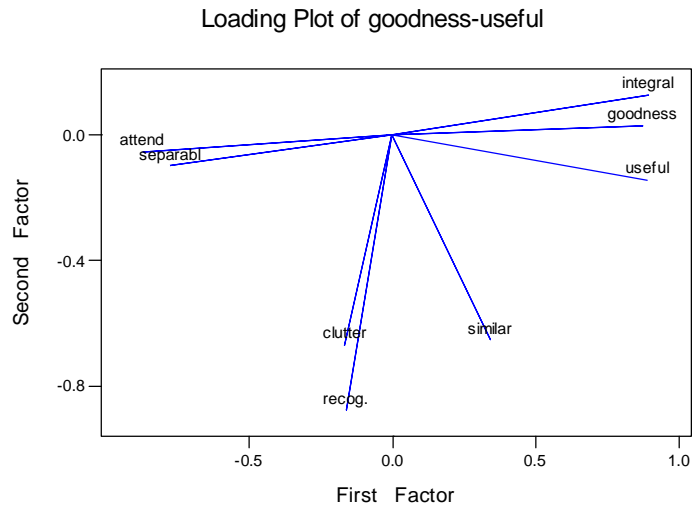


Figure 121. Loadings plot for the trees-four-low factor analysis.

Table 126. Factor analysis table for the trees-six-high conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>-0.948</b>	-0.053	0.902
integral	-0.833	0.173	0.724
attend	<b>0.853</b>	-0.001	0.728
clutter	<b>0.892</b>	0.180	0.829
separable	0.692	-0.029	0.480
similar	-0.540	0.634	0.693
recog.	0.193	<b>0.913</b>	0.871
useful	-0.871	0.192	0.795
Variance	4.6823	1.3385	6.0209
% Var	0.585	0.167	0.753

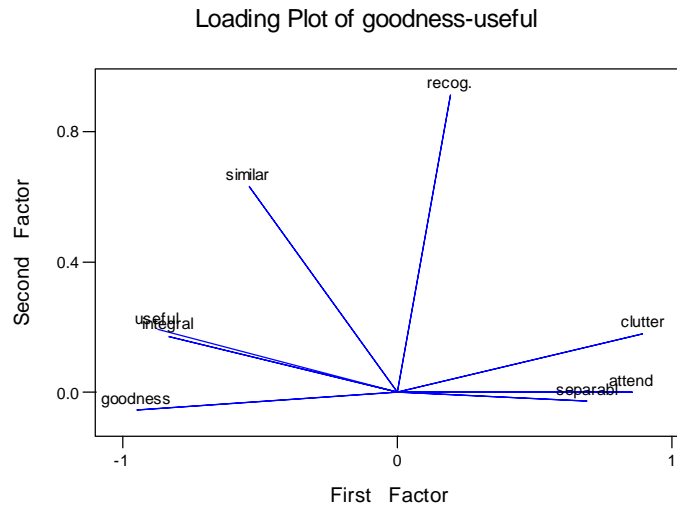


Figure 122. Loadings plot for the trees-six-high factor analysis.

Table 127. Factor analysis table for the trees-six-low conditions.

Variable	Factor1	Factor2	Communality
goodness	<b>-0.840</b>	-0.310	0.802
integral	-0.712	-0.558	0.819
attend	<b>0.907</b>	0.013	0.823
clutter	0.687	0.021	0.473
separable	0.508	<b>0.588</b>	0.604
similar	-0.276	-0.711	0.581
recog.	0.157	<b>-0.863</b>	0.769
useful	<b>-0.868</b>	-0.203	0.795
Variance	3.6204	2.0443	5.6647
% Var	0.453	0.256	0.708

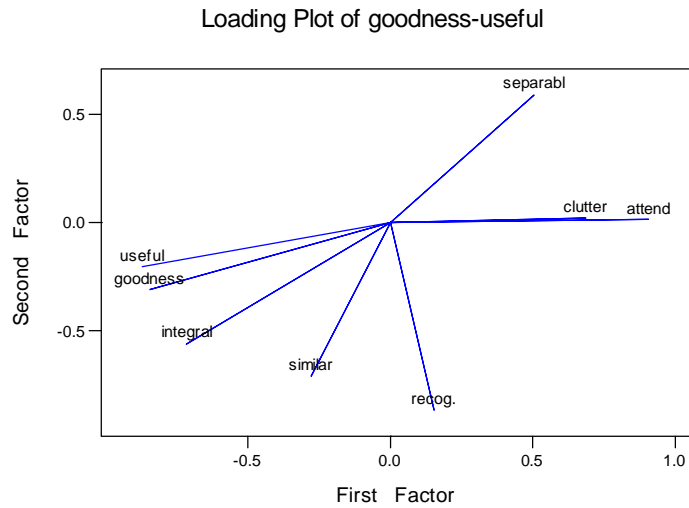


Figure 123. Loadings plot for the trees-six-low factor analysis.

Table 128. Factor analysis table for the trees-eight-high conditions.

Variable	Factor1	Factor2	Communality
goodness	0.775	-0.532	0.883
integral	<b>0.820</b>	-0.324	0.778
attend	<b>-0.824</b>	-0.171	0.709
clutter	<b>-0.856</b>	-0.223	0.783
separable	-0.581	<b>0.408</b>	0.504
similar	0.320	<b>-0.766</b>	0.689
recog.	-0.162	-0.665	0.469
useful	<b>0.824</b>	-0.319	0.781
Variance	3.8310	1.7648	5.5958
% Var	0.479	0.221	0.699

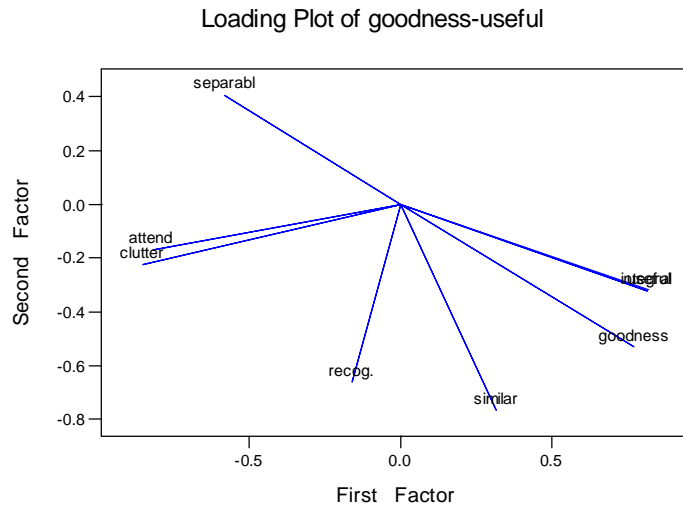


Figure 124. Loadings plot for the trees-eight-high factor analysis.

Table 129. Factor analysis table for the trees-eight-low conditions.

Variable	Factor1	Factor2	Communality
goodness	-0.545	0.618	0.679
integral	<b>-0.746</b>	0.547	0.855
attend	0.578	-0.433	0.521
clutter	<b>0.916</b>	0.138	0.858
separable	0.858	-0.255	0.801
similar	-0.318	0.632	0.501
recog.	0.165	<b>0.756</b>	0.599
useful	-0.573	0.712	0.836
Variance	3.2198	2.4300	5.6498
% Var	0.402	0.304	0.706

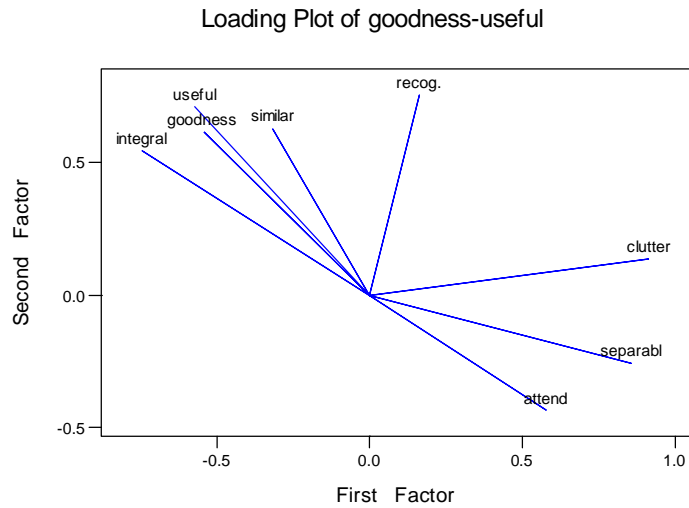


Figure 125. Loadings plot for the trees-eight-low factor analysis.

### 6.3.5. Correlations

Correlations and corresponding significance p-values were calculated to find the degree of relationship between the tested variables from experiment two. Table 130 lists and describes the various variables that were included in the correlation computations. Table 131 shows the correlations and corresponding p-values for the tested variables. Significant correlations and their p-values are highlighted in **bold** in Table 131.

Table 130. A list and description of the variables from the correlation tests for experiment two.

Variable	Variable Description
q1	What is the goodness rating of the augmented scene (1 high, 7 low).
q2	How separable were graphics from the real scene rating of the augmented scene (1 very, 7 not at all).
q3	How integral were graphics with the real scene rating of the augmented scene (1 very, 7 not at all).
q4	Tendency to focus on graphics or real scene rating of augmented scene (1 graphics, 7 real).
q5	How cluttered did graphics make the whole scene rating of the augmented scene (1 very, 7 not at all).
q6	How similar were graphics and real objects rating of the augmented scene (1 very, 7 not at all).
q7	Recognition of scene affected by graphics rating of the augmented scene (1 very, 7 not at all).
q8	Were the graphics useful to the overall scene rating of the augmented scene (1 very, 7 not at all).
rotate	Mental rotation test score (higher is better).
age	Age of the subjects (years).
gender	Gender of subjects (1 male, 2 female).
comp	Self-rating for time spent working on computers each week (1 low, 6 high).
games	Self-rating of time spent playing computer games (1 low, 7 high).
mental	Self-assessment of the ability to mentally manipulate objects within the mind (1 high, 7 low).

Table 131. Correlations and correlation p-values for tested variables from experiment two.

	q1	q2	q3	q4	q5	q6	q7	q8	rotate	age	gender	comp	games
q2	0.908												
	0.000												
q3	-0.782	-0.620											
	0.008	0.056											
q4	-0.855	-0.743	0.704										
	0.002	0.014	0.023										
q5	-0.939	-0.908	0.572	0.827									
	0.000	0.000	0.084	0.003									
q6	0.217	0.351	0.280	-0.126	-0.324								
	0.547	0.319	0.433	0.729	0.361								
q7	-0.479	-0.303	0.236	0.549	0.477	-0.239							
	0.161	0.395	0.511	0.100	0.164	0.506							
q8	0.843	0.846	-0.516	-0.674	-0.823	0.299	-0.505						
	0.002	0.002	0.127	0.033	0.003	0.402	0.136						
rotate	0.143	0.078	-0.185	-0.023	-0.170	-0.619	-0.056	0.071					
	0.694	0.831	0.610	0.950	0.638	0.056	0.879	0.846					
age	-0.426	-0.397	0.387	0.097	0.353	0.051	0.367	-0.265	-0.545				
	0.220	0.256	0.269	0.789	0.317	0.888	0.297	0.459	0.103				
gender	0.416	0.689	0.055	-0.217	-0.584	0.609	-0.029	0.466	0.046	-0.294			
	0.232	0.028	0.880	0.548	0.076	0.062	0.937	0.175	0.900	0.409			
comp	-0.336	-0.509	0.183	0.275	0.331	-0.529	0.291	-0.439	0.521	0.166	-0.319		
	0.342	0.133	0.613	0.442	0.350	0.116	0.414	0.204	0.122	0.646	0.368		
games	-0.311	-0.025	0.214	0.235	0.227	-0.092	0.803	-0.280	-0.125	0.413	0.151	-0.096	
	0.381	0.946	0.553	0.514	0.528	0.800	0.005	0.433	0.731	0.235	0.678	0.791	
mental	0.525	0.433	-0.393	-0.524	-0.500	-0.084	-0.350	0.761	0.117	0.163	0.074	0.024	-0.303
	0.119	0.211	0.261	0.120	0.141	0.817	0.322	0.011	0.747	0.652	0.838	0.948	0.395

## **6.4. Discussion**

### *6.4.1. Order*

As in the first experiment, each participant in experiment two made eight subjective assessments for 64 randomly ordered augmented scenes (comprising 32 unique scenes presented twice). Subjects were given demonstrations and asked to practice making assessments, however, some subtle practice effects were likely to occur during the experiment itself. Between the large number of conditions and the random order, these effects were hoped to be mitigated. The ANOVAs for both overall order (1-64) and individual scene presentation order (1,2) showed no significant effects for order as measured by any of the eight subjective assessments.

### *6.4.2. Augmented Environment, Integralness of Graphics, and Quantity of Graphics*

The second experiment continued the theoretical and practical investigation into integral perception in augmented reality. The main goal of experiment two was to examine the effects of including multiple graphics into augmented environments, specifically the adaptation of integral/separable perception research. As with experiment one, there was an abundance of significant results across the three independent variables of augmented environment (houses, computer lab, planets, trees), quantity (2, 4, 6, 8), and integralness (high, low). Each of the main effects were significant for at least six of the dependent measures. The interactions were significant in at least two of the ANOVAs performed on the dependent measures.

The first independent variable to be discussed is the augmented environment scene (houses, computer lab, planets, trees). However, unlike experiment one where the augmented scene was a single object, the scenes for experiment two were multiple object augmented environments. In each of the augmented scenes, graphic objects were added that closely resembled real objects already in the scene. As was described in the methodology section, the four scenes were a partially completed housing development (with added graphic houses), a half empty computer lab (with added graphic computer workstations), a picture of Jupiter and its moons (with added graphic planets and moons), and an outdoor field and trees scene (with added graphic trees). This

independent variable showed a significant main effect with six of the eight dependent measures (goodness, integral, attend, separable, similar, and usefulness). Recall that the dependent measures changed from experiment one, including some of the wording of the same questions (e.g., see the integral questions from the two questionnaires of experiments one and two).

The general pattern of significance for the independent variable augmented environment showed a consistent disparity between the planets and trees conditions, with the houses and lab conditions somewhere in the middle. The trees condition was consistently rated the worst (least integral, most separable, etc.) and the planets condition the best. One subject comment highlighted what might have occurred with the worse tree ratings in noting how obviously wrong even high integral graphics looked in a nature scene. It seems the purity of naturalistic scenes does not blend well with even slightly obvious artificially created graphics.

The planets are technically natural objects as well, however, the situation is much different. The planet graphics best embody the shape characteristics of the real planets, even in the low integralness conditions. A round circle with some coloration or pattern is going to appear at some level as being a planet when placed among other 'correctly' colored and shaped planets. The other three augmented environment conditions have a greater loss of shape information when the graphics are degraded to a low integral format.

The next main effect discussed is the differences between the four levels (2, 4, 6, 8) of the quantity independent variable. For the main effect of quantity, there were seven significant results, all but the similarity of graphic and real objects dependent measure. The expected result was a clear and pronounced slope in the direction of more negative ratings from two objects to eight. In actuality the prevailing pattern was a very flat quantity graph with only the two ends showing any practical differences, and these were not very large. This leads to a mixed interpretation of the impact of increased number of graphics in augmented reality applications. While increased quantity led to strong statistical evidence for decreased integral perception, the practical differences were not large.

The one quantity graph that did show the expected steep line was, not surprisingly, the clutter

measure shown in Figure 80. This figure shows a classic perceptual pattern, complete with a possible threshold between the six and eight levels of the quantity variable. The significant interactions involving the quantity independent variable provide more detail on the impact of increasing the number of graphics in an augmented environment scene.

The last main effect of the second study was a confirmation of earlier results in the first experiment. The graphics used to create the low and high integral conditions of experiment two followed the same procedures as those graphics created for experiment one that yielded the lowest and highest integral ratings. The ensuing variable in the second experiment, integrality, was significant for six of the eight dependent measures. Goodness, integrality, attend, separable, similarity, and usefulness had significant integrality main effects. Clutter and recognition did not. In all cases the high integrality condition was rated more favorably on integrally related dependent measures, and less so on separable type dependent measures. This result was completely expected. The lack of significance for integrality would have caused much concern about the overall soundness of this work.

The interactions for experiment two were not nearly as extensive as experiment one, and therefore not quite as informative. However, there were a total of four significant interactions to consider, two each for the interaction between the augmented environment and quantity independent variables, and that of augmented environment and integrality. As in experiment one, the augmented reality independent variable played a central role in the significant interactions.

The first interaction to be discussed will be augmented environment and quantity, significant in the clutter and recognition analyses. Quantity was previously shown to have only limited implications despite its significance in seven dependent measures. The underlying meaning behind the graphs of this interaction are difficult to discern as the values are all very close. However, there is one apparent anomaly with the houses-six condition.

The houses-six condition appears to contradict the overall increasingly negative effect of the larger number of graphics. This is clearer in the clutter graph (Figure 81), but can also be seen in the recognition graph (Figure 90). One possible explanation involves the graphics in the houses

condition having the most variability of all the augmented scene graphics.

The computer lab graphics were identical. It was already mentioned that the circular planet/moon objects were very similarly shaped circles. The more distant view of the trees scene caused all the trees graphics to generally look more similar than they actually were. But the house images, presented closer to the viewer, were all stylistically different houses. Therefore, the houses condition had the most perceptually varying graphics of all the augmented environments. The possible implication is that the two additional graphic houses added from the four to six condition were perhaps less intrusive than the graphic houses added for the other levels of the quantity variable. This possibility would lead to an increased subjective integral perception at the houses-six condition and an anomaly in the graph for the interaction between augmented environment and quantity.

Recall that a possible threshold was mentioned in the discussion on the main effect of quantity. There was a possible leveling off of the negative impact of increasing number of graphics between six and eight. The just discussed anomaly of the houses-six condition, leading to better integral perception, would actually mask this threshold effect. Considering the quantity effect without the houses-six condition would actually make the threshold effect more pronounced and, thus, be more likely to be a real phenomenon.

The last significant effect to be considered was the interaction between the independent variables augmented environment and integralness. The two significant interactions for this effect as measured by separable and similarity show a clear ordering of the augmented environments in the high integralness condition. In both measures the order from most similar to least similar, and least separable to most separable, is planets, houses, lab, and then trees. The high rating for planets is consistent with the main effect of augmented environments. The remainder of the order seems to represent the overall perceived quality of the high integral graphics for the respective augmented environments.

As in experiment one, the nonparametric tests added little to the overall data analysis. They did in all cases support the ANOVA main effect results.

### 6.4.3. *Subjective Rating Scales*

In the results section for experiment two it was explained that the procedures used to conduct the factor analyses would be slightly different than those used for experiment one. The main difference was that the data could not be reduced to eight factor analyses, rather 32 had to be completed. This potential complication was realized in more difficult to analyze results than those of the first experiment. While there was one clear pattern in all the loadings plots of experiment one, there are many different patterns in the loadings plots of experiment two. Factor analyses by nature are subject to interpretation and depend on ‘expert’ knowledge of the data. The discussion here focuses on finding patterns in the factor analyses, following the example of several other data analysis sections in this dissertation.

The 32 factor analyses can be reasonably divided into five groups. The groups were selected based on consistent features between sets of loadings plots that appeared relatively unique to each group. In practice this led to dependent measures that appeared frequently on factor one or factor two with a consistent opposite dependent measure (indicating a dependent measure relationship). It proved to be a difficult task to select groups based on consistent clusters of dependent measures. Factor one was first considered when making the groupings (as in all cases it describes that largest percentage of the variance), and then factor two. None of the groups had as unique and clear features as the factor analyses of experiment one. Nevertheless, this discussion will consider the patterns that did show in the groups chosen.

The first group seemed to represent the most ‘powerful’ of the five groups. It consisted of Figures 96, 98, 100, 108, 110, 111, 113, and 115. It was the only group that had a clear distinction among the independent variables that were used to input data into the factor analyses. This consistency was the lack of the trees augmented reality condition in any of the loadings plot factor analyses. It is not clear what impact the lack of trees variable had in this group.

The loadings plots for the first group show three possible straight line pairs of dependent measures across factors one and two. The more obvious interpretation appears to be in factor two with the recognition and similarity dependent measures. This factor appears to be a

recognition dominated factor. As the graphics are more similar, it is easier to recognize the scene as being a collection of some objects. Factor one is a much more complicated and interesting situation.

The goodness and attend dependent measures seem to be paired on factor one, as do integral and separable. It would seem intuitive to label factor one as the integral/separable factor as that has been the primary focus of this dissertation. However, several of the other groups will show that the most consistent variable pairing across the five groups is goodness and attend.

This pairing of goodness and attend, the most consistent across all five groups, is an interesting pair of dependent measures. Factor analyses are in part formed through the comparison of relative strengths of correlation between dependent measures. As shown in the questionnaire, the goodness measure has lower ratings for 'excellent' goodness and higher ratings for 'terrible' goodness. The attend measure has lower ratings for 'attending to the graphics' and higher ratings for 'attending to the real scene'. So, when the real scene tends to be attended to more than graphics, subjects have a lower goodness perception of the whole augmented scene. And vice versa, when the graphics are attended to, the subjects have a higher goodness perception of the scene. This is an interesting correlation, with the possible explanation that when the graphics are 'good', the subjects had a tendency to look at them, and when the graphics were 'not good', subjects relied on the overall context of the real scene to figure out exactly what the graphics were supposed to represent.

The second grouping comprised Figures 94, 103, 104, 109, 114, 116, 123, and 124. The loadings plots show clutter and attend at one extreme of factor one and usefulness and goodness at the other end. On factor two, similarity and separable can reasonably be considered consistent opposite partners. Goodness and attend again show as an almost straight line pair, however, the coupling is not as distinct as in the first group. The similar and separable measures are a pair on the second factor, meaning that more similar objects are less separable. In two of the eight figures in group two, useful and clutter are paired on a straight line. That is not enough conclusive evidence to say they are a matched pair.

Considering the four factor one variables together for group two, they appear to have an underlying meaning of “attending to quality graphics,” as was partially discussed previously. Clutter seems to add a reasonable element to this mix by revealing that an un-cluttered scene is correlated with usefulness and goodness.

The third group considers the Figures 95, 97, 99, 105, 106, 118, 121, and 122. A clear difference in the loadings plots of group three is the lack of a dependent measure in the positive direction of factor two. Clutter and recognition are loosely grouped only in the negative direction of factor two with no opposing dependent measures. This factor seems to indicate a very cluttered scene that is harder to recognize as a meaningful set of objects. In the first factor the straight line pairing of goodness and attend once again shows up. As in the first group the integral/separable match is also on factor one. The last addition to factor one is the usefulness measure, often appearing alongside goodness and integral. In the factor analyses results from experiment one, goodness and usefulness were often clustered, so this addition to factor one seems quite sensible.

The fourth group is the last group that has clear members on the first factor. It also represents a slight shift from the first three groups, as the goodness and attend pairing now appears on factor two. This is also the first group that has less than eight members, only Figures 102, 107, and 119 comprise the group. Consequently, the relative strength of the interpretations for the classifications begins to decrease with this group.

Factor one for the fourth group has the separable measure on one end and the usefulness and similar measures on the other. The pairing of separable and similarity was seen in group two. Usefulness and separable was evidenced in group three. The grouping here seems to be a combination of these earlier pairings. The goodness and attend measures on factor two were already mentioned. However, they do appear in relative isolation on these three graphs.

The fifth group lacks a definitive character, particularly on factor one. The group is more defined by one dominant feature of factor two, the recognition dependent measure. The graphs in this group are Figures 97, 112, 117, 120, and 125. Generally, these figures make up group five because they lacked a good fit with the other four groups, and thus warrant no discussion for

patterns that do not exist.

#### *6.4.4. Correlations*

As in experiment one, there were a number of expected correlations between the dependent measures that were more appropriately analyzed by factor analyses. There were three additional significant correlations, one involving gender and the integral dependent measure, one between computer game playing and the recognition dependent measure, and one between self-assessment of mental manipulation ability and the usefulness dependent measure.

The correlation involving gender will not be elaborated on for the same reasons as were discussed in experiment one. There were only two female subjects for a binary gender variable. That is not enough information to make general conclusions.

The second correlation between game playing and recognition of the augmented scene is mildly interesting. The correlation indicates that a frequent computer game player was more affected by the graphics when trying to recognize the visual augmented scene. Not a surprising result considering 3D games that are played today constantly require players to recognize scenes based on graphic objects.

The last correlation between self-assessment of mental rotation ability and the usefulness subjective dependent measure seems to lack validity. No reasonable explanation readily exists based on the data and experimenter observation of these studies.

#### *6.4.5. Design Guidelines*

As in the first experiment, there was an abundance of useful information in the data analysis and discussion that can be summarized in design guidelines. Table 132 states design guidelines as they pertain to either human factors, perception, and graphic design in augmented reality that came from experiment two.

Table 132. Design guidelines from experiment two.

Guideline	Implication
1. Use as few graphics as possible when designing integrally perceived augmented environments.	Human Factors, Graphic Design
2. For augmented applications that must use low-integral computer-generated graphics, maintain as much shape information as possible.	Graphic Design
3. When using multiple graphics in augmented applications, limit the total number of computer-generated graphics to six.	Human Factors, Graphic Design
4. For even minimal integral perception in nature scenes use extremely realistic computer-generated graphics, particularly for landscapes.	Graphic Design
5. For a comprehensive assessment of integral perception in multiple graphic augmented applications, ask direct subjective questions on both integral perception of the whole scene and attention paid to individual items.	Human Factors, Perception

As in the first experiment, these guidelines are based on the premise that integrally perceived graphics are scenes composed of both real and computer-generated imagery that is perceived as a unified whole. The guidelines presented here focus on augmented applications with multiple graphics. The underlying premise is that integral perception can occur when multiple computer-generated graphics are used, however, the number of graphics should be kept to a minimum to maintain integral perception.

## CHAPTER 7. EXPERIMENT THREE

### 7.1. Methodology

#### 7.1.1. Experimental Design

The third experiment investigated the effects of integrally developed augmented scenes on learning and task performance. The main goal of experiment three was to establish that the theoretical constructs formed in experiments one and two made a practical difference in representative real-world tasks. For this study, two phases of experimentation were completed with each participant: the first involved learning information from augmented scenes; and, the second examined the use of augmented scenes as instructional tools for assembly tasks.

The first phase, the learning tasks, presented subjects with three augmented images selected from the first two experiments. After viewing each image, participants made subjective assessments on the nature of the augmented image and took a written ‘quiz’ about the content of the scene.

For the learning phase, there was one between subjects independent variable, level of graphic integralness (high, low) and one within subjects variable, augmented scene (houses, computer lab, brain). The dependent measures were quiz scores, subjective assessments about the content and quality of the augmented images, and drawings of the augmented images. The quiz scores asked questions on the content and placement of objects within the augmented scenes. The subjective assessments were taken from the first two experiments (goodness, integral, and usefulness). Appendix E shows the questionnaires used to gather quiz and subjective assessment information for each of the three augmented images.

For the first phase, two-factor mixed design ANOVAs were completed for the subjective assessments of the real scene. Nonparametric analyses were used to confirm the subjective measure parametric analyses, to examine the quiz score information, and evaluate the re-creation drawings.

In the second phase of experiment three subjects completed eight assembly tasks based on verbal and augmented video instructions. The subjects first heard and viewed the instructions, then completed the individual assemblies. After each assembly, subjective assessments of the augmented instructions were made.

As in the first phase, there was one between subjects variable, level of graphic integralness (high, low), and one within subjects variable, assembly task (1,2,3,4,5,6,7,8). Dependent measures were time to complete task, errors made during task, and subjective assessments on the nature of the augmented video clip instructions. Time was recorded in seconds from the start of assembly to completion of the task as directed by the instructions, or to a point where the participant perceived assembly was correctly completed. A maximum time of 300 seconds was allowed for assembly. Critical incident errors were not specifically recorded during each assembly task; however, notations were listed if the participant made an obvious error during the assembly, and/or could not complete the assembly. The error information, for the purposes of this study, was recorded in the data analysis in binary format ('1' for errors, '0' for no errors). The subjective assessments were taken from the first two experiments (goodness, integral, and usefulness), with one additional question on the overall usefulness of the augmented instructions for completing the assembly task. Appendix F shows the questionnaire of subjective assessments used for each augmented video clip instruction.

For the second phase of experiment three, two-factor mixed ANOVAs were completed for time to complete task, and subjective assessments of the augmented instructions. Nonparametric analyses were performed to confirm the significant parametric subjective assessment effects and to analyze errors.

Practice and order effects were not controlled in either phase of experiment three, unlike the previous two experiments. The practice effect, if any, was controlled for by maintaining the exact same order of tasks between subjects. Thus, if there was an order effect, it would be the same for all participants. It was possible a practice effect existed for the assembly tasks, however, order was completely confounded with the assembly task independent variable. Therefore, no specific analyses were performed on order.

In addition to the analyses performed on the main independent variables for both phases of experiment three, supplementary correlations were calculated for various experimental parameters. The correlation analysis considered variables that had shown significance in either the first or second phase of experiment three in addition to general participant information.

### *7.1.2. Participants*

Twenty participants, aged 19-38, came from the general Virginia Tech university population. To be eligible for the study, participants needed to correctly perceive at least 7 out of 8 items on the 20/40 line of an eye chart and correctly identify 7 out of 8 color vision trials. No participant failed to meet these requirements.

### *7.1.3. Experimental Apparatus*

7.1.3.1. Experimental Setup. The equipment used to run the augmented reality system in experiment three was the exact same as used in experiments one and two, previously described in the equipment setup section of experiment one. The one addition was a Watec WAT-205A Color NTSC miniature video camera attached to the top of the head-mounted display, shown in Figure 4. The second phase of the third experiment involved a disassembled Sony Model No. HCD-C50U Compact Disc Deck Receiver shelf stereo system. Figures 126 through 129 show examples of four of the stereo components used in the assembly experimental tasks.

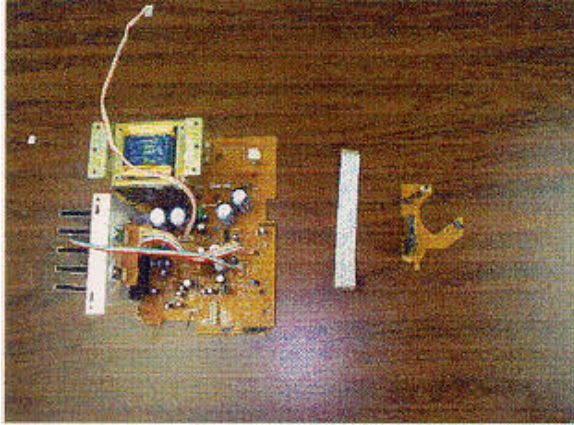


Figure 126. Stereo components used in assembly task two of experiment three.

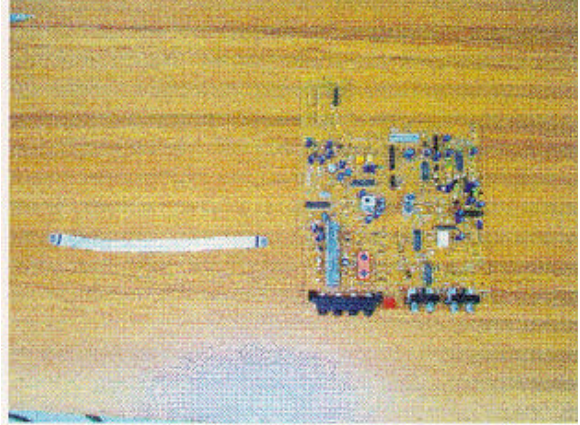


Figure 127. Stereo components used in assembly task five of experiment three.

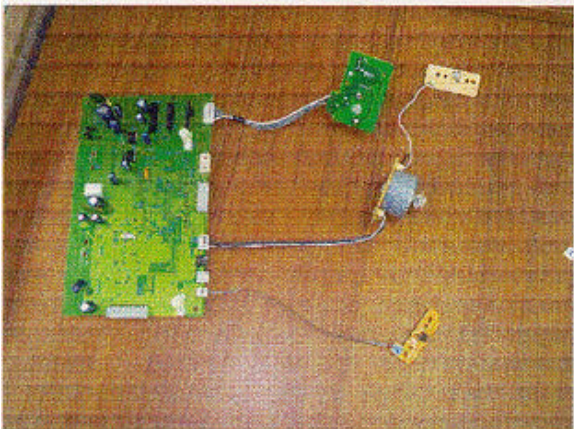


Figure 128. Stereo components used in assembly task seven of experiment three.

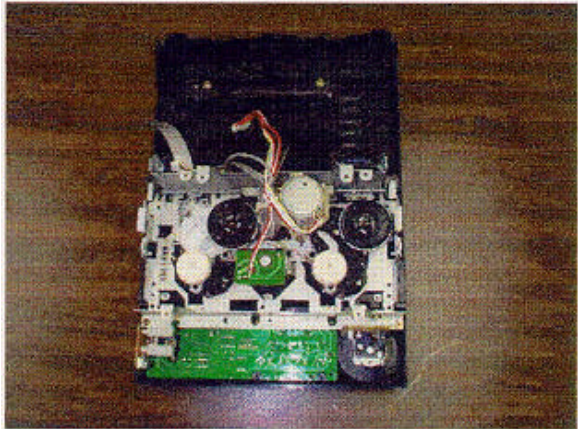


Figure 129. Stereo components used in assembly task eight of experiment three.

7.1.3.2. Augmented Environments. The equipment used to create the augmented environments in experiment three was the same as used in experiments one and two, detailed earlier in the augmented environments section of experiment one. The augmented reality stimuli for the first phase of experiment three were taken from experiments one and two. The augmented video clip instructions for the second phase were new stimuli created by the same methods as the previous two studies.

The first phase of experiment three utilized three images that were originally used in the first two studies. The first image was the housing development scene from experiment two with four graphic houses added to the development. The second image was the computer lab scene from experiment two with eight graphic computers added to the lab. The third image was the brain scene from experiment one showing an overhead view of a seated person with a superimposed brain image on top of the head. The exact composition of these images was described in the augmented environments sections of the respective experiments.

For the second phase of experiment three, augmented video clip instructions were created showing step-by-step actions needed to accomplish the eight assembly tasks. The video clip instructions were comprised of a series of two to four 10 second static augmented scenes that depicted how various stereo parts would need to be manipulated to complete assembly. The augmented scenes were created by first videotaping selected components of the disassembled stereo, then adding digital photographs of additional stereo components needed for each particular assembly task. The digital photographs of the additional stereo components were altered to meet the needs of the experimental design of high and low integral graphics depending on condition. To complete the set of step-by-step instructions, the position of the real objects and graphics were changed in each scene to represent needed actions to accomplish the given assembly task.

Prior to viewing the augmented video clip instructions, the experimenter read a verbal instruction to supplement the augmented video clips. Table 133 lists the eight verbal instructions for each assembly task.

Table 133. The eight verbal instructions for the second phase of experiment three.

<b>Verbal Task Instruction</b>
Task 1. Insert the two electronic strips into the sockets as indicated in the augmented video clip instructions.
Task 2. Attach the two components with the electronic strip as indicated in the augmented instructions.
Task 3. Adjust the stereo piece as shown in the augmented video clip instructions. Visual feedback will accompany successful activation of that stereo piece.
Task 4. Put the plug into the hole in the plastic shell as shown in the augmented instructions. Pull the plug cord through the hole until an attached plate is flat on the plastic shell. Turn the whole unit over and attach the screws as indicated.
Task 5. Turn the circuit board over and attach the electronic strip into a slot as shown in the augmented reality instructions.
Task 6. Slide the indicated piece forward to raise the CD cartridge. After raising the CD cartridge all the way, open the lowest CD drawer and remove the CD.
Task 7. Attach the mini electronic strip plugs into the correct sockets as shown in the augmented instructions.
Task 8. Spin the indicated disk in the direction shown on the augmented instructions until a faint clicking noise is heard. perception of whole scene rating of the augmented scene (1 unified, 7 distinct).

An example task in Figure 130 shows images from the low integral condition of assembly task number one. It can be clearly seen that the objects to the right of the top picture of Figure 130 are strips of some sort, and in the following pictures they attach to the shown positions on the real circuit board as indicated. For comparison, Figure 131 shows an image of all the real objects as the subject saw them when actual assembly started.

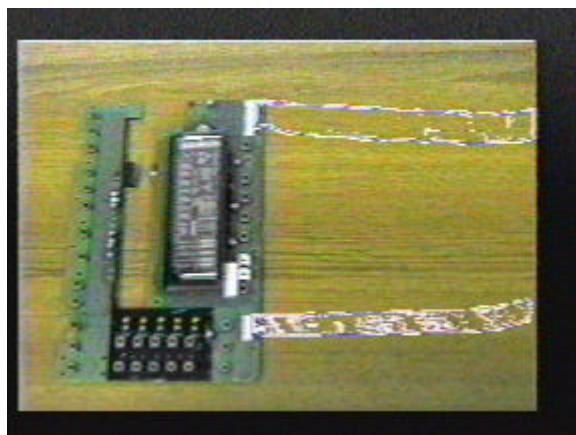
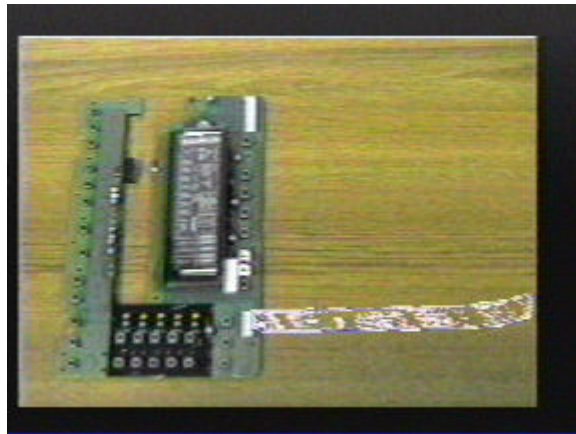
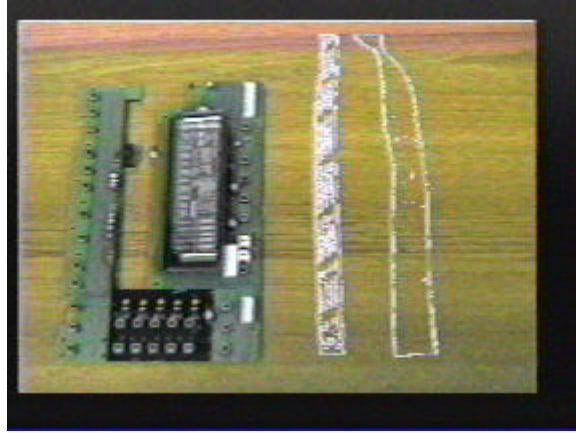


Figure 130. The three steps in the augmented video clip instructions for assembly task one of experiment three.

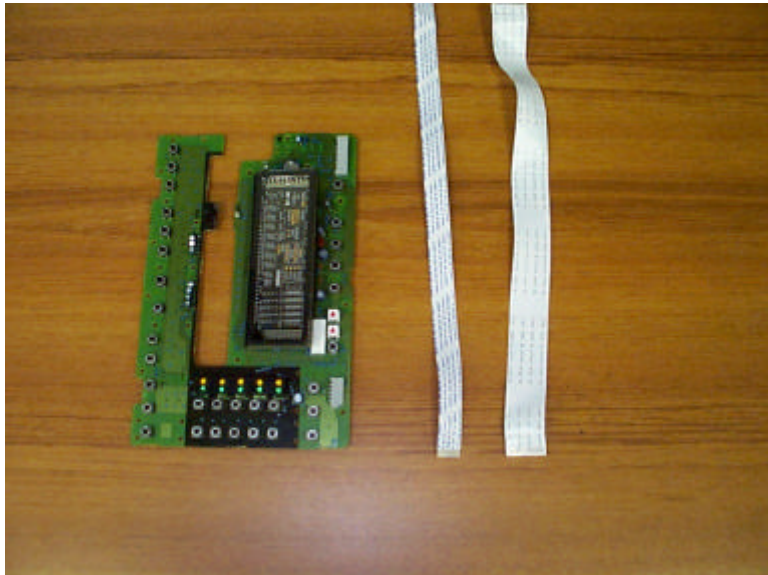


Figure 131. A digital picture of what the subjects saw in reality at the beginning of assembly task one of experiment three.

7.1.3.3. Questionnaires. There were five questionnaires, or quizzes, used in experiment three. The first was the same general information questionnaire that was used in the previous two experiments (shown in Appendix B). The next three questionnaires were administered after each of the three augmented images were viewed by subjects in the first phase of experiment three. The last questionnaire was given after each of the assembly tasks was completed during the second phase of experiment two.

Each of the phase one questionnaires contained four ‘quiz’ questions about the content of the augmented image, three subjective assessment questions on the nature of the augmented image, and a re-creation of the augmented scene drawing section. The three subjective assessments were: how integral the graphics were with the whole scene, what was the goodness of the whole scene, and how useful and appropriate were the graphics in adding information to the overall scene. Appendix E shows the three quiz questionnaires used for the first portion of experiment three.

The questionnaires given after each assembly task in the second phase had subjects rate the quality of the augmented video clip instructions with four questions. The questionnaire, shown in Appendix F, asked a goodness question, an integral question, and a usefulness question that were the same as used in the previous experiments. In addition, a last question was asked on the overall usefulness of the augmented video clip instructions for completing the task.

7.1.3.4. Vision Tests. The vision tests and criteria for participation were the same as used in the first two experiments. The two tests, acuity and color vision, were described in the vision tests section of experiment one.

#### *7.1.4. Experimental Procedures*

Participants were first greeted and asked to read and sign the consent form, shown in Appendix B. Upon agreeing to participate, subjects completed the general questionnaire and the two visual tests. Augmented reality in general, and the two phases of the study were then explained to the subjects. After the introduction to the studies, the first phase of experiment three began.

Once the subjects were comfortably and correctly wearing the head-mounted display, the experimenter played the first videotape for phase one. The video clip lasted 20 seconds, then the subject was instructed to remove the head-mounted display and complete a questionnaire as indicated. The subject viewed a total of three images and completed three questionnaires in the same fashion.

After completing the first phase of experiment three, subjects received more instruction on the nature of the assembly tasks. The procedures were explained, two representative demonstrations were shown, subject practiced example demonstration assemblies, the miniature video camera on top of the head-mounted display was explained, and any questions that the participant had were answered. The subject then was instructed to put on the head-mounted display once again and the assembly tasks began.

For each assembly task the subject began by viewing a blank screen (inside the head-mounted display). While watching the blank screen the appropriate verbal assembly task instruction was read by the experimenter. Next, the corresponding augmented video clip instructions were shown to the subject. Following the augmented video clip instructions, the subject once again briefly viewed a blank screen while the verbal instruction was repeated. Then the real assembly items were placed in front of the subject on a clear desk and the miniature video camera mounted on the head-mounted display was turned on. The experimenter began timing the assembly and the subject assembled the task as instructed. An upper limit of five minutes was instituted if the subject failed to complete any of the experimental tasks. Upon completion of the assembly, subjects removed the head-mounted display to answer the verbal assessments of the augmented video clip instructions. This set of procedures was repeated for each of the eight assembly task trials. At the completion of the last assembly task, the participants were informed the experiment was over and any questions were answered, and lastly they were thanked for their time.

## ***7.2. Expected Results***

The first phase of experiment three was conducted to assess whether augmented reality scenes created with more integral graphics would provide a better learning environment than less integral

graphics. It was expected that after viewing augmented scenes with integral graphics subjects would be better enabled to answer questions about the visual scene. Some of the 'quiz' questions contained one right answer, and others were purely judgements based on visual content. Subjects who viewed more integral augmented realities were expected to provide more accurate and consistent answers. Furthermore, it was hypothesized that subjects who viewed more integral graphics in their augmented scenes would draw the visual scene better than subjects with less integral graphics.

The second phase of experiment three was expected to show that subjects who viewed augmented video clip instructions with more integral graphics would complete assembly tasks faster and with fewer errors than subjects that had less integral graphics in their augmented instructions. This expectation was true for seven of the eight assembly tasks. In one condition, task seven, augmented information was purposely shown incorrectly. In this condition it was hypothesized that subjects with more integral graphics would make more errors and take longer time to complete the assembly.

The correlations for experiment three were expected to show some similar relationships between assessment measures as in the previous two experiments. In addition to the subjective measure significant relationships, it was hypothesized that quiz scores and time to complete assembly would vary significantly with one or more of the subjective assessment measures. Another correlation expected was self-assessment on mentally visualizing objects and time on task. While other significant correlations are possible, no specific correlation results have been designed for in experiment three.

### 7.3. Results

For the first phase of experiment three, the analyses were similar to the previous two studies for the subjective assessment rating scale dependent measures. Both ANOVAs and corresponding nonparametric tests were performed. The quiz questionnaire data was analyzed with nonparametric tests only as the majority of the questions could not be construed as continuous information. In addition to these analyses, a rank order test of the re-creation drawings of the augmented scenes was performed.

For the second phase of experiment three, the subjective assessments were analyzed as in the other studies with both ANOVAs and corresponding nonparametric tests. The continuous variable time was analyzed with an ANOVA only. In contrast, the binary error data was analyzed with only a nonparametric test.

In both phases of experiment three, the between-subjects testing afforded the use of  $p < 0.05$  significance consideration for all the ANOVA results. The use of the between-subjects design reduces subject confounding. Significant effects in the summary tables are shown in **bold**. The figures showing significant results use trend lines as in the previous two experiments.

#### 7.3.1. Subjective Measure ANOVAs: Quiz Augmented Scenes

Mixed-factor ANOVAs were performed for the independent variables augmented reality scene (houses, lab, trees) and level of integralness of the graphics within the overall scene (high, low), as measured by three subjective assessments of the augmented scene (integral, goodness, usefulness).

Table 134 shows the ANOVA summary table for the two independent variables as measured by integral assessments. There was one significant interaction. Figure 132 shows the separation in the brain augmented image between the two integralness conditions.

Table 134. ANOVA summary table for augmented reality scene and integralness, as measured by integral ratings.

Source	DF	SS	MS	F	P
ar	2	0.933	0.467	0.22	0.800
integralness	1	12.150	12.150	2.98	0.102
subject(integralness)	18	73.500	4.083	1.97	
<b>ar*integralness</b>	<b>2</b>	<b>19.600</b>	<b>9.800</b>	<b>4.72</b>	<b>0.015</b>
Error	36	74.800	2.078		
Total	59	180.983			

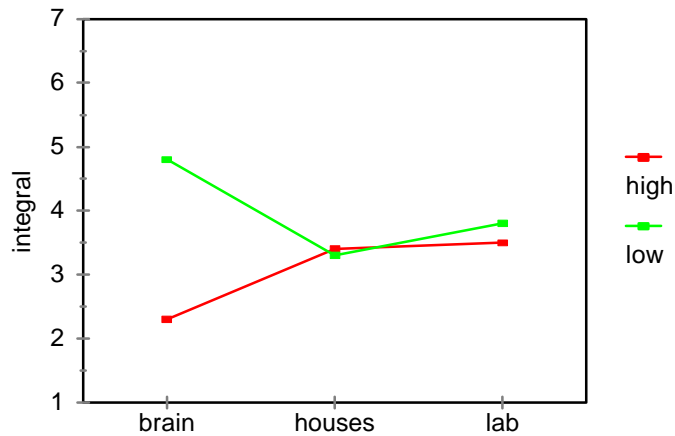


Figure 132. Graph of the interaction between augmented reality scene and integralness, significant at the  $p < 0.05$  level ( $p = 0.015$ ) as measured by integral ratings.

Table 133 shows the ANOVA summary table for the two independent variables as measured by goodness assessments. There were no significant effects.

Table 135. ANOVA summary table for augmented reality scene and integralness, as measured by goodness ratings.

Source	DF	SS	MS	F	P
ar	2	2.233	1.117	0.72	0.493
integralness	1	8.817	8.817	2.26	0.150
subject(integralness)	18	70.167	3.898	2.52	
ar*integralness	2	8.033	4.017	2.59	0.089
Error	36	55.733	1.548		
Total	59	144.983			

Table 134 shows the ANOVA summary table for the two independent variables as measured by usefulness assessments. There were two significant main effects and one significant interaction. Figure 133 shows the lab condition with lower usefulness ratings (indicating more usefulness perception). Figure 134 displays the high integralness condition with lower usefulness ratings. Figure 135 indicates a separation between the brain augmented conditions high and low integralness usefulness ratings.

Table 136. ANOVA summary table for augmented reality scene and usefulness, as measured by usefulness ratings.

Source	DF	SS	MS	F	P
<b>ar</b>	<b>2</b>	<b>18.633</b>	<b>9.317</b>	<b>4.58</b>	<b>0.017</b>
<b>integralness</b>	<b>1</b>	<b>26.667</b>	<b>26.667</b>	<b>17.31</b>	<b>0.001</b>
subject(integralness)	18	27.733	1.541	0.76	
<b>ar*integralness</b>	<b>2</b>	<b>23.433</b>	<b>11.717</b>	<b>5.76</b>	<b>0.007</b>
Error	36	73.267	2.035		
Total	59	169.733			

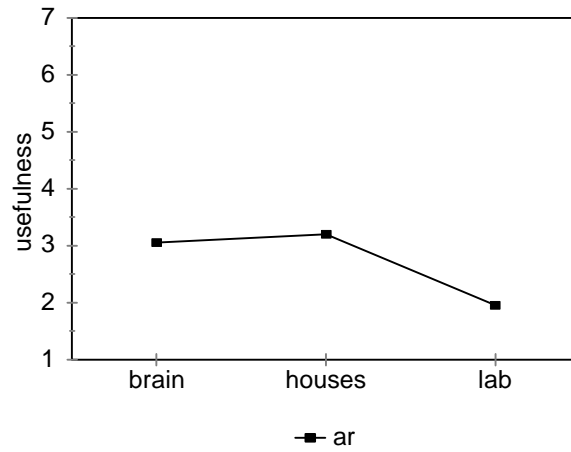


Figure 133. Graph of the main effect of augmented reality scene, significant at the  $p < 0.05$  level (0.017) as measured by usefulness ratings.

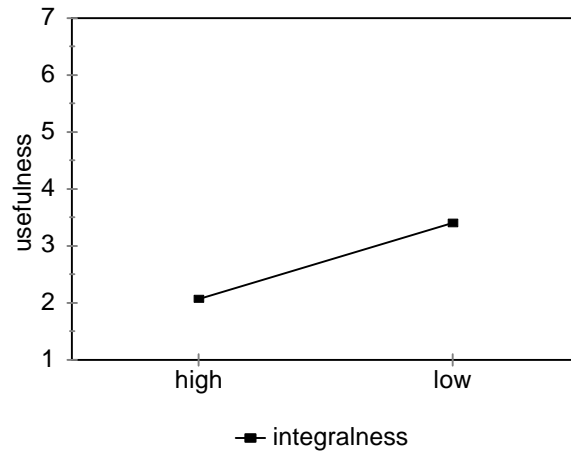


Figure 134. Graph of the main effect of integralness, significant at the  $p < 0.05$  level ( $p = 0.001$ ) as measured by usefulness ratings.

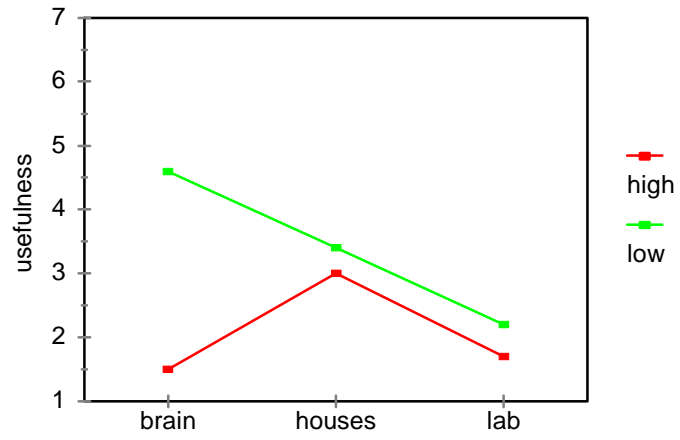


Figure 135. Graph of the interaction between augmented reality scene and integralness, significant at the  $p < 0.05$  level ( $p = 0.007$ ) as measured by usefulness ratings.

7.3.2. *Quiz Answers and Subjective Measure Nonparametric Tests: Quiz Augmented Scenes*

The answers given for the quiz scores in the learning phase of experiment three were ordinal scale data appropriate for nonparametric analysis. Tables 137 through 139 show the summary tables for the Kruskal-Wallis tests of the houses augmented scene. Table 137 shows a significant difference in additional houses estimates between the two integralness conditions. Table 136 shows a significant difference in the number of estimated floors in the closest house between the two integralness conditions. Table 137 does not show any significant differences.

Table 137. Kruskal-Wallis test for integralness as measured by number of additional houses fitting on back hill in the houses augmented scene.

<u>integralness</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	2.500	6.7	-2.87
2	10	4.000	14.3	2.87
Overall	20		10.5	

H = 8.25 DF = 1 P = 0.004  
H = 8.81 DF = 1 P = 0.003 (adjusted for ties)

Table 138. Kruskal-Wallis test for integralness as measured by number of floors in the closest house in the houses augmented scene.

<u>integralness</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	1.000	7.4	-2.34
2	10	2.000	13.6	2.34
Overall	20		10.5	

H = 5.49 DF = 1 P = 0.019  
H = 6.52 DF = 1 P = 0.011 (adjusted for ties)

Table 139. Kruskal-Wallis test for integralness as measured by perceived average prices of the houses in the houses augmented scene.

<u>integralness</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	112500	9.5	-0.79
2	10	112500	11.6	0.79
Overall	20		10.5	

H = 0.63 DF = 1 P = 0.427  
H = 0.66 DF = 1 P = 0.417 (adjusted for ties)

Tables 140 through 142 show the summary tables for the Kruskal-Wallis tests of the computer lab augmented scene. Table 140 shows a significant difference in the perceived number of graphic computers shown in the computer lab between the two integralness conditions. Tables 141 and 142 do not show any significant differences.

Table 140. Kruskal-Wallis test for integralness as measured by number of graphically presented computers in the computer lab augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	8.000	13.4	2.15
2	10	6.000	7.7	-2.15
Overall	20		10.5	

H = 4.64 DF = 1 **P = 0.031**  
H = 5.06 DF = 1 **P = 0.025** (adjusted for ties)

Table 141. Kruskal-Wallis test for integralness as measured by the total number of computers shown in the computer lab augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	12.00	8.5	-1.55
2	10	14.00	12.6	1.55
Overall	20		10.5	

H = 2.40 DF = 1 P = 0.121  
H = 2.84 DF = 1 P = 0.092 (adjusted for ties)

Table 142. Kruskal-Wallis test for integralness as measured by the monitor size of the computers shown in the computer lab augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	17.00	10.4	-0.08
2	10	17.00	10.6	0.08
Overall	20		10.5	

H = 0.01 DF = 1 P = 0.940  
H = 0.01 DF = 1 P = 0.932 (adjusted for ties)

Tables 143 through 147 show the summary tables for the Kruskal-Wallis tests of the brain augmented scene. None of the brain augmented scene summary tables show any significant differences.

Table 143. Kruskal-Wallis test for integralness as measured by the row that did not have graphically presented computers in the computer lab augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	1.000	10.1	-0.34
2	10	1.000	11.0	0.34
Overall	20		10.5	

H = 0.12 DF = 1 P = 0.734  
H = 0.20 DF = 1 P = 0.654 (adjusted for ties)

Table 144. Kruskal-Wallis test for integralness as measured by the number of brain structures in the brain augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	4.000	9.0	-1.13
2	10	4.000	12.0	1.13
Overall	20		10.5	

H = 1.29 DF = 1 P = 0.257  
H = 1.37 DF = 1 P = 0.241 (adjusted for ties)

Table 145. Kruskal-Wallis test for integralness as measured by the number of layers of the brain shown in the brain augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	3.000	11.9	1.02
2	10	2.000	9.2	-1.02
Overall	20		10.5	

H = 1.04 DF = 1 P = 0.307  
H = 1.10 DF = 1 P = 0.295 (adjusted for ties)

Table 146. Kruskal-Wallis test for integralness as measured by whether there were eyes represented in the brain augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	1.000	8.5	-1.51
2	10	2.000	12.5	1.51
Overall	20		10.5	

H = 2.29 DF = 1 P = 0.131  
H = 3.04 DF = 1 P = 0.081 (adjusted for ties)

Table 147. Kruskal-Wallis test for integralness as measured by whether there was a brain indent represented in the brain augmented scene.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	10	1.000	10.5	0.00
2	10	1.000	10.5	0.00
Overall	20		10.5	

H = 0.00 DF = 1 P = 1.000  
H = 0.00 DF = 1 P = 1.000 (adjusted for ties)

In addition to the quiz answers for the first phase, there were two significant main effect results in the previous section involving the subjective assessments of the quiz augmented scenes that were corroborated with nonparametric tests. Tables 148 and 149 show the summary tables for the results of the Kruskal-Wallis Tests for the two significant main effects of usefulness.

Table 148. Kruskal-Wallis test for augmented reality scene as measured by usefulness ratings.

<u>ar</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
brain	20	2.000	31.3	0.26
houses	20	3.000	36.9	2.01
lab	20	2.000	23.3	-2.27
Overall	60		30.5	

H = 6.15 DF = 2 P = 0.046  
H = 6.51 DF = 2 P = 0.039 (adjusted for ties)

Table 149. Kruskal-Wallis test for level of graphic integralness as measured by usefulness ratings.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
high	30	2.000	24.4	-2.71
low	30	3.000	36.6	2.71
Overall	60		30.5	

H = 7.36 DF = 1 P = 0.007  
H = 7.79 DF = 1 P = 0.005 (adjusted for ties)

### 7.3.3. Wilcoxon-Mann-Whitney Test: Quiz Augmented Scene Drawings

The drawings made by subjects for the three quiz augmented scenes were analyzed with the Wilcoxon-Mann-Whitney nonparametric test. The Wilcoxon-Mann-Whitney test takes ordinal measurement data and tests whether two independent groups have been drawn from the same population (Siegal and Castellan; 1988). The test accomplishes this through the use of rank orderings; therefore, the drawings were first ranked by two judges using the same criteria as the subjects had when drew them: focus on content and placement of objects rather than artistic quality. Table 150 shows the two judges ranks for the three quiz augmented scenes.

Table 150. The rank values for the three quiz augmented scenes by the two expert judges.

Subject	Houses 1	Houses 2	Lab 1	Lab 2	Brain 1	Brain 2
1	5	7	7	17	1	5
2	2	2	1	1	5	4
3	19	20	14	15	9	3
4	20	18	8	19	12	14
5	1	1	4	5	2	1
6	13	8	16	6	19	17
7	8	6	2	2	18	16
8	14	9	20	20	7	13
9	3	3	9	4	6	8
10	9	10	17	13	4	9
11	16	14	12	9	10	7
12	15	15	19	18	13	12
13	6	17	15	16	11	6
14	4	5	11	10	14	11
15	7	12	3	11	20	20
16	12	4	18	14	17	18
17	17	11	6	3	8	10
18	11	13	5	7	3	2
19	18	16	10	8	16	15
20	10	19	13	12	15	19

The rank scores from the two judges were then averaged and used as the scores in the Wilcoxon-

Mann-Whitney test. Table 151 shows the rank-sum statistics for the three quiz augmented scenes and resulting p-values. None of the drawing conditions reached significance, however, in two cases a trend towards significantly better drawings for the high integralness conditions was shown.

Table 151. Rank-sum statistics and p-values for the three quiz augmented scene drawings.

<b>Rank-Sum Statistics</b>	<b>P-Value</b>
Houses <sub>high</sub> = 1+2+3+5+6+8+10+12+19+20 = 86	0.0827
Houses <sub>low</sub> = 4+7+9+11+13+14+15+16+17+18 = 124	
Lab <sub>high</sub> = 1+2+3+6+11+12+14+15+16+20 = 100	0.3697
Lab <sub>low</sub> = 4+5+7+8+9+10+13+17+18+19 = 110	
Brain <sub>high</sub> = 1+3+4+5+6+7+11+14+16+19 = 86	0.0827
Brain <sub>low</sub> = 2+8+9+10+12+13+15+17+18+20 = 124	

### 7.3.4. Time and Subjective Measure ANOVAs: Assembly Tasks

Mixed-factor ANOVAs were performed for the independent variables integralness of the graphics with the overall scene (high, low), and task (1 through 8), as measured by time on task, and four subjective assessments of the augmented video instructions (integral, goodness, usefulness, overall effectiveness).

Table 152 shows the ANOVA summary table for the two independent variables as measured by time to complete task. There was one significant main effect and one significant interaction. Figure 136 shows the varying times it took to complete the eight assembly tasks. Figure 137 indicates how some of the tasks were affected by the high or low integralness conditions.

Table 152. ANOVA summary table for integralness and task, as measured by time to complete task.

Source	DF	SS	MS	F	P
<b>task</b>	<b>7</b>	<b>350960</b>	<b>50137</b>	<b>13.07</b>	<b>0.000</b>
integralness	1	5130	5130	0.55	0.468
subject(integralness)	18	167954	9331	2.43	
<b>task*integralness</b>	<b>7</b>	<b>96564</b>	<b>13795</b>	<b>3.60</b>	<b>0.001</b>
Error	126	483282	3836		
Total	159	1103890			

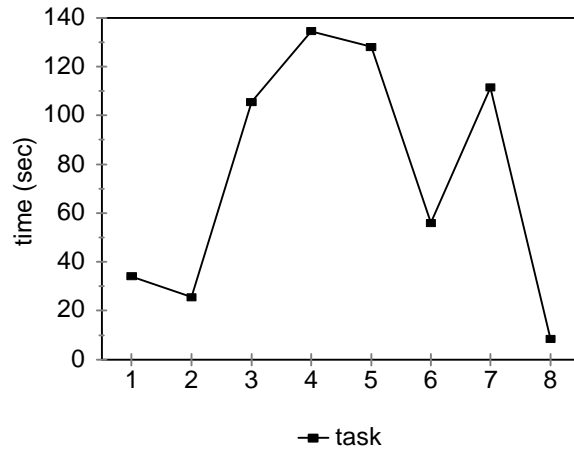


Figure 136. Graph of the main effect of task, significant at the  $p < 0.05$  level ( $p = 0.000$ ) as measured by time to complete task.

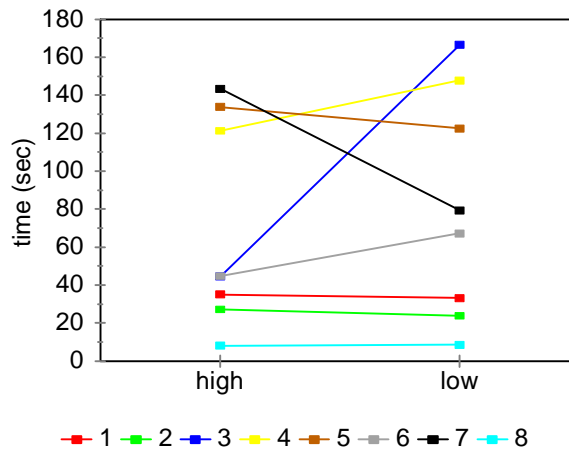


Figure 137. Graph of the interaction between task and integralness, significant at the  $p < 0.05$  level ( $p = 0.001$ ) as measured by time to complete task.

Table 153 shows the ANOVA summary table for the two independent variables as measured by the integral subjective assessments. There were two significant main effects. Figure 138 shows the varying integral assessments for the eight assembly tasks. Figure 139 displays lower integral ratings (indicating more integral perception) for the high integralness condition.

Table 153. ANOVA summary table for integralness and task, as measured by integral ratings of the graphics in the augmented instructions.

Source	DF	SS	MS	F	P
<b>task</b>	<b>7</b>	<b>73.375</b>	<b>10.482</b>	<b>8.64</b>	<b>0.000</b>
<b>integralness</b>	<b>1</b>	<b>57.600</b>	<b>57.600</b>	<b>5.34</b>	<b>0.033</b>
subject(integralness)	18	194.125	10.785	8.89	
task*integralness	7	17.000	2.429	2.00	0.060
Error	126	152.875	1.213		
Total	159	494.975			

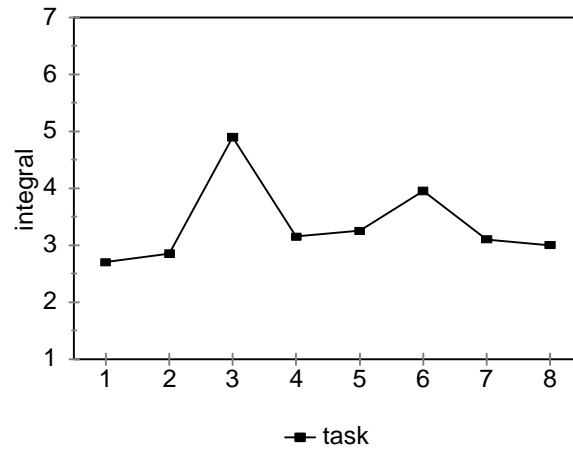


Figure 138. Graph of the main effect of task, significant at the  $p < 0.05$  level ( $p = 0.000$ ) as measured by ratings of the integral level of the graphics in the augmented instructions.

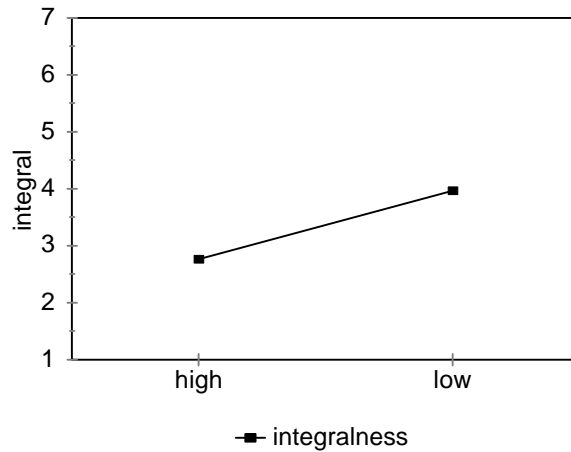


Figure 139. Graph of the main effect of integralness, significant at the  $p < 0.05$  level ( $p = 0.033$ ) as measured by ratings of the integral level of the graphics in the augmented instructions.

Table 154 shows the ANOVA summary table for the two independent variables as measured by the goodness subjective assessments. There was one significant main effect. Figure 140 shows the varying goodness assessments for the eight assembly tasks.

Table 154. ANOVA summary table for integralness and task, as measured by goodness ratings of the graphics in the augmented instructions.

Source	DF	SS	MS	F	P
<b>task</b>	<b>7</b>	<b>45.400</b>	<b>6.486</b>	<b>4.98</b>	<b>0.000</b>
integralness	1	22.500	22.500	2.87	0.108
subject(integralness)	18	141.150	7.842	6.02	
task*integralness	7	15.100	2.157	1.65	0.126
Error	126	164.250	1.304		
Total	159	388.400			

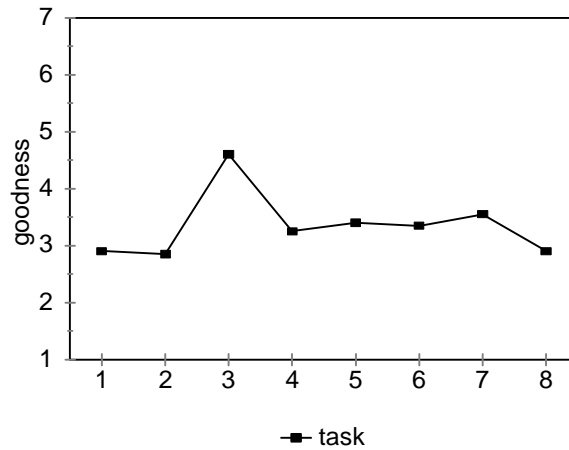


Figure 140. Graph of the main effect of task, significant at the  $p < 0.05$  level ( $p = 0.000$ ) as measured by ratings of the goodness of the graphics in the augmented instructions.

Table 155 shows the ANOVA summary table for the two independent variables as measured by the usefulness subjective assessments. There was one significant main effect and one significant interaction. Figure 141 shows the varying usefulness assessments for the eight assembly tasks. Figure 142 indicates how some of the tasks usefulness assessments were affected by the high or low integralness conditions.

Table 155. ANOVA summary table for integralness and task, as measured by usefulness ratings of the graphics in the augmented instructions.

Source	DF	SS	MS	F	P
<b>task</b>	<b>7</b>	<b>92.900</b>	<b>13.271</b>	<b>11.15</b>	<b>0.000</b>
integral	1	32.400	32.400	4.29	0.053
subject(integral)	18	136.000	7.556	6.35	
<b>task*integral</b>	<b>7</b>	<b>32.600</b>	<b>4.657</b>	<b>3.91</b>	<b>0.001</b>
Error	126	150.000	1.190		
Total	159	443.900			

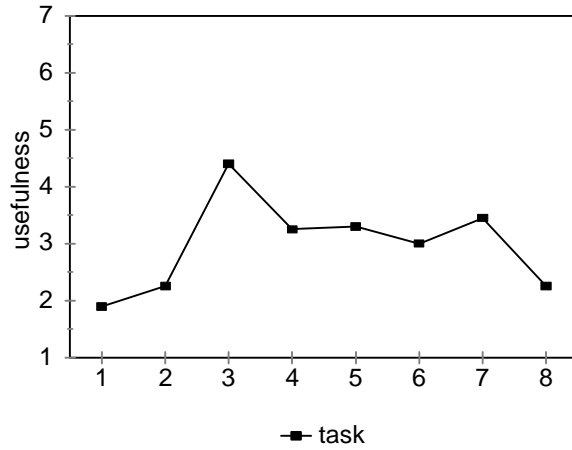


Figure 141. Graph of the main effect of task, significant at the  $p < 0.05$  level ( $p = 0.000$ ) as measured by ratings of the usefulness of the graphics in the augmented instructions.

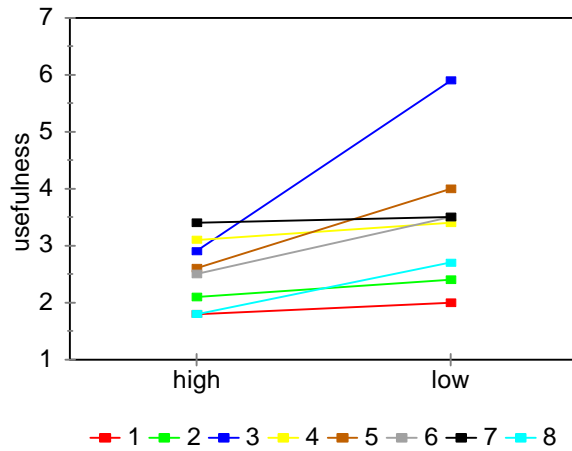


Figure 142. Graph of the interaction between task and integrality, significant at the  $p < 0.05$  level ( $p = 0.001$ ) as measured by ratings of the usefulness of the graphics in the augmented instructions.

Table 156 shows the ANOVA summary table for the two independent variables as measured by the overall effectiveness subjective assessments. There was one significant main effect and one significant interaction. Figure 143 shows the varying overall assessments for the eight assembly tasks. Figure 144 indicates how some of the tasks overall assessments were affected by the high or low integralness conditions.

Table 156. ANOVA summary table for integralness and task, as measured by the overall effectiveness ratings of the augmented instructions.

Source	DF	SS	MS	F	P
<b>task</b>	<b>7</b>	<b>107.675</b>	<b>15.382</b>	<b>9.85</b>	<b>0.000</b>
integralness	1	13.225	13.225	1.65	0.215
subject(integralness)	18	144.000	8.000	5.12	
<b>task*integralness</b>	<b>7</b>	<b>31.275</b>	<b>4.468</b>	<b>2.86</b>	<b>0.008</b>
Error	126	196.800	1.562		
Total	159	492.975			

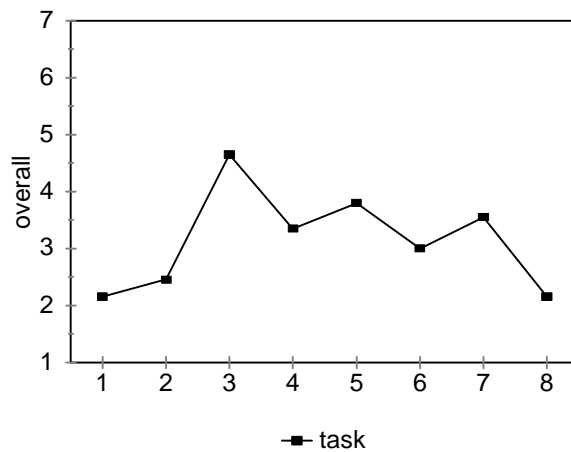


Figure 143. Graph of the main effect of task, significant at the  $p < 0.05$  level ( $p = 0.000$ ) as measured by ratings of the overall effectiveness of the augmented instructions.

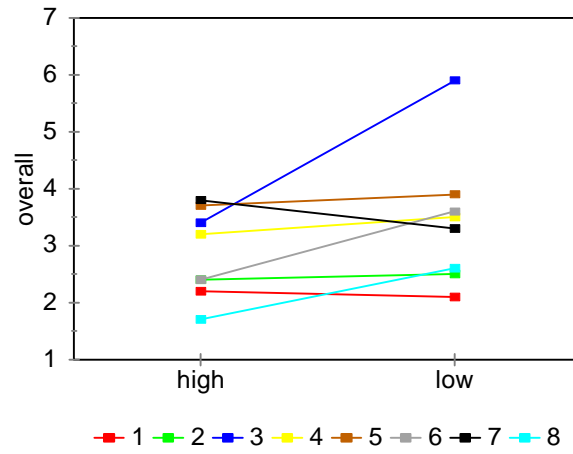


Figure 144. Graph of the interaction between task and integralness, significant at the  $p < 0.05$  level ( $p = 0.008$ ) as measured by ratings of the overall effectiveness of the augmented instructions.

### 7.3.5. Error and Subjective Measure Nonparametric Tests: Assembly Tasks

Nonparametric analyses were performed to confirm main effect significant results found in the subjective measure ANOVAs, and on the dependent measure errors in task. Tables 157 through 161 show the summary tables for the subjective measure Kruskal-Wallis tests.

Table 157. Kruskal-Wallis test for task as measured by ratings of the integral level of the graphics in the augmented instructions.

<u>task</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	20	3.000	64.1	-1.69
2	20	3.000	68.5	-1.24
3	20	5.500	115.9	3.65
4	20	3.000	76.5	-0.42
5	20	3.000	79.2	-0.14
6	20	4.000	97.5	1.75
7	20	3.000	73.3	-0.75
8	20	2.500	69.0	-1.18
Overall	160		80.5	

H = 20.07 DF = 7 **P = 0.005**

H = 20.73 DF = 7 **P = 0.004** (adjusted for ties)

Table 158. Kruskal-Wallis test for integralness as measured by ratings of the integral level of the graphics in the augmented instructions.

<u>integral</u>	<u>N</u>	<u>Median</u>	<u>Ave Rank</u>	<u>Z</u>
1	80	3.000	66.2	-3.91
2	80	4.000	94.8	3.91
Overall	160		80.5	

H = 15.25 DF = 1 **P = 0.000**

H = 15.75 DF = 1 **P = 0.000** (adjusted for ties)

Table 159. Kruskal-Wallis test for task as measured by ratings of the goodness of the graphics in the augmented instructions.

task	N	Median	Ave Rank	Z
1	20	3.000	69.4	-1.15
2	20	3.000	68.1	-1.27
3	20	5.000	109.3	2.98
4	20	3.000	79.5	-0.10
5	20	3.000	81.7	0.13
6	20	3.000	83.4	0.30
7	20	3.500	86.0	0.57
8	20	2.500	66.4	-1.45
Overall	160		80.5	

H = 12.56 DF = 7 P = 0.084

H = 13.12 DF = 7 P = 0.069 (adjusted for ties)

Table 160. Kruskal-Wallis test for task as measured by ratings of the usefulness of the graphics in the augmented instructions.

task	N	Median	Ave Rank	Z
1	20	2.000	49.4	-3.21
2	20	2.000	61.5	-1.96
3	20	4.000	111.1	3.16
4	20	3.000	90.3	1.01
5	20	3.000	93.5	1.34
6	20	3.000	83.4	0.30
7	20	3.000	95.8	1.57
8	20	2.000	59.1	-2.21
Overall	160		80.5	

H = 30.12 DF = 7 **P = 0.000**

H = 31.33 DF = 7 **P = 0.000** (adjusted for ties)

Table 161. Kruskal-Wallis test for task as measured by ratings of the overall effectiveness of the augmented instructions.

task	N	Median	Ave Rank	Z
1	20	2.000	55.0	-2.64
2	20	2.000	63.3	-1.78
3	20	5.000	114.1	3.47
4	20	3.500	88.4	0.82
5	20	4.000	99.2	1.93
6	20	3.000	79.1	-0.15
7	20	3.000	91.1	1.09
8	20	2.000	53.9	-2.75
Overall	160		80.5	

H = 30.90 DF = 7 **P = 0.000**

H = 31.95 DF = 7 **P = 0.000** (adjusted for ties)

For the categorical binary errors on task data, a chi-square test was performed. The summary table for the independent variable level of integralness is shown in Table 162. The test did not reach significance.

Table 162. Chi-square analysis of integralness as measure by errors in task.

	<u>Columns: integral</u>		
	<u>Rows: errors</u> high	low	All
no	71	62	133
	71	62	133
yes	9	18	27
	9	18	27
All	80	80	160
	80	80	160

Chi-Square = 3.609, DF = 1, P-Value = 0.057

The same test was attempted for the task independent variable and failed due to insufficient data points, specifically ‘hits’ in the yes error section across all the tasks. A Kruskal-Wallis was then performed on the integralness independent variable as measured by errors to see how close the approximation would be for a technically incorrect test for categorical data. As can be seen from Tables 162 and 163, the two resulting p-values differ only by 1/1000th. Based on this evidence, it was concluded that a Kruskal-Wallis test could serve as an approximate substitute test for the independent variable task as measured by errors. The summary table for the Kruskal-Wallis test for the independent variable task as measured by errors is shown in Table 164. This test reached significance, as expected from the previous varying dependent measures across the eight assembly tasks.

Table 163. Kruskal-Wallis test for integralness as measured by number of errors on task.

integral	N	Median	Ave Rank	Z
1	80	0.00E+00	76.0	-1.23
2	80	0.00E+00	85.0	1.23
Overall	160		80.5	

H = 1.51 DF = 1 P = 0.219  
H = 3.59 DF = 1 P = 0.058 (adjusted for ties)

Table 164. Kruskal-Wallis test for task as measured by errors in task.

task	N	Median	Ave Rank	Z
1	20	0.00E+00	67.0	-1.39
2	20	0.00E+00	71.0	-0.98
3	20	0.00E+00	83.0	0.26
4	20	5.00E-01	107.0	2.73
5	20	0.00E+00	75.0	-0.57
6	20	0.00E+00	71.0	-0.98
7	20	0.00E+00	103.0	2.32
8	20	0.00E+00	67.0	-1.39
Overall	160		80.5	

H = 16.68 DF = 7 P = 0.020  
H = 39.63 DF = 7 P = 0.000 (adjusted for ties)

### 7.3.6. Correlations

Correlations and corresponding significance p-values were calculated to find the degree of relationship between the tested variables from experiment three. Table 165 lists and describes the various variables that were included in the correlation computations. Table 166 shows the correlations and corresponding p-values for the tested variables. Significant correlations and their p-values are highlighted in **bold** in Table 166.

Table 165. A list and description of the variables from the correlation tests for experiment three.

Variable	Variable Description
fit	How many additional houses could fit on the back hill of the houses quiz augmented scene.
floors	How many floors did the closest house have in the houses quiz augmented scene.
graphic	How many graphic computers were in the computer lab quiz augmented scene.
int_q	How integral were graphics with the real scene rating of the quiz augmented scenes (1 very, 7 not at all).
use_q	Were the graphics useful to the overall scene rating of the quiz augmented scenes (1 very, 7 not at all).
time	Time in seconds to complete assembly tasks.
errors	Number of errors in the assembly tasks.
int_a	How integral were graphics with the real scene rating of the assembly augmented scenes (1 very, 7 not at all).
good_a	What is the goodness rating of the assembly augmented scenes (1 high, 7 low).
use_a	Were the graphics useful to the overall scene rating of the assembly augmented scenes (1 very, 7 not at all).
over_a	Were the assembly augmented instructions overall useful for completing the task rating of the assembly augmented scenes (1 very, 7 not at all).
age	Age of the subjects (years).
gender	Gender of subjects (1 male, 2 female).
comp	Self-rating for time spent working on computers each week (1 low, 6 high).
games	Self-rating of time spent playing computer games (1 low, 7 high).
mental	Self-assessment of the ability to mentally manipulate objects within the mind (1 high, 7 low).

Table 166. Correlations and correlation p-values for tested variables in experiment three.

	fit	floors	graphic	int_q	use_q	time	errors	int_a	good_a	use_a	over_a	age	gender	comp	games
<b>fit</b>															
	<b>0.517</b>														
	<b>0.019</b>														
<b>floors</b>															
<b>graphic</b>	-0.180	-0.416													
	0.447	0.068													
<b>int_q</b>	0.114	0.082	-0.180												
	0.632	0.732	0.448												
<b>use_q</b>	<b>0.616</b>	0.435	0.057	<b>0.463</b>											
	<b>0.004</b>	0.055	0.813	<b>0.040</b>											
<b>time</b>	0.090	0.387	-0.002	-0.297	0.090										
	0.705	0.092	0.992	0.203	0.705										
<b>errors</b>	0.278	0.042	0.060	0.154	<b>0.507</b>	0.391									
	0.235	0.861	0.801	0.518	<b>0.023</b>	0.088									
<b>int_a</b>	0.279	0.319	-0.096	<b>0.648</b>	<b>0.500</b>	0.075	0.207								
	0.233	0.171	0.689	<b>0.002</b>	<b>0.025</b>	0.754	0.381								
<b>good_a</b>	0.298	0.270	0.015	<b>0.474</b>	0.434	0.170	0.157	<b>0.859</b>							
	0.202	0.249	0.951	<b>0.035</b>	0.056	0.475	0.508	<b>0.000</b>							
<b>use_a</b>	0.319	0.205	0.062	0.384	<b>0.584</b>	0.103	0.333	<b>0.819</b>	<b>0.895</b>						
	0.170	0.385	0.796	0.094	<b>0.007</b>	0.666	0.151	<b>0.000</b>	<b>0.000</b>						
<b>over_a</b>	0.103	-0.016	-0.056	0.315	0.425	0.065	0.228	<b>0.617</b>	<b>0.700</b>	<b>0.820</b>					
	0.666	0.946	0.814	0.177	0.062	0.785	0.333	<b>0.004</b>	<b>0.001</b>	<b>0.000</b>					
<b>age</b>	0.387	0.161	0.141	0.003	0.329	0.082	0.260	0.075	0.145	0.180	0.246				
	0.092	0.497	0.555	0.990	0.156	0.730	0.268	0.754	0.543	0.447	0.296				
<b>gender</b>	<b>-0.632</b>	-0.247	-0.182	0.040	<b>-0.485</b>	-0.015	-0.266	-0.178	-0.314	-0.352	-0.167	-0.298			
	<b>0.003</b>	0.293	0.441	0.866	<b>0.030</b>	0.950	0.257	0.452	0.178	0.128	0.480	0.202			
<b>comp</b>	0.036	-0.151	0.191	-0.188	-0.283	0.288	-0.018	0.099	0.206	0.017	-0.184	-0.128	0.158		
	0.879	0.525	0.419	0.429	0.227	0.218	0.940	0.679	0.383	0.943	0.437	0.591	0.505		
<b>games</b>	-0.429	0.091	-0.156	-0.017	-0.186	0.317	0.056	-0.172	-0.274	-0.323	-0.230	-0.165	0.434	-0.165	
	0.059	0.703	0.510	0.943	0.432	0.173	0.816	0.468	0.243	0.165	0.328	0.488	0.056	0.486	
<b>mental</b>	-0.345	0.096	-0.348	-0.372	-0.429	0.272	-0.219	-0.340	-0.304	-0.338	-0.094	0.047	<b>0.629</b>	0.043	0.441
	0.136	0.686	0.132	0.107	0.059	0.245	0.353	0.143	0.192	0.145	0.695	0.843	<b>0.003</b>	0.858	0.051

## **7.4. Discussion**

### *7.4.1. Quiz Augmented Scenes*

One of the main goals of the third experiment was to establish whether integrally created graphics would play a role in applied learning and instructional tasks. The first phase of experiment three, an investigation on learning in augmented reality, began this examination.

In this first phase, subjects viewed the scenes, then were asked a variety of inquiries about various aspects of the scenes. The questions were detail neutral, meaning the questions could be answered equally by the integral and non-integral groups. There was no explicit information contained in the way the integral graphics were made that would specifically aid the answering of the questions. The drawing instructions speak to this motivation directly by stating that re-creating content and placement of objects is most important, rather than artistic presentation.

The first main experimental question asked whether integrally created graphics make a difference in the learning that occurred from the target augmented scenes. Some of the dependent measures showed this to be the case, most notably the usefulness of the graphics subjective rating. Figure 134 showed the significant effect measured by usefulness for the independent variable integralness. Compared to previous significant effects in the first two experiments, this result is quite strong with a large difference between the two means for high and low integralness.

The results showed that the integral graphics mattered most to the brain augmented reality condition. Figure 135 indicated that the low integral brain graphics were perceived much worse than all other conditions on the usefulness measure and the high integral brain graphics received the best usefulness ratings. In addition, Figure 132 showed the same result for the integral subjective rating with both brain conditions making up the extreme ends of the ratings spectrum. The most useful graphics overall were found in the computer lab, as shown in Figure 133.

There are two possible reasons why the computer lab received the best usefulness ratings overall. First, the low integral graphics created for the computer lab condition maintained a large amount

of shape information for relatively simply shaped items. The graphics still looked like computer terminals even though they were obviously low quality graphics. Second, the computer lab was readily recognized as being filled with uniform computers. The real computers were identical, and the graphics were all of the same shape. This consistency of shape likely aided the ready identification of the whole room as containing numerous computer terminals. In comparison, the houses scene, also a view of similar real and graphic objects, did not maintain either of these conditions. The graphics maintained general house form, although they were somewhat anonymous looking rectangular objects lacking the distinctive computer terminal monitor and keyboard profile. In addition, unlike the computer lab, the houses scene had more variability in graphics shape and could not take advantage of multiple identical real items for overcoming inadequacies in the graphic items.

Other than corroborating the augmented scene and integralness interaction, the integral subjective rating had no other significant information to add. Neither did any of the conditions as measured by the goodness ratings.

The nonparametric tests for the quiz questions showed a similar although slightly different pattern of results. The independent variable integralness showed significant differences for two houses condition quiz questions and one lab condition quiz question. Interestingly, the brain questions which were rated the worst in the subjective ratings of integralness and usefulness showed no significant differences in their quiz questions. While these results were significant, interpretation was much clearer when they are discussed later in the context of other results for experiment three.

The drawings for the quiz augmented scenes were more of a statistical exercise for the experimenter, although there were some expectations of significant differences. While the Wilcoxon-Mann-Whitney tests approached significance for both the houses and brain scenes, neither did actually reach significance. These nearly significant results do indicate that such an analysis could be useful for similar studies in the future.

There are several possible confounds for the quiz augmented scene results, some of which have

been discussed previously. As in the other two experiments, the exact augmented scenes used are certain to be impacted differently by the inclusion or exclusion of integrally created graphics. This is considered a strength rather than a weakness of the present study. Using multiple augmented scenes, as was mentioned earlier, was a specific choice to cover a broader spectrum of augmented reality applications rather than focusing on one in particular. Therefore the results should be more generalizable to wider audience of augmented reality users.

As with the augmented reality scenes, different quiz questions could have provided different results. Again, this is not considered a negative in the present study. Questions were asked that provided no inherent advantage to integral graphics, and in some conditions there were significant results. The intent was not to design the perfect questions, rather to demonstrate that in some conditions questions will be answered differently depending on whether graphics were presented integrally with the overall scene.

Lastly, as with the other two confounds discussed, the exact method used to develop integral and non-integral graphics could have impacted the results. The graphics were made following the same procedures as defined by the previous two experiments. Furthermore, all the graphics for each respective condition in experiment three were made following the exact same procedures as all the other graphics. This means that the low integral brain image was made the same way as the low integral houses scene, as was made the same way as the low integral assembly augmented instructions, etc. While differently designed graphics may provide other results, the graphics designed here, the same way for each subject, did demonstrate significant effects.

#### *7.4.2. Assembly Tasks*

The assembly tasks represented the most direct comparison between different conditions in the three experiments studied in this work. The intent was to see if the theory investigated in experiments one and two definitively applied to a real-world task. To ascertain this information, a straightforward experimental design tested high and low integral graphics across a range of assembly tasks.

The first result, a significant difference in time on task as shown in Figure 136, merely indicated that the set of tasks were of varying difficulty. A previous possible confound between order and task was discussed earlier in the experiment three methodology section. There is no detectable indication of an order effect in these results. Furthermore, from actual observation of the subjects, there is no indication that order played any definitive role.

From Figure 136 it can be seen that tasks 3, 4, 5, and 7 had larger task times than the other tasks. Task three was purposely centered on a more complicated and subtle area of the stereo assembly. Subjects in both the integral and non-integral conditions had difficulty completing the task correctly without some searching to conceptualize what exactly was needed to accomplish the task. Task four involved multiple steps and was no surprise to take longer than all the other tasks. Task five involved a particularly difficult task because of the limited resolution of the helmet. The plug that needed to be found on the back of the circuit card was very small and difficult, although possible, to see with the head-mounted display. Task five, like task three, usually required some time to look around and figure out exactly how the task was to be accomplished, although for different reasons. Task seven purposely contained mistakes in the augmented video instructions. These were included to see if subjects with more integral graphics would be more prone to follow mistaken instructions and take longer times to complete the assembly. This result did actually occur.

While integrality did not specifically show statistical significance, the interaction between task and integrality is more important to understanding what occurred. Figure 137 shows that two tasks had large differences between the high and low integral graphics conditions, task three and seven. Task three, the subtle task, showed the strongest result for a high integral condition obtaining faster assembly times. Task seven showed the expected result of the high integral participants taking longer to complete the task due to the more obvious errors in the instructions. Four of the tasks essentially show no difference between high and low integral conditions, tasks five, one, two and eight. Tasks one, two, and eight were the fastest conditions overall as shown in Figure 136 and did not appear to be affected by low integral graphics. It appears that the easier the task, the less likely a subject is affected by non-integrally created augmented scenes. Task five was difficult, but this task appeared overwhelmed by the technical limitation of the head-mounted

display to clearly show the backside of a circuit board. Low integralness subjects did appear to have more variability in the areas that they searched on the back of the board, but the time did not reflect this experimenter observation.

Tasks four and six appear to have slightly better times with the integral augmented scenes over the non-integral scenes. The practical significance can be found when comparing the actual average times between the two integralness conditions. The times for task four were 121.31 seconds for the high integralness condition and 147.77 for the low integralness condition, a difference of 26.46 seconds or nearly 18%. The times for task six were 44.69 seconds for the high integralness condition and 67.30 seconds for the low integralness condition, a difference of 22.61 seconds or almost 34%. These numbers would be quite 'significant' in a manufacturing plant, such as Boeing, concerned about productivity on their assembly line. So, while the integralness factor itself did not show statistical significance, the interaction between task and integralness revealed many practical implications for integrally created augmented environments.

The subjective assessments of the augmented instructions showed significant results for task in all four ANOVAs. The pattern of results is similar, as shown in Figures 138, 140, 141, and 143. Task three consistently receives the worst ratings across conditions, followed by task six. Task three involved the subtle yet complicated task, which was universally disliked by subjects. Task three was the first task of any difficulty in the set of eight tasks, which might have affected its ratings. This may be a partial order effect because the first two tasks were generally completed so quickly, that no matter which of the harder tasks came next, it might have been negatively rated. Task six had three distinct steps needed to accomplish the task, however, no obvious explanation exists for the universal slightly worse subjective ratings than the other tasks.

Integralness does appear as a significant effect in the integral subjective rating ANOVA. This result if absent would have been a significant finding, however, the failure of integralness to appear as a significant effect in more of the subjective ratings is a bit surprising. It nearly reaches significance in two of the other measures, but is perhaps best explained, as before, with the interaction results.

The interaction between task and integralness showed significance in the subjective ratings ANOVAs for usefulness of the graphics in the augmented instructions, and the overall effectiveness of the augmented instructions for accomplishing the task. The graphs for these interactions, Figures 142 and 144, appear to confirm the main effect of integralness. Task three clearly dominates the figures as it has much worse ratings in both graphs. Task six also is discernible as having worse ratings for both graphs in the low integralness condition. The rest of the tasks have little to no differences between the two levels of integralness.

The nonparametric testing involving the time and subjective assessments very nearly paralleled the parametric information. The one additional measure that was not covered in the parametric analysis was errors on task. Errors would have been more appropriately measured through videotaping the assembly sessions and cataloguing predefined errors as discovered from viewing the tapes. The assembly sessions in this experiment were not taped; however, notations were made as to whether the subject made errors or had problems completing the task. The error information failed to reach significance for integralness, similar to many of the previous tests in the third experiment.

Errors did show significance for the task variable, albeit in an approximate Kruskal-Wallis analysis rather than a chi-square test. Examining the raw data in conjunction with this Kruskal-Wallis test shows that task three (4 errors), task four (10 errors), and task seven (9 errors) had the most errors. From previous discussion, it is not surprising tasks three and seven showed up on the most errors list. Task three was the complicated yet subtle task. Task seven included the purposely incorrect plug instructions. Task four contained the most steps needed to complete the task, thus afforded the most opportunity to make errors. Since the tasks were not intended to be functionally equivalent, not much more credence can be given to the error information obtained from the statistical analysis.

#### *7.4.3. Correlations*

While none of the correlations analyzed were intended to lead to predictor information, there were several results that warrant discussion. As with the some sets of correlations, some were

expected, others brought up interesting implications, and others were not entirely explainable.

There were several correlations between the various subjective ratings of either the quiz augmented scenes or the assembly augmented instructions. The subjective ratings were purposely trying to examine a similar issue from several aspects, so it is of no surprise that there were many correlations among the subjective assessment questions. The errors measure also indicated a significant relationship with the subjective assessment of usefulness of the graphics in the augmented quiz scene. This is an incongruous relationship since the errors were collected for the assembly task and should not have been related to any subjective assessment variable from the first phase of experiment three.

It is of minor interest that the two houses quiz augmented scene significant results were significantly correlated. In the fit question, a larger rating would imply the houses were small, being able to fit on the hill. In the floors question, a larger rating would imply the houses were large, having more floors. Seemingly these are opposite lines of thought, yet the two variables are correlated.

It is interesting to see that gender was significantly correlated with self-assessment of mental manipulation ability, with females rating their ability worse than males. It is an often found relationship that females tend to perform worse than males on mental manipulation tests, although that was not the case in the first two studies where a mental rotation test was used to assess exactly that ability. Nor did this specific correlation appear in the other studies. There were more females in the 20 subjects of experiment three than either of the first two experiments, which may have been enough evidence for this relationship to show.

Two other gender relationships showed significance, one with the houses fitting on the back hill quiz question, and a second with one of the subjective assessment measures. Neither of these correlations are worthy of serious consideration as gender is only a two level variable, and neither of these relationships make any intuitive sense.

The last significant correlation provides some interesting analysis. The number of houses fitting

on the back hill quiz question correlated significantly with the usefulness rating of the quiz augmented scenes. This implies that the perception that the scenes are not as useful correlates with the perception that more houses can fit on the back hill. Considering this result with the other significant results involving the usefulness question in parametric and nonparametric analysis, this could indicate that less integral graphics could lead to an overestimation in tasks similar to the houses estimation. This highlights a potentially notable limitation of poorly integrated graphics in an augmented reality scene, although it is only one significant correlation.

#### *7.4.4. Design Guidelines*

As in the first two experiments, there was an abundance of useful information in the data analysis and discussion that can be summarized in design guidelines. Table 167 states design guidelines as they pertain to either human factors, perception, and graphic design in augmented reality that came from experiment three.

Table 167. Design guidelines from experiment three.

Guideline	Implication
1. Maintaining shape information is a key determinant of graphic intergralness, particularly for learning tasks involving multiple computer-generated graphics.	Human Factors, Graphic Design
2. Computer-generated objects that are intended to represent the same type of object should use as consistently identical subcomponents as feasibly possible. Cross-identification between the similar objects will increase integral perception in multiple object augmented environments.	Human Factors, Graphic Design
3. Medical imagery, nature imagery, or any other photorealistic imagery must be very highly realistic to achieve an integral environment appropriate for a learning environment.	Graphic Design
4. Simple assembly task performance is not enhanced by the use of integrally designed computer-generated graphics. For these tasks, use high integral graphics only as aesthetically desired.	Human Factors
5. Complex assembly task performance can be considerably enhanced by the use of integrally designed computer-generated graphics. For these tasks, use as highly integrally designed graphics as possible.	Human Factors
6. The capabilities of the display used in an augmented system (especially resolution) must be considered when designing integral graphics. Many display systems will not be capable of displaying graphics with the full subtleties necessary for high integral perception.	Human Factors, Graphic Design
7. Tasks involving estimation and assessment of various aspects of an augmented environment should involve as integrally designed graphics as possible. Furthermore, estimation and assessment in general may be a poor task for augmented environments.	Human Factors

The guidelines for experiment three are generally based on the different tasks that were conducted by participants. Different guidelines resulted from the learning and assembly tasks, as well as some overall recommendations. The idea that shape information is important to integral perception that was alluded to in the previous two experiments was very apparent in the task-oriented third experiment. The tasks also showed variation as related to complexity, with some tasks (estimation and assessment) possibly being identified as tasks not appropriate for augmented reality systems at all.

## CHAPTER 8. OVERALL DISCUSSION AND CONCLUSIONS

### *8.1. Nature of Integral Graphics in Augmented Environments*

One of the main goals throughout this research was to capture and describe the essence of integrally perceived graphics in augmented reality applications. This proved to be a much more subtle and intricate phenomenon than was originally anticipated. Perhaps it should have been expected considering the vast body of psychological research that discusses the concept in a number of different contexts: integral/separable research, selective attention, object and scene recognition, whole and part perception, contextual influence, parallel processing, and partial object identification. This research aimed to provide a significant contribution to this rich topic, from both a psychological theorist and augmented reality designer point of view.

In essence, the whole body of this dissertation describes the nature of integral perception in augmented reality. However, several key ideas and guidelines can be gleaned from the overall work. This phenomenon, the perception that computer-generated graphics and the real scene are one unified view, was originally thought to be mostly a function of realistic graphic design. While realism in graphics certainly plays a significant role, the context of the specific augmented reality scene takes primary importance. This was one of the fundamental revelations of this work, that the idea of integrality is important, but even more so is the exact situation that it is utilized in.

Beyond the expanded role of scene context, the greater lesson is the impact of the overall augmented system on integral perception. More than integrally designed graphics and scene context, the whole augmented setup must be considered to truly comprehend integral perception. The display system hardware, the graphics computer engine, the software algorithms used, and a myriad of other system variables can all impact integral perception in augmented reality. In the end, there is not one single or two main contributors to integral perception, but a holistic contribution of the entire system that plays the crucial role for achieving integral perception in augmented reality.

## ***8.2. Design Guidelines***

In practical terms relating to this dissertation, each of the many different augmented scenes tested in the three experiments seemed to have its own special elements that affected, or were affected by, integral perception in augmented reality. It is the specific situations and the understanding of the nature of integral perception in augmented reality that formed the basis of the design guidelines that is one of the key contributions of this work.

The design guidelines were based on the significant results of the three studies and the observations of the experimenter. Certainly more research needs to be done to verify the guidelines and generate more additions to the list. The specific guidelines are compiled from the three design tables of the three experiments.

Table 168. Overall design guidelines.

Guideline	Implication
1. For more simplistic augmented applications, greater attention must be paid to creating more integral graphics.	Graphic Design
2. Use integral graphics to avoid the perception that the graphics are distracting to the overall scene.	Perception
3. All things being equal, transparent graphics should be used for greater integral perception in augmented reality applications.	Perception, Graphic Design
4. Three-dimensional augmented scenes should be designed with transparent graphics to aid the illusion of depth.	Graphic design
5. Dynamic augmented reality applications need very integrally designed graphics to maintain acceptable levels of integral perception.	Perception, Graphic Design
6. To achieve higher levels of integral perception with outline graphics, they should be displayed clearly, although high quality graphics are not necessary.	Graphic Design
7. Use color textures for increased integral perception only if color adds meaningful information to the augmented scene.	Perception, Graphic Design
8. See-through display systems should be avoided when high integral perception is desired in augmented applications.	Human Factors
9. The quality of the display system used to view the augmented scenes plays a key role in the ultimate integral perception of augmented graphics.	Human Factors
10. For assessing integral perception, use subjective measures defined as integral, blending, or unified scene.	Human Factors
11. Use as few graphics as possible when designing integrally perceived augmented environments.	Human Factors, Graphic Design
12. For augmented applications that must use low-integral computer-generated graphics, maintain as much shape information as possible.	Graphic Design
13. When using multiple graphics in augmented applications, limit the total number of computer-generated graphics to six.	Human Factors, Graphic Design
14. For even minimal integral perception in nature scenes use extremely realistic computer-generated graphics, particularly for landscapes.	Graphic Design
15. For a comprehensive assessment of integral perception in multiple graphic augmented applications, ask direct subjective questions on both integral perception of the whole scene and attention paid to individual items.	Human Factors, Perception
16. Maintaining shape information is a key determinant of graphic intergralness, particularly for learning tasks involving multiple computer-generated graphics.	Human Factors, Graphic Design
17. Computer-generated objects that are intended to represent the same type of object should use as consistently identical subcomponents as feasibly possible. Cross-identification between the similar objects will increase integral perception in multiple object augmented environments.	Human Factors, Graphic Design
18. Medical imagery, nature imagery, or any other photorealistic imagery must be very highly realistic to achieve an integral environment appropriate for a learning environment.	Graphic Design
19. Simple assembly task performance is not enhanced by the use of integrally designed computer-generated graphics. For these tasks, use high integral graphics only as aesthetically desired.	Human Factors
20. Complex assembly task performance can be considerably enhanced by the use of integrally designed computer-generated graphics. For these tasks, use as highly integrally designed graphics as possible.	Human Factors
21. The capabilities of the display used in an augmented system (especially resolution) must be considered when designing integral graphics. Many display systems will not be capable of displaying graphics with the full subtleties necessary for high integral perception.	Human Factors, Graphic Design
22. Tasks involving estimation and assessment of various aspects of an augmented environment should involve as integrally designed graphics as possible. Furthermore, estimation and assessment in general may be a poor task for augmented environments.	Human Factors

The guidelines list comprises a collection of general and specific recommendations for creating integrally perceived augmented environments. Several general themes can be extracted from the group that have a global impact on designing integral perceived augmented reality systems.

The first theme to be discussed is the importance of shape information in computer-generated graphics for integral perception. Preserving the shape information of computer-generated graphics is important for any augmented environment, but especially for graphics that have a commonly recognizable form. In the studies described, the conditions that had multiple objects of real and computer-generated graphics provided participants with direct comparisons of the objects. These direct comparisons were used by participants to judge whether the additional graphics matched the expected shape and other visual characteristics of the real images. In the low integral conditions, the shape information was specifically cited by participants as differing from the real objects. In high integral conditions, shape information appeared to be a necessary precursor to achieve the intended perception. Shape information was a more obvious visual factor in the low integral conditions which caused a greater impact on poor integral perception.

Beyond direct comparisons, the type of object visualized was also important for the concept of preserving shape information. The stereotypical mental model for objects influences the perception that multiple similar graphic and real items are integrally perceived. Examining the four multiple object conditions, the planets were the easiest to represent by shape, despite their different sizes, because they were all circles. Both the real and graphic images matched the expected circle shape without any difficulty and were perceived as highly integral. The computer room generally matched expectations of items with a computer shape, although the direct comparison likely played a larger role in these identically sized objects. The houses conditions was perhaps the most interesting in that the participants seemed to have a stereotypical mental model of what a house should look like as opposed to comparing the various real and graphic houses within an augmented scene. The degraded computer-generated graphic images lost the edges that defined the stereotypical house, creating instead a general amorphous outline of a house-sized object. The liquid dynamic model had no preconceived expected shapes for participants. The dynamic nature of this condition also prevented any exact direct comparisons between objects.

The next overall theme is the necessity of highly realistic computer-generated graphics for naturalistic scenes. Generally, this theme is the idea that context plays a large role in the expected realism of computer-generated graphics. Specifically, this idea was highlighted in the scenes that involved natural scenes (the houses and trees) and to a lesser degree the human body (the brain). These visually unadulterated environments required much realistic computer-generated graphics to be perceived integrally. Scenes with hard edges that did not contain naturalistic elements (for example the computer outline, computers, the stereo components, even the liquid dynamic model) did not require as realistic graphics to be perceived as integral augmented environments.

The participants in the houses and trees conditions expressed that these scenes were not as integrally perceived as others. The lack of realism in the computer-generated graphics was cited as a main reason. To a lesser degree, the human brain condition received similar comments. However, participants usually expected the brain to look more like an artificial medical image rather than a real brain.

The third overall theme discusses the main weakness in augmented reality for producing integrally perceived scenes. While the technical aspects of the field have been rapidly developing, augmented reality system hardware still needs to advance several levels before truly indistinguishable integral scenes can be made. Computer artefacts such as processing speed and graphic accelerators have been improving steadily over the years and should soon reach acceptable levels for integrally perceived augmented environments. However, the performance of displays for augmented systems needs to dramatically improve before truly engaging integral environments can be created. Much experimentation with larger field-of-views, crisper resolution, higher contrast, and truer color displays needs to be completed.

The last main theme to be discussed is the implication of different types of tasks on integral perception in augmented environments. A variety of task environments were used in the studies described. The first two experiments were simple perceptual subjective task environments of varying complexity. Depending on the exact nature of the task environment and other factors, participants rated integral perception accordingly. The most interesting case of the first two experiments was the liquid dynamic model. Similar to the last issue of naturalistic scenes needing

more realistic graphics to be perceived with high integral perception, the liquid dynamic conditions showed that augmented environments including a dynamic component substantially increased the need for correspondingly suitable graphics for high integral perception. In dynamic situations, not only does the graphic realism need to be maintained, but the expected dynamic movements must be matched.

More than the first two experiments, the theme of task type was eminently showcased in the third experiment. The third experiment was divided into two parts, a learning phase and an assembly phase. First, the tasks in the learning phase were relatively unaffected by differently designed integral or non-integral graphics. There were specific conditions that were affected, but the vast majority of situations did not differ as much as expected between high and low integrally designed graphics. The assembly tasks on the other hand provided a wealth of differently complex task environments to examine.

The assembly tasks were purposely varied from one task to another to test a range of activities. The simple assembly tasks showed no benefit from increased integral perception. The complex tasks, however, were greatly affected by low or high integral computer-generated graphics. The impact of integral perception in the complex tasks was so fundamental that in some cases the tasks could not be completed by a few of the subjects due to frustration with trying to figure out the low integral instructions. The low integral augmented environments exacerbated the difficult circumstances in the task environments. Whereas the high integral environments greatly aided task performance.

The last interesting element of the assembly tasks was the condition that was purposely setup incorrectly to examine whether integral graphics could hinder performance. This very interesting result revealed the effects of integral perception in a novel way that was not apparent in the rest of the studies. The intuitiveness that integral perception helps in augmented tasks was amply demonstrated in this dissertation, but showing that integral perception works in reverse depending on the task exhibits the true need to attune to proper design of integral perception in augmented environments.

### **8.3. Future Research**

The present work represented an opening into a new area of research in augmented reality. As with much research, it introduced far more questions than it answered. The design guidelines are presented as a listing of concrete suggestions for designers, however, there are many more questions to be considered in future research projects on this topic. Selected relevant topics are briefly discussed here.

One of the main conclusions of the present research is the need to expand the investigation of integral graphics to as many different augmented reality applications as possible. The nature of integral graphics seems to be very dependent on the specific application use. Research on additional augmented reality applications, specifically applied examples, would be a very valuable extension of the present work.

Augmented reality technology is rapidly improving. While every effort was made to use high-end equipment in the present study, there were known limitations with better alternatives either already existing or which will shortly be available. Two of the more pertinent technologies is the use of head-tracking, and a higher field-of-view and resolution head-mounted display. Head-tracking has been successfully demonstrated in augmented reality applications, and intense effort was expended in an attempt to incorporate it into the present work. The challenge is not in achieving head-tracking, but in maintaining correct registration of the graphics to their real-world counterparts. Many new methods and algorithms are being presented in annual conferences and journals to achieve accurate registration. Hopefully soon one method will be straightforward enough to use in future integral graphics research.

The head-mounted display, and its technical parameters, surely affect the integral perception of graphics. Direct comparisons could be made between different display mediums, for example a simple monitor and head-mounted display could have easily been compared in the present work. An alternate approach would be to vary specific technical parameters within the same display device (e.g., field of view, stereopsis, resolution, etc.). There has been a variety of work in related research areas that have investigated display parameters, but none yet in augmented reality

or specifically integral perception research.

Graphic design is far more varied than the simple color, grayscale, solid white, and wireframe levels of the texture variable used in experiment one. The higher integral perception of transparent graphics over opaque graphics was a revelation. However, the design of graphics is a hugely complex issue with nearly infinite variations. Several subjects mentioned a barely visible shimmer that was present in all the graphics from an artefact of their creation with the video mixing equipment. Even the simple inclusion or exclusion of this shimmer would likely affect integral perception. There are many different aspects of graphic design and presentation that would be interesting to study in the context of integral perception in augmented reality.

In the dependent measures, overall, only a mid-range of integral perception ratings were given to the augmented scenes used in this dissertation. There appears to be plenty of room for improvement. However, the measurement methods themselves may have impacted how the graphics and augmented scenes were judged. Other subjective assessment methods, or tailored objective measures, may provide additional insight into integral perception in augmented environments.

#### ***8.4. Conclusions***

The stated goals for this dissertation were to bridge the theoretical ideas from integral related psychology research with the nascent efforts to study the human component in augmented reality systems. Specifically the perceptual phenomenon of integral perception was investigated in augmented reality. The three progressive experiments completed for this dissertation laid very substantial groundwork in an attempt to meet these goals.

This dissertation addressed the void of human-centered investigations in augmented reality by applying theories and principles from integral perception related research to augmented environments. The dissertation also expanded the research methodology of integral perception related research areas by demonstrating that augmented environments can serve as an excellent experimental testing resource. Furthermore, an extensive list was created of human factors,

perception, and graphic design guidelines for augmented applications based on the experimental results. Lastly, the sheer mass of data collected has left a large foundation in place for any number of different studies to tie into.

Augmented reality technology is intended to help the human user better understand the environment they are dealing with by superimposing meaningful graphics onto the real-world. This dissertation showed that in order to meet this objective, it is imperative that we as designers must consider integral perception in augmented reality.

## CHAPTER 9. REFERENCES

- Azuma, R. T. (1997). A survey of augmented reality. *Presence*, 6 (4), 355-385.
- Baird, K. M. (1999). *Evaluating the effectiveness of augmented reality and wearable computing for a manufacturing assembly task*. Unpublished thesis, Virginia Tech, Blacksburg, Virginia.
- Barfield, W., and Baird, K. (1998). Issues in the design and use of wearable computers, *Virtual Reality: Research, Development, and Applications*, (3), 157-166.
- Barfield, W., Rosenberg, C., and Lotens, W. A. (1995). Augmented-reality displays. In W. Barfield and T. Furness (Eds.), *Virtual Environments and Advanced Interface Design* (pp. 542-576). New York: Oxford University Press.
- Barfield, W., Sanford, J., and Foley, J. (1988). The mental rotation and perceived realism of computer-generated three-dimensional images. *International Journal of Man-Machine Studies*, 29, 669-684.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94 (2), 115-147.
- Caudell, T. P. (1994). Introduction to augmented and virtual reality. In *Proceedings of the SPIE Vol. 2351 - Telemanipulator and Telepresence Technologies* (pp. 272-281).
- Caudell, T. P., and Mizell, D. W. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing process. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences* (pp. 659-669).
- Crawford, H. J., and Christensen, L. (1995). *Developing research skills: A laboratory manual* (3rd ed.). Boston: Allyn & Bacon.
- Drascic, D., and Milgram, P. (1996). Perceptual issues in augmented reality. In *Proceedings of the SPIE Vol. 2653 - Stereoscopic Displays and Virtual Reality Systems III* (pp. 123-134).
- Ellis, S. R. (1998). Localization of virtual objects in the near visual field. *Human Factors*, 40 (3), 415-433.
- Ellis, S. R. (1994). What are virtual environments? *IEEE Computer Graphics & Applications*, January, 1994.
- Feiner, S., MacIntyre, B., Höllerer, T., and Webster, A. (1997). A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. In *Proceedings of the International Symposium on Wearable Computing*, Cambridge, MA, October 13-14, 1997 (pp. 74-81).
- Feiner, S., MacIntyre, B., and Sellgmann, D. (1993). Knowledge-based augmented reality. *Communications of the ACM*, 36 (7), 53-62.
- Foard, C. F. and Nelson, D. G. K. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, 113 (1), 94-111.
- Furness, T. A., and Barfield, W. (1995). Introduction to virtual environments and advanced interface design. In T.A. Furness and W. Barfield (Eds.) *Virtual Environments and Advanced Interface Design* (pp. 3-13). New York: Oxford University Press.
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, 8, 98-123.

- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- Garner, W. R., and Felfody, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225-241.
- Gescheider, G. A. (1985). *Psychophysics: Method, theory, and application*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Goldstein, I. L. (1993). *Training in organizations*. Pacific Grove, California: Brooks/Cole Publishing Company.
- Gottwald, R. L. and Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception and Psychophysics*, 18 (1), 26-28.
- Grau, J. W. and Nelson, D. G. K. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, 117 (4), 347-370.
- Greenberg, D. P. (1999). A framework for realistic image synthesis. *Communications of the ACM*, 42 (8), 44-53.
- Handel, S. and Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception and Psychophysics*, 12 (1B), 108-116.
- Helft, M. (1997). Augmented reality scientists want respect. *Wired News*. Available on-line at <http://www.wired.com/news/technology/story/4179.html>
- Hoffmann, J. (1996). Visual object recognition. In W. Prinz and B. Bridgeman (Eds.), *Handbook of Perception and Action, Volume One, Perception* (pp. 297-344). San Diego, CA: Academic Press, Inc.
- Jakob, R. J. K., and Sibert, L. E. (1992). The perceptual structure of multidimensional input device selection. In *Proceedings of SIGCHI 1992* (pp. 211-218).
- Kakez, S., Figue, J., and Conan, V. (1998). A generic model for immersive documentation environments. *Virtual Reality*, 3, 176-186.
- Kancherla, A. R., Rolland, J. P., Wright, D. L., and Burdea, G. (1995). A novel virtual reality tool for teaching dynamic 3D anatomy. *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine 1995, Nice* (pp. 163-169).
- Kaplan, K. (1997). 3-d technology blends fact and fantasy. In *Los Angeles Times, Monday, March 3, 1997*. Accessible on-line at <http://tracker.usc.edu/~ykcho/star-LATimes030397.html>
- Kehrer, L. and Meinecke, C. (1996). Perceptual organization of visual patterns: The segmentation of textures. In W. Prinz and B. Bridgeman (Eds.), *Handbook of Perception and Action, Volume One, Perception* (pp. 25-70). San Diego, CA: Academic Press, Inc.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Lion, D., Rosenberg, C., and Barfield, W. (1993). Overlaying three-dimensional computer graphics with stereoscopic live motion video: Applications for virtual environments. *Society of Information Display International Symposium*, May 18-29, Seattle, WA (pp. 483-486).
- McCann, R. S. and Foyle, D. C. (1994). Superimposed symbology: Attentional problems and design solutions. *SAE Transactions: Journal of Aerospace*, 103, 2009-2016.
- McGee, M. K. (1998). *Assessing negative side effects in virtual environments*. Unpublished thesis, Virginia Tech,

Blacksburg, Virginia.

- McKenna, M. and Zeltzer, D. (1992). Three dimensional visual display systems for virtual environments. *Presence*, 1 (4), 421-458.
- Meister, D. (1985). *Behavioral analysis & measurement methods*. New York, NY: John Wiley & Sons.
- Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *SPIE Vol. 2351 Telemanipulator and Telepresence Technologies (1994)*.
- Nash, J. (1997). Wiring the jet set. *Wired Digital Inc*. Accessible on-line at <http://www.wired.com/wired/5.10/wiring.html>
- Neumann, O. (1996). Theories of attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of Perception and Action, Volume Three, Attention* (pp. 389-446). San Diego, CA: Academic Press, Inc.
- Neumann, U., and Majoros, A. (1998). Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *Proceedings for the Virtual Reality Annual Symposium, 1998* (pp. 4-11).
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13 (1), 87-108.
- Prinz, W. and Bridgeman, B. (1996). Introduction. In W. Prinz and B. Bridgeman (Eds.), *Handbook of Perception and Action, Volume One, Perception* (pp. 1-8). San Diego, CA: Academic Press, Inc.
- Rekimoto, J. (1997). NaviCam: A magnifying glass approach to augmented reality. *Presence*, 6 (4), 399-412.
- Rosch, E. and Lloyd, B. B. (1978). Process. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 73-77). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Sanders, M. S., and McCormick, E. J. (1987). *Human factors in engineering and design*. New York, NY: McGraw-Hil Publishing Company.
- Sharma, R., and Molineros, J. (1997). Computer vision-based augmented reality for guiding manual assembly. *Presence*, 6 (3), 292-317.
- Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In G. R. Lockhead and J. R. Pomerantz (Eds.), *The perception of structure* (pp. 53-71). Washington, DC: American Psychological Association.
- Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, 115 (1), 58-61.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54-87.
- Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Siegel, S. and Castellan Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill, Inc.
- Smith, E.P. (1998). *Class notes for applied multivariate methods: Spring 1998*. Virginia Tech University.

- Snow, M. P. (1996). *Charting presence in virtual environments and its effects on performance*. Unpublished dissertation, Virginia Tech, Blacksburg, Virginia.
- Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R. W., and Pentland, A. (1997). Augmented reality through wearable computing. *Presence*, 6 (4), 386-398.
- Stokes, A., Wickens, C., and Kite, K. (1990). *Display technology: Human factors concepts*. Warrendale, PA: Society of Automotive Engineers, Inc.
- Sutherland, I. E. (1965). The ultimate display. *International Federation of Information Processing*, 2, 506-508.
- Swezey, R. W., and Llaneras, R. E. (1997). Models in training and instruction. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 514-577). New York: Wiley.
- Vallino, J. (1997). Introduction to augmented reality. Available on-line at <http://www.cs.rochester.edu:80/u/vallino/research/AR/introduction.html>
- van der Heijden, A. H. C. (1996). Visual attention. In O. Neumann and A. F. Sanders (Eds.), *Handbook of Perception and Action, Volume Three, Attention* (pp. 5-42). San Diego, CA: Academic Press, Inc.
- Vandenberg, S. and Kuse, A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.
- Webster, A., Feiner, S., MacIntyre, B., Massie, W., And Krueger, T. (accessed March, 1999). Augmented reality in architectural construction, inspection, and renovation. Available on-line at <http://www.cc.columbia.edu/cu/gsap/BT/RESEARCH/PAPERS/ar-asce.htm>
- Weintraub, D. J., and Ensing, M. (1992). *Human factors issues in head-up display design: The book of HUD*. University of Michigan: Crew System Ergonomics Information Analysis Center.
- Wickens, C. D. (1992). *Engineering psychology and human performance*. New York, NY: HarperCollins Publishers, Inc.
- Wickens, C. D., Baker, P. (1995). Cognitive issues in virtual reality. In W. Barfield and T. Furness (Eds.), *Virtual Environments and Advanced Interface Design* (pp. 514-541). New York: Oxford University Press.

## Appendix A: Experiment One Trial Questionnaire

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Indicate appropriate response for the scene just viewed

1. What is the goodness of the scene just viewed?

Excellent   1   2   3   4   5   6   7   Terrible

2. How separable were the graphics from the real scene?

Very separable   1   2   3   4   5   6   7   Not at all separable

3. How integral were the graphics with the real scene?

Very integral   1   2   3   4   5   6   7   Not at all integral

4. Were the graphics distracting to the overall scene?

Very much so   1   2   3   4   5   6   7   Not at all

5. How distinguishable were the graphics from the real scene?

Very distinguishable   1   2   3   4   5   6   7   Not at all distinguishable

6. How well did the graphics and real scene blend into one overall view?

Very well blended   1   2   3   4   5   6   7   Not at all blended

7. Were the graphics useful and appropriate to the overall visual scene?

Very appropriate   1   2   3   4   5   6   7   Not at all appropriate

8. Did you perceive the whole scene as one unified view or distinct graphic and real components?

Unified   1   2   3   4   5   6   7   Distinct components

## Appendix B: Universal General Questionnaire

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Indicate appropriate response

1. Age: \_\_\_\_\_

2. Gender

Male Female

3. Handedness

Right Left Ambidextrous

4. On average how much time per week do you use a computer?

Less than 4 hrs. 4 to 10 hrs. 10 to 20 hrs. 20 to 30 hrs. 30 to 40 hrs. More than 40 hrs.

5. How often do you play computer games?

Frequently 1 2 3 4 5 6 7 Never

6. How would rate your ability to mentally visualize and manipulate objects in your head?

Excellent 1 2 3 4 5 6 7 Terrible

## **Appendix C: Universal Informed Consent Form**

### VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

#### Informed Consent for Participants of Investigative Projects

Title of Project: An Investigation into the Separable/Integral Dimensions of Augmented Reality

Investigator: Mike McGee

#### I. The Purpose of this Research

The purpose of this research is to examine the various parameters of graphics as they relate to the perception of augmented environments as an integral scene and performance in realistic tasks.

#### II. Procedures

You will be participating in a simulated augmented reality study. Augmented reality is the superimposing of computer-generated graphics onto a real-world scene typically presented in a head-mounted display. In this experiment, you will be viewing a series of simulated augmented reality scenes in a head-mounted display. After instruction and demonstration, you will be presented different simulated scenes and then making a set of judgements about each scene concerning how discernable, distracting, integrated, etc., the graphics are with the real scene. You will view each scene for approximately 20 seconds. Depending on your the experimental setup, you may be doing real tasks based on the simulated augmented scenes. Prior to viewing the augmented scenes, you will fill out a small questionnaire on general information, take a short mental rotation test, and have both your acuity and color vision tested. The experiment will take no more than 1 hour and 30 minutes. The testing area will be the Virtual Environments Laboratory, directed by Woodrow Barfield, in room 526 on the fifth floor of Whittemore Hall.

#### III. Risks

There are minimal risks to you in this study. There is no inherent danger, unpleasant experiences, or emotional distress in any of the experimental testing involving the simulated augmented environments.

#### IV. Benefits of this Project

This research will assist ongoing efforts to determine if augmented environments are perceived as one integrated visual scene, or separately. Understanding this phenomenon will be useful both to future researchers and designers who wish to utilize systems with high levels of perceived integrality. For example, medical training using an augmented reality system would take advantage of computer-generated graphics strongly perceived as being part of a real person. While this research should yield benefits to the above stated goals, no promise or guarantee is offered. Participation in this project should not depend on a guarantee of benefits.

V. Extent of Anonymity and Confidentiality

You will be identified only by a subject number in data analysis. No written results of this study will be traceable to any participant by name.

VI. Compensation

You will not be compensated for your voluntary participation in this study.

VII. Freedom to Withdraw

You are free to withdraw from this study at any time without penalty. You are free not to answer any questions or respond to any experimental situations that you choose without penalty. There may be circumstances under which the investigator may determine that you should not continue the experiment.

VIII. Approval of Research

This research project has been approved, as required, by the Institutional Review Board for Research Involving Human Subjects at Virginia Polytechnic Institute and State University.

IX. Subject's Responsibilities

I voluntarily agree to participate in this study. I agree to undergo the procedures of this experiment as described above.

X. Subject's Permission

I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project.

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

Should I have any questions about this research or its conduct I may contact:

Mike McGee  
Investigator  
ISE Department  
540 552-8546

Woodrow Barfield  
Faculty Advisor  
ISE Department  
540 231-2547

H.T. Hurd  
Chair, IRB  
Research Division  
540 231-5281

## Appendix D: Experiment Two Trial Questionnaire

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Indicate appropriate response for the scene just viewed

1. What was the goodness of the scene just viewed?

Excellent    1    2    3    4    5    6    7    Terrible

2. Were the graphics integral to the visual scene or distracting?

Very integral    1    2    3    4    5    6    7    Very distracting

3. Did you have a tendency to visually attend to the graphics or the real scene?

Completely on graphics    1    2    3    4    5    6    7    Completely on real scene

4. How cluttered do the graphics make the overall scene?

Very cluttered    1    2    3    4    5    6    7    Not at all cluttered

5. Did you perceive the scene as separate parts (graphics and real) or unified?

Totally separable    1    2    3    4    5    6    7    Totally unified

6. How similar were the graphic objects to the real objects in the visual scene?

Completely similar    1    2    3    4    5    6    7    Not at all similar

7. How much was your recognition of the whole visual scene affected by the graphic objects?

Very affected    1    2    3    4    5    6    7    Not at all affected

8. Were the graphics useful and appropriate to the overall visual scene?

Very appropriate    1    2    3    4    5    6    7    Not at all appropriate

## Appendix E: Experiment Three Quiz Questionnaires

Experiment Three Trial Questionnaire: Houses

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Circle best response to the questions below.

1. How many additional houses could have fit on the back hill in front of the forest?

1      2      3      4      5      6      7      8

2. How many floors would you estimate the closest house had?

1      2      3      4      5      6      7      8

3. What would you say the average price was of the houses in the scene (in 1000's of \$, assume Blacksburg area)?

50      75      100      125      150      200      250      300      400      500

4. How integral were the graphics with the real scene in the image just viewed?

Very integral    1    2    3    4    5    6    7    Not at all integral

5. What was the goodness of the scene?

Excellent    1    2    3    4    5    6    7    Terrible

6. Were the graphics useful and appropriate to the overall visual scene?

Very appropriate    1    2    3    4    5    6    7    Not at all appropriate

7. Draw the image just presented to the best of your recollection on the next page (content and placement of objects, not artistic quality, is the main goal). Indicate on the drawing which house you believed to be the closest with the letter 'A.'

Experiment Three Trial Questionnaire: Lab

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Circle best response to the questions below.

1. How many graphic computers were in the scene?

1      2      3      4      5      6      7      8      9      10

2. How many total computers were in the scene?

4      6      8      10      12      14      16      18      20      22

3. What was the diagonal size of the monitors (in.)?

11      12      13      14      15      17      19      20      21      23

4. There were four rows of computers. Which row did not have graphic computers?

Back row      Second to last row      Second closest row      Front row

5. How integral were the graphics with the real scene in the image just viewed?

Very integral    1    2    3    4    5    6    7    Not at all integral

6. What was the goodness of the scene?

Excellent    1    2    3    4    5    6    7    Terrible

7. Were the graphics useful and appropriate to the overall visual scene?

Very appropriate    1    2    3    4    5    6    7    Not at all appropriate

8. Draw the image just presented to the best of your recollection on the next page (content and placement of objects, not artistic quality, is the main goal).

Experiment Three Trial Questionnaire: Brain

Subject number: \_\_\_\_\_

Date: \_\_\_\_\_

Instructions: Circle best response to the questions below.

1. Roughly how many major structures were showing in the brain?

1      2      3      4      5      6      7      8      9      10

2. Brains are composed of layers. How many layers would you say were presented in the image just viewed?

1      2      3      4      5      6      7      8      9      10

3. Were eyes represented in the brain image?

yes      no

4. Brains typically have a small indent near the far back of the skull. Did the presented image have such an indent?

yes      no

5. How integral were the graphics with the real scene in the image just viewed?

Very integral    1    2    3    4    5    6    7    Not at all integral

6. What was the goodness of the scene?

Excellent    1    2    3    4    5    6    7    Terrible

7. Were the graphics useful and appropriate to the overall visual scene?

Very appropriate    1    2    3    4    5    6    7    Not at all appropriate

8. Draw the image just presented to the best of your recollection on the next page (content and placement of objects, not artistic quality, is the main goal).

## Appendix F: Experiment Three Assembly Questionnaire

### Experiment Three Trial Questionnaire: Assemblies

Instructions: Circle best response to the questions below.

1. In the augmented instructional video clips, how integral were the graphics with the real scene?

Very integral   1   2   3   4   5   6   7   Not at all integral

2. What was the goodness of the video clips?

Excellent   1   2   3   4   5   6   7   Terrible

3. Were the graphics useful and appropriate to the overall visual scene of the video clips?

Very appropriate   1   2   3   4   5   6   7   Not at all appropriate

4. Overall, rate the usefulness of the instructional video clips for accomplishing the tasks.

Excellent   1   2   3   4   5   6   7   Terrible

## **Vita**

Michael K. McGee has been a human factors engineer at IBM and Computing Devices Canada, Communications Systems Division. He holds a M.S. in Industrial and Systems Engineering from Virginia Tech, and a B.S. in Psychology from the University of Illinois at Urbana/Champaign. His primary interests are in designing and evaluating interfaces for advanced information display systems.