

Can police officers foresee the future? Predicting outcomes from thin
slices of police-public encounters

by

Ariane-Jade Khanizadeh

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in
partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

Carleton University
Ottawa, Ontario

© 2019

Ariane-Jade Khanizadeh

Abstract

Thin slice studies are studies that examine judgments based on brief exposure to expressive behaviours or still images. Only one previous study has examined the prediction of outcomes within a law enforcement context from thin slices of a police-public encounter, and it demonstrated that experienced officers outperformed less experienced officers in terms of the quality, appropriateness, and accuracy of their predictions (Suss & Ward, 2012). The present study extends this research by examining how a range of factors – including operational years of experience and training, familiarity with the encounter, confidence in the prediction, and thin slice length – impact prediction accuracy. Participants with varying levels of police experience and training were recruited. Participants viewed 16 randomly ordered videos (half of these were 10 seconds and half were 30 seconds in length) depicting a thin slice of a police-public encounter. After each video, the participant was asked to predict whether the subject would harm or attempt to harm the officer(s). My results demonstrated that higher levels of training, greater familiarity, and greater confidence in one's predictions was associated with greater odds of providing an accurate response; operational years of policing experience was not associated with this outcome. My results also demonstrated that most of these variables' relationships with prediction accuracy disappear when examining longer thin slices (i.e., 30-second videos), and have slightly larger effects when examining shorter thin slices (i.e., 10-second videos). Finally, specialized police training, years of experience, and familiarity were, in turn, found to predict greater confidence in one's predictions. These findings and their implications are discussed.

Keywords: anticipation, outcome prediction, thin slice, police-public encounter, expertise, training, schema, use-of-force

Acknowledgements

I would first like to express my gratitude to my supervisor, Dr. Craig Bennell. You introduced me to a field that I had been interested in for a long time. I am forever grateful for having had the opportunity to work with you and to pursue research in a field that I am truly passionate about. Thank you for allowing me to grow as a researcher throughout these two years. I have learned more than I could have imagined from you; thank you for your support and guidance throughout this whole process.

I would also like to extend my gratitude to the police collaborators who assisted me with developing the concept for this project, which I was fortunate to be able to see grow and make my own. I am also thankful for their help with participant recruitment, and a special thank you to Kevin Cyr for your dedication in recruiting police participants and the time you took to discuss my thesis.

I would also like to thank the law enforcement professionals who provided valuable feedback, and everyone who took the time to participate in this study.

Thank you to Colin Kaiser and Mirza Karimullah who helped with the methodology and design of this study. I am incredibly grateful for the time you took out of your schedules to provide me with your expertise. This project would not have been possible without your help.

Thank you to my committee member, Dr. Andrew Smith, and my examiner, Dr. Diana Majury, for taking the time to review and critique my research. I would also like to extend my gratitude to our graduate administrator, Etelle Bourassa; thank you for all your assistance and guidance throughout my degree.

I would like to thank Dr. Andrew Smith, Dr. Cecilia Jorgensen, Logan Ewanation, and Simon Baldwin for their statistical guidance and advice. Thank you for answering my questions

and helping me navigate my data. Also, a special thank you to Simon Baldwin who has supported me both academically and professionally, by providing me with opportunities to learn and grow as a researcher, and for continuously supporting me in my endeavours throughout my degree. Thank you to Heather McGale, for reading drafts and providing me with both academic and moral encouragement and support. I am truly grateful for all of your guidance.

I would also like to thank Andrew, Brittany, Heather, Simon, Audrey, Tori, Bryce, and Jamie – lab-mates who I am also very fortunate to be able to call my friends. I am grateful I got to share these experiences with all of you. Thank you for all the laughs.

Thank you to my aunt Rachel, my uncle Ian, and my cousins, Amiel and Rhiannon, who have supported me since my move to Ottawa. To Nancy and Stavros, thank you for your continued support and for always being there; you have become my second family. To Raquel, thank you for your friendship and all your support. And to my best friend, Sam - thank you for being there, always.

Dedication

I would like to dedicate this work to my parents, Shahrokh and Elise, and my brother, Tristan. I cannot thank you enough for your continued support. Thank you for encouraging me through the ups and downs and for inspiring me to take a step back and enjoy every moment. I couldn't have done this without you.

Table of Contents

Abstract..... **i**

Acknowledgements **ii**

Dedication **iv**

Table of Contents **v**

List of Figures..... **viii**

List of Tables **ix**

List of Appendices..... **xiii**

Introduction..... **1**

Thin Slice Research..... **2**

 Using Thin Slices to Make Predictions 3

 Thin Slice Research Related to Police Use-of-Force..... 7

Factors that Influence the Accuracy of Thin Slice Predictions **14**

 Participant Characteristics 14

 Expertise and pattern recognition..... 14

 Confidence in thin slice predictions. 18

 Characteristics of the Thin Slice Stimuli 20

 Familiarity of the situation. 20

 Length of the thin slice..... 21

Purpose of the Proposed Study..... **24**

Methods..... **25**

Participants 25

Materials/Measures..... 27

 Demographic questionnaire..... 27

 Instructions..... 28

 Video stimuli..... 30

 Temporal occlusion paradigm..... 31

 Follow-up questions..... 31

 Training scores..... 32

Procedure..... 33

Analyses 34

 Demographic data..... 34

 Hypothesis testing..... 35

Results 36

 Demographic Data..... 36

 Descriptive Trial Data 38

 Prediction Response Accuracy 39

 Regression Analyses: Prediction Accuracy 40

 Officers vs. civilians..... 42

 Training..... 43

 ERT/SWAT training..... 46

 Years of operational police experience..... 51

 Familiarity..... 52

 Length of the thin slice..... 54

 Confidence..... 54

 Regression Analyses: Confidence 56

Training..... 57

 ERT/SWAT training 58

Years of operational police experience. 59

Familiarity..... 60

Discussion..... 61

 Predictors of Accuracy 63

 Predictors of Confidence 66

 The Role of Thin Slice Length 67

 Limitations and Future Directions 69

 Conclusion..... 72

References 74

List of Figures

Figure 1. The National Use of Force Framework. Reprinted from the CACP, 2000, Retrieved

from https://www.cacp.ca/policiesguidelines.html?asst_id=214..... 11

List of Tables

Table 1. The Four Conditions for All Trials Presented to Participants 34

Table 2. Breakdown of Participants’ Prediction Responses Based on Ground Truth. 39

Table 3. Correlation Matrix for All Independent (Predictor) Variables. 40

Table 4. Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer)
 Predicting Accuracy (with Civilian as the Reference Group). 42

Table 5. Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer)
 Predicting Accuracy for Short Thin Slices (with Civilian as the Reference Group). 43

Table 6. Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer)
 Predicting Accuracy for Long Thin Slices (with Civilian as the Reference Group). 43

Table 7. Binary Logistic Regression Results of Specialized Training Predicting Accuracy. 44

Table 8. Binary Logistic Regression Results of Specialized Training Predicting Accuracy for
 Short Thin Slices. 44

Table 9. Binary Logistic Regression Results of Specialized Training Predicting Accuracy for
 Long Thin Slices. 45

Table 10. Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers. 47

Table 11. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy
 (With Civilians as the Reference Group). 47

Table 12. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy
 (With ERT/SWAT Officers as the Reference Group). 47

Table 13. Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers for Short Thin
 Slices. 48

Table 14. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Short Thin Slices (With Civilians as the Reference Group). 48

Table 15. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Short Thin Slices (With ERT/SWAT Officers as the Reference Group). 49

Table 16. Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers for Long Thin Slices. 50

Table 17. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Long Thin Slices (With Civilians as the Reference Group). 50

Table 18. Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Long Thin Slices (With ERT/SWAT Officers as the Reference Group). 50

Table 19. Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy. 51

Table 20. Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy for Short Thin Slices. 52

Table 21. Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy for Long Thin Slices. 52

Table 22. Binary Logistic Regression Results of Familiarity on Each Trial Predicting Accuracy on That Trial. 53

Table 23. Binary Logistic Regression Results of Familiarity on Each Trial Presenting a Short Thin Slice in Predicting Accuracy on That Trial. 53

Table 24. Binary Logistic Regression Results of Familiarity on Each Trial Presenting a Long Thin Slice in Predicting Accuracy on That Trial. 54

Table 25. Binary Logistic Regression Results of The Length of The Thin Slice Presented in Each Trial Predicting Accuracy on That Trial. 54

Table 26. Binary Logistic Regression Results of Confidence on Each Trial Predicting Accuracy on That Trial. 55

Table 27. Binary Logistic Regression Results of Confidence on Each Trial Presenting Short Thin Slices Predicting Accuracy on That Trial. 55

Table 28. Binary Logistic Regression Results of Confidence on Each Trial Presenting Long Thin Slices Predicting Accuracy on That Trial. 56

Table 29. Multilevel Linear Regression Results of Training Predicting Confidence on Each Trial. 57

Table 30. Multilevel Linear Regression Results of ERT/SWAT Training Predicting Confidence on Each Trial (With Civilians as the Reference Group). 59

Table 31. Multilevel Linear Regression Results of ERT/SWAT Training Predicting Confidence on Each Trial (With ERT/SWAT Officers as the Reference Group). 59

Table 32. Multilevel Regression Results of Years of Operational Police Experience Predicting Confidence on Each Trial. 60

Table 33. Multilevel Linear Regression Results of Familiarity on Each Trial Predicting Confidence on That Trial. 61

Table 34. Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy. 95

Table 35. Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy in Short Thin Slices. 95

Table 36. Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy in Long Thin Slices. 95

Table 37. Multilevel Linear Regression Results of Recoded Aggregate Training Score Predicting Confidence Ratings. 96

List of Appendices

Appendix A . Demographic Questionnaire for Police Officers 82

Appendix B . Demographic Questionnaire for Civilians..... 85

Appendix C . Informed Consent Form 88

Appendix D . Debriefing Form..... 91

Appendix E . Binary Logistic Regression Results Following the Recoding of Civilian Training
 Responses to Police-Specific Training Options..... 94

Appendix F . Binary Logistic Regression Results Following the Recoding of Civilian Training
 Responses to Police-Specific Training Options..... 95

Introduction

Every day, police officers are faced with situations that can be unpredictable. In every police-public encounter, police officers must conduct risk assessments and decide how to respond to what they are facing based on a variety of factors (e.g., environmental, behavioural, tactical, and perceptual considerations; Royal Canadian Mounted Police, 2017). Imagine a situation in which officers are responding to a call involving a person in crisis. In addition to considering tactical information (e.g., options for concealment and cover, the availability of backup, distance from the subject) and various perceptions (e.g., presence of bystanders, one's own abilities, possession of a weapon), officers need to consider the subject's behaviour (Canadian Association of Chiefs of Police [CACCP], 2000). In order to fully assess a subject's behaviour, an officer might rely on the subject's nonverbal and verbal cues to determine whether they are likely to engage in violent behaviour, and whether the subject will be responsive to the officer(s) on scene. Based on these predictions, the officer may intervene in a certain manner to ensure that subject, public, and officer safety is maintained. Having the ability to predict the outcome of an encounter will likely allow the officer to choose the best course of action when responding (Suss & Ward, 2012).

The proposed study will seek to examine whether officers can predict the outcomes of encounters based on short segments (i.e., thin slices) of police-public encounters. The study will also explore a variety of factors that may be associated with the accuracy of anticipated outcomes (i.e., the experience level and training of the observing officer, the familiarity of the observing officer with the situation they are assessing, the length of exposure to the stimuli, and the confidence of the observing officer's predictions). Developing a better understanding of these issues is important, not only for theoretical reasons (e.g., it will contribute to our understanding

of how schemas might assist in making outcome predictions), but because such an understanding may contribute to practical matters, such as making improvements to police training (e.g., by determining what factors are associated with accurate predictions so that these factors can potentially be focused on in training). In the remainder of this document, a review of the thin slice research will be provided, followed by a discussion of a thin slice study related to police use-of-force. Then, I will present an overview of factors that may influence the accuracy of thin slice predictions in the policing context. Finally, the methods and statistical analyses used in the thesis will be described, along with the results and a discussion of the findings.

Thin Slice Research

Thin slice studies are studies that examine judgments based on brief exposure (usually less than five minutes) to expressive behaviours (e.g., Ambady, 2010; Fowler, Lilienfeld, & Patrick, 2009; Nguyen & Gatica-Perez, 2015) or to still images (e.g., Ambady, Hallahan, & Conner, 1999; Rule & Ambady, 2008). Studies have demonstrated that humans can be quite accurate in predicting future outcomes from only brief exposures to an individual's current behaviour (e.g., Ambady & Rosenthal, 1993; Blanck & Rosenthal, 1984; Tskhay, Zhu, & Rule, 2017), as well as predicting the traits of individuals simply by being exposed to an image of an individual's face for a short period of time (Rule & Ambady, 2008). Examples of outcome-related thin slice research include studies that have examined negotiation outcomes based on conversational dynamics (Curhan & Pentland, 2007), predictions of governor elections in the United States based on short videos of candidates in which sound was manipulated (Benjamin & Shapiro, 2009), and predictions of soccer plays based on videos taken from above the soccer net (Ward, Ericsson, & Williams, 2013). Examples of social situations that have been investigated include examinations of the accuracy of leadership judgments based on perceptual charisma

ratings (Tskhay et al., 2017), judgments of sexual orientation based on nonverbal behaviour (Ambady et al., 1999), and judgments of high and low expectancy in children (i.e., whether a counselor expects a child to be “high” or “low” on a given trait, such as athleticism) based on ratings of camp counselor tone of voice (in terms of warmth and hostility; Blanck & Rosenthal, 1984).

Interestingly, predictions made from thin slices of behaviour are often as accurate as predictions made from behaviour sampled over longer periods of time (Ambady & Rosenthal, 1992). In addition, thin-slice predictions appear to be intuitive (i.e., based on an automatic process), similar to making social judgments (Ambady, 2010). In fact, research has suggested that deliberation may hinder the accuracy of one’s predictions (see Ambady, 2010, for review). The disadvantage of deliberating has been demonstrated in the field of deception detection. For example, Albrechtsen, Meissner, and Susa (2009) demonstrated that participants in their study were more accurate in detecting deception when watching a five-second video compared to a three-minute video; participants were also more accurate when identifying deception while they completed a concurrent task (i.e., were forced to use automatic processing) and were not as accurate when they were asked to provide a rationale before making a decision.

Using Thin Slices to Make Predictions

What is it about thin slices that allow them to be used to make accurate predictions of future outcomes or personal traits? The answer to this question is not entirely clear yet, but it seems likely that thin slices include cues that provide just enough information to allow the viewer to adequately perceive and process the stimuli in order to make adequate predictions about what is being observed (Abernethy, 1993). The important cues likely vary depending on the stimuli, and the specific prediction task. For example, when making predictions about

psychopathy ratings, body posture and facial expressions may be key to making accurate predictions (Fowler et al., 2009), whereas in sporting predictions, the essential information may include movement sequences of an opponent's limbs (Abernethy, Gill, Parks, & Packer, 2001), and relational information (e.g., locations of middle offensive players in relation to the rest of the players; Williams, Hodges, North, & Barton, 2006).

With respect to the predictive value of facial expressions and bodily cues, “nonverbal leakage” has been studied in thin slice research as it appears to be at the root of many of the cues that can be used to make accurate predictions. Nonverbal leakage refers to the idea that individuals will display behavioural indicators with parts of the body that are not consciously controlled. Early research has suggested that nonverbal leakage may be diagnostically useful, for example when predicting whether someone is lying (Ekman & Friesen, 1969), or whether someone is experiencing anxiety (Waxer, 1977). More recent thin slice research has also demonstrated how nonverbal behaviour can contribute to accuracy in outcome predictions. For example, as highlighted above, it appears that governor election outcomes can be predicted from nonverbal behaviour (Benjamin & Shapiro, 2009) and leadership can be predicted from judgments of an individual's charisma based on nonverbal behaviour (Tskhay et al., 2017).

Static images, which may be considered a type of nonverbal cue, have also been examined in the context of thin slice research. For example, in addition to judgments of charisma being made from nonverbal behaviour, Tskhay et al. (2017) demonstrated that charisma can also be predicted from ratings of physical attractiveness based on still pictures of faces. This is consistent with earlier research that focused on thin slice predictions based on static cues. Rule and Ambady (2008), for instance, demonstrated that a company's financial success could be predicted from a greyscale image of a CEO's face. After the researchers had controlled for age,

attractiveness, and emotion of CEOs, they observed that characteristics related to power (i.e., competence, dominance, facial maturity) and leadership, predicted from the CEO's facial characteristics, were associated with the company's financial success (i.e., the company's profits).

In addition, the utility of verbal cues for making accurate predictions based on thin slices of information has been the focus of some research, with a complex pattern of results emerging from existing studies. For example, thin slice studies have demonstrated that verbal cues can be useful in predicting one's intelligence (Borkenau & Liebler, 1993; Fowler et al., 2009), in predicting whether one suffers from antisocial personality disorder (Fowler et al., 2009), and in predicting whether someone is hireable (Nguyen & Gatica-Perez, 2015). However, other research has demonstrated that verbal cues may hinder thin slice predictions, such as judgments of student excellence based on videos of their teachers (Babad, Bernieri, & Rosenthal, 1991), or election outcomes (Benjamin & Shapiro, 2009). Still other research has suggested that the predictive value of verbal information may depend on the circumstances. For instance, based on the findings of their meta-analysis, Ambady and Rosenthal (1992) suggested that verbal cues may be helpful when viewers can observe one's behaviour over a longer period of time, but may not contribute to accuracy when viewing thin slices of behaviour.

Finally, extensive research has examined patterns of movement cues or kinematic information to show how this information can be used to accurately anticipate outcomes in thin slice research (e.g., Abernethy et al., 2001; Sebanz & Shiffrar, 2009; Williams et al., 2006). In fact, motion may improve prediction accuracy, compared to motionless stimuli, as was observed in a study by Ambady et al. (1999), where judgments of sexual orientation were less accurate when based on still images versus dynamic stimuli.

By removing information related to colour, texture, contour, location, and spacing (Abernethy et al., 2001), studies using point light displays¹ have demonstrated that individuals with more expertise in a particular domain may rely on relational or kinematic information to predict situational outcomes. Abernethy and colleagues (2001), for instance, demonstrated that experts in squash outperformed novice squash players when anticipating the end location of a squash ball following an opponent's stroke displayed in a point light format, even when provided with very little information (i.e., the ball's location 160 milliseconds prior to the ball hitting the racket on the screen). Experts outperformed novices by making fewer perceptual mistakes related to depth and direction perception.

In another study by Sebanz and Shiffar (2009), basketball experts outperformed novices in predicting deceptive intentions of a player displayed on a screen (i.e., whether they would make a real or a fake pass) when watching a thin slice dynamic video, but not when a static picture of the last frame of the video was presented. This suggested to the researchers that experts likely use kinematic information to make their predictions. Experts in this study also appeared to rely less on postural and facial expressions compared to novices, and could anticipate movement deception from various viewpoints, although performing best when they viewed the player from the front.

Similarly, in yet another thin slice study, experts and novices were examined to see if they could identify whether they had previously seen a soccer play from thin slices of game play (Williams et al., 2006). Consistent with previous research, experts outperformed novices and

¹ Point light displays consist of placing points of light over joints of the body to depict basic biological motion (Johansson, 1973).

were able to utilise relational information between players to make relatively accurate recognition judgments.

The challenges with using many of the cues discussed above to make predictions include the fact that different individuals will likely have different baselines with respect to the frequency with which they exhibit certain behaviours (Johnson, 2006, 2007, 2019; Vrij, Akehurst, & Morris, 1997), and these baselines may be affected by a range of factors. These factors can include ethnicity, culture, and emotional agitation (Johnson, 2006, 2007), the setting within which a person finds themselves (Johnson, 2019), as well as personality traits, such as self-consciousness and behavioural control (Vrij et al., 1997). In the words of Ekman and Friesen (1967), individuals may have their own “basal position” (p. 719). Differences can also be observed in an individual’s “style” of expression, also known as “expressive behaviour” (i.e., the way a person behaves, as a whole), and this can be influenced by factors such as sex, femininity and masculinity, personality traits, and body type (Gallaher, 1992). Thus, it may be difficult to identify specific behavioural patterns that can be expected of different individuals, especially without prior exposure to these individuals, which increases the complexity of predicting accurate outcomes from thin slices of information.

Thin Slice Research Related to Police Use-of-Force

In the context of policing on the street, it is sometimes important for predictions to be made quickly to help officers respond more appropriately to the situations they face. Indeed, because there is a high degree of uncertainty in the field of policing and some decisions have to be made in a split-second, being able to predict the outcome of an encounter (i.e., the future behaviour of a subject) early in an interaction may help officers better prepare for and respond to that situation. Thus, thin slice studies are likely to be beneficial in a law enforcement context so

that researchers can determine if it is possible to make accurate predictions based on thin slices of information, and if so, to identify any factors that are associated with superior predictive abilities (Suss & Ward, 2018).

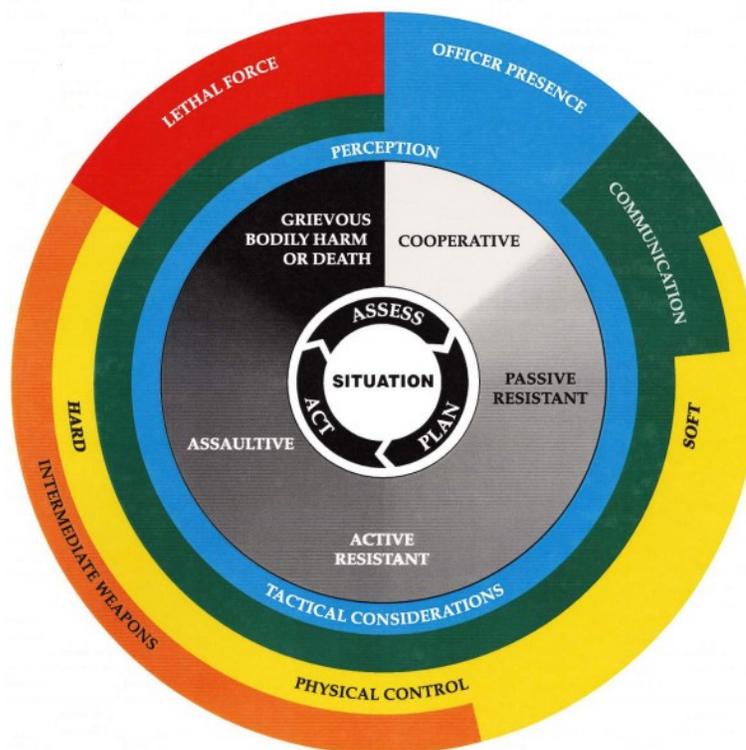
To the best of my knowledge, only one study has been conducted that has examined the prediction of outcomes made by police officers based on thin slices of police-public interactions (Suss & Ward, 2012). This study involved two groups of participants: experienced officers (e.g., officers with special weapons and tactics [SWAT] training) and less experienced officers (i.e., police recruits). Participants were exposed to short video simulation exercises in which the officers interacted with different situations that were characterized as being highly likely to occur or not (e.g., domestic assault vs. suicide bomber), and that were either threatening (e.g., the subject was swinging a baseball bat), lethal (e.g., the subject was drawing a firearm), or benign (e.g., the subject was complying with the officer's requests). Videos were preceded with an audio dispatch message describing the situation that would be displayed on the screen. Occlusion points were identified for each video (i.e., the earliest point in a video at which enough information has been presented that may indicate the outcome of the scenario). At the pre-determined occlusion point, the video on the screen disappeared and the officers had to make various predictions about the interaction. More specifically, following the occlusion point, Suss and Ward asked participants to predict the outcome(s) of the scenario, indicate how they might respond if they were faced with that situation, and rate the likelihood of each response they provided for both outcome and response predictions.

Suss and Ward (2012) observed that, compared to less-skilled officers, experienced officers: (1) provided fewer low-quality options and more high-quality options for both outcomes and responses (ratings of quality were provided by use-of-force and police decision-making

experts, i.e., subject matter experts [SMEs]); (2) predicted accurate outcomes more frequently and provided a response option that was deemed more appropriate by SMEs more frequently; and (3) provided higher likelihood ratings for the correct outcome option and the more appropriate response option more frequently. Finally, experienced officers who had provided the correct outcome and response options first (in their list of options) also chose these options as their final decisions more frequently.

What might explain the ability of certain officers to make accurate outcome predictions based on thin slices of police-public interactions? As is the case in other thin slice studies, it is likely that behavioural (and other) cues exist within police-public encounters that allow some officers, especially experts, to do this (Ward, Suss, Eccles, Williams, & Harris, 2011). In fact, through training, officers are typically taught to focus on certain cues in order to make accurate assessments of risk. For example, within Canada's National Use of Force Framework (NUFF; see Figure 1), specific cues are highlighted to help officers assess risk in any given situation. Not only does this assessment help officers determine the appropriate level of force to use in a specific encounter, the framework also helps the public understand the circumstances under which officers may be permitted to use force to maintain public and officer safety, in part by assisting officers with their post-event articulation (i.e., explaining their decision-making process after a use-of-force encounter; CACP, 2000). The framework includes three main components for assessment: (1) a consideration of situational or environmental factors ("assess-plan-act"), (2) a consideration of subject behaviour (i.e., cooperative, passive or active resistant, assaultive, grievous bodily harm or death), and (3) a consideration of perceptions and tactical information (CACP, 2000).

More specifically, the innermost circle of the NUFF represents how an officer must continuously “assess-plan-act” in a given situation. The next circle (moving outwards) represents the various subject behaviours that an officer might encounter: cooperative (i.e., the subject is complying to the officer’s requests), passive resistance (i.e., the subject is resisting to comply with an officer’s requests, such as not moving when being asked to move), active resistance (i.e., the subject is physically resisting to comply with an officer’s requests, such as physically avoiding arrest), assaultive (i.e., the subject is threatening to or is assaulting any individual), and grievous bodily harm or death (i.e., the subject is engaging in behaviour that could cause serious harm or death to any individual). The next circle contains officer perceptions (e.g., strength of the subject compared to personal strength) and tactical considerations (e.g., whether backup is on the way), and this circle is therefore unique to the individual officer. This section of the NUFF is used to both assess the situation and gauge the subject’s behaviour, which will consequently impact the intervention/level of force that is used when responding. The final, outermost circle represents various interventions or levels of force an officer can employ: officer presence, communication (both verbal and non-verbal), physical control soft (i.e., techniques used to control a subject that have a low probability of resulting in injury to either the subject or the officer), physical control hard (i.e., techniques used to stop the actions of a subject that have a higher probability of resulting in injury to either the subject or the officer), intermediate weapons (i.e., the use of less-lethal weapons, such as conducted energy weapons or oleoresin capsicum spray [pepper spray]), and lethal force (CACP, 2000).



The officer continuously assess the situation and acts in a reasonable manner to ensure officer and public safety.

Figure 1. The National Use of Force Framework. Reprinted from the CACP, 2000, Retrieved from https://www.cacp.ca/policiesguidelines.html?asst_id=214

The CACP (2000) has identified certain behaviours that subjects may display before they physically attack an officer. These include not complying with or ignoring the officer, hiding from the officer, standing in an aggressive manner, being verbally aggressive, becoming completely still, asking many questions, and venting one’s emotions. Based on empirical research, other cues have also been deemed important. For example, environmental cues have been associated with an increased likelihood of an officer being attacked, such as the location of the incident (e.g., different districts of a police department; Stetser, 1999), the location of the call (e.g., in an alley; Bierie, 2015), and the type of call an officer is responding to (e.g., disorderly conduct; Stetser, 1999). The subject being male and being under the influence of alcohol or

drugs are also subject characteristics that have been suggested to increase the likelihood of an officer being attacked (Bierie, 2015; Covington, Huff-Corzine, & Corzine, 2014; Stetser, 1999). Finally, certain perceptual and tactical cues have been associated with an increased likelihood of an officer being attacked, such as the time of day at which the incident takes place (e.g., evening/night time; Brandl, 1996; Covington et al., 2014; Stetser, 1999), and the presence of more than one officer on scene (Covington et al., 2014).

Numerous challenges emerge when attempting to use these sorts of cue categories to predict the outcome of interactions. For example, after a review of the literature, Johnson (2011) came to the conclusion that subject characteristics, which are often thought to be highly associated with assaults on an officer, may simply be characteristics of the general population that frequently interact with the police (e.g., individuals of lower socioeconomic status). In addition, cues that may signal a future attack on an officer may vary as a function of a wide range of factors, such as the type of call the officer is responding to (Johnson, 2011). What is needed is empirical research that determines the features of police-public interactions that are predictive of aggression and violence. Unfortunately, little research has been conducted on this topic. The studies by Johnson and Aaron (e.g., Johnson, 2011, 2015, 2019; Johnson & Aaron, 2013) represent much of the research that has been carried out.

Johnson (2011) used a series of regression analyses to study whether certain subject characteristics (e.g., unemployed, using drugs, married, living with the victim, etc.) were predictive of an assault on a police officer in 1,951 domestic assault cases. This study observed that having a “hostile demeanor” (toward the police arriving on scene) was one of the most important cues for signalling an impending attack against the responding officer(s) (i.e., the odds of an attack were 12.6 times higher). In addition to exhibiting hostility towards officers, other

factors that seemed to be related to attacks on officers included: the offender had consumed alcohol, was living with the victim, had damaged property, and was unemployed.

Building off of Johnson and Aaron's (2013) study, which examined perception of cues for impending violence in a sample of university students, Johnson (2015) sampled individuals who had experienced violence first-hand; that is, police officers who had experienced at least one physical assault while on duty. These officers were asked to watch a video clip of a police-public encounter, depicting a subject being verbally aggressive toward an officer. They were then asked to subsequently indicate their concern for 23 behaviours (i.e., as being indicative of impending violence) on a seven-point Likert scale ranging from "no concern" to "very concerned."

Three behaviours appeared to be associated with very high concerns, including instances when the subject "assumes a boxer's stance," "invades personal space," and "places hands in pockets." Other behaviours that were strongly associated with high concerns of impending violence included instances when the subject "clenches hands," "looks around the area," and "makes threats." Behaviours that were perceived as moderately concerning included instances when the subject displayed the following: "head roll or neck stretches," "jaw muscle tenses," "paces back and forth," "sweats profusely," "stretches arms or shoulders," "tenses up whole body/becomes rigid," "yells," "removes excess clothing," "face becomes flushed red," "breathes more rapidly," "angry expression," and "stares into your eyes." Behaviours associated with low concern for potential violence included: "makes exaggerated hand gestures," "avoids eye contact," "places hands on hips," and "blinks eyes rapidly," along with "crying," which was associated with the least concern to participants.

Although this study did not examine which behaviours were predictive of violence (i.e., it examined cues that participants *thought* would be predictive of violence), a more recent study by

Johnson (2019) examined nine behaviours, most of which overlapped with those examined by Johnson (2015), to assess whether these could in fact be predictive of violence. He found that “fighter’s stance,” “invades personal space,” “hands in pockets,” and “clenched hands” – all behaviours that officers reported to be of very high or high concern for impending violence in Johnson (2015) – *were* associated with violent interactions between the officer(s) and subject in the videos (Johnson, 2019). On the other hand, “looks around area” and “pacing” – behaviours that officers reported to be of high and moderate concern, respectively, in Johnson (2015) – were found to be displayed *less* often for subjects who were violent compared to non-violent subjects.

Although the current thesis will not specifically examine cues associated with accurate predictions of impending violence, it will attempt to examine what factors might lead to better odds of making accurate outcome predictions. These will be discussed in the following section.

Factors that Influence the Accuracy of Thin Slice Predictions

Research suggests that a variety of factors influence an individual’s ability to anticipate various outcomes based on thin slices of information. The following section reviews individual characteristics, specifically expertise and pattern recognition, characteristics of thin slice stimuli, including familiarity with the scenario and the length of the thin slice, as well as confidence in one’s predictions and their relation to prediction accuracy. These factors are focused on because they are central to the proposed study.

Participant Characteristics

Expertise and pattern recognition. As illustrated already in this literature review, participant expertise is one of the major factors that has been examined in thin slice studies. Generally speaking, this research has suggested that experts in a given domain are able to pick up on cue patterns more efficiently than individuals without such experience (e.g., Chase &

Simon, 1973; Gabbett & Abernethy, 2013), which typically allows them to make more accurate predictions based on thin slices of information (e.g., Sebanz & Shiffrar, 2009; Ward et al., 2013). Expertise is usually thought to be the result of accumulated time spent in deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993) and the development of expertise is believed to take several years (reviewed in Ericsson et al., 1993). In the law enforcement context, officers may develop expertise through various training opportunities as well as through experience in the field (Ericsson et al., 1993).

The impact of training and experience on cue detection (and enhanced prediction performance) has been clearly shown in both non-police and police settings. For example, in addition to the studies discussed previously, Gabbett and Abernethy (2013) conducted a study in which 88 rugby players of differing levels of expertise (“high-,” “intermediate-,” and “low-skilled”) and differing age groups (“senior,” “under 20,” and “under 16”) were tasked with responding to a life-size video of a “ruck play” (i.e., a type of rugby play), where a rugby player can be seen kicking the ball on screen. Participants had to respond by running forward, right, or left in response to this play. Overall, the researchers observed that more highly skilled players were quicker and more accurate in their responses compared to less skilled players, and that older players were also quicker and more accurate compared to younger players. The authors suggested that this was likely due to “task-specific experience,” and that extensive practice and experience in the player’s domain is what contributed to superior performance. The observed expert-novice differences were suggested to depend on experts’ superior ability to recognize cues relevant to the task at hand.

In another study, Williams, Ward, Knowles, and Smeeton (2002) provided relatively unskilled tennis players (playing recreationally for an average of 3.8 years) with perceptual

training to help them anticipate the location of the ball and to plan future behaviour, either by identifying relevant regions on their opponent (e.g., midriff region) or identifying specific cues (e.g., hip orientation). This training was demonstrated to benefit performance, in terms of prediction speed. In fact, the authors suggested that, following the perceptual training, the players who had been playing recreationally performed in a similar manner to skilled players (i.e., in terms of response accuracy and decision time), who had participated in an average of 500 tournaments and had been playing tennis for an average of 11.9 years. Similar sorts of results have been found in other studies of sports training (Abernethy, Schorer, Jackson, & Hagemann, 2012; Murgia et al., 2014).

Within the policing context, expertise differences in cue detection have also been demonstrated, and this has been shown to relate to enhanced anticipation and performance. For example, Vickers and Lewinski (2012) exposed novice and expert police officers, armed with a standard pistol loaded with one round of Simunition,² to seven trials of a scenario in which they were told to guard a supposed government office. In the scenario, the officer encountered a subject, also equipped with a standard pistol loaded with one round of Simunition, who became very upset, ultimately pulling out a cellphone (on two trials) or a gun (on five trials) from their jacket. The officer's task was to make a "shoot/don't shoot" decision. The researchers measured shot accuracy, response accuracy, speed of response, and eye gaze (i.e., visual fixations and saccades from a mobile eye tracker).

² Simunition is artificial ammunition that is used in simulation training. Simunition is loaded into a firearm and will emit a coloured cartridge that marks a target once the firearm has been shot ("Simunition: Non-lethal training ammunition," n.d.).

Consistent with other non-policing research, experts in Vickers and Lewinski's (2012) study performed better than novices; not only did they exhibit better shooting accuracy (i.e., they hit the subject more frequently), their responses were also more accurate (i.e., they shot the subject on the cellphone trials less frequently) and quicker (i.e., they were faster at unholstering their firearm and aimed and fired their pistol quicker). Importantly, expert officers were also more likely to visually fixate on regions in which a weapon was more likely to be hidden (e.g., the inside of the subject's jacket), and spent more time fixating on the subject's weapon than on their own firearm. The researchers concluded that better performance was associated with having longer visual fixations while the assailant was drawing their weapon/cellphone and turning to face the officer, followed by shorter fixations during the time the officer could see the firearm/cellphone the subject was holding. Overall, this study suggested that experts may have the skills to efficiently identify relevant cues of interest in order to anticipate an outcome (i.e., the presence of a lethal threat) of a police-public encounter.

Ward and colleagues (2011) also demonstrated expert-novice differences in anticipation accuracy within a police setting. The purpose of their study was to examine how decisions are made (i.e., option generation) by officers with varying levels of experience. Participants completed three practice trials, followed by 20 test trials (including nine non-lethal and 11 lethal trials) on a use-of-force simulator. The videos participants were interacting with were occluded if they fired their weapon; but they were played in full if participants did not discharge their firearm. Participants provided retrospective verbal reports of their thought process for half of the test trials, and these statements were separated into assessment or intervention statements (i.e., statements that reflected either the assessment or the intervention phase of decision-making). Each statement was coded as "critical" if it would help the situation or "non-critical" if it would

not. The authors also analysed response time and shot accuracy on the first shot fired. They observed that experts provided significantly more critical options in both assessment and intervention phases compared to novices, however no differences were observed in terms of the number of non-critical options. Expert-novice differences were also observed in the type of statements participants made. For example, in the assessment phase, experts made more “monitoring” and “prediction” statements, and in the intervention phase, experts made more “decision” statements.³ Finally, the authors observed that skilled officers responded more quickly and were more accurate than less-skilled officers.

Confidence in thin slice predictions. Another participant characteristic that will be examined in the current study relates to prediction confidence (not only how various observer characteristics relate to confidence, but how confidence relates to accuracy). Confidence has also been examined in a thin slice context, but only to a limited degree. A complex pattern of findings emerges from existing literature.

In an early study, Smith, Archer, and Costanzo (1991) asked participants to watch a 30- to 60-second video and respond to a question specific to the individuals or the situation being depicted in that video (e.g. “who won the game,” “which of two women is the mother of a boy” [p. 6]). Participants were then asked to rate their degree of confidence in their response. Overall, Smith et al. (1991) found a strong positive relationship between confidence and the accuracy of

³ Monitoring statements were those that were made in real-time during an event, or in Ward et al.’s (2013) words “actual on-line thoughts” (p. 240). Prediction statements encompassed those that presumed how the encounter might end (Ward et al., 2013). Finally, decision statements included statements in which a course of action was chosen (Ward et al., 2011).

participant responses. A positive relationship between confidence and accuracy has also been supported by other thin slice research (Patterson, Foster, & Bellmer, 2001).

A more recent study by Ames, Kammrath, Suppes, and Bolger (2010) revealed different results. They conducted three studies in which they asked participants to predict an individual's personality on a seven-point Likert scale. The person was depicted in either a still image (in the first study) or a 60-second video (in the second and third studies) in which the verbal content was removed (only different aspects of speech, such as tone, could be heard). In their third study, participants were divided into three groups – a control condition (no experimental manipulation), an “intuition” condition (this group was provided information on intuitive decision-making), and a “reason” condition (this group was provided information on analytical decision-making). All participants were asked to rate their confidence for each personality trait prediction they made on a 7-point Likert scale and, in all three studies, they examined variables that might predict confidence ratings.

Across all three studies, Ames et al. (2010) found that confidence varied by personality trait (e.g., higher confidence ratings were found for predictions of extraversion) and, importantly, that confidence was generally unrelated to accuracy. However, in both the second and third study, the authors found that participants who reported no confidence in their predictions generally predicted that their responses would be inaccurate. Ames and his colleagues also found that confidence and accuracy increased when participants were shown videos (second study) instead of still images (first study). In their third study, accuracy did not differ across experimental groups.

In terms of the variables that were examined to see if they predicted confidence ratings, Ames et al. (2010) found, among other things, that “judgment extremity” (i.e., reporting “strong”

ratings, such as “strongly agreeing” or “strongly disagreeing”) predicted one’s degree of confidence, but did not moderate the relationship between confidence and accuracy (Study 1, 2, and 3); that participants who reported higher self-efficacy in person perception reported higher confidence scores, and that higher confidence scores were provided “when the target seemed to fit a type” and “when the target reminded her or him of someone she or he knew” (Study 2 and 3); that higher confidence ratings were given by participants “who think they are very good at judging people,” “who have high faith in intuitive decision-making,” and “who have a low need for analytical decision making” (Study 2); and that specificity ratings of the individual depicted in the video (i.e., whether something specific about the individual prompted the participant to make a given personality prediction) was associated with higher confidence, as well as higher accuracy (Study 3; although this finding varied across experimental groups).

Characteristics of the Thin Slice Stimuli

In addition to personal characteristics, it is possible that the characteristics of thin slice stimuli may impact one’s accuracy in predicting outcomes. Consistent with this, a number of thin slice features have been investigated by previous researchers and have been shown to impact prediction accuracy. The video features that will be examined in the current study include: the type of encounter (or, more specifically, an individual’s familiarity with the type of encounter) and the length of the thin slice.⁴

Familiarity of the situation. Research studies from settings outside policing have indicated that the specific circumstances displayed within a thin slice can alter one’s ability to

⁴ The length of the thin slice will be examined as a proxy for the number of cues present within the thin slice, with the assumption that, the longer the thin slice, the more environmental and behavioural cues will be available to make an outcome prediction.

make accurate judgments and predictions about it, particularly if the observer is familiar with the circumstances they are being exposed to (e.g., Klein, Calderwood, & Clinton-Cirocco, 1986). Likewise, with respect to police-public encounters, an officer's familiarity with a given scenario has been shown to improve their performance in simulated encounters, both in terms of response time and shot accuracy (Ward et al., 2011). It is known that different types of police-public encounters consist of different risk factors for officers (Ellis, Choi, & Blaus, 1993); being aware of this, and having experienced these risks in training or when responding to different calls, appears to impact how officers perceive and respond to the various situations they encounter. Given that individuals with less experience will not have acquired these schemas yet, they will likely be unable to draw on past situations to infer the outcome of a current situation (Gonzalez, Lerch, & Lebiere, 2003). The magnitude of differences in outcome prediction accuracy between individuals who are familiar (versus less familiar) with the type of encounter they are exposed to will be explored in the present study.

Length of the thin slice. The length of thin slices and its relationship with prediction accuracy has also been examined. For example, a study by Nguyen and Gatica-Perez (2015) demonstrated that the length of the observation period may indeed impact prediction accuracy as well as the cues one uses to make predictions. These researchers examined the predictive accuracy of ratings of how hireable participants were based on thin slices of a job interview. They observed that thin slices led to accurate predictions, however full interviews (rather than thin slices of these interviews) led to even higher accuracy in predictions. With respect to the use of cues, they found that certain cues (e.g., "visual features," such as head movement) contributed to prediction accuracy in the full videos, but did not in the thin slices. They hypothesized that these types of cues need to be observed over longer periods of time, which is not possible when

viewing thin slices of interactions. Similar results have been reported in a thin slice study examining personality, affect, and intelligence predictions (Carney, Colvin, & Hall, 2007).

Interestingly, despite results like these, other studies have found that slice length does not have an impact on prediction accuracy (e.g., Ambady & Rosenthal, 1993; Tskhay et al., 2017). In addition, meta-analytic research (based on older studies) does not support the view that thin slice length impacts the accuracy of predictions that can be made from thin slices. For example, Ambady and Rosenthal's (1992) meta-analysis of 44 studies (38 sets of results) found that predictions made from shorter thin slices (of less than 30 seconds) were no more accurate than predictions made from longer thin slices (up to five minutes). It is not entirely clear from the meta-analytic results why no difference in prediction accuracy emerges. It may be that such a difference would emerge if various moderator variables were taken into account. One possibly important moderator might be the expertise of the individual making the predictions. More specifically, it could be that more experienced individuals may be able to better anticipate outcomes with less information compared to novices.

For example, both Abernethy et al. (2001) and Ward et al. (2013) have demonstrated this by examining prediction accuracy when a video is occluded at different points in time (the earlier the occlusion point, the less information the participant has to make their prediction). Abernethy et al. (2001), for instance, examined expert and novice squash players. In their first experiment, participants viewed videos and point-light displays of a player hitting a squash ball. Each of these videos/point-light displays had five different occlusion points, ranging from 160 milliseconds before the player hit the ball to when the player had hit the ball and the ball was no longer in contact with the player's racket. Participants were asked to predict the end location of the ball. The researchers observed an interaction between expertise and occlusion time, which

was driven by the fact that experts made fewer prediction mistakes across various time points early in the videos, but novices did not. This finding suggested that experts (but not novices) can identify relevant cues for prediction purposes very early in a scenario (Abernethy et al., 2001).

Ward et al. (2013) examined expert-novice soccer player differences in outcome anticipation (i.e., choosing between three final ball locations after a pass) and decision-making strategies in soccer players. Participants observed a video of a soccer play, which was occluded at three different time points: after the opposing player had kicked the ball (labelled as “post”), at the point when the opposing player was making contact with the ball (labelled as “contact”), or 120 milliseconds before the player touched the ball (labelled as “pre”). Each participant underwent three practice trials, followed by 24 test trials (eight for each “pre,” “contact,” and “post” conditions). Participants were asked to report on what the opposite player might do to generate possible options, and then they ranked each option “in an order that reflected the greatest threat posed to their defense” (p. 236). Analyses demonstrated greater expert-novice differences when the thin slice video was occluded before the opponent made contact with the ball or at the exact moment of contact, compared to after the opponent had already kicked the ball, suggesting that more experienced players could make more accurate predictions with less information (Ward et al., 2013).

Extending the results of these studies to the domain of policing, it is possible that more experienced/trained individuals will be able to predict the outcome of a police-public encounter based on more limited information (i.e., a shorter thin slice), despite the possibility of being presented with fewer environmental, situational, and subject behaviour cues. This hypothesis will be tested in the current study.

Purpose of the Proposed Study

The proposed study will examine outcome prediction accuracy of police-public encounters, when police officers (with varying levels of expertise and training) and civilians (with no expertise in policing and no police-specific training) are tasked with viewing a thin slice of the encounter under varying conditions (when encounter familiarity and thin slice length varies). I also examined how confidence varies across participants that possess different levels of training and experience, and how confidence in the thin slice predictions relates to prediction accuracy. More specifically, the following hypotheses will be examined:

1. Participants with specialized training (e.g., emergency response team [ERT] or SWAT⁵) will demonstrate superior performance, compared to officers and civilians without this training, in:
 - a. anticipating whether the subject will harm or attempt to harm the officer(s), and
 - b. anticipating impending/possible harm when presented with shorter thin slices (no such differences will emerge for longer thin slices).
2. Officers with more experience (i.e., more years of police service) will demonstrate superior performance in:
 - a. anticipating whether the subject will harm or attempt to harm the officer(s), and

⁵ The main difference between ERT and SWAT is that ERT is the term used in Canada and SWAT is the term used in the United States. Both ERT and SWAT involve very similar training practices.

- b. anticipating impending/possible harm when presented with shorter thin slices (no such differences will emerge for longer thin slices).
3. Participants who have greater familiarity with the type of encounter illustrated in the thin slice will demonstrate superior performance in:
 - a. anticipating whether the subject will harm or attempt to harm the officer(s), and
 - b. anticipating impending/possible harm when presented with shorter thin slices (no such differences will emerge for longer thin slices).
4. Participants with specialized training (e.g., ERT or SWAT) and officers with more experience (i.e., more years of police service), as well as participants reporting greater familiarity with the type of encounter illustrated in the thin slice, will report higher confidence scores.

The following research question will also be examined:

1. Will confidence in one's predictions be related to the odds of providing an accurate response?

Methods

The methods for this study were based loosely on the methods used by previous researchers who have studied similar issues (e.g., Suss & Ward, 2012; Ward et al., 2011).

Participants

The police participants were recruited with the help of various collaborators who work within policing (e.g., Royal Canadian Mounted Police [RCMP], Ontario Provincial Police [OPP], Federal Bureau of Investigation [FBI]). These collaborators sent an ethics-approved recruitment announcement containing a link to the online experimental package (in Qualtrics) to their police

contacts via email. Social media was also used to recruit these participants (i.e., a Twitter announcement). Carleton University students (hereafter referred to as civilians), who had no experience working as a law enforcement officer, correctional officer, or military officer were recruited through the SONA system and consisted of students registered in SONA-eligible psychology courses.

I collected data from 228 participants, of which 57 either did not consent to participate or chose to withdraw while completing the study. These participants were removed from the dataset. A further 45 participants did not respond to any of the test trials and, therefore, they too were removed from the dataset. The final sample size was of 126 participants. It is important to note that not all of these participants completed all 16 test trials, nor did they all have useable data for each test trial (trials in which participants reported they had already seen the video footage or indicated that they had experienced technical difficulties viewing the video were excluded). The completed test trials were relied on for all analyses.

The sample was made up of both civilians ($n = 62$, 49.2%) and law enforcement officers ($n = 64$, 50.8%). Most of the civilian sample was female ($n = 40$, 64.5%). The most frequently reported ethnicity⁶ in the civilian sample was Caucasian ($n = 27$, 43.5%), followed by Southeast Asian ($n = 17$, 27.4%), and Arab or West Asian ($n = 9$, 14.5%). Civilian participants had an average age of 22.11 ($n = 61$, $SD = 5.28$), ranging from 17 to 59 years old.⁷ Almost all civilian participants reported residing in Canada ($n = 60$, 96.8%; the civilian sample was only made up of

⁶ Participants (both civilian and law enforcement) could select multiple ethnicities.

⁷ Data for the age of one participant was excluded due to its implausibility.

Carleton University students [Ottawa, ON, Canada]), and two participants reported another response ($n = 2, 3.2\%$).

Most of the law enforcement sample was male ($n = 55, 85.9\%$). The most frequently reported ethnicity in the law enforcement sample was Caucasian ($n = 59, 92.2\%$); the next most frequently reported ethnicities were Black or African American ($n = 3, 4.7\%$) and Hispanic ($n = 3, 4.7\%$). Law enforcement participants had an average age of 35.95⁸ ($n = 64, SD = 11.16$), ranging from 20 to 59 years old. Law enforcement participants mostly reported residing in the United States ($n = 47, 73.4\%$); one quarter reported residing in Canada ($n = 16, 25.0\%$) and one participant reported another response ($n = 1, 1.6\%$). The average number of operational years served by the police participants was 7.92 ($n = 64, SD = 8.86$), ranging from 0 to 32 years of operational police experience.

Materials/Measures

Demographic questionnaire. When participants clicked on the link to the experimental package on Qualtrics, they were asked whether they were currently an operational police officer or a Carleton University student. Based on their response, they were directed to either the police demographic questionnaire (see Appendix A) or the civilian questionnaire (see Appendix B). In the police questionnaire, various demographic characteristics were assessed, including: country in which the participant resides, age, gender, ethnicity, years of police experience, and years of operational police experience. The questionnaire also assessed whether police participants thought they had “less” training, the “same amount of” training, or “more” training than other

⁸ The average age and gender split of our sample is similar to the demographic profile of Canadian police officers more generally (Conor, 2017).

police officers from their agency in various areas (e.g., ERT or SWAT training, active shooter training, de-escalation training, etc.). I also collected information on the participant's self-reported frequency of being involved in encounters with a person in which they were at risk of being physically harmed by that person, and the participant's self-reported ability to anticipate outcomes from interactions with members of the public.

The civilian demographic questionnaire was similar to that of the police questionnaire with a few modifications. First, if a participant selected that they were a Carleton University student, they were then asked if they had ever worked as a police, correctional, or military officer. If the participant responded "yes" to this question, they would be directed to the debriefing form, and therefore excluded from this study. I applied this exclusion criterion because I was looking for participants with no policing (or related) experience. Second, civilian participants were not asked how many years of police experience and years of operational police experience they had. Finally, civilian participants were asked whether they thought they had "less" training, the "same amount of" training, or "more" training *compared to police officers* (rather than police officers from their agency). The list of training options was the same in the civilian demographic questionnaire as in the police demographic questionnaire. All other questions were the same as in the police demographic questionnaire.

Instructions. Following the demographic questionnaire, participants read a set of instructions before beginning the prediction task. Participants were told that they would begin with four practice trials, which would vary in content and length, and would be randomised. They were also told that data from these trials would *not* be analysed. Participants were then told that following these practice trials, they would complete 16 test trials, which would also vary in content and length, and data from these test trials *would* be analysed. Participants were told that,

on each trial, they would be presented with a brief message indicating the nature of the interaction that would be displayed in the body-worn camera footage (which will be referred to as a “dispatch message”), and that, once they clicked the arrow to proceed to the next page, the video would begin to play automatically, and then would proceed to the prediction question once it was done playing.

Participants were informed that they would have 10 seconds to respond “yes” or “no” with their computer mouse to the prediction question: “Based on a balance of probabilities, do you believe that the subject will harm or attempt to harm the officer(s)?” I asked participants to respond as quickly, but as accurately as possible, and told participants that if they did not respond within 10 seconds they would automatically proceed to the next trial. Participants were told that if they responded to the prediction question within 10 seconds, they would be asked subsequent questions related to that video in the same order for each trial. Specifically, they were told to: (1) provide a confidence rating on a scale ranging from 0 (“not at all confident”) to 100 (“extremely confident”) regarding their prediction, (2) list the cues they used to make their prediction in a text box, (3) rate their familiarity with the type of encounter depicted in the footage on a scale from 0 (“not at all familiar”) to 100 (“extremely familiar”), (4) indicate whether they had previously seen the footage that was played in the trial (if answered “yes” or “maybe/not sure”, this trial would be marked as missing), and (5) indicate whether they had any technical difficulties viewing the video (if answered “yes” or “other”, this trial would be marked as missing).

Finally, I asked participants to ensure that the volume on their computers was turned on for the duration of the study and to take caution when adjusting their sound as some videos could be much louder than others.

Video stimuli. Real-life videos depicting an encounter between a police officer(s) and a civilian(s) were presented to participants online. These videos were obtained online, in the public domain. All videos were displayed from a first-person perspective (i.e., from the officer's body-worn camera footage). These videos depicted different types of encounters (e.g., disturbances, suspicious person, traffic stop, etc.). The length of the thin slices were either 10 seconds long, which is consistent with stimuli used in previous thin slice research (e.g., Benjamin & Shapiro, 2009; Fowler et al., 2009; Raab & Johnson, 2007), or 30 seconds long. Prior to each video playing, a brief written message was presented to participants to provide information on the nature of the situation the participants would be watching in the trial (e.g., "You have responded to a call for service; an unwanted person is being asked to leave a bar/restaurant."). These "dispatch" messages were written in collaboration with two SMEs (experienced police instructors). This aspect of the study was consistent with previous research (Suss & Ward, 2012, 2013), reflected real-world conditions where dispatch messages would be available to the responding officer, and provided some degree of context to the footage.

Each participant underwent four practice trials in random order, in which two 10-second and two 30-second videos were presented (two in which the subject did *not* harm or attempt to harm the officer[s], and two in which the subject *harmed* or *attempted to harm* the officer[s]). Data from these practice trials were not used in the analyses. The remaining 16 videos were used as test trials (and, therefore, used in the analyses) and were also presented in a randomised order. There were eight test trial videos in which the subject did *not* harm or attempt to harm the officer(s). Half of these videos were 10 seconds in length, whereas the other half were 30 seconds in length. The other eight test trial videos involved a subject that *harmed* or *attempted to*

harm the officer(s). Again, half of these videos were 10 seconds in length and the remainder were 30 seconds in length.

As discussed more thoroughly below, two SMEs – experienced police instructors – identified the optimal occlusion point in each video based on ecologically valid environmental cues that could theoretically be used to accurately predict the outcome of the interaction (loosely based on the methods used in Suss and Ward [2013]). The videos were then edited to stop at the point in the video that corresponded to the occlusion point identified by the SMEs (this procedure is described below). After the end of the video, participants were prompted with a series of questions (see below “Follow-up questions”). The SMEs also identified “ground truth” for each video (i.e., whether or not the subject harmed or attempted to harm the officer(s) in the video) in order to determine whether the predictions made by participants were correct (i.e., accurate) or incorrect (i.e., inaccurate).

Temporal occlusion paradigm. Loosely based on the protocol described in Suss and Ward (2013), two SMEs identified an occlusion point in each video based on available environmental, situational, and subject-related cues. The occlusion point represented the earliest point in the video at which there was enough information that should allow the participant to predict the outcome of the situation (Suss & Ward, 2012, 2013). Both SMEs came to a consensus on the occlusion point for each video. Videos were subsequently edited to display either 10 seconds preceding this occlusion point or 30 seconds preceding this occlusion point.

Follow-up questions. As discussed briefly above, participants were asked a prediction question following each video that sought to identify whether they could predict the outcome of the situation depicted in the thin slice (i.e., video) that was presented to them. The following question was asked: “Based on a balance of probabilities, do you believe that the subject will

harm or attempt to harm the officer(s)?" Participants had 10 seconds to click "yes" or "no" on the screen. This time limit was imposed to help ensure that participants responded quickly to the prediction questions. Thus, a countdown starting at 10 seconds was displayed on the screen for the prediction question in each trial. After 10 seconds, the survey would automatically proceed to the follow-up questions (if the participant provided a response to the prediction question) or to the following trial (if the participant did not provide a response to the prediction question in the allotted time).

Next, participants were asked to provide a confidence rating on a scale ranging from 0 ("not at all confident") to 100 ("extremely confident") with regard to their prediction, to provide a list of cues they used to make their prediction in a text box (e.g., subject characteristics, tactical considerations, environmental factors, postural/facial/vocal indicators, etc.), and to rate their familiarity with the type of encounter depicted in the footage on a scale from 0 ("not at all familiar") to 100 ("extremely familiar").

On each trial, I also asked participants to indicate whether they had previously seen the footage that was played and to indicate whether they had any technical difficulties viewing the video. If a participant indicated that they had or that they were not sure if they had previously seen the footage in a given trial, this trial was excluded (i.e., considered missing data). Similarly, if a participant indicated that they had technical difficulties viewing the video on a given trial, this trial was excluded (i.e., considered missing data).

Training scores. For questions regarding experience with training, participants had to select whether they thought they had "less" training, the "same amount of" training, or "more" training compared to police officers (for civilian participants) or other officers in their agency (for police participants) for a variety of training options (e.g., ERT or SWAT training, active

shooter training, de-escalation training). A score of “1” was given for the selection of “less” training, a score of “2” was given for the selection of “same amount of” training, and a score of “3” was given for the selection of “more” training. Thus, it was possible to calculate an aggregate training score for each participant (i.e., the higher the score, the more training an individual reported having, relative to other police officers [in general, for civilian participants, or within one’s agency, for police participants]). The aggregate training variable was treated as a continuous variable. I also specifically examined ERT or SWAT training.

Procedure

Participants accessed the study online through a link to Qualtrics. This link was provided by email or through social media (to recruit police officers) or through the SONA system (to recruit civilians, i.e., Carleton University students). Participants were first asked to complete an informed consent form (Appendix C). Participants were then asked whether they were currently an operational police officer or a current Carleton University student. Based on this response, participants were then asked to respond to one of two questionnaires (one for police officers and another for civilians) that assessed demographic information and various life experiences (described above). After completing the demographic questionnaire, participants were provided with the instructions, as outlined above. Recall that these instructions informed participants about completing four practice trials and 16 test trials, and provided an overview of the questions that would be asked (in addition to providing information related to the time-sensitive nature of the prediction question).

The practice trials presented two videos depicting a subject who would *not* harm or attempt to harm the officer(s) (one of 10 seconds in length and one of 30 seconds in length) and two videos depicted a subject who *would* harm or attempt to harm the officer(s) (one of 10

seconds in length and one of 30 seconds in length). These were presented in a randomised manner across participants to prevent order effects. The 16 test trials also included eight videos depicting a subject who would *not* harm or attempt to harm the officer(s) and eight videos depicted a subject who *would* harm or attempt to harm the officer(s), with half of each type of video being 10 seconds in length and the other half being 30 seconds in length. The test trials were also presented in a randomised order to prevent order effects. Please refer to Table 1 for a breakdown of each condition for the videos. Following each video in the practice and test phase, participants were asked to respond to the prediction question and the five follow-up questions described above. Following the completion of all test trials, participants were debriefed (see Appendix D) and thanked for their time. All participants who clicked on the “withdraw” option during the study were also debriefed and their data were not used in the analyses.

Table 1

The Four Conditions for All Trials Presented to Participants

	Did the subject harm or attempt to harm the officer(s)?	
	Yes	No
10-second videos	Practice trials: 1 Test trials: 4	Practice trials: 1 Test trials: 4
30-second videos	Practice trials: 1 Test trials: 4	Practice trials: 1 Test trials: 4

Analyses

Data was downloaded from Qualtrics using IBM SPSS Statistics, version 25. All analyses were conducted using this statistical package.

Demographic data. Descriptive statistics are presented for all participants’ demographic data. A Pearson Chi Square analysis was conducted to assess whether differences existed in gender across officers and civilians. Independent samples *t*-tests were used to examine whether differences existed between these two samples for all continuous demographic data (i.e., age,

self-reported frequency of being in situations in which the participant was at risk of being harmed by another individual, and self-reported ability to anticipate the outcomes of encounters with members of the public). A Pearson Chi Square analysis was conducted to assess whether differences existed between officers and civilians on prediction accuracy. Finally, I conducted independent samples *t*-tests to examine whether officers and civilians differed on familiarity with the type of encounter being presented on each trial (hereafter referred to as familiarity) and confidence in their prediction.

Hypothesis testing. To examine the hypotheses, I considered using multilevel models to account for the lack of independence between observations, given the repeated measures design of this study, with all trials being nested within participant clusters (Heck, Thomas, & Tabata, 2012; Sommet & Morselli, 2017). To determine whether using a multilevel procedure was appropriate, I examined the variability of intercepts across participants by calculating the intraclass correlation (ICC; Sommet & Morselli, 2017), which yielded a value of 0.0009. Given the negligible between participant variation this value suggested, I concluded, based on Sommet and Morselli's (2017) advice, that it was unnecessary to proceed with a multilevel procedure. Thus, to examine the first, second, and third hypotheses, I conducted binary logistic regressions, with prediction accuracy as the outcome variable. Predictor variables (i.e., aggregate score for training, ERT/SWAT training, years of operational police experience, and familiarity) were examined individually in separate models. I also explored whether confidence was associated with the odds of providing an accurate response by including it as a predictor in a separate model. All continuous variables (i.e., aggregate score for training, years of operational police experience, familiarity, and confidence) were grand mean centered.

I was also interested in examining the models described above, for Hypotheses 1 to 3, by trial type (i.e., whether the trial presented a video that was 10 seconds or a video that was 30 seconds in length). Thus, I also conducted these binary logistic regression analyses (described above) for 10-second trials and 30-second trials separately. In addition, I explored whether thin slice length was associated with the odds of providing an accurate response by including it as a predictor in a separate model.

To examine the fourth hypothesis, I conducted a multilevel linear regression analysis, with confidence as the outcome variable. Again, to determine whether using a multilevel procedure was appropriate, I examined the variability of intercepts across participants by calculating the ICC, which yielded a value of 0.988. I therefore concluded I should continue with a multilevel procedure (Sommet & Morselli, 2017). Predictor variables (i.e., aggregate score for training, ERT/SWAT training, years of experience, and familiarity) were examined individually in separate models. The aggregate score for training, as well as years of experience, were grand mean centered, given that these are “level 2” variables (i.e., participant-level variables), and familiarity was group (or cluster) mean centered, given that it is a “level 1” variable (i.e., trial-level variable; Enders & Tofighi, 2007).

Results

Demographic Data

A Pearson’s Chi Square test was conducted to determine whether the officer and civilian samples differed in their gender composition. This analysis revealed a significant difference ($\chi^2(1, n = 126) = 33.73, p < .001$). Specifically, as one might expect, the officer sample consisted mostly of males ($n = 55, 85.9\%$), whereas the civilian sample consisted mostly of females ($n = 40, 81.6\%$). There was also a different composition of reported ethnicities for officers and

civilians. Again, as expected, the vast majority of officers reported being Caucasian ($n = 59$, 92.2%), whereas only approximately half of the civilian sample reported being Caucasian ($n = 27$, 43.6%), with South or East Asian also frequently being reported by civilians ($n = 17$, 27.4%).

A series of independent samples t -tests were conducted to assess whether differences existed between officers and civilians on additional demographic variables. Levene's test for age was significant ($p < .001$) indicating heteroscedasticity between groups; thus, Welch's t -statistic was used for this variable. Unsurprisingly, a significant difference was observed between groups on age ($t(93.27) = 9.00$, $p < .001$), with officers being significantly older ($M = 35.95$, $SD = 11.16$) than civilians ($M = 21.87$, $SD = 5.58$).

Participants were asked to report the frequency, on a scale from 0 ("never") to 100 ("extremely often"), with which they have been involved in encounters with a person where they were at risk of being physically harmed by that person. Levene's test was significant ($p = .010$) indicating heteroscedasticity between groups; thus, Welch's t -statistic was used. Again, unsurprisingly, a significant difference was observed between groups ($t(105.18) = 3.95$, $p < .001$), with officers reporting significantly higher frequencies ($M = 42.07$, $SD = 28.89$) compared to civilians ($M = 22.91$, $SD = 22.31$).

Participants were also asked to rate their ability to anticipate outcomes of interactions with members of the public on a scale from 0 ("extremely poor") to 100 ("excellent"). Interestingly, no significant differences were observed on this measure, suggesting that officers self-report similar abilities in anticipating outcomes of interactions with members of the public ($M = 61.47$, $SD = 25.07$) as civilians ($M = 57.15$, $SD = 24.40$).

Descriptive Trial Data

A Pearson's Chi Square test was conducted to examine whether officers and civilians differed on prediction accuracy for each trial.⁹ A significant difference was observed ($\chi^2(1, n = 1698) = 23.21, p < .001$), such that officers were more frequently accurate in their predictions ($n = 552, 63.6\%$) compared to students ($n = 442, 52.0\%$) across trials. An independent samples t -test was conducted to examine whether differences existed between officers and students in terms of their reported familiarity on each trial. Levene's test for familiarity was significant ($p < .001$) indicating heteroscedasticity between groups; thus, Welch's t -statistic was used for this variable. A significant difference was observed ($t(1323.73) = 13.403, p < .001$), such that officers ($M = 53.41, SD = 31.00, n = 708$) had significantly higher familiarity scores across trials compared to civilians ($M = 32.80, SD = 25.16, n = 628$). Finally, an independent samples t -test was conducted to examine whether differences existed between officers and students regarding reported confidence on each trial. Levene's test for confidence was significant ($p = .001$) indicating heteroscedasticity between groups; thus, Welch's t -statistic was used for this variable. Again, a significant difference was observed ($t(1607.02) = 5.183, p < .001$), such that officers ($M = 61.68, SD = 22.07, n = 825$) reported significantly higher confidence in their predictions across trials compared to civilians ($M = 55.69, SD = 24.622, n = 810$).

⁹ Hereafter, n values reflect the number of trials that were completed overall, rather than the number of participants.

Prediction Response Accuracy

Approximately half of the trials presenting videos in which the subject harmed or attempted to harm the officer were responded to correctly (i.e., the “hit”¹⁰ rate was 52.0%), and over half the trials presenting videos in which the subject *did not* harm or attempt to harm the officer(s) were responded to correctly (i.e., the “correct rejection”¹¹ rate was 63.7%). It should be noted that, when examining all trials, civilians were accurate on 52.0% of trials, and officers were accurate on 63.6% of trials. Overall, over half of participants’ responses (58.0%) were accurate (see Table 2).

Table 2

Breakdown of Participants’ Prediction Responses Based on Ground Truth

			Ground Truth		
			Yes	No	Total
			Count (%)	Count (%)	Count
Prediction: Based on a balance of probabilities, will the subject harm or attempt to harm the officer(s)?	Yes	Civilian	166 (39.9%)	148 (35.7%)	314 (37.8%)
		Officer	268 (64.0%)	165 (36.7%)	433 (49.9%)
		Total	434 (52.0%)	313 (36.3%)	747 (44.0%)
	No	Civilian	250 (60.1%)	266 (64.3%)	516 (62.2%)
		Officer	151 (36.0%)	284 (63.3%)	435 (50.1%)
		Total	401 (48.0%)	550 (63.7%)	951 (56.0%)
Total	Civilian	416 (49.8%)	414 (48.0%)	830 (48.9%)	
	Officer	419 (50.2%)	449 (52.0%)	868 (51.1%)	
	Total	835 (100%)	863 (100%)	1698 (100%)	

¹⁰ A “hit” corresponds to a trial in which the participant responded “yes” when there was indeed harm or attempted harm by the subject in the video. The hit rate is calculated as follows: hits/(hits + misses). Misses are when a participant *incorrectly* identifies that subject *would not* harm or attempt to harm the officer(s). The miss rate is calculated as follows: 1- hit rate.

¹¹ A “correct rejection” corresponds to a trial in which the participant responded “no” when, in fact, there was no harm or attempted harm by the subject in the video. The correct rejection rate is calculated as follows: 1 – false alarm rate. False alarms are when a participant *incorrectly* identifies that a subject *would* harm or attempt to harm the officer(s). The false alarm rate is calculated as follows: false alarms/(false alarms + correct rejections).

Regression Analyses: Prediction Accuracy

For exploratory purposes, I conducted a series of Pearson’s correlations to examine the relationship between the continuous predictor variables that were included in the upcoming regression models. As can be seen in Table 3, years of operational police service was significantly correlated with measures of training, as well as familiarity, and confidence (to a lesser degree). The aggregate training score was significantly correlated with familiarity. The aggregate training score was also correlated with confidence, but to a lesser degree. Finally, familiarity and confidence were also significantly correlated. Each of these predictors were individually examined within the regression models.

Table 3

Correlation Matrix for All Independent (Predictor) Variables

Independent variables (predictors)	Years of operational police service	Training (aggregate)	Familiarity	Confidence
Years of operational police service	1 (n = 1024)	.534** (n = 688)	.532** (n = 708)	.275** (n = 825)
Training (aggregate)	–	1 (n = 1376)	.503** (n = 882)	.131** (n = 1102)
Familiarity		–	1 (n = 1336)	.340** (n = 1318)
Confidence			–	1 (n = 1635)

Note. ‘***’ corresponds to significance at $p < .01$ (two-tailed).

I collected data using a repeated measures design (i.e., each participant responded to multiple trials), which violated the assumption of independence of observations required to conduct traditional logistic regression analyses. Therefore, I first examined whether multilevel logistic regression was necessary by examining prediction accuracy on each trial, nested within each participant. I followed the procedures stipulated by Sommet and Morselli (2017), beginning with the examination of the unconditional mean model (i.e., null model) to determine whether

intercepts varied between participants. This model demonstrated that intercepts did not significantly vary between participants ($B = 0.003, p = .928$). I also saved the estimated intercepts for each participant as predicted probabilities. Upon visual inspection, the predicted probabilities were observed to remain fairly constant across participants, indicating that the intercepts do not appear to be random.

As a further test of whether multilevel regression modelling was necessary, I examined the variability of intercepts across participants by calculating the ICC, as recommended by Sommet and Morselli (2017). This was done to investigate the proportion of between participant variability relative to the total amount of variability. I obtained an ICC of 0.0009, suggesting that 0.9% of the odds of providing an accurate response was explained by between participant differences. In other words, this ICC value suggested that the odds of providing an accurate response on the trials did not differ between participants (i.e., there was negligible between participant variation), and almost all of the variation in accuracy could be explained by within participant differences. Given this negligible between participant variation, it was unnecessary to proceed with a multilevel procedure (Sommet & Morselli, 2017).

As such, I conducted a series of binary logistic regressions to determine whether training, experience, familiarity, and length of the thin slice could predict the odds of providing an accurate response to the question: “Based on a balance of probabilities, do you believe that the subject will harm or attempt to harm the officer(s)?” The assumption of linearity between continuous independent variables and the logarithmic odds was tested using the Hosmer and Lemeshow test for goodness of fit, as well as the methods described in Hosmer and Lemeshow (1989; as cited by Field, 2013). The Hosmer and Lemeshow test for goodness of fit was non-significant ($p > .05$) for all models with continuous predictors. In addition, the interaction

between each independent variable and its logarithmic transformation when included in the regression model (Hosmer & Lemeshow, 1989, as cited by Field, 2013) was non-significant ($p > .05$). Thus, I concluded that my data met this assumption.

Officers vs. civilians. Before testing the formal hypotheses, an exploratory binary logistic regression analysis was conducted to examine whether being a law enforcement officer (compared to being a civilian) was related to the odds of providing an accurate response. The model was significant ($\chi^2(1, n = 1698) = 23.26, p < .001$); specifically, the odds that officers provided an accurate response increased by 60.9% compared to civilians (see Table 4). The -2 log likelihood of this model was 2287.55.

Table 4

Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer) Predicting Accuracy (with Civilian as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for	
							Exp(B)	
							Lower	Upper
Officer	0.476	0.099	23.095	1	0.000	1.609	1.325	1.954
Constant	0.082	0.069	1.392	1	0.238	1.085		

When examining trials in which participants were presented with *short* thin slices (i.e., 10 second videos), this model remained significant ($\chi^2(1, n = 837) = 17.29, p < .001$). In fact, the size of the effect became larger. Specifically, on short trials, the odds that officers provided an accurate response increased by 78.9% compared to civilians (see Table 5). The -2 log likelihood of this model was 1,132.24.

When examining trials in which participants were presented with *long* thin slices (i.e., 30 second videos), this model also remained significant ($\chi^2(1, n = 861) = 7.07, p = .008$). However, the size of the effect was smaller. Specifically, on long trials, the odds that officers provided an

accurate response increased by 44.9% compared to civilians (see Table 6). The -2 log likelihood of this model was 1,150.71.

Table 5

Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer) Predicting Accuracy for Short Thin Slices (with Civilian as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Officer	0.582	0.141	17.105	1	0.000	1.789	1.358	2.357
Constant	-0.063	0.099	0.411	1	0.521	0.939		

Table 6

Binary Logistic Regression Results of Participant Type (i.e., Civilian vs. Officer) Predicting Accuracy for Long Thin Slices (with Civilian as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Officer	0.371	0.140	7.040	1	0.008	1.449	1.102	1.906
Constant	0.225	0.098	5.250	1	0.022	1.253		

Training. To investigate Hypothesis 1, a binary logistic regression analysis was conducted to examine whether aggregate training was related to the odds of providing an accurate response (recall, that a higher aggregate score indicated a greater degree of specialized training). The model was significant ($\chi^2(1, n = 1144) = 10.00, p = .002$). Overall, the more training one had, the better one was at making accurate predictions. Specifically, for every additional unit increase in grand-mean-centered aggregate training, the odds of making an accurate prediction increased by 2.2% (see Table 7). The -2 log likelihood was 1,552.91.

Table 7

Binary Logistic Regression Results of Specialized Training Predicting Accuracy

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Exp(B)	
							Lower	Upper
Training (aggregate)	0.022	0.007	9.813	1	0.002	1.022	1.008	1.036
Constant	0.323	0.061	27.675	1	0.000	1.382		

Note. "Training (aggregate)" was grand mean centered.

When specifically examining trials in which participants were presented with *short* thin slices, this model remained significant ($\chi^2(1, n = 565) = 7.03, p = .008$) and the effect size remained similar: for every unit increase in grand-mean-centered aggregate training, the odds of making an accurate prediction on short trials increased by 2.6% (see Table 8). The -2 log likelihood of this model was 771.25.

When specifically examining trials in which participants were presented with *long* thin slices (i.e., 30 second videos), this model was no longer significant ($\chi^2(1, n = 579) = 3.37, p = .067$; see Table 9). The -2 log likelihood of this model was 778.66.

Table 8

Binary Logistic Regression Results of Specialized Training Predicting Accuracy for Short Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Exp(B)	
							Lower	Upper
Training (aggregate)	0.025	0.010	6.883	1	0.009	1.026	1.006	1.045
Constant	0.231	0.087	7.083	1	0.008	1.260		

Note. "Training (aggregate)" was grand mean centered.

Table 9

Binary Logistic Regression Results of Specialized Training Predicting Accuracy for Long Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Training (aggregate)	0.018	0.010	3.313	1	0.069	1.018	0.999	1.038
Constant	0.414	0.087	22.564	1	0.000	1.512		

Note. "Training (aggregate)" was grand mean centered.

It should be noted that some civilian participants reported that they had "the same amount of" training as police officers, or even "more" training than police officers, for a variety of training options. Given that it seems very unlikely that civilians would have had police-specific training (e.g., ERT/SWAT training, active shooter training, instructor training in use-of-force/firearms, crisis negotiation training, police defensive tactics training, scenario-based training for use-of-force, and simulator training for use-of-force), the analysis presented below was based on recoded civilian data where responses of having "more" training or "the same amount of" training for these specific training options were recoded to indicate "less" training.

In general, results from these analyses were similar to those reported above. The overall model remained significant ($\chi^2(1, n = 1698) = 20.48, p < .001$), with every additional unit in grand-mean-centered training being associated with a 2.6% increase in the odds of making an accurate prediction. The -2 log likelihood for this model was 2,290.33.

When examining *short* thin slices, the recoded model was also significant ($\chi^2(1, n = 837) = 17.28, p < .001$), and was associated with a similar effect size (approximately 1% higher). Specifically, with every additional unit in grand-mean-centered training, the odds of making an accurate prediction response on short thin slices increased by 3.4%. The -2 log likelihood for this model was 1,132.24.

On the other hand, I obtained a different result when examining *long* thin slices. This recoded model was now significant ($\chi^2(1, n = 861) = 5.07, p = .024$). However, the effect size was similar to the effect size that was obtained prior to having recoded the data. For every additional unit in grand-mean-centered training, the odds of making an accurate prediction response increased by 1.8%. The -2 log likelihood for this model was 1,152.71. Findings based on the recoded data are presented in Appendix E.

ERT/SWAT training. As a further test of Hypothesis 1, a binary logistic regression was also conducted to examine whether ERT or SWAT training, specifically, predicted response accuracy. However, given that some of the civilian responses indicated they had “more” ERT/SWAT training or “the same amount of” ERT/SWAT training compared to police officers (as described above), I was concerned with the validity of these responses. As such, I categorised training responses for this variable as follows: (1) *civilians*, regardless of whether they had selected “less,” “the same amount of,” or “more” ERT/SWAT training compared to a police officer, (2) *police officers* who had selected “less” or “the same amount of” ERT/SWAT training compared to officers within their agency, and (3) *ERT/SWAT officers* who had selected that they had “more” ERT/SWAT training compared to officers from their agency.

As depicted in Table 10, ERT/SWAT officers were accurate most frequently (67.6% of trials), followed by police officers (62.0% of trials), and civilians (52.0% of trials). The binary logistic regression model was significant ($\chi^2(2, n = 1698) = 25.75, p < .001$). Compared to civilians, the odds that police officers provided an accurate response increased by 50.0%, and the odds that ERT/SWAT officers provided an accurate response increased by 92.1% (see Table 11). No significant differences were observed between police officers and ERT/SWAT officers (see Table 12). The -2 log likelihood of this model was 2,285.07.

Table 10

Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers

	Not accurate	Accurate	Total
Civilian	398 (48.0%)	432 (52.0%)	830
Police Officer	234 (38.0%)	381 (62.0%)	615
ERT/SWAT Officer	82 (32.4%)	171 (67.6%)	253
Total	714 (42.0%)	984 (58.0%)	1698

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 11

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy (with Civilians as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Lower	Upper
Civilian (reference group)			25.335	2	0.000			
Police officer	0.406	0.108	14.024	1	0.000	1.500	1.213	1.855
ERT/SWAT officer	0.653	0.151	18.643	1	0.000	1.921	1.428	2.584
Constant	0.082	0.069	1.392	1	0.238	1.085		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 12

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy (with ERT/SWAT Officers as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Lower	Upper
ERT/SWAT officers (reference group)			25.335	2	0.000			
Civilian	-0.653	0.151	18.643	1	0.000	0.520	0.387	0.700
Police officer	-0.247	0.158	2.455	1	0.117	0.781	0.573	1.064
Constant	0.735	0.134	29.936	1	0.000	2.085		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

When specifically examining trials in which participants were presented with *short* thin slices, ERT/SWAT officers’ accuracy was maintained (67.7% of trials) and police officers’ accuracy decreased by just under 2% (60.6% of trials). Civilians’ accuracy also decreased by just under 4% (48.4% of trials; see Table 13). The regression model remained significant ($\chi^2(2, n = 837) = 19.23, p < .001$). Compared to civilians, the odds that police officers provided an accurate response increased by 63.8%, and the odds that ERT/SWAT officers provided an accurate response increased by 123.7% on short trials (see Table 14). Still, no significant differences were observed between police officers and ERT/SWAT officers on short trials (see Table 15). The -2 log likelihood of this model was 1,130.29.

Table 13

Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers for Short Thin Slices

	Not accurate	Accurate	Total
Civilian	212 (51.6%)	199 (48.4%)	411
Police Officer	119 (39.4%)	183 (60.6%)	302
ERT/SWAT Officer	40 (32.3%)	84 (67.7%)	124
Total	371 (44.3%)	466 (55.7%)	837

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 14

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Short Thin Slices (with Civilians as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Lower	Upper
Civilian (reference group)			18.810	2	0.000			
Police officer	0.494	0.154	10.321	1	0.001	1.638	1.212	2.214
ERT/SWAT officer	0.805	0.216	13.900	1	0.000	2.237	1.465	3.416
Constant	-0.063	0.099	0.411	1	0.521	0.939		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 15

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Short Thin Slices (with ERT/SWAT Officers as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for	
							Exp(B)	
							Lower	Upper
ERT/SWAT officers (reference group)			18.810	2	0.000			
Civilians	-0.805	0.216	13.900	1	0.000	0.447	0.293	0.683
Police officers	-0.312	0.225	1.912	1	0.167	0.732	0.471	1.139
Constant	0.742	0.192	14.916	1	0.000	2.100		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

When specifically examining trials in which participants were presented with *long* thin slices, ERT/SWAT officers’ accuracy was again maintained (67.4% of trials). On the other hand, police officers’ accuracy increased by almost 3% (63.3% of trials) and civilians’ accuracy increased by approximately 5% (55.6% of trials; see Table 16). This regression model also remained significant ($\chi^2 (2, n = 861) = 7.77, p = .021$), however the effects were smaller. Specifically, compared to civilians, the odds that police officers provided an accurate response increased by 37.4%, and the odds that ERT/SWAT officers provided an accurate response increased by 65.4% on long trials (see Table 17). Again, no significant differences were observed between police officers and ERT/SWAT officers on long trials (see Table 18). The -2 log likelihood of this model was 1,150.01.

Table 16

Accuracy Rates for Civilians, Police Officers, and ERT/SWAT Officers for Long Thin Slices

	Not accurate	Accurate	Total
Civilians	186 (44.4%)	233 (55.6%)	419
Police Officer	115 (36.7%)	198 (63.3%)	313
ERT/SWAT Officer	42 (32.6%)	87 (67.4%)	129
	343 (39.8%)	518 (60.2%)	861

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 17

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Long Thin Slices (with Civilians as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Civilian (reference group)			7.690	2	0.021			
Police officer	0.318	0.153	4.320	1	0.038	1.374	1.018	1.855
ERT/SWAT officer	0.503	0.212	5.625	1	0.018	1.654	1.091	2.506
Constant	0.225	0.098	5.250	1	0.022	1.253		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 18

Binary Logistic Regression Results of ERT/SWAT Training Predicting Accuracy for Long Thin Slices (with ERT/SWAT Officers as the Reference Group)

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
ERT/SWAT officers (reference group)			7.690	2	0.021			
Civilians	-0.503	0.212	5.625	1	0.018	0.605	0.399	0.916
Police officers	-0.185	0.221	0.697	1	0.404	0.831	0.538	1.283
Constant	0.728	0.188	15.022	1	0.000	2.071		

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Years of operational police experience. To investigate Hypothesis 2, a binary logistic regression analysis was conducted to examine whether years of operational police experience was related to the odds of providing an accurate response (only officers were examined in this analysis to avoid skewing the data, given that all civilians would have zero years of operational police experience). The model approached significance ($\chi^2(1, n = 868) = 3.56, p = .059$) and therefore these results should be interpreted with caution. The analysis suggested that for every additional year of grand-mean-centered years of operational police service, the odds of making an accurate prediction increased by 1.5%. Thus, for every additional five years of grand-mean-centered years of operational police service, the odds of making an accurate prediction increased by 7.5% (see Table 19). The -2 log likelihood of this model was 1,134.76.

Table 19

Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Years of operational police service	0.015	0.008	3.492	1	0.062	1.015	0.999	1.032
Constant	0.563	0.071	63.170	1	0.000	1.756		

Note. “Years of operational police service” was grand mean centered.

When specifically examining trials in which participants were presented with *short* thin slices, the regression model became significant ($\chi^2(1, n = 426) = 6.81, p = .009$). Specifically, for every additional year of grand-mean-centered years of operational police service, the odds of making an accurate prediction on a short trial increased by 3.1%. Thus, for every additional five years of grand-mean-centered years of experience, the odds of making an accurate prediction increased by 15.4% (see Table 20). The -2 log likelihood of this model was 556.07.

When specifically examining trials in which participants were presented with *long thin* slices, this model remained non-significant ($\chi^2(1, n = 442) = 0.008, p = .931$; see Table 21). The -2 log likelihood of this model was 575.13.

Table 20

Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy for Short Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Years of operational police service	0.030	0.012	6.472	1	0.011	1.031	1.007	1.055
Constant	0.534	0.102	27.627	1	0.000	1.705		

Note. “Years of operational police service” was grand mean centered.

Table 21

Binary Logistic Regression Results of Years of Operational Police Experience Predicting Accuracy for Long Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Years of operational police service	0.001	0.011	0.008	1	0.931	1.001	0.979	1.023
Constant	0.596	0.099	35.994	1	0.000	1.816		

Note. “Years of operational police service” was grand mean centered.

Familiarity. To investigate Hypothesis 3, a binary logistic regression was conducted to examine whether familiarity with the type of encounter being presented was related to the odds of providing an accurate response. The regression model was significant ($\chi^2(1, n = 1327) = 9.26, p = .002$). The more familiar one reported being with the type of encounter, on a scale from 0 to 100, the greater the odds they would provide an accurate prediction response. Specifically, for every 10 unit increase in grand-mean-centered familiarity, the odds of providing an accurate response increased by 5.7% (see Table 22). The -2 log likelihood of this model was 1,784.11.

Table 22

Binary Logistic Regression Results of Familiarity on Each Trial Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Familiarity	0.006	0.002	9.200	1	0.002	1.006	1.002	1.009
Constant	0.379	0.056	45.683	1	0.000	1.461		

Note. "Familiarity" was grand mean centered.

When specifically examining trials in which participants were presented with *short* thin slices, the regression model remained significant ($\chi^2(1, n = 656) = 10.81, p = .001$). In fact, the effect became slightly more pronounced: for every 10 unit increase in grand-mean-centered familiarity, the odds of providing an accurate response increased by 8.8% (see Table 23). The -2 log likelihood of this model was 881.40.

When specifically examining trials in which participants were presented with *long* thin slices, the regression model was no longer significant ($\chi^2(1, n = 671) = 1.06, p = .304$; see Table 24). The -2 log likelihood of this model was 899.30.

Table 23

Binary Logistic Regression Results of Familiarity on Each Trial Presenting a Short Thin Slice in Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Familiarity	0.009	0.003	10.641	1	0.001	1.009	1.003	1.014
Constant	-0.049	0.139	0.125	1	0.724	0.952		

Note. "Familiarity" was grand mean centered.

Table 24

Binary Logistic Regression Results of Familiarity on Each Trial Presenting a Long Thin Slice in Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Familiarity	0.003	0.003	1.056	1	0.304	1.003	0.998	1.008
Constant	0.310	0.138	5.033	1	0.025	1.363		

Note. "Familiarity" was grand mean centered.

Length of the thin slice. For exploratory purposes, a binary logistic regression analysis was conducted to examine whether the length of the thin slice (30 seconds, compared to 10 seconds) was related to the odds of providing an accurate response. The regression model approached significance ($\chi^2(1, n = 1698) = 3.508, p = .061$); these results should therefore be interpreted with caution. These results suggest that the odds of participants providing an accurate response increased 20.2% on 30-second thin slice trials, compared to 10-second thin slice trials (see Table 25). The -2 log likelihood of this model was 2,307.30.

Table 25

Binary Logistic Regression Results of The Length of The Thin Slice Presented in Each Trial Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
30-second thin slice	0.184	0.098	3.505	1	0.061	1.202	0.991	1.458
Constant	0.228	0.070	10.736	1	0.001	1.256		

Note. 10-second thin slice is the reference group.

Confidence. To investigate the Research Question, a binary logistic regression analysis was conducted to examine whether confidence in one’s prediction was associated with the odds of making an accurate prediction. This model was significant ($\chi^2(1, n = 1625) = 11.22, p = .001$). Individuals reporting higher levels of confidence in their prediction, on a scale ranging from 0 to

100, had greater odds of making an accurate prediction. Specifically, for every 10 unit increase in grand-mean-centered confidence, the odds of providing an accurate response increased by 7.2% (see Table 26). The -2 log likelihood of this model was 2,195.43.

Table 26

Binary Logistic Regression Results of Confidence on Each Trial Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Confidence	0.007	0.002	11.155	1	0.001	1.007	1.003	1.011
Constant	0.291	0.052	30.908	1	0.000	1.338		

Note. "Confidence" was grand mean centered.

When specifically examining trials in which participants were presented with *short* thin slices, the regression model remained significant ($\chi^2(1, n = 800) = 12.90, p < .001$). The effect size became slightly more pronounced: for every 10 unit increase in grand-mean-centered confidence, the odds of providing an accurate response increased by 10.9% (see Table 27). The -2 log likelihood of this model was 1,083.09.

When specifically examining trials in which participants were presented with *long* thin slices, the regression model was no longer significant ($X(1, n = 825) = 1.27, p = .260$; see Table 28). The -2 log likelihood of this model was 1,106.73.

Table 27

Binary Logistic Regression Results of Confidence on Each Trial Presenting Short Thin Slices Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Confidence	0.011	0.003	12.691	1	0.000	1.011	1.005	1.017
Constant	0.184	0.074	6.123	1	0.013	1.202		

Note. "Confidence" was grand mean centered.

Table 28

Binary Logistic Regression Results of Confidence on Each Trial Presenting Long Thin Slices Predicting Accuracy on That Trial

	B	S.E.	Wald	df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower	Upper
Confidence	0.003	0.003	1.266	1	0.260	1.003	0.997	1.009
Constant	0.397	0.074	28.705	1	0.000	1.487		

Note. “Confidence” was grand mean centered.

Regression Analyses: Confidence

As stated in the previous section, I collected data using a repeated measures design. The data therefore violated the assumption of independence of observations required to conduct traditional linear regression analysis. As described above, I examined whether multilevel linear regression was necessary by examining confidence ratings for each trial nested within each participant. I followed the procedures stipulated by Sommet and Morselli (2017), beginning with the examination of the unconditional mean model (i.e., null model) to determine whether intercepts varied between participants, with confidence as the outcome variable.

Unlike the analysis with accuracy as the outcome variable, this null model (with confidence as the outcome variable) demonstrated that intercepts significantly varied between participants ($B = 275.51, p < .001$). To examine variability of intercepts across participants, I also calculated the ICC, for which I obtained a value of 0.988. In contrast to the accuracy analyses, this suggested that 98.8% of the variability in confidence ratings could be explained by between participant differences, and only a negligible amount of this variation could be explained by within participant differences. I therefore proceeded with a multilevel linear regression procedure (Heck et al., 2012; Sommet & Morselli, 2017).

The assumptions underlying multilevel linear regression were examined. The assumption of linearity between the dependent variable and the independent variables was assessed visually

with a scatter plot. Visual inspection suggested a linear association between the variables. The assumption of normally distributed residuals was examined by visual inspection of normal probability plots for each of the independent variables (i.e., aggregate training, years of operational police experience, and familiarity). This assumption was satisfied for all the independent variables.

Training. To investigate Hypothesis 4, I conducted a multilevel linear regression analysis to examine whether the amount of training one had (in aggregate) could predict one’s confidence ratings. The aggregate score for training was not significant (see Table 29). The -2 log likelihood was 9599.23.

Table 29

Multilevel Linear Regression Results of Training Predicting Confidence on Each Trial

	Estimate	Std. Error	Sig.	95% CI	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	58.97	1.70	0.000	55.58	62.37
Training (aggregate) (B ₂₀)	0.33	0.19	0.080	-0.04	0.71
Random effects					
Residual	300.61	13.31	0.000	275.62	327.87
Intercept [subject = id]	204.37	36.60	0.000	143.88	290.29

Note. “Training (aggregate)” was grand mean centered (given that it is a level-2 variable, i.e., a participant-level variable).

As described above, some civilian responses appeared unlikely (i.e., some reported having the “same amount of” or “more” training on police-specific training options). As such, I recoded some of this data to determine if the models differed (specifically, I recoded “more” training or the “same amount of” training for police-specific training options to “less training” compared to other officers, as described above). I conducted a multilevel linear regression analysis to examine the relationship between training and confidence with the recoded data. I found that the relationship became significant ($p = .001$), and the slope was slightly larger

suggesting a steeper increase in confidence associated with a unit increase in training.

Specifically, for every unit increase in grand-mean-centered aggregate training, confidence was expected to increase by 0.57 units (please refer to Appendix F for these results).

ERT/SWAT training. To also examine Hypothesis 4, I conducted a multilevel linear regression analysis to examine whether ERT/SWAT training could predict one's confidence ratings. As described above, because of concerns regarding the validity of some civilian responses to the police-specific training options, I categorised training responses for this variable as: (1) *civilians*, regardless of whether they had selected "less," "the same amount of," or "more" ERT/SWAT training compared to a police officer, (2) *police officers* who had selected "less" or "the same amount of" ERT/SWAT training compared to officers from their agency, and (3) *ERT/SWAT officers* who had selected that they had "more" ERT/SWAT training compared to officers from their agency.

Compared to civilians, being an ERT/SWAT officer was associated with an expected 9.21 unit increase in confidence ($p = .046$; see Table 30). No significant differences were observed between police officers and civilians ($p = .179$; see Table 30) or between police officers and ERT/SWAT officers with regard to confidence ($p = .328$; see Table 31). The -2 log likelihood of this model was 14,227.11.

Table 30

Multilevel Linear Regression Results of ERT/SWAT Training Predicting Confidence on Each Trial (With Civilians as the Reference Group)

	Estimate	Std. Error	Sig.	95% CI	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	55.00	2.27	0.000	50.51	59.49
ERT/SWAT officers (B ₂₁)	9.21	4.57	0.046	0.15	18.26
Police officers (B ₂₂)	4.59	3.39	0.179	-2.13	11.30
Civilians (reference group; B ₂₃)	0 ^b	0			
Random effects					
Residual	292.53	10.63	0.000	272.42	314.12
Intercept [subject = id]	261.72	37.35	0.000	197.85	346.19

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Table 31

Multilevel Linear Regression Results of ERT/SWAT Training Predicting Confidence on Each Trial (With ERT/SWAT Officers as the Reference Group)

	Estimate	Std. Error	Sig.	95% CI	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	64.21	3.97	0.000	56.34	72.07
Civilians (B ₂₁)	-9.21	4.57	0.046	-18.26	-0.15
Police officers (B ₂₂)	-4.62	4.70	0.328	-13.93	4.70
ERT/SWAT officers (reference group; B ₂₃)	0 ^b	0.00			
Random effects					
Residual	292.53	10.63	0.000	272.42	314.12
Intercept [subject = id]	261.72	37.35	0.000	197.85	346.19

Note. Civilians included all civilians, regardless of the level of training they reported for ERT/SWAT training. Police officers included all officers who reported having “less” or “the same amount of” ERT/SWAT training compared to officers in their agency. ERT/SWAT officers reported having “more” ERT/SWAT training compared to officers in their agency.

Years of operational police experience. To further examine Hypothesis 4, a multilevel linear regression analysis was also conducted to examine whether years of operational police

experience can predict one’s confidence in their outcome prediction. This model demonstrated that as one’s operational years of service increased, their level of confidence was also expected to increase. Specifically, for every additional year of grand-mean-centered years of operational police service, confidence was expected to increase by 0.65 units (see Table 32). The -2 log likelihood was 7,051.93.

Table 32

Multilevel Regression Results of Years of Operational Police Experience Predicting Confidence on Each Trial

	Estimate	Std. Error	Sig.	95% CI	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	60.87	1.91	0.000	57.05	64.70
Years of operational police service (B ₂₀)	0.65	0.22	0.004	0.22	1.08
Random effects					
Residual	250.48	12.86	0.000	226.51	277.00
Intercept [subject = id]	210.19	42.17	0.000	141.85	311.46

Note. “Years of operational police service” was grand mean centered (given that it is a level-2 variable, i.e., a participant-level variable).

Familiarity. As a final test of Hypothesis 4, a multilevel linear regression analysis was conducted to examine whether familiarity on each trial predicted one’s confidence in their prediction on that trial. The -2 log likelihood was 11,378.25. This model demonstrated that as one’s familiarity with the type of scenario being presented on that trial increased, one’s confidence in their response on that trial was also expected to increase. Specifically, for every additional unit of group-mean-centered (i.e., participant-centered) familiarity, confidence was expected to increase by 0.23 units (see Table 33).

Table 33

Multilevel Linear Regression Results of Familiarity on Each Trial Predicting Confidence on That Trial

	Estimate	Std. Error	Sig.	95% CI	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	58.53	1.70	0.000	55.17	61.90
Familiarity (B ₁₀)	0.23	0.03	0.000	0.17	0.29
Random effects					
Residual	267.92	10.92	0.000	247.35	290.21
Intercept [subject = id]	278.00	43.60	0.000	204.43	378.06

Note. “Familiarity” was *group* mean centered (given that it is a level-1 variable, i.e., a trial-level variable).

Discussion

The present study sought to examine how well police officers (with varying levels of expertise and training) and civilians could predict the outcome of a police-public encounter when viewing a thin slice of the encounter under varying conditions (e.g., when encounter familiarity, confidence in prediction, and thin slice length varies). In this section, the various findings related to these issues will be discussed, and study limitations and directions for future research will be described.

Before discussing the main findings related to the current study, which relate to predictors of accuracy and confidence when anticipating the outcome of thin slices, I would first like to briefly discuss some of the more general findings related to the accuracy results obtained for the police officers and civilian participants I sampled. Interestingly, the civilians who took part in the current study exhibited significantly lower levels of accuracy (52.0%) compared to police officers (63.6%) when predicting whether the subject seen in the thin slices would harm or attempt to harm the officer(s) depicted in the videos. This was despite the fact that officers and civilians expressed similar self-perceived abilities for anticipating outcomes of interactions with members of the public. Evidently, these perceptions appear to be somewhat flawed given that

officers were capable of making accurate predictions more frequently than civilians, and civilians demonstrated accuracy at around chance-levels in the present study.

It is also interesting to note the distribution of different types of predictions across participant groups. For example, civilians more frequently predicted that they did *not* believe the subject would harm the officer(s) (62.2%), whereas officers made approximately equal predictions of impending harm (49.9%) and no harm (50.1%). With regard to accuracy, both civilians and officers made “correct rejections” (i.e., correctly identifying that the subject would *not* harm or attempt to harm the officer[s]) in approximately two thirds of “no harm” trials¹², suggesting that both officers and civilians possess similar abilities to correctly identify when subjects will not harm or attempt to harm the officer(s) in the videos. In contrast, officers made “hits” (i.e., correctly identifying that a subject *would* harm or attempt to harm the officer[s]) on approximately 24% more “harm” trials¹³, compared to civilians, suggesting that officers might be better able to anticipate impending harm compared to civilians. Presumably this difference reflects the fact that officers are more frequently involved in (or exposed to) encounters in which they are at risk of being physically harmed (as was demonstrated in this study through self-report). This likely increases an officer’s ability to identify cues associated with harm in such situations (as was demonstrated in, e.g., Johnson’s [2015, 2019] studies) and increases vigilance to behavioural indicators that one believes are likely to lead to violence (Riggs, Rothbaum, & Foa, 1995).

¹² Trials in which the subject *did not* harm or attempt to harm the officer(s) in the video.

¹³ Trials in which the subject harmed or attempted to harm the officer(s) in the video.

Predictors of Accuracy

In previous research, expertise has been associated with an ability to make accurate predictions based on thin slices of information (e.g., Ward et al., 2013). In law enforcement, expertise is likely developed through various training opportunities and field experience (Ericsson et al., 1993). Supporting this, and concurring with my hypotheses, the amount of police training one reported having in the current study was associated with greater odds of participants providing an accurate response when predicting whether the subjects in the thin slices they viewed would harm or attempt to harm the officer involved in the encounter. I also found that officers who reported having “more” ERT/SWAT training had significantly greater odds of providing an accurate response (on 67.6% of trials) than those having “the same amount of” or “less” ERT/SWAT training than other police officers (on 62.0% of trials) and civilians (on 52.0% of trials). Increased accuracy associated with having SWAT training has also been supported by previous research examining outcome predictions based on thin slices of police-public encounters (Suss & Ward, 2012). It will be valuable in future research to explore the potentially important role that other forms of specific police training play in allowing officers to make accurate outcome predictions based on thin slices of information, such as de-escalation training or training related to interactions with subjects who are perceived to have a mental illness.

In the present study, familiarity was highly (positively) correlated with both police training and years of operational police service. In other words, the longer one had been an operational police officer and the more police training one possessed, the more familiar they tended to be with a situation being depicted on a given trial. Supporting my hypotheses, I found that greater familiarity with the type of encounter being presented on a given trial was associated

with greater odds of providing an accurate response. Given the relationship between familiarity, level of training, and years of operational police experience, one could reason that exposure to various scenarios, both in training (e.g., scenario-based training) and in the field (i.e., by being exposed to a given type of encounter), would encourage task-relevant, adaptive schemas to be formed, which might allow individuals to more effectively and efficiently process information from thin slices in order to make accurate predictions from that information. Certainly, other research would support this possibility (Ames et al., 2010; Gonzalez et al., 2003), including research from the policing field (e.g., Boulton & Cole, 2016). This provides additional support for the potential value of realistic police training.

Given the strong correlation between the years of operational police service, one's level of training, and familiarity I observed in the current study, one might also expect years of operational police service to be associated with the odds of providing accurate predictions based on thin slices of police-public encounters. However, contrary to this finding, and to my hypotheses, the number of operational years of police experience was not significantly associated with the odds of providing an accurate prediction. My results therefore suggest that both training and familiarity might be more important in constituting the definition of "expertise" in the field of policing rather than years of experience. In support of this, non-policing literature also appears to emphasize one's training and field experience (e.g., familiarity) in the definition of expertise, rather than simply quantifying the number of years of experience one has. For example, studies in the field of sports that have found expert-novice differences in anticipation accuracy, quantified experience by examining the amount of practice (e.g., playing in tournaments) and the level of the players (e.g., "club level performers or above"; Williams et al., 2002) rather than relying solely on the years of experience one had playing the sport. Furthermore, research

examining police prediction accuracy has also categorised “skilled officer” groups (i.e., “expert” officers in their study, compared to novice officers who were officer recruits or rookies) as those possessing SWAT training (Suss & Ward, 2012; Vickers & Lewinski, 2012).

However, before I discuss the value of using “years of operational experience” as a proxy for expertise, it is important to consider that “years of operational police experience” may be a more complicated construct than how it was measured in the current study. For example, my measure of operational police experience did not take into account *when* that operational experience was obtained. An officer in the present study with 20 years of operational police experience may have acquired these years of operational experience earlier in their career and may have spent the last couple of years working in a non-operational capacity (i.e., working outside of the field). These officers may therefore be “rustier” compared to other officers who might have fewer years of operational experience, but who are currently operational. The timing of operational experience (and potentially other variables) would have to be addressed in future research before firm conclusions about the value of years of operational experience can be drawn.

Finally, in the present study, greater confidence in one’s prediction on a given trial was associated with greater odds of providing an accurate response on that trial. These findings concur with the findings of previous research (Patterson et al., 2001; Smith et al., 1991); however they are not in line with the research conducted by Ames et al. (2010), which found that confidence in one’s predictions of personality based on thin slices of information was largely unrelated to accuracy (with the exception that zero confidence ratings were associated with inaccurate responses). It is possible that differences in the confidence-accuracy relationship that are found between studies stem from the nature of the prediction. Specifically, predictions in

Ames et al.'s (2010) study were made on a 7-point Likert scale, whereas those in the present study, as well as the studies conducted by Patterson et al. (2001) and Smith et al. (1991), were not scale-based ratings. It is possible that the relationship between confidence and accuracy in thin slice studies is more likely to be found when the outcome prediction is made in absolute terms (e.g., yes or no) rather than along a gradient (e.g., 1 to 7).

Predictors of Confidence

Beyond exploring the relationship between confidence and accuracy, I also examined variables that might predict the level of confidence one has in making a prediction across thin slice trials. Unfortunately, given the dearth of research looking at issues of confidence in thin slice settings, it is difficult to know how to interpret some of the findings related to this construct. However, a number of important things should be mentioned.

In the present study, the recoded aggregate score for training, increased familiarity on a given trial, and years of operational police service (to a lesser degree) were associated with expected increases in levels of confidence on a given trial. In addition, officers who reported having "more" ERT/SWAT training than other officers were also found to have higher expected levels of confidence across trials (compared to civilians). These findings make sense given that one might expect highly specialized police training (i.e., ERT/SWAT training) and operational police experience to provide a participant with increased confidence in responding to situations that are highly relevant to their job.

With regard to familiarity, previous research by Ames et al. (2010) supports my finding that greater familiarity with the stimuli encountered in a trial is likely to be associated with greater confidence in one's predictions about that stimuli. These researchers found in their study of personality predictions that, if the individual depicted in the stimuli corresponded with a

“certain type of person” or “reminded [the participant] of someone,” confidence ratings in the participant’s personality judgments were higher. Given that these factors are likely to be related to one’s schemas about people (Ames et al., 2010), and that familiarity is likely associated with the schemas one has formed (Boulton & Cole, 2016), one would expect familiarity to be a predictor of one’s degree of confidence. As an alternative explanation, given the positive relationship between confidence and accuracy, familiarity and accuracy, and familiarity and confidence in the current study, it is possible that familiarity and confidence (which are instilled through training) make the task easier, allowing for individuals who are higher on these variables to be more accurate.

The Role of Thin Slice Length

My results suggested that longer thin slices were associated with slightly greater odds of providing an accurate response, however this relationship was not significant. Despite this, I cannot confidently conclude that thin slice length has no effect on prediction accuracy. One possibility that might explain why I did not observe a significant effect of thin slice length could be the magnitude of the difference in video lengths I selected (i.e., 10-seconds vs. 30-seconds). In other words, the 20-second difference between the 10- and 30-second videos might have been too small for significant prediction accuracy differences to be detected. This is further supported by my differing findings related to thin slice length, which are discussed next.

As hypothesised, training, years of operational experience, and familiarity were related to the accuracy of outcome predictions when participants were presented with short thin slices (i.e., 10-second videos), which is also consistent with previous research. For example, research has suggested that experts are able to identify relevant cues for prediction purposes very early in scenarios (i.e., with access to very little information), whereas novices are not (Abernethy et al.,

2001; Ward et al., 2013). As argued above, familiarity can also be interpreted as being an element of expertise. My results therefore suggest that possessing superior expertise in policing (in this case, by having more training and years of operational experience, and greater familiarity with the types of stimuli encountered in the current study) is associated with greater odds of providing an accurate response for shorter thin slices, despite being presented with fewer environmental, situational, and subject behaviour cues. This is further supported by my finding that, while officers with “more” ERT/SWAT training maintained their level of accuracy for both short and long trials, police officers with “the same amount of” or “less” ERT/SWAT training, and civilians, became slightly less accurate on short trials (where there were fewer cues and therefore less information) and slightly more accurate on long trials (where there were more cues and more information).

When examining long thin slices (i.e., 30-second videos), the relationship between training (the aggregate score), familiarity, years of operational experience, and the odds of providing an accurate prediction became non-significant, as I hypothesized. One possibility for the finding that expertise differences in prediction accuracy exist for short thin slices, but disappear in long thin slices, is that only police experts have the sorts of schemas that are required to make sense of police-public encounters based on very thin (i.e., short) slices of information. For those without this expertise, much more information might be required to understand the encounter and develop a “detailed representation” of the situation at hand (Ward et al., 2011), thereby allowing them to improve their ability to make accurate predictions. This thinking appears to be consistent with existing literature, which has suggested that certain cues that can be sampled over longer periods of time, might benefit one’s prediction accuracy (Carney et al., 2007; Nguyen & Gatica-Perez, 2015). In the current study, this may be especially true for

people with less specialized training in the encounters being assessed or without any police training at all (i.e., civilians).

Interestingly, officers with “more” ERT/SWAT training did not have significantly greater odds of providing an accurate response than police officers with the “same amount of” or “less” ERT/SWAT training, for all trials, including both short and long thin slices. However, my results still demonstrated a small reduction in group differences in longer thin slices. On the other hand, and contrary to my hypotheses, officers with “more” ERT/SWAT training had significantly greater odds of providing an accurate response compared to civilians, for *both* short and long thin slices. However, again, my results demonstrated a reduction in group differences in longer thin slices. Specifically, in long trials, the difference in accuracy rates between civilians and officers with “more” ERT/SWAT training was 11.8%, which is approximately half of the difference that was observed between these two groups in short trials (i.e., 19.3%).

Limitations and Future Directions

There are a number of limitations associated with the current study that warrant discussion. Linked to these limitations are also potential opportunities for future research. In this section, I will describe both the major limitations of this study, and some potentially fruitful lines of future research.

Arguably, the most important limitation of the current study is the use of body-worn camera (BWC) footage as the source of the thin slices. While this technology is certainly able to capture elements of police-public encounters, and its use represents a step forward in this area of research given the previous reliance on simulated (i.e., potentially artificial) scenarios, BWC footage is also known to be problematic (Force Science Institute, 2014). For example, such footage does not necessarily capture all elements of an encounter (e.g., because the lens is

obstructed). Also, BWC footage only provides a first-person perspective; therefore, other potentially important sources of information (e.g., the responding officer's [who is wearing the camera] body language) cannot be taken into consideration when predicting the outcome of that situation. In addition to these issues, a serious limitation was the fact that the BWC footage used in this study was all taken from the internet. Not only does this leave open the possibility that the footage was altered in some way when it was uploaded to the internet, which could have impacted the original recording (e.g., edited or formatted), but variation in footage quality or type (e.g., due to different cameras and lenses being used to record events) might have existed across thin slice conditions. In the future, it would be valuable to replicate the current study using BWC footage from one police service. This would prevent researchers from having to rely on footage from the internet, while also ensuring that camera specifications (and, thus, the resulting footage) are more consistent.

Another serious limitation is that it is not entirely clear based on the data analysed in this thesis why accurate and inaccurate predictions were made. While the sorts of factors examined in this thesis (e.g., training) seem to account for some of the variation, it would have been valuable to examine the cues that participants used to make their decisions. Not only would this shed light on how people might accurately discriminate between different types of use-of-force encounters (i.e., those in which the subject will harm or attempt to harm the officer(s) and those in which the subject will not), this would have real value in a police training context if reliable predictors of the outcome could be uncovered. As discussed in the Methods section, I collected this information by having participants list the cues they used to make their prediction following each trial. There was simply not enough time to fully examine this data within the current thesis. This will be a priority in the near future. When such work is completed it will be important, however,

to realize that such an examination is also limited in that it assumes participants are able to correctly determine what cues they relied on to make their decisions. Research in psychology has raised some doubts about this (e.g., Nisbett & Wilson, 1977; Smith et al., 1991); this will have to be taken into consideration when presenting this research.

An additional limitation is that there were no “validity checks” in the demographic questionnaire participants were asked to complete, making it possible that inaccurate data was included in my analyses. While this might be a problem for all of the questions asked, the most obvious example of this related to some of the civilian responses with regard to training (described in the Results section). While I do not believe that this limitation prevented meaningful conclusions to be reached in the current study (e.g., recoding of this training data generally did not significantly impact the results of the regression models, suggesting that the potentially questionable data had a minimal impact), it would be beneficial in future research to use procedures that might minimize this problem. For example, in-person (rather than online) testing might allow researchers to probe participants’ responses, like those related to training, that appear to be questionable. Carrying out this research in person may also encourage participants to be more truthful if this is in fact contributing to some of the surprising demographic data that was encountered in the current study.

Finally, my study was limited in the conclusions it could draw from my results with regard to the various factors that were manipulated in the current study. For example, findings suggested that training and experience have a potential impact on how well people can determine outcomes of police-public encounters. However, these variables closely align with officer-student distinctions, and officers and students in the present study also varied along other variables as well (e.g., age, gender, ethnicity, etc.). Thus, I cannot rule out the importance of

these other variables as potential explanations for the results. Future research that attempts to match students and officers on some of these variables might prove useful in this regard.

Likewise, it is difficult to know how to interpret some of the findings related to thin slice length because the 10- and 30-second slices varied in ways beyond simply time. For example, the amount of information increases in the 30-second slices, as does the potential for conflicting information. In the future, it would be useful to design a study to deal with these kinds of issues (e.g., by including 10-second conditions that overlap with various portions of the 30-second thin slices).

Conclusion

Policing requires split-second decisions to be made in the field and being able to predict the future behaviour of a subject early in an interaction may help officers better prepare for and respond to that situation. My study examined the ability of officers and civilians to accurately anticipate the outcomes of police-public interactions based on thin slices of the encounters. The findings demonstrated that there are a variety of factors that are indicative of whether one will be able to predict possible harm to an officer(s) in these incidents. These include the training that police officers receive, the degree to which officers are familiar with the types of encounters depicted in the BWC footage used in this study, as well as the confidence officers have in their outcome predictions. Not only do these findings have potential theoretical importance, for example in terms of the role of schemas in making accurate outcome predictions in the field of policing based on very short exposures to police-public encounters, they also have potential practical implications. Indeed, if future research that deals with some of the limitations discussed above replicated the findings reported in this thesis, the results would speak to the importance of police training, especially realistic training that relies on the types of scenarios that officers

regularly encounter on the streets (given that this would enhance familiarity with these encounters). If research can help increase one's ability to accurately, and quickly, predict the outcomes of potential use-of-force encounters they have with the public, this should enhance the safety of both the public and police officers.

References

- Abernethy, B. (1993). Searching for the minimal essential information for skilled perception and action. *Psychological Research*, 55(2), 131–138. <https://doi.org/10.1007/BF00419644>
- Abernethy, B., Gill, D. P., Parks, S. L., & Packer, S. T. (2001). Expertise and the perception of kinematic and situational probability information. *Perception*, 30(2), 233–252. <https://doi.org/10.1068/p2872>
- Abernethy, B., Schorer, J., Jackson, R. C., & Hagemann, N. (2012). Perceptual training methods compared: The relative efficacy of different approaches to enhancing sport-specific anticipation. *Journal of Experimental Psychology: Applied*, 18(2), 143–153. <https://doi.org/10.1037/a0028452>
- Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance? *Journal of Experimental Social Psychology*, 45(4), 1052–1055. <https://doi.org/10.1016/j.jesp.2009.05.017>
- Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry*, 21(4), 271–278. <https://doi.org/10.1080/1047840X.2010.524882>
- Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, 77(3), 538–547. <https://doi.org/10.1037/0022-3514.77.3.538>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behaviour as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2), 256–274. <https://doi.org/10.1037/0033-2909.111.2.256>
- Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social*

Psychology, 64(3), 431–441. <https://doi.org/10.1037/0022-3514.64.3.431>

Ames, D. R., Kammrath, L. K., Suppes, A., & Bolger, N. (2010). Not so fast: The (not-quite-complete) dissociation between accuracy and confidence in thin-slice impressions.

Personality and Social Psychology Bulletin, 36(2), 264–277.

<https://doi.org/10.1177/0146167209354519>

Babad, E., Bernieri, F., & Rosenthal, R. (1991). Students as judges of teachers' verbal and nonverbal behavior. *American Educational Research Journal*, 28(1), 211–234.

<https://doi.org/10.3102/00028312028001211>

Benjamin, D. J., & Shapiro, J. M. (2009). Thin-slice forecasts of gubernatorial elections. *Review of Economics and Statistics*, 91(3), 523–536. <https://doi.org/10.1162/rest.91.3.523>

Bierie, D. M. (2015). Assault of police. *Crime and Delinquency*, 63(8), 899–925.

<https://doi.org/10.1177/0011128715574977>

Blanck, P. D., & Rosenthal, R. (1984). Mediation of interpersonal expectancy effects: Counselor's tone of voice. *Journal of Educational Psychology*, 76(3), 418–426.

<https://doi.org/10.1037/0022-0663.76.3.418>

Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology*, 65(3), 546–553. <https://doi.org/10.1037/0022-3514.65.3.546>

Boulton, L., & Cole, J. (2016). Adaptive flexibility: Examining the role of expertise in the decision making of authorized firearms officers during armed confrontation. *Journal of Cognitive Engineering and Decision Making*, 10(3), 291–308.

<https://doi.org/10.1177/1555343416646684>

- Brandl, S. G. (1996). In the line of duty: A descriptive analysis of police assaults and accidents. *Journal of Criminal Justice*, 24(3), 255–264. [https://doi.org/10.1016/0047-2352\(96\)00007-4](https://doi.org/10.1016/0047-2352(96)00007-4)
- Canadian Association of Chiefs of Police. (2000). The National Use of Force Framework [Diagram]. Retrieved January 9, 2019, from https://www.cacp.ca/policiesguidelines.html?asst_id=214
- Canadian Association of Chiefs of Police [CACP]. (2000). A National Use of Force Framework. Ottawa, ON: Canadian Association of Chiefs of Police. Retrieved from https://www.cacp.ca/cacp-use-of-force-advisory-committee-activities.html?asst_id=199
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41(5), 1054–1072. <https://doi.org/10.1016/j.jrp.2007.01.004>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Conor, P. (2017). Police resources in Canada, 2017. Retrieved August 26, 2019, from <https://www150.statcan.gc.ca/n1/pub/85-002-x/2018001/article/54912-eng.htm>
- Covington, M. W., Huff-Corzine, L., & Corzine, J. (2014). Battered police: Risk factors for violence against law enforcement officers. *Violence and Victims*, 29(1), 34–52. <https://doi.org/10.1891/0886-6708.VV-D-12-00022>
- Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92(3), 802–811. <https://doi.org/10.1037/0021-9010.92.3.802>
- Ekman, P., & Friesen, W. V. (1967). Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills*, 24(3), 711–724.

<https://doi.org/10.2466/pms.1967.24.3.711>

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88–106. <https://doi.org/10.1080/00332747.1969.11023575>

Ellis, D., Choi, A., & Blaus, C. (1993). Injuries to police officers attending domestic disturbances: An empirical study. *Canadian Journal of Criminology*, 35(2), 149–168. Retrieved from <https://heinonline.org/HOL/P?h=hein.journals/cjccj35&i=155>

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. <https://doi.org/10.1037/1082-989X.12.2.121>

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. <https://doi.org/10.1037/0033-295X.100.3.363>

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. (M. Carmichael, Ed.) (4th editio). London: Sage Publications Ltd.

Force Science Institute. (2014). 10 limitations of body cams you need to know for your protection. Retrieved August 28, 2019, from <https://www.policeone.com/police-products/body-cameras/articles/7580663-10-limitations-of-body-cams-you-need-to-know-for-your-protection/>

Fowler, K. A., Lilienfeld, S. O., & Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment*, 21(1), 68–78. <https://doi.org/10.1037/a0014938>

Gabbett, T. J., & Abernethy, B. (2013). Expert–novice differences in the anticipatory skill of rugby league players. *Sport, Exercise, and Performance Psychology*, 2(2), 138–155. <https://doi.org/10.1037/a0031221>

- Gallaher, P. E. (1992). Individual differences in nonverbal behavior: Dimensions of style. *Journal of Personality and Social Psychology*, *63*(1), 133–145.
- Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635. [https://doi.org/10.1016/S0364-0213\(03\)00031-4](https://doi.org/10.1016/S0364-0213(03)00031-4)
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*. New York, NY: Routledge/Taylor & Francis Group.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. <https://doi.org/10.3758/BF03212378>
- Johnson, R. R. (2006). Confounding influences on police detection of suspiciousness. *Journal of Criminal Justice*, *34*(4), 435–442. <https://doi.org/10.1016/j.jcrimjus.2006.05.009>
- Johnson, R. R. (2007). Race and police reliance on suspicious non-verbal cues. *Policing*, *30*(2), 277–290. <https://doi.org/10.1108/13639510710753261>
- Johnson, R. R. (2011). Predicting officer physical assaults at domestic assault calls. *Journal of Family Violence*, *26*(3), 163–169. <https://doi.org/10.1007/s10896-010-9346-0>
- Johnson, R. R. (2015). Perceptions of interpersonal social cues predictive of violence among police officers who have been assaulted. *Journal of Police and Criminal Psychology*, *30*(2), 87–93. <https://doi.org/10.1007/s11896-014-9145-2>
- Johnson, R. R. (2019). Exploring the validity of behavioral cues predictive of physically resisting arrest. *Journal of Police and Criminal Psychology*, *34*(2), 134–144. <https://doi.org/10.1007/s11896-018-9280-2>
- Johnson, R. R., & Aaron, J. L. (2013). Adults' beliefs regarding nonverbal cues predictive of violence. *Criminal Justice and Behavior*, *40*(8), 881–894.

<https://doi.org/10.1177/0093854813475347>

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society Annual Meeting*, 30(6), 576–580.

<https://doi.org/10.1177/154193128603000616>

Murgia, M., Sors, F., Muroi, A. F., Santoro, I., Prpic, V., Galmonte, A., & Agostini, T. (2014).

Using perceptual home-training to improve anticipation skills of soccer goalkeepers.

Psychology of Sport and Exercise, 15(6), 642–648.

<https://doi.org/10.1016/j.psychsport.2014.07.009>

Nguyen, L. S., & Gatica-Perez, D. (2015). I would hire you in a minute: Thin slices of nonverbal

behavior in job interviews. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15* (pp. 51–58). Seattle, WA: ACM.

<https://doi.org/10.1145/2818346.2820760>

Nisbett, R. E., & Wilson, T. D. (1977). Psychological Review. *Psychological Review*, 84(3), 231–259.

Patterson, M. L., Foster, J. L., & Bellmer, C. D. (2001). Another look at accuracy and confidence in social judgments. *Journal of Nonverbal Behavior*, 25(3), 207–219.

<https://doi.org/10.1023/A:1010675210696>

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied*, 13(3), 158–170.

<https://doi.org/10.1037/1076-898X.13.3.158>

Riggs, D. S., Rothbaum, B. O., & Foa, E. B. (1995). A prospective examination of symptoms of posttraumatic stress disorder in victims of nonsexual assault. *Journal of Interpersonal*

Violence, 10(2), 201–214. <https://doi.org/10.1177/0886260595010002005>

- Royal Canadian Mounted Police. (2017). Incident Management/Intervention Model. Retrieved from <http://www.rcmp-grc.gc.ca/en/incident-management-intervention-model-imim>
- Rule, N. O., & Ambady, N. (2008). The face of success. *Psychological Science, 19*(2), 109–111. <https://doi.org/10.1111/j.1467-9280.2008.02054.x>
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review, 16*(1), 170–175. <https://doi.org/10.3758/PBR.16.1.170>
- Simunition: Non-lethal training ammunition. (n.d.). Retrieved December 31, 2018, from <https://simunition.com/en/>
- Smith, H. J., Archer, D., & Costanzo, M. (1991). “Just a hunch”: Accuracy and awareness in person perception. *Journal of Nonverbal Behavior, 15*(1), 3–18. <https://doi.org/10.1007/BF00997764>
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology, 30*(1), 203–218. <https://doi.org/10.5334/irsp.90>
- Stetser, M. (1999). *The use of force in police control of violence: Incidents resulting in assaults on officers* [Doctoral dissertation]. University of Hawai’i, Honolulu, Hawai’i.
- Suss, J., & Ward, P. (2012). Use of an option generation paradigm to investigate situation assessment and response selection in law enforcement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 297–301. <https://doi.org/10.1177/1071181312561069>
- Suss, J., & Ward, P. (2013). Investigating perceptual anticipation in a naturalistic task using a temporal occlusion paradigm: A method for determining optimal occlusion points. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 304–

308. <https://doi.org/10.1177/1541931213571067>

Suss, J., & Ward, P. (2018). Revealing perceptual-cognitive expertise in law enforcement: An iterative approach using verbal-report, temporal-occlusion, and option-generation methods. *Cognition, Technology & Work*, 20(4), 585–596. <https://doi.org/10.1007/s10111-018-0493-z>

Tskhay, K. O., Zhu, R., & Rule, N. O. (2017). Perceptions of charisma from thin slices of behavior predict leadership prototypicality judgments. *Leadership Quarterly*, 28(4), 555–562. <https://doi.org/10.1016/j.leaqua.2017.03.003>

Vickers, J. N., & Lewinski, W. (2012). Performing under pressure: Gaze control, decision making and shooting performance of elite and rookie police officers. *Human Movement Science*, 31(1), 101–117. <https://doi.org/10.1016/j.humov.2011.04.004>

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior*, 21(2), 87–102. <https://doi.org/10.1023/A:1024951902752>

Ward, P., Ericsson, K. A., & Williams, A. M. (2013). Complex perceptual-cognitive expertise in a simulated task environment. *Journal of Cognitive Engineering and Decision Making*, 7(3), 231–254. <https://doi.org/10.1177/1555343412461254>

Ward, P., Suss, J., Eccles, D. W., Williams, A. M., & Harris, K. R. (2011). Skill-based differences in option generation in a complex task: A verbal protocol analysis. *Cognitive Processing*, 12(3), 289–300. <https://doi.org/10.1007/s10339-011-0397-9>

Waxer, P. H. (1977). Nonverbal cues for anxiety: An examination of emotional leakage. *Journal of Abnormal Psychology*, 86(3), 306–314. <https://doi.org/10.1037/0021-843X.86.3.306>

Williams, A. M., Hodges, N. J., North, J. S., & Barton, G. (2006). Perceiving patterns of play in

dynamic sport tasks: Investigating the essential information underlying skilled performance.

Perception, 35(3), 317–332. <https://doi.org/10.1068/p5310>

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, 8(4), 259–270. <https://doi.org/10.1037/1076-898X.8.4.259>

Appendix A

Demographic Questionnaire for Police Officers

1. What country do you reside in? _____
2. What is your age (years old)? _____
1. What is your gender?
 - Male Female Other Prefer not to answer
2. What is your ethnicity (select all that apply)?
 - African American/Black
 - Arab or West Asian (e.g., Iranian, Armenian, Lebanese, Moroccan, etc.)
 - Caucasian/White/Western European
 - Eastern European
 - First Nation, Inuit, or Alaska Native
 - Hispanic/Latino
 - Métis
 - Native Hawaiian or Pacific Islander
 - South or East Asian
 - Other (please specify): _____
 - Prefer not to answer
3. How long have you been working as a law enforcement officer? _____ years

4. How many of these years were you operational (i.e., on the streets)? _____ years

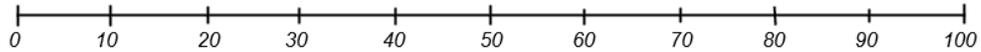
5. Compared to other officers in your agency, do you believe you have "less training", the "same amount of training", or "more training" in the following training areas?

	Less training	Same amount of training	More training
Emergency response team (ERT) or Special Weapons and Tactics (SWAT) training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Immediate Action/Rapid Deployment (IARD) training or Active Shooter training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

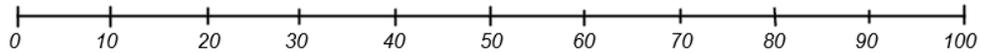
Instructor training in use of force/firearms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Martial arts training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De-escalation training (consider all types of de-escalation training for this question)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Communication-based de-escalation training (e.g., verbal judo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mental health training for de-escalation purposes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Crisis negotiation training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conflict resolution training for de-escalation purposes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other de-escalation training (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Use-of-force training (consider all types of use-of-force training for this option)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Police Defensive Tactics (PDT) training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Scenario-Based training (SBT) for use of force	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Simulator training for use of force (i.e., virtual training)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shooting range training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other use of force training (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- 6. How often have you been involved in encounters with a person where you were at risk of being physically harmed by that person (0 = never; 100 = extremely often)?



- 7. How would you rate your ability to anticipate outcomes of interactions with members of the public (0 = extremely poor; 100 = excellent)?



Appendix B

Demographic Questionnaire for Civilians

3. Have you ever worked as a police, correctional, or military officer?¹⁴

- Yes
- No

4. What country do you reside in? _____

5. What is your age (years old)? _____

6. What is your gender?

- Male
- Female
- Other
- Prefer not to answer

7. What is your ethnicity (please select all that apply)?

- African American/Black
- Arab or West Asian (e.g., Iranian, Armenian, Lebanese, Moroccan, etc.)
- Caucasian/White/Western European
- Eastern European
- First Nation, Inuit, or Alaska Native
- Hispanic/Latino
- Métis
- Native Hawaiian or Pacific Islander
- South or East Asian
- Other (please specify): _____
- Prefer not to answer

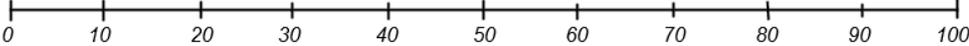
8. Compared to police officers, do you believe you have "less training", the "same amount of training", or "more training" in the following training areas?

	Less training	Same amount of training	More training
Emergency response team (ERT) or Special Weapons and Tactics (SWAT) training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

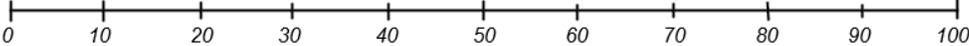
¹⁴ If participants respond “yes” to this question, participants will be thanked for their time and the study will stop (data from these individuals will not be collected).

Immediate Action/Rapid Deployment (IARD) training or Active Shooter training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Instructor training in use of force/firearms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Martial arts training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
De-escalation training (consider all types of de-escalation training for this question)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Communication-based de-escalation training (e.g., verbal judo)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mental health training for de-escalation purposes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Crisis negotiation training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conflict resolution training for de-escalation purposes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other de-escalation training (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<hr/>			
Use-of-force training (consider all types of use-of-force training for this option)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Police Defensive Tactics (PDT) training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Scenario-Based training (SBT) for use of force	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Simulator training for use of force (i.e., virtual training)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shooting range training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other use of force training (please specify)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

9. How often have you been involved in encounters with a person where you were at risk of being physically harmed by that person (0 = never; 100 = extremely often)?



10. How would you rate your ability to anticipate outcomes of interactions with members of the public (0 = extremely poor; 100 = excellent)?



Appendix C

Informed Consent Form

This informed consent form is designed to explain to you the study's purpose, the required tasks, and provide additional information to allow you to decide whether or not you wish to participate in the study. Please take the time to read this information carefully.

Title of Study: Can police foresee the future? Predicting outcomes from thin slices of police-public encounters

Research Personnel. The following people are involved in this study and may be contacted at any time if you have questions or concerns: Ariane Khanizadeh (Lead Researcher – ariane.khanizadeh@carleton.ca), Dr. Craig Bennell (Faculty Advisor – craig.bennell@carleton.ca).

Further Information, Questions, Concerns: If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4080 or via email at ethics@carleton.ca).

Purpose:

The study will examine how well police officers (with varying levels of expertise and training) and civilians (with no expertise in policing and no training) can make predictions about the outcomes of police-public encounters based on short video segments of the encounters. We are also interested in how a variety of factors (when familiarity with the type of encounter varies and when the length of the video varies) impacts prediction accuracy.

Task Requirements: If you agree to participate in this study, you will first be asked to respond to a demographic questionnaire. Following this, you will be asked to complete 4 practice trials (data from these trials will not be recorded) and 16 test trials (data from these trials will be recorded). In each of these trials, you will view a short video segment of body-worn camera footage of a police-public encounter and will be asked to answer a series of questions related to the video. Each trial will take between 1 minute and 2 minutes to complete.

Compensation: Participation in this study is completely voluntary. If you are a police professional who is participating in this study, no compensation will be provided. If you are a Carleton University student participating in this study through SONA, you will receive course credit (0.5%) for your participation.

Duration and Locale: The study will take approximately 45 minutes to complete. This study will take place completely online.

Potential Risk/Discomfort: There are possible emotional/psychological risks associated with viewing police officer body-worn camera footage as some footage may contain violent and/or sensitive content. Your participation in this study is completely voluntary, and you may withdraw at any time by clicking on the withdraw button at the bottom of each page. For Carleton University students: a decision to withdraw will in no way impact course credit you receive from participating in this study.

Anonymity/Confidentiality:

All data will be collected through the online survey tool Qualtrics. Qualtrics employs multiple layers of security to ensure that data remains private and secure. All surveys created are placed in a Secure Survey Environment (SSE) and the web pages are encrypted with secure socket layer (SSL). Only persons with authorized access to a survey account can download the data from this server. Qualtrics is SAS 70 certified and meets the rigorous privacy standards imposed on health care records by the Health Insurance Portability and Accountability Act (HIPAA). All Qualtrics accounts are protected by password-access, and Qualtrics employees will not access the protected accounts without express permission by the account owner.

Responses collected will be confidential and no identifying information is requested. Although Qualtrics will initially have access to Internet Protocol (IP) addresses when you submit your responses, the IP addresses are discarded and cannot be connected to survey responses. At the end of the survey, you are encouraged to close the browser window. Throughout the survey, a "Withdraw" option will be included, which you can click at any time, and which will lead you to the debriefing form.

The data will remain on the Qualtrics account until the end of the study and will then be deleted. No backups will be kept on the Qualtrics server after the deletion has been processed. The data will be stored in a statistical database and kept on a secure computer in Dr. Bennell's research laboratory as well as on the lead researcher's secure personal computer. The only individuals who will be able to access these files are Dr. Bennell, Ms. Khanizadeh, and graduate students of Dr. Bennell. The data will be kept in locked cabinets or in password protected computer files for five years, as recommended by the American Psychological Association (for publication purposes), before being deleted.

The research findings will be published and/or presented for educational purposes at Carleton University and within the community. Given the anonymity of responses, no personal or identifying information will be linked to any data in any of the documents, presentations, or publications

Right to Withdraw: Your participation in this study is entirely voluntary. At any point during the study, you have the right to withdraw completely without any explanation or penalty. If you withdraw from the study your data will be deleted from the study.

For Carleton University students: withdrawal from this study will have no impact on the course credit you earn from participating in this study.

Given the anonymity in this study, you will not be able to withdraw from the study once you have completed it.

This study has received clearance by the Carleton University Research Ethics Board B and assigned # 110407.

Consent: Checking the "I consent to participate in this study" will act as your online signature, indicating that you have fully read and understood the above statement and freely consent to participate in this study. It also indicates that you are aware that the data collected during the study may be used for educational presentations and research purposes, that the data collected in this study will be kept strictly confidential, and that you may withdraw from the study at any

time during the study, however withdrawal will *not* be possible after you have completed the study. By checking the “I consent to participate in this study” box, you are indicating that you understand the above and wish to participate in this study.

I consent to participate in this study

I do not consent to participate in this study

Appendix D
Debriefing Form

Thank you for participating in our study. We truly appreciate your time.

If you have any questions about the research, you may contact one of the researchers associated with the project:

Lead Researcher:

Ariane Khanizadeh, Carleton University, Department of Psychology

Email: ariane.khanizadeh@carleton.ca

Faculty Supervisor:

Dr. Craig Bennell, Carleton University, Department of Psychology

Tel.: (+1) 613-520-2600 x1769

Email: craig.bennell@carleton.ca

We would like to provide you with a bit of additional information about our study in case you are interested.

What are we trying to learn in this research?

The study is examining how well police officers (with varying levels of expertise and training) and civilians (with no expertise in policing and no training) can make predictions about the outcomes of police-public encounters based on short video segments of the encounters. We are also interested in how a variety of factors (when familiarity with the type of encounter varies and when the length of the video varies) impacts prediction accuracy.

Why is this important?

Previous research has demonstrated that expertise can result in better prediction accuracy when those predictions are based on short segments of a given event, or what we call “thin slices” (e.g., short segments of observing one’s behaviour). Only one study of this nature has been conducted in a policing environment (Suss & Ward, 2012). This study also showed that police officers with more experience exhibit higher levels of prediction accuracy when they are asked to predict the outcome of police-public encounters based on brief exposures to those encounters.

The current study will add to the growing body of research examining outcome anticipation based on short segments of a given event and will further our understanding of how certain variables – police expertise, thin-slice length, and familiarity with the type of encounter you were exposed to – affect these predictions.

This study will have potential implications for police training. Providing novice officers with the knowledge an experienced officer has (e.g., cues associated with more accurate outcome anticipation) may help improve officer responses in police-public encounters. In addition, identifying environmental cues (e.g., presence of weapons) that can be incorporated into training may help officers learn to act more rapidly in a given situation, based on these cues (Ward et al., 2013).

What are our hypotheses and research questions?

In general, we hypothesize that police officers with specialized training and more years of experience will be better at predicting impending violence (and exhibit more confidence in their predictions), particularly when viewing shorter videos. We also hypothesize that individuals with greater familiarity with the type of encounter being presented in the video will be better at predicting impending violence, and exhibit greater confidence in their predictions.

We will also explore the cues that are associated with accurate prediction, and whether these cues differ based on training, years of experience, familiarity with the type of encounter, and the length of the video.

Where can I learn more?

If you are interested in learning more about this research, we recommend the following articles:

Suss, J., & Ward, P. (2012). Use of an option generation paradigm to investigate situation assessment and response selection in law enforcement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 297–301.
<https://doi.org/10.1177/1071181312561069>

Suss, J., & Ward, P. (2018). Revealing perceptual-cognitive expertise in law enforcement: An iterative approach using verbal-report, temporal-occlusion, and option-generation methods. *Cognition, Technology & Work*. <https://doi.org/10.1007/s10111-018-0493-z>

Ward, P., Ericsson, K. A., & Williams, A. M. (2013). Complex perceptual-cognitive expertise in a simulated task environment. *Journal of Cognitive Engineering and Decision Making*, 7(3), 231–254. <https://doi.org/10.1177/1555343412461254>

Is there anything I can do if I found this experiment to be emotionally upsetting?

Yes. If you feel any distress or anxiety after participating in this study, please feel free to contact any of the following services (different numbers are listed for different locations):

In Canada:

- Carleton University Health and Counseling Services: **613-520-6674**
- Distress Centre of Ottawa and Region: **613-238-3311** (<http://www.dcottawa.on.ca>).
- Ontario Crisis Line: **1-866-531-2600**
- British Columbia Crisis Line: **1-800-SUICIDE**
- Manitoba Crisis Line: **1-877-435-7170**
- Saskatchewan Crisis Line: **1-306-525-5333**
- Alberta Crisis Line: **403-266-4357**
- Quebec National Crisis Line: **1-866-277-3553**
- New Brunswick Crisis Line: **1-800-667-5005**
- Nova Scotia Crisis Line: **1-888-429-8167**
- Newfoundland and Labrador Line: **1-888-737-4668**
- Prince Edward Island Crisis Line: **1-800-218-2885**
- First Nations and Inuit Hope for Wellness: **1-855-242-3310**
- Northwest Territories: **1-800-661-0844**
- Nunavut Line: 7 pm-11 pm (EST): **1-800-265-3333**

- Yukon Crisis Line: 7 pm-12 am (PST) :**1-844-533-3030**
- You can also text **CONNECT** (English) or **PARLER** (French) to 686-868 to reach the Crisis Text Line.

In the United States:

- Crisis Call Center: **(775) 784-8090** or text **ANSWER** to **839863**
- You can also text **HOME** to 741-741 to reach the Crisis Text Line.

What if I have questions later?

If you have any remaining concerns, questions, or comments about the experiment, please feel free to contact Ariane Khanizadeh (Lead Researcher), at: ariane.khanizadeh@carleton.ca , or Dr. Craig Bennell (Faculty Sponsor), at: craig.bennell@carleton.ca ((+1) 613-520-2600, ext. 1769).

If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B (by phone at 613-520-2600 ext. 4080 or via email at ethics@carleton.ca).

Carleton University Project Clearance:

Clearance #: 110407

Date of Clearance: May 08, 2019

Appendix E

Binary Logistic Regression Results Following the Recoding of Civilian Training Responses to Police-Specific Training Options

Table 34

Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Training (aggregate) – Recoded	0.026	0.006	19.980	1	0.000	1.026	1.015	1.038
Constant	0.335	0.050	45.575	1	0.000	1.398		

Table 35

Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy in Short Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Training (aggregate) – Recoded	0.034	0.008	16.705	1	0.000	1.034	1.018	1.051
Constant	0.246	0.071	12.181	1	0.000	1.279		

Table 36

Binary Logistic Regression Results of Recoded Aggregate Training Score Predicting Accuracy in Long Thin Slices

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Training (aggregate) – Recoded	0.018	0.008	4.990	1	0.026	1.018	1.002	1.035
Constant	0.422	0.070	36.330	1	0.000	1.525		

Appendix F

Binary Logistic Regression Results Following the Recoding of Civilian Training Responses to
Police-Specific Training Options

Table 37

*Multilevel Linear Regression Results of Recoded Aggregate Training Score Predicting
Confidence Ratings*

	Estimate	Std. Error	Sig.	95% Confidence Interval	
				Lower	Upper
Fixed effects					
Intercept (B ₀₀)	58.15	1.51	0.000	55.15	61.14
Training (aggregate) - Recoded (B ₁₀)	0.57	0.17	0.001	0.23	0.92
Random effects					
Residual	292.50	10.63	0.000	272.40	314.09
Intercept [subject = id]	247.70	35.49	0.000	187.05	328.01