The European Journal of International Law Vol. 30 no. 4

Outcome Bias and Expertise in Investigations under International Humanitarian Law

Tomer Broude* and Inbar Levy**

Abstract

Many international law decisions are made by individuals, often possessed with expertise, legal or otherwise. We examine individual international humanitarian law (IHL) decision-making on two levels: military decisions made ex ante *regarding real-time operational questions under conditions of uncertainty and imperfect information, and subsequent* ex post *evaluations of the propriety of military decisions in the context of military investigations regarding legal responsibility with respect to proportionality and reasonableness. IHL requires* ex post *investigators to consider only information available at the time decisions were made. Through an experimental vignette study conducted with laypersons, legal experts and people with field experience, we test whether they are susceptible to cognitive 'outcome bias', specifically the extent to which the knowledge of operational outcomes, especially regarding incidental civilian harm, influences* ex post *normative evaluations. Our results demonstrate a general tendency towards outcome bias, which is somewhat tempered by expertise. Individuals with operational decision-making experience may be less prone to outcome bias than legal experts. We discuss possible implications for the design of military investigations relating to IHL.*

1 Introduction: Individual Decision-Making in International Humanitarian Law

One objection to the application of psychology to international law relates to the 'methodological individualism' underlying behavioural research, which contradicts

* Bessie and Michael Greenblatt, QC Chair in International and Public Law, Faculty of Law and Department of International Relations; Academic Director, Minerva Center for Human Rights, Hebrew University of Jerusalem, Israel. Email: tomerbroude@gmail.com. The authors thank Henry Bantick, Noa Hirsh and Dana Kleiman for diligent research assistance, as well as numerous colleagues for helpful comments at several workshops and conferences; Christoph Engel, Tom Noah, William Saban, Adi Leibovitch, Tom Nachbar, Yahli Shereshevsky and Raanan Sulitzeanu-Kenan were particularly helpful. Special thanks to Bruce Oswald and the Asia Pacific Centre for Military Law at Melbourne Law School for generous support.
** DPhil (Oxon), Lecturer, Melbourne Law School, Australia. Email: inbar.levy@unimelb.edu.au.

statist and institutional characteristics of international law.[1] Individual behavioural observations cannot simply be transposed to other levels of analysis,[2] even if they suggest falsifiable hypotheses on state conduct.[3] Yet many international law decisions are made by individuals (or by small 'elite' decision-making groups),[4] often possessed with expertise, legal or otherwise. States incur international responsibility for individual acts,[5] and personal responsibility for violations of international criminal law is well established.[6] Individual interactions with international law can be studied with the same methods as other legal systems while raising correlative methodological issues – for example, external validity.[7]

We provide an indicative research example focusing on individual international humanitarian law (IHL) decision-making on two levels. The first level involves military decision-makers – soldiers, commanders and legal counsel[8] 'calling the shots' *ex ante* on real-time operational questions before an incident occurs (or, rather, is initiated) – in conditions of uncertainty and imperfect information.[9] The second decision-making level comprises subsequent *ex post* evaluations of the propriety of military decisions in the context of military investigations regarding legal responsibility. The latter cannot directly affect operational outcomes, although they may indirectly influence future decisions, through mechanisms of deterrence and other behaviour-forming mechanisms. Importantly, in this article, we are mainly concerned with the second level, particularly the extent to which investigators take into account operational outcomes,

1 The term is used loosely to refer to cognitive psychology's focus on the individual. L. Udehn, *Methodological Individualism: Background, History and Meaning* (2001). For a stylized non-psychological approach in international law, see Megiddo, 'Methodological Individualism', 60 *Harvard Journal of International Law* (*HJIL*) (2019) 601.

2 Broude, 'Behavioral International Law', 163 *University of Pennsylvania Law Review* (2015) 1099, at 1124; van Aaken, 'Behavioral International Law and Economics', 55 *HJIL* (2014) 425, n. 50; but see Engel, 'The Behaviour of Corporate Actors: How Much Can We Learn from the Experimental Literature?', 6(4) *Journal of Institutional Economics* (2010) 445.

3 See, e.g., Galbraith, 'Treaty Options: Towards a Behavioral Understanding of Treaty Design', 53(2) *Virginia Journal of International Law* (2013) 309.

4 Saunders, 'No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making', 71(1) *International Organization* (2017) 219.

5 International Law Commission, Articles on Responsibility of States for Internationally Wrongful Acts, UN Doc A/56/83, 3 August 2001, ch. II.

6 Rome Statute of the International Criminal Court 1998, 2187 UNTS 90, Art. 25.

7 For early discussion of external validity in behavioral research in law, see Konečni and Ebbesen, 'External Validity of Research in Legal Psychology', 3(1–2) *Law and Human Behavior* (1979) 39; in the context of international law, see Dunoff and Pollack, 'Experimenting with International Law', 28(4) *European Journal of International Law* (*EJIL*) (2017) 1317.

8 Military advocates general or judge advocates general play crucial roles in operational decision-making; see, e.g., Lohr and Gallotta, 'Legal Support in War: The Role of Military Lawyers', 4(2) *Chicago Journal of International Law* (2003) 465.

9 Military decision-makers may not be directly concerned with international law but, rather, with their particular military organization's rules and manuals, which ideally reflect international humanitarian law (IHL) principles, and play a (not uncontroversial) role in the establishment of customary IHL, extensively using military manuals. J.-M. Henckaerts and L. Doswald-Beck, International Committee of the Red Cross, *Customary International Humanitarian Law* (2006); *contra*, Garraway, 'The Use and Abuse of Military Manuals', 7 *Yearbook of International Humanitarian Law* (*YIHL*) (2004) 425.

even though these are not part of the information available to military decision-makers at the time of the decision and, in IHL, are not formally part of the investigatory calculus at all.

We examine combat situations construed to be governed by *jus in bello*. Broadly stated, IHL imposes standards of proportionality and (even less determinately) reasonableness on soldiers and commanders regarding military attacks that may incur 'incidental harm' upon civilians. In addition to these substantive requirements, IHL and other areas of international law (for example, human rights law) require investigations of combat incidents following their occurrence, with potential international legal implications, whether at the individual (criminal) or state levels of responsibility.[10] We will primarily consider the effects of a cognitive 'outcome bias' regarding IHL-related investigations.[11] As we will elaborate below, outcome bias causes the evaluation of the quality of a decision to be influenced by the knowledge of the decision's outcomes.

First, we hypothesize that, although standards of operational reasonableness and proportionality require investigators to consider only information available at the time decisions were made, knowledge of outcomes, especially regarding civilian damage, can have significant effects on the normative evaluation of the actions taken. Second, we hypothesize about classes of individuals who may be assigned to investigate operational actions and establish legal responsibility, depending on their expertise. We explore distinctions between laypersons, experts with military field experience and international law experts. Legal and military experts have different training and develop different skills as well as mindsets. These different background factors can influence perceptions of risk, evaluations of propriety and the comprehension of conditions faced by first-level operational decision-makers.

To our knowledge, no similar studies have been conducted regarding outcome bias and expertise in IHL. Experimental research on proportionality in operational decision-making has hitherto focused on ideological biases influencing numerical assessments of outcomes.[12] Our focus is on non-ideological cognitive biases. Part 2 sketches elements of IHL proportionality and the duty to investigate. Part 3 discusses outcome bias and the implications for military investigations. Part 4 describes and explains our experimental approach and presents our results. Part 5 concludes with some remarks on both normative and prescriptive implications and on the utility (and limits) of experimental research on individual decision-making for public international law.

[10] See Cohen and Shany, 'Beyond the Grave Breaches Regime: The Duty to Investigate Alleged Violations of International Law Governing Armed Conflicts', 14 *YIHL* (2011) 37.

[11] See Baron and Hershey, 'Outcome Bias in Decision Evaluation', 54(4) *Journal of Personality and Social Psychology* (1988) 569. This is distinct from the 'hindsight bias' (see subpart 3.A below), in which people overvalue, *ex post*, the foreseeability of an action's outcome. See Blank, Musch and Pohl, 'Hindsight Bias: On Being Wise after the Event', 25(1) *Social Cognition* (2007) 1; E. Zamir and D. Teichman, *Behavioral Law and Economics* (2018), at 336.

[12] See Sulitzeanu-Kenan, Kremnitzer and Alon, 'Facts, Preferences, and Doctrine: An Empirical Analysis of Proportionality Judgment', 50(2) *Law and Society Review* (2016) 348.

2 *Ex Ante* and *Ex Post* Assessments of the Proportionality and Reasonableness of Military Actions

The international legal stage for our experimental research comprises two elements reflecting the double individual focus discussed above. First, there is the application of the substantive standards of proportionality and reasonableness under IHL. Because our research is indicative rather than comprehensive, and its main concern is decision-making, we deliberately simplify and avoid otherwise important technical legalities (such as the characterization of IHL in relation to international human rights law, the classifications of conflict and combatants, territorial application and so forth). Rather, we focus on the *ex ante* nature of operational decision-making under conditions of risk and uncertainty. Second, and most importantly for this study, we discuss the scope of the duty to investigate after an attack to determine its conformity to these substantive parameters, concentrating on *ex post* evaluation.

A Proportionality and Reasonableness in IHL

IHL balances between military advantage and humanitarian considerations through the principle of proportionality expressed, *inter alia*, in Article 51(5)(b) of Additional Protocol I to the Geneva Conventions relating to international armed conflict, which prohibits any attack 'which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, … which would be excessive in relation to the concrete and direct military advantage anticipated'.[13] This principle is considered customary international law, applicable to both international armed conflict and non-international armed conflict, *mutatis mutandis*.[14] Proportionality also applies with respect to the right to life in general human rights law, with a comparable, yet different, formulation.[15]

In any case, proportionality is very much a value judgment with both objective and subjective elements.[16] It borders on the principle of distinction between civilian and military targets because disproportionate attacks may be considered indiscriminate.[17] Thus, reasonableness has been raised as a standard for assessing the quality

[13] Additional Protocol I to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts 1977, 1125 UNTS 3.

[14] On the customary status of this and similar principles, see, e.g., Trial Judgment, *Prosecutor v. Kupreškić et al* (IT-96-16-T), Trial Chamber, 14 January 2000, § 524.

[15] See United Nations (UN) Human Rights Committee, General Comment no. 36 (2018) on Article 6 of the ICCPR on the Right to Life, Doc. CCPR/C/GC/36, 30 October, 2018. *Stricto sensu*, proportionality in this regard is formulated differently in general international human rights law ('the amount of force applied cannot exceed the amount strictly needed for responding to the threat'; para. 12), but violations of the IHL principle of proportionality would constitute violations of international human rights law as well (para. 64).

[16] See International Committee of the Red Cross (ICRC), *The Principle of Proportionality in the Rules Governing the Conduct of Hostilities under International Humanitarian Law*, prepared and edited by Laurent Gisel (2018), at 52.

[17] *Ibid.*, at 53.

of proportionality decisions – the 'reasonable military commander' standard,[18] which assumes that 'many military commanders will agree on whether incidental harm is clearly disproportionate to the military advantage'.[19] Beyond these margins, however, there may be significant differences in views on what is proportionate and reasonable. The fluidity of proportionality and reasonableness stems not only from the difficulty in assigning quantitative values to human lives but also to military gains or necessity of response to a threat, as the case may be.

In the formulation of proportionality, the words 'expected' and 'anticipated' are the most significant for present purposes. The standard of conduct is independent of the actual outcome of an attack. Rather, it relates to the quality of the decision made on the basis of the information available, within reasonable efforts, to the military decision-maker at the time of the decision – that is, on an *ex ante* basis and 'not with the benefit of hindsight'.[20] More incidental civilian harm than the attacker expected in these circumstances may be the mistaken result of flawed information or unanticipated intervening factors, but the attack might still be viewed as proportionate and reasonable under IHL. The same could be said of a military advantage that was not ultimately achieved, although reasonably anticipated.

B The Duty to Conduct Military Investigations on the Basis of Ex Ante Information

In order to monitor and avoid such mistakes and promote accountability, international law generally requires states to investigate, *ex post*, civilian deaths in military actions.[21] The formal standard of review remains, however, the quality of the decisions *ex ante*. Thus, for example, in the 2009 Kunduz fuel tankers case (on which we base one of our experiments), the Prosecutor-General at Germany's Federal Court of Justice investigated whether international or domestic crimes had been committed during an airstrike against two NATO-associated fuel tankers, stolen by the Taliban. The airstrike, conducted by the US Air Force at the call of German ground forces, resulted in the deaths of dozens of civilians, raising concerns that proportionality and perhaps distinction had been transgressed. The Prosecutor-General correctly stated that proportionality must be assessed from the attacker's perspective at the time of the attack, without retrospection according to the actual outcomes,[22] ultimately determining that, even if the civilian harm had been anticipated, it would be considered proportionate and reasonable.

[18] International Criminal Tribunal for the former Yugoslavia, *Final Report to the Prosecutor Reviewing the NATO Bombing Campaign in the FRY* (2004), para. 50, available at www.icty.org/en/press/final-report-prosecutor-committee-established-review-nato-bombing-campaign-against-federal; see also Judgement and Opinion, *Prosecutor v. Galić* (IT-98-29-T), Trial Chamber I, 5 December 2003, para. 58.

[19] ICRC, *supra* note 16, at 53.

[20] *Ibid.*, at 9.

[21] See, e.g., Israel's Turkel Commission, *The Public Commission to Examine the Maritime Incident of 31 May, 2010*, Second Report, February 2013, ch. B.

[22] Public Prosecutor-General of the Federal Court of Justice, Einstellungsvermerk, 3BJs 6/10–4, Karlsruhe, 16 April 2010, at 65, available at www.generalbundesanwalt.de/docs/einstellungsvermerk20100416offen.pdf.

The Israeli High Court of Justice in the *Targeted Killings* case (which is the basis for another one of our experiments) recommended a slightly different approach.[23] Justice Aharon Barak suggested that 'targeted killings' must be subjected to both *ex ante* and *ex post* scrutiny. Prior to the attack, a thorough examination must be conducted of every scenario potentially causing incidental harm. After the attack, an *ex post* review should be conducted, subject to judicial review, still based on the military commander's *ex ante* knowledge. Investigations *ex post* ostensibly aim to ensure accountability of *ex ante* decisions. Consequently, they may influence *ex ante* decision-making. Commanders and other individual decision-makers, aware that their decisions will be investigated *ex post*, are more likely to give due consideration to all possibilities when making operational decisions.[24]

3 Hypotheses: Military Investigations, Outcome Bias and Expertise

This sketch of proportionality and investigation issues in IHL raises significant questions. For example, who is best qualified to evaluate military decisions *ex post* and how? We consider two empirical questions in this regard: is outcome bias a factor in such evaluations and are decision-makers differentially affected by this bias, depending on relevant experience and expertise? On this basis, we develop experimental hypotheses, which are set out below.

A Outcome Bias and Military Investigations

Outcome bias is the tendency to evaluate the quality of decisions in light of their outcome. In particular, people tend to evaluate decisions with poor outcomes more harshly than decisions with favourable outcomes, even when the probabilities in real time were identical and known to the evaluators.[25] In experiments concerning either medical matters or monetary gambles (that is, in settings that are both non-military and non-legal), participants were provided with descriptions and outcomes of decisions made under conditions of uncertainty and asked to rate the quality of decisions. The results showed that knowledge of outcomes influenced the evaluation of the decision.[26] Other studies have shown that people's perceptions of probability shift with

[23] High Court of Justice (Israel) 769/02, *Public Committee Against Torture in Israel et al. v. Govt. of Israel et al.*, 11 December 2005, PD 62(1) 507 (2006) (Isr.). For analysis, see Cohen and Shany, 'A Development of Modest Proportions: The Application of the Principle of Proportionality in the *Targeted Killings* Case', 5 *Journal of International Criminal Justice* (2007) 310, especially at 317–318.

[24] See Cohen and Shany, *supra* note 23, at 319; and more generally Livnat, Sulitzeanu-Kenan and Kogut, 'Foresighted Outcome Effect: A Micro-foundation of Agents' Risk Aversion in Principal-Agent Relations', 1(1) *Journal of Behavioral Public Administration* (2018) 1.

[25] Teichman, 'The Hindsight Bias and the Law in Hindsight', in E. Zamir and D. Teichman (eds), *The Oxford Handbook of Behavioral Economics and the Law* (2014) 3.

[26] See Baron and Hershey, *supra* note 11.

knowledge of outcomes, though non-cognizant of these effects.[27] Outcome bias is related to, but distinct from, 'hindsight bias' – the inclination, after an event's outcomes are known, to view them as having been predictable.[28] Both outcome bias and hindsight bias have implications for law[29] and, in some circumstances, may overlap, but outcome bias is of greater interest insofar as investigations are concerned, given its focus on quality rather than probability.

What might outcome bias mean for IHL investigations? We do not aim to identify new cognitive biases but only to connect them to international law.[30] We consider that outcome bias is reflected in IHL investigations of military events and could distort their process and results. Investigators assess the conduct of first-level military decision-makers, usually after exposure to outcomes; investigations are generally launched only when incidental civilian harm is high or if outcomes have otherwise been excessive. While such outcomes may be indirectly relevant to the legality of the decisions – for example, as evidence of extreme cases of indiscriminate attacks,[31] we focus on situations in which investigators must evaluate proportionality and reasonableness only according to information that decision-makers had *ex ante* under conditions of uncertainty. This leads us to our first hypothesis:

Hypothesis 1: *ex post* evaluations of military decision-making under IHL will be influenced by knowledge of outcomes. If outcomes are poor/favourable, evaluators will deem decisions disproportionate and unreasonable or proportionate and reasonable.

B Expertise and Outcome Bias

If knowledge of outcomes affects investigations, might investigators' operational and/or legal expertise be a modifying factor in assessments of decision-making quality? In psychological literature, expertise is the ability to develop skilled intuition, acquired when people store experience-based information, to be accessed quickly, when relevant.[32] In other words, experts access 'system 1' (fast, intuitive) thinking instead of

[27] Fischhoff, 'Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment under Uncertainty', 12(4) *Quality and Safety in Health Care* (2003) 304.

[28] Wood, 'The Knew-It-All-Along Effect', 4(2) *Journal of Experimental Psychology: Human Perception and Performance* (1978) 345; Zamir and Teichman, *supra* note 11 at 337: '[H]indsight bias focuses on evaluators' ex-post assessments of probability … outcome bias focuses on evaluators' assessments of the *quality* of decisions.'

[29] Rachlinski, 'A Positive Psychological Theory of Judging in Hindsight', 65(2) *University of Chicago Law Review* (1998) 571.

[30] In comparison, for a recent study on outcome/hindsight biases and criminal adjudication, see Kern Griffin, 'Criminal Adjudication, Error Correction, and Hindsight Blind Spots', 73 *Washington and Lee Law Review* (2016) 165.

[31] See, e.g., Fellmeth *et al.*, 'Targeting Decisions and Consequences for Civilians in the Colombian Civil Strife', 26 *Minnesota Journal of International Law* (2017) 501; but see Appeal Judgement, *Prosecutor v. Blaškić* (IT-95-14-A), Appeals Chamber, 29 July 2004, §§ 463–464, where the Appeals Chamber reproached the Trial Chamber's reasoning, which regarded civilian casualties (that is, the outcome) as evidence of illegality.

[32] Expertise is 'the possession of domain-specific knowledge … acquired through experience or training … that leads to superior performance'. Zamir and Teichman, *supra* note 11, at 115. See Larrick and Feiler, 'Expertise in Decision-Making', in G. Keren and G. Wu (eds), *Wiley Blackwell Handbook of Judgement and Decision-Making* (2015) 696, at 697.

'system 2' (slow, deliberative).[33] Two conditions for acquiring such intuitive judgment skills that reflect true expertise are: (i) an environment that is sufficiently regular to be predictable and (ii) an opportunity to learn these regularities through prolonged practice.[34] A prime example of expertise is chess players. It is estimated that chess masters and grandmasters store about 50,000 chess positions in their memory, to be applied when relevant. It takes about 10 years to develop this kind of expertise.[35] Chess, however, is an extreme example of regularity – a set board, players and clear rules. Does the military legal environment allow for this type of learning? What makes a military investigator an expert? Is it accumulated field experience or legal knowledge? And is it at all possible to develop such expertise in the context of military investigations? These are the questions we are trying to address, indicatively, in this article, through the perspective of outcome bias.[36]

Moreover, we are not concerned with the speed of the *ex post* decision, which usually enjoys slow 'system 2' conditions but, rather, with its quality/qualities, in relation to the IHL requirement that first-level decisions be assessed on the basis of the information available at the time. While no expert can be expected to anticipate every outcome of military operations – given the overarching conditions of uncertainty – we postulate that there is something about practical experience, real-life training and perhaps also doctrinal training that could establish expertise that is beyond mere academic knowledge. These could, potentially, mitigate outcome bias.[37]

In our current empirical examination, we test whether field commanders and legal advisors possess relevant decision-making expertise that enables them to discount *ex post* knowledge of outcomes when irrelevant from a formal legal perspective, and we form our second hypothesis:

> Hypothesis 2: individuals with field experience and IHL training will be less susceptible to outcome bias in *ex post* quality evaluations of military decisions under IHL than inexperienced and untrained individuals.

[33] D. Kahneman, *Thinking, Fast and Slow* (2011).

[34] Kahneman and Klein, 'Conditions for Intuitive Expertise: A Failure to Disagree', 64(6) *American Psychologist* (2009) 515, at 519.

[35] Charness *et al.*, 'The Role of Deliberate Practice in Chess Expertise', 19 *Applied Cognitive Psychology* (2005) 151; Gobet and Simon, 'Recall of Random and Distorted Chess Positions: Implications for the Theory of Expertise', 24 *Memory and Cognition* (1996) 493.

[36] For more on intuition and legal decision-making, see Griffin and Kahneman, 'Judgmental Heuristics: Human Strengths or Human Weaknesses?', in L.G. Aspinwall and U.M. Staudinger (eds), *A Psychology of Human Strengths: Fundamental Questions and Future Directions for a Positive Psychology* (2003) 319; Guthrie, Rachlinski and Wistrich, 'Blinking on the Bench: How Judges Decide Cases', 93 *Cornell Law Review* (*CLR*) (2007) 1; Guthrie, Rachlinski and Wistrich, 'Inside the Judicial Mind', 86 *CLR* (2001) 777.

[37] On professionals acting in real-life situations (firefighters), see G.A. Klein, *Sources of Power: How People Make Decisions* (1998). Our concern is not the expertise of field decision-makers but, rather, of *ex post* evaluators. In the international legal context, one path-breaking study conducted with legal academics suggests that international law experts may be better at discounting irrelevant information in the process of treaty interpretation (Shereshevsky and Noah, 'Does Exposure to Preparatory Work Affect Treaty Interpretation? An Experimental Study on International Law Students and Experts', 28(4) *EJIL* (2017) 1287), which is similar to the hypothesis that IHL experts will be better at discounting outcomes.

Within this hypothesis, we add another element relating to potential differences in evaluations by experts with decision-making experience in the field and those whose expertise is based only on legal training who do not have field experience. As noted above, the psychological literature affords great weight to experience-based expertise. Our third hypothesis holds that, regardless of legal instructions, people with field experience and expertise will display smaller tendencies towards *ex post* judgment influenced by revealed outcomes than experts with legal training without field experience:

Hypothesis 3: individuals with field experience will be more consistent in their *ex post* evaluations, independent of outcome knowledge, than individuals with legal training only.

4 Experimenting with Expertise: Assessing Outcome Bias

We now present our research design and results with respect to these hypotheses. The results are displayed in Table 1. Additional information regarding design and data can be found in this article's Supplementary Appendices.[38]

A Methodology and Research Design

In order to test the effects of outcome bias in the IHL context, regarding these three hypotheses, we conducted a series of vignette experiment studies[39] in Israel, Australia[40]

Table 1: *McCann* scenario

	Use of force reasonable (%)	Use of force not reasonable (%)	Use of force lawful (%)	Use of force not lawful (%)
Negative outcome				
MLS students	26 (34.2)	50 (65.8)	37 (48.7)	39 (51.3)
Israeli students	50 (32.7)	103 (67.3)	81 (52.9)	72 (47.1)
ADF officers	15 (42.9)	20 (57.1)	17 (48.6)	18 (51.4)
International law experts	17 (37)	29 (63)	19 (41.3)	27 (58.7)
IDF field experience	18 (41.9)	13 (58.1)	18 (58.1)	13 (41.9)
Positive outcome				
MLS students	43 (62.3)	26 (37.7)	47 (68.1)	22 (31.9)
Israeli students	89 (54.3)	75 (45.7)	116 (70.7)	48 (29.3)
ADF officers	21 (77.8)	6 (22.2)	21 (77.8)	6 (22.2)
International law experts	22 (64.7)	12 (35.3)	27 (79.4)	7 (20.6)
IDF field experience	17 (50)	17 (50)	26 (76.5)	8 (23.5)

[38] The Supplementary Appendices for this article are available at https://academic.oup.com/ejil/article-lookup/doi/10.1093/ejil/chaa005#supplementary-data.

[39] On vignette studies, see Steiner *et al.*, 'Experiments for Survey Research: A Case Study on the Fair Gender Research Gap', 7(2) *Journal of Methods and Measurement in the Social Sciences* (2016) 52.

[40] Note, relatedly, the differing articulations of proportionality in Israeli and Australian military manuals. See ICRC, IHL Database, available at https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule14.

and online.[41] Experiments were conducted with university students (some of the latter with military experience, as explained below), Australian Defence Force (ADF) officers and others. Participants were given descriptions of three different operational military cases and were asked to make legal determinations as *ex post* evaluators of the proportionality, reasonableness and legality of the operational decisions made in each case. Each participant in the study was randomly assigned a different outcome for each case (positive or negative outcome). The sequence of cases for evaluation was standard, not randomized, fixing possible ordinal effects. Assessments under a 'negative outcome' (for example, a large number of civilian casualties with a small military advantage gained)[42] were then compared with assessments under a 'positive outcome' (for example, a small number of civilian casualties with a significant military advantage).[43]

The sample (numbers detailed below) was selected with all hypotheses in mind, particularly in order to compare experts and non-experts and legal and military experts. As may be imagined, access to expert surveyees was not easy, reflecting a significant limitation on the execution of this kind of research. Soldiers, commanders and military/judge advocates-general (MAGs/JAGs) are bound by organizational rules on public exposure that create chilling effects even with respect to anonymous surveys and even when ethics approvals are received well in advance (as in the case of the Melbourne Law School (MLS), Hebrew University of Jerusalem (HUJ) and the ADF). Moreover, with future studies in mind, we restricted expert surveys to a minimum in order to avoid 'survey fatigue' in this elite group.[44]

The ADF sample covered two expertise bases (legal and operational). The first group included the staff of the office of the inspector-general of the Australian Defence Force (IGADF): '[T]he staff of the Office of IGADF comprises multidisciplinary teams of permanent and Reserve military personnel, and Australian Public Service employees who have knowledge and experience of Service life and the military justice system.'[45] This group comprises no more than a few dozen officers. The second group included the ADF legal officers ('Defence Legal'). According to the Military Law Center's website,[46] Defence Legal is staffed by 127 permanent ADF legal officers.

In addition, in the Israeli students group, we identified participants with relevant military training by a series of short questions. In particular, we asked the students who had served in the Israeli Defense Forces (IDF) about the nature of their service – that

[41] Through the 'Qualtrics' online survey software. See Qualtrics, available at www.Qualtrics.com.

[42] We did not manipulate military advantage and incidental civilian harm separately. This would require a much larger sample, and, in any case, the current research does not hypothesize whether decision-makers are more/less sensitive to one side of the proportionality calculus.

[43] We did not include in this indicative study a survey group either exposed to 'no outcome' or asked to make an *ex ante* decision for two reasons: first, access to military experts for surveys is very limited, demanding a parsimonious survey, and, second, in substance, gaps between positive/negative outcome groups are sufficient for establishing outcome bias and expertise effects.

[44] See Olson, 'Survey Burden, Response Rates, and the Tragedy of the Commons', 34(2) *Journal of Continuing Education in the Health Professions* (2014) 93.

[45] Inspector-General of the Australian Defense Forces, Annual Report, 1 July–30 June 2017, at 5, available at www.defence.gov.au/mjs/_Master/docs/IGADFAnnualReport2016-17.pdf.

[46] Defence Legal, available at www.defence.gov.au/Legal/.

is, whether they had any field experience and whether they were involved in making decisions of the kind presented in the survey. We also collected basic demographic information and asked several questions about the cases at the end of the survey. A pilot study on Amazon MTurk was conducted before running the full study in order to confirm clarity and understanding.[47] Finally, we collected responses from different public international law groups (for example, the American Society of International Law's Lieber Society and the Association for the Promotion of International Humanitarian Law in Israel). Here as well, we identified experts with practical military experience by a series of questions at the beginning of the survey about the nature of their service as well as its duration.

The first vignette was loosely based on the *McCann* case, where security forces tailed Irish Republican Army suspects in Gibraltar and used lethal force when they believed they intended to detonate a car bomb.[48] In technical legal terms, this can be understood as more of a human rights law case than one of IHL or one of distinction, but, for our purposes, it displays very similar elements of proportionality and reasonableness. The second vignette was based on the Kunduz fuel tankers case, already mentioned above. The third case was hypothetical and included a much simplified factual scenario of targeted killing. We asked subjects whether they were familiar with the *McCann* case and/or the fuel tankers case to control for any bias in this respect. Tables 1–3 present the survey results, which are discussed in more detail below and in Supplementary Appendix B online.

Table 2: Kunduz fuel tankers scenario

	Order to attack proportional (%)	**Order to attack not proportional (%)**	**Order to attack lawful (%)**	**Order to attack not lawful (%)**
Negative outcome				
MLS students	23 (31.9)	49 (68.1)	36 (50)	36 (50)
Israeli students	65 (41.9)	90 (58.1)	78 (50.3)	77 (49.7)
ADF officers	14 (46.7)	16 (53.3)	12 (40)	18 (60)
International law experts	21 (58.3)	15 (41.7)	22 (61.1)	14 (38.9)
IDF field experience	13 (50)	13 (50)	15 (57.7)	11 (42.3)
Positive outcome				
MLS students	53 (72.6)	20 (27.4)	55 (75.3)	18 (24.7)
Israeli students	115 (71)	47 (29)	131 (80.9)	31 (19.1)
ADF officers	24 (77.4)	7 (22.6)	27 (87.1)	4 (12.9)
International law experts	34 (77.3)	10 (22.7)	38 (86.4)	31 (79.5)
IDF field experience	30 (76.9)	9 (23.1)	31 (79.5)	8 (23.5)

[47] Amazon MTurk, available at www.mturk.com.

[48] *McCann v. United Kingdom*, Appl. no. 18984/91, Judgment of 27 September 1995. For the full vignettes, see Supplementary Appendix A online.

Table 3: *Targeted killing* scenario

	Order to attack proportional (%)	Order to attack not proportional (%)	Order to attack lawful (%)	Order to attack not lawful (%)
Negative outcome				
MLS students	43 (53.1)	38 (46.9)	44 (54.3)	37 (45.7)
Israeli students	88 (55.7)	70 (44.3)	104 (65.8)	54 (34.2)
ADF officers	23 (79.3)	6 (20.7)	25 (86.2)	4 (13.8)
International law experts	26 (72.2)	10 (27.8)	25 (69.4)	11 (30.6)
IDF field experience	21 (53.8)	18 (46.2)	24 (61.5)	15 (38.5)
Positive outcome				
MLS students	46 (71.9)	18 (28.1)	47 (73.4)	17 (26.6)
Israeli students	118 (74.2)	41 (25.8)	122 (76.7)	37 (23.3)
ADF officers	27 (90)	3 (10)	27 (90)	3 (10)
International law experts	36 (81.8)	8 (18.2)	35 (79.5)	9 (20.5)
IDF field experience	16 (61.5)	10 (38.5)	18 (69.2)	8 (30.8)

B Results: Students

In Australia, 145 juris doctor students from MLS, considered to be laypersons, took part in the study (63 females and 82 males; N = 145). The results show a distinct outcome bias across all three cases (all results are set out in detail in Supplementary Appendix B online). We asked participants several background questions as well as questions about any previous familiarity with the cases. Aside from a few minor exceptions, the participants did not have any military experience and were not exposed to the cases described in the survey prior to taking part in the study. The results demonstrate that when presented with a negative outcome respondents were more likely to determine that the attack was unreasonable, disproportionate and unlawful than when presented with a positive outcome. This difference was shown to be statistically significant.

In Israel, participants were recruited from the entire student body of HUJ and Ben-Gurion University (BGU) in the Negev. In total, 192 students from HUJ (113 females and 79 males) and 125 students from BGU (78 females and 47 males) took part in the study (N = 317). The survey included the same three cases above, translated into Hebrew. As opposed to the MLS students, a majority of participants had military backgrounds (157 students from HUJ and 114 students from BGU), and some of the participants had relevant field experience. The general results showed a statistically significant difference between the positive outcome group and the negative outcome group across all three cases. Here as well, respondents were more likely to determine that the decision was unreasonable, disproportionate and unlawful when presented with a negative outcome than when presented with a positive outcome.

C Results: Legal Experts

In Australia, 62 ADF officers (from Defense Legal and from the IGADF, including 18 females and 44 males) took part in the study (N = 62). In this group of experts, the results showed a statistically significant difference between the positive outcome group and the negative outcome group in the first two cases but not in the third case. In the third case, a brief hypothetical case of a targeted killing of terrorists, the expert participants were not influenced by the result, and the responses to both the negative and positive result scenarios were almost identical. In the public international law interest groups, we received 80 responses from international law experts (27 females and 53 males) (N = 80). The majority of participants served in the military (55 in total, 38 in operational field roles). Similarly to the ADF officers, outcomes significantly influenced evaluation in the first and second cases, though less so in the third case.

D Results: Field Experience

Finally, we performed a segmentation of data received from the Israeli students and identified 65 students who had self-identified as possessing operational field military experience in the IDF. We then found that for this subgroup of participants (n = 65), the results displayed some outcome bias, but at a lower degree than in other groups. Out of the three expert groups (ADF officers, international law experts and the group of Israeli students with IDF field experience), the respondents of the latter group were the least affected by the normatively irrelevant outcome factor.

E Discussion of the Results

The experimental results generally affirm the existence of outcome bias in *ex post* evaluations of operational decisions but only partially confirm advantages to expertise and experience. With respect to Hypothesis 1, it is clear that most subjects were influenced by knowledge of outcomes in their decisions regarding reasonableness and proportionality, deeming operational decisions disproportionate and unreasonable when outcomes were poor. Regarding Hypothesis 2, experienced and expert subjects were also influenced by outcome knowledge, albeit to a lesser extent and not *vis-à-vis* the third case. With respect to Hypothesis 3, the results demonstrate that individuals with operational decision-making experience may be, to some extent, less prone to outcome bias than legal experts without field experience.

Beyond these results, directly relating to the hypotheses, one additional noteworthy point emerged from the data. In the surveys, we also asked subjects whether they considered the operational decisions to be lawful. Without exception, in all groups of subjects, the consideration of lawfulness exceeded the findings of reasonableness and/or proportionality. In other words, people (both lay and expert) considered operational decisions to be lawful even when they thought of them as unreasonable and/or disproportionate. There can be several explanations for this discrepancy that would require further research. The simplest explanations relate directly to individual responsibility. Subjects may have understood 'un/lawfulness' as international criminality. At an analytical level, the subjects may have considered the legal framework

presented to them and distinguished between violations of IHL as such and criminal acts that incur personal responsibility; they would not be mistaken in this respect. At a more intuitive level, the subjects may have been 'forgiving' towards acts that they considered unreasonable, but tolerable in terms of lawfulness. In any case, outcome bias was displayed quite clearly across all groups as well as with respect to lawfulness.

5 Conclusions: Can the Outcome Bias in Military Investigations Be Debiased?

Our results indicate that outcome bias in evaluations of operational decisions is prevalent in non-expert groups. It exists in expert groups as well but to a lesser extent.[49] People have difficulty ignoring information about the outcomes of the decisions whose quality they evaluate *ex post*, even when this information is irrelevant from a formal legal perspective. These conclusions are consistent with experimental analysis in other areas of law.[50] For the purposes of this special issue on the psychology of international law, it is clear that the cognitive biases of individual decision-makers engaging with international law can have significant normative and prescriptive implications.

Actual awareness of the effects of outcome knowledge can be found in different military-related legal materials. This arises not only in the context of IHL proportionality. For example, in the Coroner's Court of Queensland's findings of the inquest into the deaths of three Australian soldiers in Afghanistan in 2012, we find an explicit reference to the impact of hindsight bias:

> The impact of hindsight bias and affected bias must also be considered when analysing the evidence. Hindsight and affected bias can occur where after an event has occurred, particularly where the outcome is serious, there is an inclination to see the event as predictable, despite there being few objective facts to support its prediction.

This reference addresses hindsight, not outcome, and relates to injury that is not incidental civilian harm but, rather, harm to one's own military forces – an issue worth addressing in future research – but it brings home the need to take into account susceptibility to outcome bias in military investigations. This is true first and foremost with respect to the international legal frameworks of proportionality and reasonableness. Military investigations (and, in this respect, we can also include criminal prosecutions) develop the normative standards of proportionality and reasonableness on a case-by-case basis. They will usually be conducted when outcomes have been poor (either in terms of severe incidental civilian harm or unattained military advantage).

[49] The comparison between the student groups and expert groups above holds some promise for an assessment of the value of expertise in overcoming outcome bias. However, outcome bias is only one element in the complex exercise of military decision-making. It should therefore be considered alongside other factors (such as professional bias or in-group bias).

[50] Regarding laypersons and judges outcome and hindsight bias in negligence cases, see, e.g., Oeberst and Goeckenjan, 'When Being Wise after the Event Results in Injustice: Evidence for Hindsight Bias in Judges' Negligence Assessments', 22(3) *Psychology, Public Policy and Law* (*PPPL*) (2016) 271.

To the extent that they are consistently influenced by outcome bias, the standard set *ex post* might be higher than the standard that IHL establishes for *ex ante* assessment, perhaps causing soldiers and commanders to exercise more caution in operational decisions. Indeed, a possible double-edged sword of the awareness of outcome bias has been demonstrated in a recent study that investigated the effect of anticipated outcome knowledge on foresight judgments and decisions. The study showed that 'people anticipate the effect of outcome judgment passed by another party to which they are accountable, and adjust their behavior by opting for a less risky alternative'.[51] This finding would suggest that soldiers might alter their behaviour to avoid risky options if they were to assume the military investigator is affected by the outcome bias.

From a humanitarian perspective, this could be considered a positive trend. However, one must also consider the potentially negative effects of *ex ante* hesitation, with respect to the attainment of military objectives, for example, the lack of which may prolong conflict, perhaps with more severe cumulative incidental civilian harm. In this context, outcome bias may impair quality assessments of military decision-making, which could perhaps prevent future civilian harm through better procedures and protocols.

Could these biases be 'debiased?' In other words, can investigators of military decisions be led to ignore or otherwise set aside the outcomes of the military decisions they are investigating? Given the high news profile of military operations with high degrees of civilian harm and otherwise negative outcomes in practice, it would be very difficult, though not impossible, to insulate *ex post* investigators from outcome knowledge. Moreover, we would presume that a 'pre-booting' or a 'pre-ruling' mechanism of reviewing military decisions (that is, when outcomes are truly unknown) is not that different from current MAG/JAG practices (in some military organizations) and would be inexpedient not only in 'hot' field combat situations but also in predetermined targeting cases, which still require last minute proportionality and reasonableness assessments.

Taking into account the limitations of our study, our results suggest that preference should be granted to appointing *ex post* investigators with field experience, who appear to have a better understanding of the *ex ante* decision-making process, which translates into lower susceptibility to outcome bias. This suggestion has three qualifications, however. First, lower outcome bias does not mean that investigators are less critical of decision-making quality; indeed, the results demonstrate that field-experienced investigators can be more critical than laypersons and legal experts, while maintaining a better balance between different outcome knowledge. Second, investigators with field experience need not be 'internal' to the military organization whose conduct and decisions are under investigation. Such self-investigations bear additional institutional-political biases that could erode the advantages of field experience. And, third, we are not claiming that the experts polled in our study make perfect decisions. This kind of statement could lead to an adverse result of reinforcing overconfidence

[51] Livnat, Sulitzeanu-Kenan and Kogut, 'Foresighted Outcome Effect: A Micro-Foundation of Agents' Risk Aversion in Principal-Agent Relations', 1(1) *Journal of Behavioral Public Administration* (2018) 1.

bias, a present phenomenon in expert decision-making.[52] Conversely, we hope that our proposal would help experts develop their expertise by constantly observing their decision-making process and improving it using a variety of tools.

Moreover, we suggest several complementary debiasing mechanisms for *ex post* military investigations. One possibility would be to mandate investigations of all similarly positioned military decisions, regardless of their actual outcome, on a regular basis, or investigations of randomly assigned cases, leading perhaps to better evaluations of the quality of decisions, both with respect to incidental civilian harm and to military advantage. This could hone evaluative expertise and desensitize investigators to outcomes.[53] In a second scenario, investigators would not know the outcome of the decision they are investigating, but they would be exposed only to a 'time capsule' of information as was available to the relevant decision-maker in real time. Finally, in another complementary scenario, investigators would rather deliberately be apprised of the outcomes of the act; with respect to positive outcomes, this could incentivize a deeper investigation of decision-making quality as well as an understanding of the consideration that led to these outcomes.

[52] See, e.g., Loftus *et al.*, 'Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes', 16(2) *PPPL* (2010) 133; Speirs-Bridge *et al.*, 'Reducing Overconfidence in the Interval Judgments of Experts', 30(3) *Risk Analysis* (2010) 512.

[53] This proposal is supported by the finding that judges make similar rulings on probable cause in foresight and in hindsight: Rachlinski, Guthrie and Wistrich, 'Probable Cause, Probability, and Hindsight', 8 *Journal of Empirical Legal Studies* (2011) 72.