

From the laboratory to the real world:

Evaluating the impact of impostors, expertise
and individual differences on human
face matching performance

Dragana Calic

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

School of Psychology
The University of Adelaide

11 January 2013

Contents

<i>Abstract</i>	<i>x</i>
<i>Declaration</i>	<i>xiii</i>
<i>Acknowledgments</i>	<i>xiv</i>
<i>Chapter 1 Understanding Applied One-to-One Face Matching Performance</i>	<i>16</i>
1.1 <i>Face Recognition: An Overview</i>	<i>22</i>
1.2 <i>Face Matching: Definition and Application</i>	<i>25</i>
1.2.1 <i>Surveillance: Identification or One-to-Many Face Matching</i>	<i>27</i>
1.2.2 <i>Access Control: Verification or One-to-One Face Matching</i>	<i>29</i>
1.3 <i>Current Research Aims: Factors Considered</i>	<i>36</i>
<i>Chapter 2 The Human Operator</i>	<i>42</i>
2.1 <i>Human Operator Performance</i>	<i>42</i>
2.1.1 <i>Face Processing</i>	<i>44</i>
2.2 <i>Factors that Impact Human Operator Face Matching Performance</i>	<i>49</i>
2.2.1 <i>Image and Environment Specific Factors</i>	<i>50</i>
2.2.2 <i>Individual Specific Factors</i>	<i>55</i>
2.2.3 <i>Human Operator Specific Factors</i>	<i>61</i>
2.2.4 <i>The Impact of Impostors</i>	<i>70</i>
2.2.5 <i>Experimental versus Live Evaluations</i>	<i>72</i>

Chapter 3	<i>Image Preparation: Acquisition, Sourcing, Look-Alike Definition and Impostor Generation</i>	77
3.1	<i>Impostor Image Sourcing</i>	79
3.2	<i>Acquisition of Target Participant (TP) Imagery</i>	81
3.2.1	Target Participant Recruitment and Participation	81
3.2.2	Experimental Conduct	81
3.3	<i>Look-Alike Selection and Impostor Generation</i>	87
3.4	<i>Editing and Normalisation of Imagery</i>	90
3.4.1	Still Imagery	90
3.4.2	Video Imagery	93
3.5	<i>Assessment of Still Imagery</i>	93
Chapter 4	<i>Experiment 1: Scoping Laboratory One-to-One Face Matching</i>	98
4.1	<i>Method</i>	100
4.1.1	Participants	100
4.1.2	Materials	101
4.1.3	Design and Procedure	102
4.2	<i>Analysis</i>	105
4.3	<i>Results</i>	108
4.3.1	The Impact of Impostor Frequency	108
4.3.2	The Impact of Impostor Types	111
4.3.3	Confidence Ratings and Decision Latency	114
4.4	<i>Discussion</i>	117
Chapter 5	<i>Experiment 2: Live One-to-One Face Matching</i>	120
5.1	<i>Method</i>	123
5.1.1	Participants	123
5.1.2	Materials	126
5.1.3	Design and Procedure	129
5.2	<i>Results</i>	137
5.2.1	Overall Performance	137

5.2.2	The Impact of Human Operator Expertise.....	138
5.2.3	The Impact of Impostor Types	142
5.2.4	Individual Differences	144
5.2.5	Confidence Ratings	148
5.2.6	The Post-Evaluation Survey	150
5.3	<i>Discussion</i>	154
Chapter 6 Experiment 3: Laboratory One-to-One Face Matching		157
6.1	<i>Human-Algorithm Performance Comparisons</i>	158
6.2	<i>Method</i>	166
6.2.1	Participants	166
6.2.2	Materials.....	167
6.2.3	Design and Procedure.....	169
6.3	<i>Results</i>	170
6.3.1	Human Operator Performance.....	171
6.3.2	Human Operator and Automated System Performance.....	179
6.4	<i>Summary of Experiment 3</i>	181
Chapter 7 Discussion and Conclusion.....		183
7.1	<i>Discussion of the Key Findings</i>	185
7.1.1	Effects of Experimental Factors and Human-Algorithm Performance Comparison.....	193
7.1.2	Comparison of Live and Laboratory One-to-One Face Matching Findings	214
7.2	<i>Concluding Remarks</i>	219
References		223
Appendices		237

List of Figures

<i>Figure 1:</i>	<i>A generic surveillance application, adapted from McLindin (2005).....</i>	<i>28</i>
<i>Figure 2:</i>	<i>A generic access control application, adapted from McLindin (2005).....</i>	<i>30</i>
<i>Figure 3:</i>	<i>The Thatcher illusion, adapted from Thompson (1980).....</i>	<i>47</i>
<i>Figure 4:</i>	<i>Five components of Image Preparation</i>	<i>78</i>
<i>Figure 5:</i>	<i>Image Preparation: Still imagery acquisition equipment set up, acquisition process and an example of a still image</i>	<i>84</i>
<i>Figure 6:</i>	<i>Image Preparation: Video imagery acquisition equipment set up and acquisition process</i>	<i>86</i>
<i>Figure 7:</i>	<i>Image Preparation: Exemplifying impostor types and how external images were modified to appear similar to internal imagery.....</i>	<i>92</i>
<i>Figure 8:</i>	<i>Image Preparation: Assessment of still imagery stimuli presentation</i>	<i>95</i>
<i>Figure 9:</i>	<i>Experiment 1: Stimuli presentation</i>	<i>104</i>
<i>Figure 10:</i>	<i>Experiment 1: Decision confidence rating scale</i>	<i>105</i>
<i>Figure 11:</i>	<i>The ROC curve for different values of d'</i>	<i>106</i>
<i>Figure 12:</i>	<i>Experiment 1: ROCs by impostor frequency conditions.....</i>	<i>111</i>

Figure 13:	<i>Experiment 2: Untrained and trained/experienced Human Operators</i>	124
Figure 14:	<i>Experiment 2: An example of Target Participant ID card</i>	126
Figure 15:	<i>Experiment 2: Testing environment showing cubicles in which Human Operators were seated and the arrows representing TP movement.....</i>	130
Figure 16:	<i>Experiment 2: Target Participant imagery presented to Human Operators.....</i>	131
Figure 17:	<i>Experiment 2: An example of Target Participant video image acquisition</i>	134
Figure 18:	<i>Experiment 2: An example of Target Participant and Human Operator interaction</i>	136
Figure 19:	<i>Experiment 2: ROCs for trained and untrained Human Operator performance.....</i>	140
Figure 20:	<i>Human and algorithm performance on easy and difficult face pairs, adapted from O’Toole, Phillips, et al., (2007)</i>	161
Figure 21:	<i>Human (black) and algorithm performance on very high and high resolution imagery, adapted from O’Toole, et al., (2008).....</i>	163
Figure 22:	<i>Experiment 3: Trained/experienced Human Operators</i>	166
Figure 23:	<i>Experiment 3: Target Participant imagery presented to Human Operators.....</i>	170
Figure 24:	<i>Experiment 3: ROC of trained Human Operator performance.....</i>	172
Figure 25:	<i>Experiment 3: ROCs of Human Operator and algorithm performance ...</i>	180
Figure 26:	<i>A summary of untrained and trained Human Operator one-to-one face matching performance across all experiments.....</i>	200

Figure 27: Live Experiment 2: Hit Rates of participating agencies' and untrained Human Operators..... 202

Figure 28: ROCs comparing live Experiment 2 and laboratory Experiment 3 findings for trained Human Operators..... 215

List of Tables

<i>Table 1:</i>	<i>Decision outcomes.....</i>	<i>32</i>
<i>Table 2:</i>	<i>External image databases used in this research</i>	<i>80</i>
<i>Table 3:</i>	<i>Impostor frequency conditions and the number of images presented as part of each condition.....</i>	<i>103</i>
<i>Table 4:</i>	<i>Response decision matrix</i>	<i>108</i>
<i>Table 5:</i>	<i>Experiment 1: Overall accuracies and SDT measures by impostor frequency</i>	<i>109</i>
<i>Table 6:</i>	<i>Experiment 1: Performance rates by impostor frequency.....</i>	<i>110</i>
<i>Table 7:</i>	<i>Experiment 1: Hit Rates for impostor types by impostor frequency.....</i>	<i>112</i>
<i>Table 8:</i>	<i>Experiment 1: Hit Rates by impostor type (frequency combined).....</i>	<i>114</i>
<i>Table 9:</i>	<i>Experiment 1: Confidence ratings overall and by impostor type.....</i>	<i>115</i>
<i>Table 10:</i>	<i>Experiment 1: Decision latency overall and by impostor type.....</i>	<i>116</i>
<i>Table 11:</i>	<i>Experiment 2: Overall performance rates.....</i>	<i>138</i>
<i>Table 12:</i>	<i>Experiment 2: Overall accuracy, performance rates and SDT indices for trained and untrained Human Operators</i>	<i>139</i>

Table 13:	<i>Experiment 2: Overall accuracies and performance rates by agency.....</i>	141
Table 14:	<i>Experiment 2: Human Operator Hit Rates by impostor type</i>	143
Table 15:	<i>Experiment 2: Pearson’s correlations between Human Operator performance and individual differences tests</i>	145
Table 16:	<i>Experiment 2: Pearson’s correlations between trained and untrained Human Operator performance and individual differences tests</i>	147
Table 17:	<i>Experiment 2: Confidence ratings overall and by impostor type</i>	149
Table 18:	<i>Experiment 3: Performance rates and SDT indices</i>	171
Table 19:	<i>Experiment 3: Overall accuracies and performance rates by agency.....</i>	173
Table 20:	<i>Experiment 3: Hit Rates by impostor type.....</i>	174
Table 21:	<i>Experiment 3: Confidence ratings overall and by impostor type</i>	175
Table 22:	<i>Experiment 3: Decision latency overall and by impostor type.....</i>	176
Table 23:	<i>Experiment 3: Pearson’s correlations between trained Human Operator performance and individual differences tests</i>	178
Table 24:	<i>Experiment 3: Automated FR system match scores by image type</i>	181
Table 25:	<i>Comparison of laboratory Experiments 1 and 3 with previous laboratory one-to-one face matching performance evaluations</i>	186
Table 26:	<i>Comparison of Experiment 2 with previous live one-to-one face matching performance evaluations</i>	188
Table 27:	<i>Human Operator Hit Rates by impostor type.....</i>	196
Table 28:	<i>Human Operator Hit Rates by hard and easy impostor types.....</i>	198

<i>Table 29:</i>	<i>Comparing live Experiment 2 and laboratory Experiment 3 findings.....</i>	<i>215</i>
<i>Table 30:</i>	<i>Comparison of live and laboratory one-to-one face matching performance: Current and Megreya and Burton's (2008) findings</i>	<i>217</i>
<i>Table 31:</i>	<i>Pearson's correlations between trained and untrained Human Operator performance and individual differences</i>	<i>252</i>

Abstract

Evaluating Human Operator face matching performance in applied settings, such as airports, surveillance and access control settings would not only be logistically difficult, but it may not be possible due to many unknowns, such as the presence of impostors. Consequently, Human Operator performance has most commonly been evaluated experimentally, in well controlled laboratory settings. However, the question is, do the results obtained in the well controlled laboratory settings sufficiently reflect, and can they explain what happens in the real world? This applied problem has motivated the principal aim of this research to evaluate the feasibility of extrapolating one-to-one face matching performance findings from laboratory to the real world access control setting, and, in the process, support the development of an ecologically motivated performance evaluation methodology that could be used for future performance assessments, beyond the research reported this thesis.

The approach taken to address this aim stemmed from the focus on identity verification or *one-to-one face matching* task, predominantly performed within access control settings. This focus helped identify numerous factors that may affect face matching performance within access control settings. As a result, this research evaluated the impact of *impostor type and frequency*, *Human Operator expertise* and *individual differences* on one-to-one face matching performance. A preliminary evaluation (Experiment 1) provided important methodological input

into subsequent experiments. To address the principal aim, Human Operator face matching performance was first assessed within a simulated *live* access control setting (Experiment 2) which was subsequently replicated within a laboratory setting (Experiment 3). Experiment 3 also assessed the performance of an automated FR system performance to evaluate the usability of the current methodology beyond only assessing Human Operator performance.

From a methodological perspective, this research emphasised the complexities associated with evaluating and understating applied face matching performance. Applied performance may be contingent on interplay of different factors, depending on the considered applied setting. Therefore, it may not be possible to assess and state one single “level” of Human Operator performance that would be relevant to all applied settings and tasks. Instead, Human Operator performance can be assessed in light of the different environmental and task constraints, with the focus on a set of factors. Applied claims need to be appropriately qualified by explaining the exact nature of the face matching task as well as any other factors that may have affected performance.

Finally, having considered the impact of frequency and type of impostors, Human Operator expertise and individual differences, the main finding of this research showed that while overall face matching performance in the live and laboratory settings was equivalent, in the live access control setting, Human Operators were more inclined to indicate that two presented stimuli were a match, suggesting a confirmation bias. These findings are discussed in light of previous work.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Dragana Calic and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

.....
Dragana Calic

11 January 2013

Acknowledgments

As this long journey comes to some form of a conclusion, I would like to take a few moments to reflect and thank the many people who have been there for me over the past many years.

Special thanks go to my supervisory panel. My principal supervisor John Dunn and co-supervisor Anna Ma-Wyatt from The University of Adelaide, School of Psychology, and my external supervisors Brett McLindin and Veneta MacLeod from National Security Systems Analysis Task, Defence Science and Technology Organisation.

John, you have always ensured I look at things differently. I have learnt an invaluable amount through our discussions and I still fail to realise the source of your many ideas – it has not ceased to amaze and at times frustrate me. I owe you my most sincere gratitude for your generous and continuous guidance and support

Anna, thank you for your help throughout this process – for your eagerness to listen and ensure that I was on track.

Brett, thank you for opening my eyes to a whole new world – the world of applied research that actually makes a difference (yes, I am still the young naïve student!) and systems engineering. Among many things, you have taught me to appreciate diagrammatic representations, shorter sentences, and realise that nothing is logistically impossible. Words cannot express thankfulness for you always being there at the shortest of notices, returning drafts in record time and always willing to discuss different approaches. I am forever grateful.

Veneta, you have been my determined and persistent supervisor that I have needed. What more can I say but to at least try to acknowledge your absolute commitment and determination to me as a student and a friend. I will never forget you insisting to see drafts and ensure that I was going well even while on maternity leave. Imogen is a lucky little girl to have a mum as wonderful, caring and dedicated as yourself.

I have been so lucky to conduct my research through DSTO and especially the NSSA task. To all members of the task – I would not have been able to do it without your help and assistance. From administrative help, constant Excel and Matlab assistance, help with various trials' preparations and conduct (ah, the days of vacuuming masses of dried up millipedes and getting rid of redbacks), photographic assistance, participant recruitment, constant FR and life related discussions and much, much more. I thank you all wholeheartedly.

To my Cog Lab group – Victoria, Rachel, Anastasia, Annemarie, Fernando and Nick. It has been wonderful to have friends to relate to during this time. You have been an invaluable support day in, day out, during all these years.

To my DSTO gang – Laura, Kendy, Piotr, Carolyn, Kathy, Maya. Our lunches, lively events (just think of Departure), many thrilling discussions about the special land of DSTO and life in many forms have enlightened my way of thinking about culture and life! My special thanks goes to Dan. Thank you for meticulously reading through my thesis and always patiently listening to my endless thesis related rants.

To my friends and family (outside of university and DSTO) who at times had to pick up the pieces and deal with my curt answers to how my research was progressing questions (I mean, seriously, who asks that ☺). I cannot thank you all enough for rescuing me from the constant Ph.D land that I was absorbed in. To my karate community – I no longer have an excuse, time to start preparing for second Dan grading!

Finally, I am forever indebted to my family – mum and Ivan. Your endless patience, encouragement and willingness to hear me out over and over and over again. Mama, thank you for bringing sister and I to this wonderful country. Where would we be without you? I love you. My darling little sis – I have so much to say but no words at all. You are my true and only rock. I love you sekice moja mala.