Journal of Police and Criminal Psychology
https://doi.org/10.1007/s11896-020-09364-4

Police Expertise and Use of Force: Using a Mixed-Methods Approach to Model Expert and Novice Use-of-Force Decision-Making

Laura Mangels[1] · Joel Suss[2] · Brian Lande[3]

Abstract

Improving police use-of-force training is methodologically difficult. By providing a method for identifying the "expert" response to any given scenario, and by triangulating multiple methods, we aim to contribute towards police departments' capacities to engage in more effective and targeted training. Forty-two police experts and 36 novices watched five scenarios taken from body-worn camera footage. The videos would pause at several points, and respondents gave both close-ended survey answers and open-ended written answers. Using a mixed-methods approach combining quantitative regression and natural-language-processing techniques, we triangulated our findings to reach conclusions regarding the differences between experts and novices. Relative to novices, expert police officers were more likely to report the importance of force mitigation opportunities to any given scenario in close-ended questions, and were more likely to use words associated with verbal de-escalation; novices were more likely to use words associated with physical control. The materials can be accessed at https://osf.io/wujkz/.

Keywords Police · Expertise, use of force · Body-worn cameras · Training · Mixed methods · Natural language processing

Introduction

While Stephon Clark, Alton Sterling, Michael Brown, and Eric Garner have now long been household names, police officer decision-making regarding the use of deadly force continues to come under scrutiny. Police departments that have been found to have violated citizens' rights—such as Baltimore, Albuquerque, Seattle, and Ferguson—have been placed under costly consent decrees by the US Department of Justice. However, since as early as the 1990s, police reformers and researchers have noted the intractable difficulty in actually tackling excessive police use of force (Klockars 1995; Skolnick and Fyfe 1993; Worden 1995). Without understanding the specific skills involved in deciding upon an appropriate use of force, as well as the concomitant training that can generate high levels of skill (i.e., expertise), interventions are ineffective (e.g., Bolger 2015; Geller and Toch 1995). Further, it is difficult to identify what is "expert" about experts because the experts themselves operate as if with a "felt sense," and are not always able to articulate their tacit knowledge (Evans 2008; Klein 2017; Rolfe 2005; Taylor et al. 2013). Nonetheless, expertise is important as expert judgment tends to be more effective, efficient, and accurate than that of novices (Ericsson et al. 1993; Klein et al. 2017). In this study, we take on this challenge, and we demonstrate how the use of relatively simple methods can make expert knowledge relatively easy to explicitly identify. Law enforcement presents a special challenge for researchers studying expertise and for developing effective training based on experts' performance. Police officers are called upon to respond to incidents that are dynamic and uncertain, requiring *naturalistic decision-making* (Klein et al. 1993). This decision-making occurs in the context of other macrocognitive activities, including sensemaking, attention allocation, and anomaly detection (Hoffman and McNeese 2009; Klein et al. 2003; Zimmerman 2008). Compared with more traditionally researched domains of expertise (e.g., chess or sports), police roles (e.g., patrol, SWAT, canine, sex crime investigations)

✉ Laura Mangels
laura.mangels@gmail.com

Joel Suss
joel.suss@wichita.edu

Brian Lande
brian.lande@polis-solutions.com

[1] University of California, Berkeley, CA, USA

[2] Wichita State University, Wichita, KS, USA

[3] Polis Solutions, Seattle, WA, USA

Published online: 24 January 2020

and tasks (e.g., interviewing, building searches, arrest and control techniques) are incredibly varied (Bittner 1990). Each may entail multiple types of expertise, and there are few objective methods, with the exception perhaps of marksmanship, for determining skill level (Suss and Boulton 2019; Vickers and Lewinski 2012). Relatively few studies have directly compared novice to experienced officers (Boulton and Cole 2016; Suss et al. 2014; Suss and Ward 2018; Zimmerman 2006). Other studies have examined the decision-making of either novices (Hine et al. 2018) or experienced officers (Harris et al. 2017; Klein et al. 2015) without comparing them, making it difficult to understand what distinguishes experts from novices. For the purposes of training, it is not surprising that there is an inherent difficulty in evaluating "correct" decision-making and thus what is a "reasonable" response for officers with different levels of training and experience (Alpert and Smith 1994; Geller and Toch 1995).

In the current study, we rely on a unique multi-method approach combining both textual analysis and regression that allows us to triangulate both qualitative and quantitative data about police officers' decision-making processes. We presented videos of real-life police–citizen interactions (i.e., from police body-worn cameras), and we asked both constrained and open-ended questions about officers' perceptions (what they saw and heard), judgements (how they made sense of the situation), and decisions as they watched the scenarios unfold. Most importantly, we use a simple, yet powerful, design that allows us to harness, trace, and distill the rich experience and cognitive processes of expert officers themselves to then create measurable, objective, situation-specific benchmarks that can be utilized to inform targeted use-of-force training on a larger scale. In other words, we can pinpoint what is "expert" about experts, which can guide the improvement of the efficiency and efficacy of use-of-force training.

Review of the Literature

Cognitive aspects of complex task performance are typically investigated using *cognitive tasks analysis techniques* (CTA; Hoffman and Militello 2008). CTA techniques encompass a range of methods that include experiments, observations, and interviews. Regardless of the method employed, their advantage is that they provide ways to understand the knowledge base (e.g., tacit knowledge, mental models) and mental processes that support expert performance. CTA techniques have been applied to law enforcement, albeit infrequently.

Zimmerman (2008) investigated how less-experienced (1–3 years) and experienced (7–30 years) police officers made sense of an actual traffic stop incident, in which it was also revealed that the vehicle was being used to transport drugs. Compared to less-experienced officers, experienced officers described the event more elaborately, identified more danger cues, provided more evaluations about what was happening, and made more predictions about what could happen. While the scenario used in Zimmerman's study did not provide any insight into officer use of force, Boulton and Cole (2016) asked less-experienced and experienced UK firearms officers to recall a challenging armed confrontation. The interview data were coded for key decision-making themes (i.e., experiential knowledge, strategies, adaptation), and the number of times each theme appeared in interviews was contrasted for the two groups. Although this study's inductive approach could possibly be relevant to understanding the use of force, in this case, the main discrepancies observed between groups were related to experienced officers' reporting lower cognitive loads and showing a greater ability to act independently. Though insightful for understanding police officer and expert cognition, the implications for use-of-force training are limited.

Ward et al. (2011) investigated the role of expertise in decision-making using interactive video scenarios of the type commonly used during police simulator training. Police recruits and experienced officers were equipped with a duty belt, holster, and laser handgun. They interacted (i.e., assumed tactical positions, gave verbal commands) with video scenarios that depicted a range of incidents (e.g., traffic stop, domestic violence call), some of which required a "shoot" response, while others could be resolved without the officer needing to use force. Retrospective verbal report data revealed that experienced officers generated more appropriate (i.e., task-relevant) options and prioritized those options better than the recruits. Suss and Ward (2018) further investigated the role of experience in decision-making by presenting police with video clips edited to end abruptly at a specific point in the action (i.e., just before a critical event), a method referred to as "temporal occlusion." At the point of occlusion, recruits and experienced officers were asked to predict what could happen next and to describe how they could respond. Similar to Ward et al. (2011), the data indicated that experienced officers evaluated scenarios differently than less-experienced officers, and selected more appropriate responses based on those evaluations. Both these studies have the advantage of being directly relevant to the use of force, and support that an officer's level of experience does in fact have important implications for the appropriate deployment of force by police officers. However, we must keep in mind that "appropriate" is an inherently tricky parameter to define, as we know that expertise itself involves complex skill sets that are not easily articulated by experts themselves.

CTA methods often provide rich, qualitative data. However, the procedures for transforming text and speech data generated during CTA is burdensome in terms of labor time and labor costs and imposes constraints on the sample sizes that researchers can handle (Basit 2003; Hoffman 1987, 1989; Hoffman 1989). Because our study treats the appropriate use of force as an emergent quality to be discovered—rather than given a priori—we are able to allow the experts themselves to show us what an expert solution looks like. Further, because we utilize the input of all the experts in our study, we have a robust “crowd-sourced” version of a solution to scenarios that corresponds to the “average” or consensus across experts—so we need not rely on just a few experts. Also, because we use decorated officers regarded by their peers to be experts (rather than simply relying on police officers with a certain number of years of experience), and because we triangulate quantitative and qualitative data sources both deductively and inductively, we can be reasonably sure of the robustness of our methodological approach, despite its simplicity and relatively low labor cost. Altogether, we believe we have an innovative and powerful tool for contributing to the growing literature on how expert and novice police officers differ. Here we will demonstrate how experts and novices differ in their understanding of the use of force.

Data and Methods

This study investigates officers’ perceptions of five scenarios depicted in footage taken from body-worn cameras from departments across the USA, and compares how groups of novice and expert police officers differ in their thinking on the appropriate use of force.

The Surveys As part of a force de-escalation training project funded by the Bureau of Justice Assistance (a component of the Department of Justice within the US Office of Justice Programs, BJA-2016-VI-BX-K005), the data were collected at two separate urban police departments in the USA. Because the original project’s focus was on experts alone, data on novices were not originally collected. The original sample of experts was surveyed as part of a collaborative effort with one of the police departments, to derive an empirically grounded model of expert decision-making. The goal was to use this expert model to facilitate training by giving officers insight into the cognitive content and processes of expert responses to situations. As such, our effort was from the beginning driven by the needs of applied research, to assist the department in grounding its facilitation guides and training rubrics. Experts ($n = 42$) in use of force were recruited (76% response rate) from a police department with over 12,000 sworn officers and almost 2000 civilian personnel, covering a city of roughly 2.7 million inhabitants. The experts were identified by the department’s commanders and comprised experienced officers who were qualified as use-of-force instructors and police academy instructors. Command staff based their selection of experts on reputation, time on the job, training, achievement of speciality assignment, and being Use-of-Force/Firearms/Defensive Tactics instructors. In addition to being identified as experts, the experts all also held positions where they had to instruct other officers. A novice sample was subsequently added once we realized the value of being able to compare expert and novice reasoning among police officers. This only became salient after modeling expertise in isolation and discussing the use of the expert model in real-world training situations. Novices ($n = 36$) were recruited from a police department that employs around 700 commissioned officers and over 200 non-sworn personnel, covering a metropolitan population of approximately 270,000. Specifically, the novices were recruited (97% response rate) from a cohort of officers in training who, by their lack of operational experience, we deemed novices. They were in the final week of a 26-week recruit-training course. Novices and experts participated voluntarily; novices received monetary compensation in the form of gift cards in return for their time.

By watching videos that were embedded in the surveys, novice and expert officers observed footage from police body-worn cameras depicting five different real-life police-citizen encounters from across the USA (1—Fort Collins, Colorado; 2—Tulsa, Oklahoma; 3—Palestine, Texas; 4—Miami-Dade, Florida; 5—Springfield, Missouri). Descriptions of the scenarios can be accessed at https://osf.io/8vjha/. To elicit participants’ thoughts with respect to decision-making, each scenario was divided into three segments. At the end of each segment—which we refer to as a *decision point*—we temporally occluded (i.e., paused playback) the scenario and participants responded to multiple-choice and text-entry questions. The questions asked the officers to describe various aspects of the scenario, as well as their recommended course of action. Officers accessed the online surveys through the Qualtrics Survey system on computer terminals or laptops with individual earbuds. All participants completed the five surveys in a fixed order. We decided to have the experts complete the surveys in a fixed order because we were initially unsure they would be able to complete all five scenarios within the time available; when we expanded our sample to include novices, we maintained the same fixed order for comparability. The videos and questions can be accessed at https://osf.io/yp6gd/. Most officers completed the five surveys in 135 min or less. Although such a lengthy period may raise concerns regarding fatigue, respondents did not report feeling bored or tired during debriefing.

At each decision point, respondents first answered a set of open-ended questions. Respondents typed their answers into text boxes, in response to the following three questions: (1) If you were handling this situation, what would you do in the next few seconds? (Describe specifically what you would do if you were actually on scene.) (2) From the start of the video until it stopped, what are the cues to which you recall paying attention? For example, what were you seeing, hearing, or paying attention to? List the three most important cues you remember and describe their significance. (3) Based on all of the information available to you at this moment (for example, what you know from the briefing and the cues you were noticing), how would you describe what is happening right now?

The open-ended questions were developed from standard probes used when conducting cognitive task analysis (Crandall et al. 2006). Specifically, we sought to elicit information about (a) course of action, (b) critical cues heeded, and (c) sensemaking/situation assessment, as these are the ways in which performance is typically assessed in law enforcement training. At any given decision point in a scenario, there are many different courses of action that are *possible*, but there may only be few that are advisable, appropriate, and feasible. The ability of experts to select a more appropriate course of action than novices has been documented in law enforcement (Johnson et al. 2014; Suss and Ward 2018), and more-experienced officers attend to different information than less-experienced officers (Vickers and Lewinski 2012; Ward et al. 2011). Finally, sensemaking is one of the key macrocognitive activities described by Klein et al. (2007): officers have to read and categorize situations so that they can bring experience—schemas and mental models—to bear on the situation. If an officer has an inaccurate "read" of the situation, they are less likely to select an appropriate course of action.

Next, respondents rated (a) the citizen's greatest level of resistance, (b) the greatest level of force used by the officer, and (c) how great of a threat the citizen was to the officer or other people. The respondents also rated the relevance of factors they may have considered in determining the appropriate level of force. An "objective reasonableness standard" for the use of force was laid out in the 1989 US Supreme Court's ruling in Graham v. Connor (1989), and the respondents rated their agreement (definitely false, probably false, neither true nor false, probably true, or definitely true) on these factors: (1) the threat was imminent; (2) the threat was severe; (3) the person had greater size, strength, or fighting skill; (4) the person was under the influence of drugs or alcohol; (5) weapons or weapons of opportunity (improvised devices) were nearby; (6) the offense was serious; (7) injury to officers was likely; (8) escape was likely; (9) presence of multiple suspects and/or hostile bystanders; (10) opportunities for force mitigation (such as time, cover, and distance) were available; and (11) backup was available.

Respondents Experts comprised 51% of our sample, while the remaining 49% were novices (see Table 1). Table 1 also shows that 78% of the expert officers surveyed were male, while 83% of the novice officers were male (overall, 80% were male across both groups).

On average, experts held 17 years of experience (Table 2). Unsurprisingly, expert officers had a much higher rate of having completed a degree of higher education (89% versus 34% for novices). Novices were much more likely to have been US military veterans (31%) when compared to expert officers (8%).

Table 1 Descriptive statistics for dependent and independent variables

	Mean	Std. dev.	Min
Sensitivity to mitigation	7.1	3.7	0
Sensitivity to backup	8.1	2.8	0
Graham score	51.8	20	− 12
Male	0.8	0.4	0
BA (no MA)	0.36	0.48	0
MA +	0.26	0.44	0
Veterans	0.19	0.4	0
Expert	0.51	0.5	0
Years of police experience	9.0	9.6	0

Analytic Strategy The analysis is a multi-method approach, unique in its reliance on both multiple-choice survey answers for deductive statistical models and on inductive exploration of written-text answers to open-ended questions. Although relatively uncommon in existing research, combining inductive and deductive approaches in a single study can improve not only analytical accuracy, but also interpretability (Greco et al. 2001).

Our regression models utilized two separate dependent variables that focus on recognizing opportunities to mitigate force, which was the focus of the BJA training development. In the first set of models, the dependent variable is the degree of relevance assigned by the officer in response to the statement that "opportunities for mitigation (such as time, cover, distance, etc.) were available." The answer options were definitely false (coded as "0"), probably false (= 1), neither true nor false (= 2), probably true (= 3), or definitely true (= 4). The second set of regression models used a dependent variable constructed in a similar fashion, though the question in this case specifically asked about the relevance of backup opportunities rather than force mitigation opportunities more generally ("backup was available"). In the case of both variables, answers were aggregated across the five scenarios to form a combined measure of officer sensitivity to force mitigation opportunities that controls for situation-specific factors. The model also includes robust standard errors, a more rigorous way of handling data that are clustered (Arellano 1987). We used a significance level of 0.05 in all of the regression models.

In the textual analysis, we took advantage of the open-ended textual prompts built into the survey design to

Table 2 Comparing experts and novices

	Experts (%)
Percent male	78
Percent BA +	89
Percent MA +	48
Percent veterans	8
Median years of experience	17

inductively identify linguistic patterns that may diverge according to expertise status. For each scenario, each officer answered close-ended survey questions (upon which the regression models were based) and wrote text in response to open-ended prompts. The prompts probed the respondent regarding the relevant cues to which s/he was directing their attention, as well as the actions the respondent would have taken if they had been the officer responding to the incident. The text responses were combined and aggregated across scenarios and decision points for experts and for novices, generating two blocks of text upon which we based our inductive text-mining analyses. These techniques are useful for analyzing text generated by open-ended prompts, such as those used in the scenarios, as the word counts and relative frequencies can be calculated automatically.

The simplest language models represent a document by its word distribution, ranking the frequency of single words. This kind of word frequency analysis, similar to the process used to generate a word cloud, can efficiently highlight linguistic trends that can then be relatively easily compared across groups (Manning and Schütze 1999). However, research on information processing has shown that two-word combinations (known as "bigrams"), or possibly even three-word combinations ("trigrams") prove to be more intuitively useful for capturing information (Kaptein et al. 2010). In our study, we analyze the results of single-word frequency, as well as both bigrams and trigrams. Unlike the case with the regression models, there is no specific outcome we are seeking in the textual analysis. Any differences that appear in how officers describe their decision-making about the use of force in the depicted scenarios are emergent and in no way conditioned on any kind of model specification.

We used additional textual analysis to examine whether we could identify linguistic patterns that more closely track with the regression results, which specifically suggest that experts are more likely to pay attention to the availability of force mitigation options like time, cover, and distance. In addition to tabulating the frequency for the words "time," "cover," and "distance," we also included their synonyms and inflected variants (for example, "mitigate," "mitigated," and "mitigation" would each be counted as "mitigation").

By combining regression techniques and text-analysis techniques, we are able to triangulate deductive (regression analysis) and inductive (text-mining) results to observe multiple patterns in our outcome of interest, thereby increasing the robustness of our analytical claims.

Results

The regression results show that experts were significantly ($p = 0.02$) more likely than novices to report that force mitigation was a relevant factor (Table 3). Even with controls for gender (Table 3, model 2), educational background (model 3), or veteran status (model 4), being an expert versus a novice was significantly predictive, with our model showing experts scoring on average between 13 and 40% higher than their novice counterparts in their evaluation of the importance of force mitigation opportunities. Except for scenario-specific effects, none of the control variables proved to be significant.

Table 3 OLS regression results for expert versus novice sensitivity to force mitigation opportunities

	(1)	(2)	(3)	(4)
Expert	0.967*	0.986*	1.21*	1.25*
	(0.383)	(0.386)	(0.523)	(0.537)
Male		0.392	0.296	0.266
		(0.397)	(0.441)	(0.435)
BA			− 0.140	− 0.125
			(0.522)	(0.527)
Advanced degree			− 0.484	− 0.433
			(0.651)	(0.675)
Veteran				0.284
				(0.570)
Scenario 1	− 1.34**	− 1.33**	− 1.34**	− 1.34**
	(0.416)	(0.418)	(0.422)	(0.423)
Scenario 2	− 5.89***	− 5.88***	− 5.89***	− 5.89***
	(0.554)	(0.555)	(0.557)	(0.557)
Scenario 3	− 4.63***	− 4.62***	− 4.63***	− 4.63***
	(0.505)	(0.510)	(0.512)	(0.511)
Scenario 4	− 0.335	− 0.326	− 0.331	− 0.340
	(0.399)	(0.400)	(0.403)	(0.401)
Constant	9.11***	8.77***	8.92***	8.85***
	(0.413)	(0.580)	(0.719)	(0.753)
Observations	346	346	346	346
R-squared	0.433	0.435	0.436	0.437

Cluster-robust standard errors in parentheses

***$p < 0.001$; **$p < 0.1$; *$p < 0.05$

Additionally, experts seem to be more likely to be aware of opportunities for backup as an important factor in a scenario, net of controls (Table 4). Here we see that educational level is in fact also relevant to officers' relative emphasis on backup opportunity (Table 4, model 3). Although gender (model 2) and veteran status (model 4) have no significant effect on reported relevance of backup opportunities, having an advanced degree (master's degree or higher) is predictive of being less likely to prioritize the availability of backup when making decisions about appropriate force ($p = 0.03$). Interestingly, this effect runs counter to the effect of expertise, which suggests that the skills experts develop over the years certainly cannot be attributed to their formal education. After controlling for the countervailing effect of education, experts scored significantly more highly on the measure of the extent to which they agreed that the availability of backup was a

Table 4 OLS regression results for expert versus novice sensitivity to backup opportunities

	(1)	(2)	(3)	(4)
Expert	0.300	0.298	1.05*	1.14*
	(0.440)	(0.446)	(0.478)	(0.482)
Male		− 0.045	− 0.388	− 0.455
		(0.441)	(0.450)	(0.468)
BA			− 0.643	− 0.618
			(0.472)	(0.472)
Advanced degree			− 1.57*	− 1.47*
			(0.641)	(0.652)
Veteran				0.618
				(0.495)
Scenario 1	1.29**	1.29**	1.29**	1.29**
	(0.421)	(0.422)	(0.427)	(0.428)
Scenario 2	− 1.59**	− 1.59**	− 1.61**	− 1.62**
	(0.446)	(0.448)	(0.455)	(0.457)
Scenario 3	− 1.28**	− 1.28**	− 1.32**	− 1.33**
	(0.448)	(0.449)	(0.454)	(0.453)
Scenario 4	0.409	0.409	0.389	0.373
	(0.403)	(0.403)	(0.406)	(0.405)
Constant	8.14***	8.18***	8.74***	8.60***
	(0.431)	(0.565)	(0.649)	(0.667)
Observations	346	346	346	346
R-squared	0.149	0.149	0.179	0.186

Cluster-robust standard errors in parentheses

***$p < 0.001$; **$p < 0.1$; *$p < 0.05$

relevant factor for evaluating the appropriate level of force for the depicted scenario ($p = 0.02$). Compared to the novices, our model shows that the experts scored, on average, 13–20% higher in their consideration of backup opportunities, net of controls.

The regression results show that the expert group, relative to the novices, have a heightened sensitivity to the presence of opportunities for force mitigation and for backup when evaluating the appropriate level of force for a given scenario. This effect cannot be explained by differences across the scenarios or by other officer traits, such as gender, educational background, or veteran status.

Our first set of text analyses (Table 5) ranks the relative frequency of single words for both experts and novices. These results show that expert officers are more likely to emphasize verbal, de-escalating engagement. In the observed hierarchy of single words for experts and novices, "verbal" emerged as one of the top ten words mentioned by experts (at position no. 4, accounting for 2% of all words); it was not mentioned frequently enough by novices to make their top ten words. Meanwhile, novices used the word "control" at a much higher rate than did experts (at position no. 7, accounting for 2% of all words).

Table 5 Top ten single words for experts and novices (count, and percentage of total)

Word rank	Experts
1	Subject (1213, 9%)
2	Officer (795, 6%)
3	Officers (733, 5%)
4	Verbal (309, 2%)
5	Offender (291, 2%)
6	Knife (215, 2%)
7	Distance (189, 1%)
8	Continue (175, 1%)
9	Weapon (173, 1%)
10	Situation (161, 1%)

These patterns again seem to emerge when we look at bigram or trigram hierarchies. Although the top two bigrams for experts refer to verbal commands ("verbal direction" and "verbal commands," together representing 8% of all two-word combinations), the top bigram for novices is "gain control" (accounting for 3%) (Table 6).

When we examine three-word combinations (Table 7), the top four instances for experts refer to verbal commands (e.g., "give verbal commands"), for a combined 7% of all trigrams represented. Meanwhile, the top two trigrams for novices are "gain control [of] suspect" and "carotid control hold" (i.e., a form of neck restraint; see Mitchell et al. 2012). Comparing across single-world, bigram, and trigram results allows us to see how seemingly neutral words like "continue," which ranked highly for experts in the single-word relative frequency ranking (Table 5), take on new significance when we allow for more textual context (i.e., longer strings of words). The bigram and trigram analyses reveal that the increased likelihood for experts to use this word can be largely attributed to the use of de-escalation-related phrasing like "continue verbal" (Table 6) and "continue give verbal" (Table 7).

One possibility for explaining novice-expert differences in the textual analysis results may be simply that experts have

Table 6 Top ten bigrams for experts and novices (count, and percentage of total)

Bigram Rank	Experts
1	Verbal direction (114, 4%)
2	Verbal commands (108, 4%)
3	One officer (59, 2%)
4	Drop knife (57, 2%)
5	Deadly force (51, 2%)
6	Give verbal (42, 2%)
7	Gain control (38, 1%)
8	Continue verbal (37, 1%)
9	Less lethal (36, 1%)
10	Both officers (33, 1%)

Table 7 Top ten trigrams for experts and novices (count, and percentage of total)

Trigram Rank	Experts
1	Continue give verbal (20, 2%)
2	Give verbal commands (19, 2%)
3	Giving verbal commands (16, 2%)
4	Giver verbal direction (14, 1%)
5	Great bodily harm (13, 1%)
6	Giving verbal direction (11, 1%)
7	Gain control subject (11, 1%)
8	Proactive pat down (11, 1%)
9	Verbal commands subject (10, 1%)
10	Continue give verbal (20, 2%)

homogenized their language over time, as they conform to professional norms surrounding word usage. For example, one might hypothesize that novices, in place of using words like "command," might instead use words like "say." However, this does not seem to be what is driving the results. Although it is true that expert officers are more likely to use words like "command" and that novices are more likely to use words like "say," it is still the case that expert officers overall make greater reference to verbal commands once we expand our dictionary of words that are counted to include standard synonyms ("speak," "say," "tell," "instruct," and "command"). These textual-analysis findings corroborate the regression results, which suggest that expert officers are more likely to emphasize force mitigation opportunities.

Table 8 shows the tabulated frequency for the words "time," "cover," and "distance," as well as their synonyms and inflected variants (for example, "mitigate," "mitigated," and "mitigation" would each be counted as "mitigation"). None of these words were among the top ten most heavily used words, each representing less than 1% of all words generated. Although experts do not have a higher rate of making reference to cover, their written responses do display a strong relative bias towards temporal awareness ("time") and spatial awareness ("distance"), relative to novices.

Although evaluating the aggregated text is a particularly powerful approach because it allows us to look for expert- and novice-specific linguistic patterns that cannot be reduced to the characteristics of any given scenario, zeroing in on individual scenarios in some cases can be quite illustrative too. As the final piece of our multi-part analysis, we examine the word frequency rank for one specific scenario, thereby allowing for the potential that scenario-specific patterns could emerge.

Table 8 Expert versus novice usage of mitigation-related words

	Experts	Novices
Cover	76	115
Time	107	76
Distance	194	103
Mitigation	9	0

In a word-hierarchy analysis of one single scenario (scenario 1) (Table 9), one of the top single words that emerges for experts, but not for novices, is "Jerry." This finding is significant because "Jerry" is the first name of the suspect. The fact that experts are significantly more likely to refer to the suspect by using his name suggests that they are emphasizing verbal de-escalation techniques, which further corroborates our aggregate-level findings. It should also be noted that in scenario 1, as was the case when we analyzed cross-scenario data, experts are more likely than novices to use de-escalation-related terms like "distance," "verbal," "taser," and "continue." These findings also suggest that a word-ranking approach can be productively applied both to identify cross-scenario and scenario-specific ways in which the groups of experts and novices differ from each other.

Discussion

We used novel tools and methods to make explicit the tacit knowledge of experts. While expert knowledge is typically gained over thousands of hours of experience, our approach allows us to identify key components of expert policing that may be useful for training novices. Eliciting expert knowledge—which is typically tacit and therefore difficult for experts to articulate—through traditional knowledge-elicitation techniques is time-consuming and expensive. It typically means looking only at experts and only at a small group of them. Taken together, the regression and textual analyses unequivocally demonstrate that our method can be used to show that one group ("experts"), relative to another ("novices"), have a heightened sensitivity to the presence of opportunities for force mitigation. These results are aligned

Table 9 Top ten single words for the scenario 1 (count, and percentage of total)

Experts	Novices
Subject (268, 7%)	Officers (269, 6%)
Officers (266, 7%)	Suspect (232, 6%)
Knife (120, 3%)	Knife (161, 4%)
Verbal (118, 3%)	Subject (100, 2%)
Distance (111, 3%)	Lethal (91, 2%)
Officer (90, 2%)	Officer (76, 2%)
Taser (87, 2%)	Drop (75, 2%)
Jerry (69, 2%)	Commands (70, 2%)
Continue (66, 2%)	Taser (70, 2%)
Direction (59, 1%)	Towards (65, 2%)

with recent findings that officers' level of experience, measured in years, is positively predictive that officers will use some de-escalation techniques, resulting in calmer citizens (Todak and James 2018). Furthermore, besides having greater awareness of cover, time, distance, or opportunities for backup, experts are also more likely to emphasize verbal strategies for engaging with citizens. This stands in direct contrast to novices, who instead prioritize gaining physical control.

Many behavioral and cognitive skills contribute to effective police work. However, we were able to identify a key cognitive skill—recognition of force mitigation opportunities—as a crucial differentiator between experts and novices. According to our analyses, experts are more likely to emphasize (1) verbal de-escalation, (2) distance as a force mitigation opportunity (3) time as a force mitigation opportunity, and (4) backup opportunities (i.e., getting more personnel on scene). These are policing skills that can be specifically targeted in training, and doing so is not merely a matter of top-down policy recommendation. Rather, these strategies emerge as the ways experts—as compared to new entrants into the police professions—tend to prioritize their actions. While demand for de-escalation training is at an all-time high among police departments in the USA, research on the efficacy of de-escalation training has not kept pace with its implementation (Engel et al. 2019). Although our study does not investigate the effects of de-escalation training directly, our results do suggest that experts systematically diverge from novices when it comes to decision-making about the use of force, favoring de-escalation tactics to a significantly greater extent. In this respect, our findings can be seen as a promising indicator that the recent demand for de-escalation training across the country is in fact warranted.

The practical implications of this work for concrete training interventions are clear. The results of this study can be used to focus future training on force mitigation as a way to accelerate novice learning to bring them more rapidly in line with expert reasoning on use of force. Not only are we able to focus our identification of learning objectives, but we are able to hone in on the critical structures of expert reasoning and specific use-of-force situations to generate training feedback (expert insights) for video-based decision-making exercises. We learned it would be wise to specifically target novices' tendency to emphasize physical control over all other techniques, i.e., novices need a broader conception of "control" than is generally taught in police academies or Field Training programs; they need a conception that places an emphasis on attention to verbal engagement, distance, time, and cover over an emphasis on physical control. Consequently, we suggest a greater emphasis on spatial-reasoning tasks that can serve as a way of deliberately training recognition of meaningful spatial patterns and action options (e.g., where to stand in relation to threats, cover, concealment, or team members). Tactical decision exercises similar to those reported here generally show promise for addressing such aspects of tactical decision-making not normally addressed in training (Wolfe et al., in press).

Because our sample of experts and our sample of novices were taken from different departments, this study cannot stand in for a true comparison of novices and expert use of force, as such a comparison would necessarily require controls for department size and culture. Although we do demonstrate that experienced officers are more likely to emphasize de-escalation, this finding could be wholly related to their particular department's culture rather than to the fact that they are experts. Although one might argue that the kinds of perceptual-cognitive aspects of performance measured here (particularly among novices, who are relatively lacking in experience and exposure) may not be as susceptible to influence by local departmental culture, we do know that organizational-level variation across police departments is significant for understanding use of force. Agency policies (Alpert and Rojek 2011) and management practices (Wolfe and Piquero 2011) influence police conduct, as do broader regional and city-specific factors. Recent experiences on the job can also significantly impact expert performance (Klham and Tillyer 2015). Although we do offer suggestive evidence regarding the relationship between expertise and the use of force, additional research is needed to determine whether this finding is truly a function of expertise, or whether it is linked to city-specific or organization-specific factors. To address these concerns, future research should endeavor to recruit both expert and novice participants from the same agency.

Rather than providing definitive evidence for expert-versus-novice approaches to use of force, our research's strength lies in its demonstration of the potential utility of a methodological approach that is both simple and powerful. Our mixed-methods inductive approach can be applied to future research to more precisely pin down these kinds of relationships, thereby informing how we research the use of force and how we build a solid foundation for improved training in the police. And the methodological strategy we have developed in this paper could be applied to study any particular group of experts, to find out what is "expert" about them. All in all, these practical interventions have the potential to make meaningful interventions with relatively low levels of new investment in training.

Funding Information This work was supported by the Department of Justice within the Office of Justice Programs (grant number BJA-2016-VI-BX-K005).

Compliance with Ethical Standards

Conflict of Interest Not applicable

Ethical Approval This research was classified as "exempt" under Title 45 of the Code of Federal Regulations, §46.104, Exempt Research, Category 3(i). Research involved collection of information from adult subjects via written responses/data entry. Subject prospectively agreed to participate in data collection and the information obtained was recorded by investigators in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects.

Informed Consent All participants gave informed consent.

References

Alpert GP & Rojek J (2011) Frontline police officer assessments of risks and decision making during encounters with offenders (briefing paper no. 5). Nathan, Australia: ARC Centre of Excellence in Policing and Security

Alpert GP, Smith WC (1994) How reasonable is the reasonable man: police and excessive force. J Crim Law Criminol 85:481–501. https://doi.org/10.2307/1144107

Arellano M (1987) Computing robust standard errors for within-groups estimators. Oxf Bull Econ Stat 49:431–434. https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x

Basit T (2003) Manual or electronic? The role of coding in qualitative data analysis. Educ Res 45:143–154. https://doi.org/10.1080/0013188032000133548

Bittner E (1990) Aspects of police work. Northeastern University Press, Boston

Bolger PC (2015) Just following orders: a meta-analysis of the correlates of American police officer use of force decisions. Am J Crim Justice 40:466–492. https://doi.org/10.1007/s12103-014-9278-y

Boulton L, Cole J (2016) Adaptive flexibility: examining the role of expertise in the decision making of authorized firearms officers during armed confrontation. J Cogn Eng Decis Making 10:291–308. https://doi.org/10.1177/1555343416646684

Graham V. Connor (1989) 490 U.S. 386

Crandall BW, Klein GA, Hoffman RR (2006) Working minds: a practitioner's guide to cognitive task analysis. MIT Press, Cambridge

Engel R, McManus H, & Herold T (2019). The deafening demand for de-escalation training: a systematic review and call for evidence in police use of force reform. Unpublished manuscript. https://www.theiacp.org/sites/default/files/IACP_UC_De-escalation%20Systematic%20Review.pdf. Accessed December 4 2019

Ericsson KA, Krampe RT, Tesch-Römer C (1993) The role of deliberate practice in the acquisition of expert performance. Psychol Rev 100: 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Evans JS (2008) Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol 59:255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Geller WA, Toch H (eds) (1995) And justice for all: understanding and controlling police abuse of force. Police Executive Research Forum, Washington

Greco S, Masciari E, Pontieri L (2001) Combining inductive and deductive tools in data analysis. AI Commun 14:69–82

Harris KR, Eccles DW, Freeman C, Ward P (2017) 'Gun! Gun! Gun!': An exploration of law enforcement officers' decision-making and coping under stress during actual events. Ergonomics 60:1112–1122. https://doi.org/10.1080/00140139.2016.1260165

Hine, K. A., Porter, L. E., Westera, N. J., Alpert, G. P., & Allen, A. (2018). What were they thinking? Factors influencing police recruits' decisions about force Polic Soc Advance online publication https://doi.org/10.1080/10439463.2018.1432612

Hoffman RR (1987) The problem of extracting the knowledge of experts from the perspective of experimental psychology. AI Mag 8(2):53–67. https://doi.org/10.1609/aimag.v8i2.583

Hoffman RR (1989) A survey of methods for eliciting the knowledge of experts. ACM SIGART Bull 108:19–27. https://doi.org/10.1145/63266.63269

Hoffman RR, McNeese MD (2009) A history for macrocognition. J Cogn Eng Decis Making 3:97–110. https://doi.org/10.1518/155534309x441835

Hoffman RR, Militello L (2008) Perspectives on cognitive task analysis: historical origins and modern communities of practice. Psychology Press, New York

Johnson RR, Stone BT, Miranda CM, Vila B, Lois J, James SM et al (2014) Identifying psychophysiological indices of expert versus novice performance in deadly force judgment and decision making. Front Hum Neurosci 8. https://doi.org/10.3389/fnhum.2014.00512

Kaptein R, Hiemstra D, Kamps J (2010) How different are language models and word clouds? In: Proceedings of the European Conference on Information Retrieval, vol 32, pp 556–568. https://doi.org/10.1007/978-3-642-12275-0_48

Klein GA (2017) Sources of power: How people make decisions, 20th edn. MIT press, Cambridge

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (Eds.). (1993). Decision making in action: models and methods. Westport: Ablex Publishing

Klein GA, Ross KG, Moon BM, Klein DE, Hoffman RR, Hollnagel E (2003) Macrocognition. IEEE Intell Syst 18(3):81–85. https://doi.org/10.1109/MIS.2003.1200735

Klein GA, Phillips JK, Rall EL, Peluso DA (2007) A data-frame theory of sensemaking. In: Hoffman RR (ed) Expertise out of context: proceeding of the sixth international conference on naturalistic decision making. Lawrence Erlbaum Associates, New York, pp 113–155

Klein G, Klein HA, Lande B, Borders J, Whitacre JC (2015) Police and military as good strangers. J Occup Organ Psychol 88:231–250. https://doi.org/10.1111/joop.12110

Klein G, Shneiderman B, Hoffman RR, Ford KM (2017) Why expertise matters: a response to the challenges. IEEE Intell Syst 32(6):67–73. https://doi.org/10.1109/MIS.2017.4531230

Klham CF, Tillyer R (2015) Rethinking the measurement of officer experience and its role in traffic stop searches. Police Q 18(4):343–367. https://doi.org/10.1177/1098611115585315

Klockars CB (1995) A theory of excessive force and its control. In: Geller WA, Toch H (eds) Police violence: understanding and controlling police abuse of force. Police Executive Research Forum, Washington, pp 1–22

Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge

Mitchell JR, Roach DE, Tyberg JV, Belenkie I, Sheldon RS (2012) Mechanism of loss of consciousness during vascular neck restraint. J Appl Physiol *112*:396–402. https://doi.org/10.1152/japplphysiol.00592.2011

Rolfe J (2005) *Decision making of police forward commanders in high-risk situations* (Unpublished doctoral dissertation). Queensland University of Technology, Brisbane

Skolnick JH, Fyfe JJ (1993) Above the law: Police and the excessive use of force. Free Press, New York

Suss J, Boulton L (2019) Expertise in law enforcement. In: Ward P, Schraagen JM, Gore J, Roth EM (eds) The Oxford handbook of expertise. Oxford University Press, Oxford, pp 765–791

Suss J, Ward P (2018) Revealing perceptual–cognitive expertise in law enforcement: an iterative approach using verbal-report, temporal-occlusion, and option-generation methods. Cogn Technol Work 20:585–596. https://doi.org/10.1007/s10111-018-0493-z

Suss J, Belling PK, & Ward P (2014) Use of cognitive task analysis to probe option-generation in law enforcement. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58, 280–284. https://doi.org/10.1177/1541931214581058

Taylor TZ, Elison-Bowers P, Werth E, Bell E, Carbajal J, Lamm KB, Velazquez E (2013) A police officer's tacit knowledge inventory

(POTKI): establishing construct validity and exploring applications. Police Pract Res 14:478–490. https://doi.org/10.1080/15614263.2013.802847

Todak N, James L (2018) A systematic social observation study of police de-escalation tactics. Police Q 21(4):509–543. https://doi.org/10.1177/1098611118784007

Vickers JN, Lewinski W (2012) Performing under pressure: gaze control, decision making and shooting performance of elite and rookie police officers. Hum Mov Sci 31:101–117. https://doi.org/10.1016/j.humov.2011.04.004

Ward P, Suss J, Eccles DW, Williams AM, Harris KR (2011) Skill-based differences in option generation in a complex task: a verbal protocol analysis. Cogn Process 12:289–300. https://doi.org/10.1007/s10339-011-0397-9

Wolfe SE, Piquero AR (2011) Organizational justice and police misconduct. Crim Justice Behav 38:332–353. https://doi.org/10.1177/0093854810397739

Worden RE (1995) The "causes" of police brutality: theory and evidence on police use of force. In: Geller WA, Toch H (eds) And justice for all: Understanding and controlling police abuse of force. Police Executive. Research Forum, Washington, pp 31–60

Zimmerman LA (2006) Law enforcement decision making during critical incidents: a three-pronged approach to understanding and enhancing law enforcement decision processes (Doctoral dissertation). Retrieved from https://digitalcommons.utep.edu/dissertations/AAI3214014/

Zimmerman LA (2008) Making sense of human behavior: explaining how police officers assess danger during traffic stops. In: Schraagen JM, Militello L, Ormerod T, Lipshitz R (eds) Naturalistic decision making and macrocognition. Ashgate Publishing, Aldershot, pp 121–140