

William & Mary Bill of Rights Journal

Volume 27 (2018-2019)
Issue 3 Symposium: Rights Protection in
International Criminal Law and Beyond

Article 4

March 2019

Understanding Crime Gravity: Exploring the Views of International Criminal Law Experts

Stuart Ford

Follow this and additional works at: <https://scholarship.law.wm.edu/wmborj>



Part of the [Criminal Law Commons](#), [Human Rights Law Commons](#), [International Law Commons](#), and the [Military, War, and Peace Commons](#)

Repository Citation

Stuart Ford, *Understanding Crime Gravity: Exploring the Views of International Criminal Law Experts*, 27 Wm. & Mary Bill Rts. J. 659 (2019), <https://scholarship.law.wm.edu/wmborj/vol27/iss3/4>

Copyright c 2019 by the authors. This article is brought to you by the William & Mary Law School Scholarship Repository.
<https://scholarship.law.wm.edu/wmborj>

UNDERSTANDING CRIME GRAVITY: EXPLORING THE VIEWS OF INTERNATIONAL CRIMINAL LAW EXPERTS

Stuart Ford*

INTRODUCTION

It is widely accepted that not all crimes have the same gravity.¹ Some crimes are simply more grave than others.² Establishing a hierarchy of crimes has been difficult, however, even in domestic contexts.³ Nevertheless, there is now a growing body of literature addressing the relative gravity of the most common domestic offenses.⁴ But crime gravity is not just a feature of domestic criminal law. It also matters in international criminal law.⁵

The concept of gravity is enormously important at the International Criminal Court (ICC).⁶ It appears prominently in the Preamble to the Rome Statute⁷ and is

* Professor of Law at the John Marshall Law School in Chicago, Illinois. This Article was made possible by a summer research grant from the John Marshall Law School. It was improved by the comments of the participants in the *William & Mary Bill of Rights Journal's* 2018 Symposium. Professor Margaret Ryznar also provided valuable comments. Any mistakes that remain are mine.

¹ For example, the Model Penal Code states that the purpose of sentencing provisions is, among other things, to “render sentences . . . proportionate to the gravity of [the] offenses . . .”). MODEL PENAL CODE § 1.02(2)(a)(i) (AM. LAW. INST., Proposed Final Draft 2017). This would not be necessary if all crimes shared the same gravity.

² In this Article, the term “crime gravity” refers generally to the seriousness of a crime. It is roughly synonymous with the concepts of crime severity and crime seriousness. This Article will use gravity because that is the word that is used by the International Criminal Court, although severity and seriousness may sometimes be used as alternatives for stylistic purposes. See Stuart Ford, *What Investigative Resources Does the International Criminal Court Need to Succeed?: A Gravity-Based Approach*, 16 WASH. U. GLOBAL STUD. L. REV. 1, 6 (2017) [hereinafter Ford, *Investigative Resources*].

³ See Paul H. Robinson & Robert Kurzban, *Concordance and Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829, 1832–35 (2007) (noting that there have been significant disagreements among criminal law theorists about whether it is possible to create a meaningful hierarchy of the gravity of different offenses).

⁴ *Id.* at 1854–67 (describing studies that use empirical techniques to identify the relative gravity of different offenses).

⁵ See Margaret M. deGuzman, *Gravity and the Legitimacy of the International Criminal Court*, 32 FORDHAM INT'L L.J. 1400, 1400–01 (2009) (noting that crime gravity is an extremely important concept in international criminal law).

⁶ See Ford, *Investigative Resources*, *supra* note 2, at 5–18 (describing the concept of crime gravity and its importance at the ICC).

⁷ For example, the Preamble references “unimaginable atrocities that deeply shock the conscience of humanity” and says that “such grave crimes threaten the peace, security and well-being of the world.” Rome Statute of the International Criminal Court, July 17, 1998,

used nine times in the body of the Statute.⁸ It is crucial at nearly every stage of the proceedings.⁹ For example, the Office of the Prosecutor (OTP) is required to consider “the gravity of the crime” when deciding whether to open a formal investigation.¹⁰ Once the Prosecutor has requested charges against an accused, the Court must determine whether the case is “of sufficient gravity” to be tried at the ICC.¹¹ And, once a person has been found guilty, the gravity of their crimes is a key factor in determining their sentence.¹²

Crime gravity may also be very important to the long-term success of the ICC.¹³ In particular, the success of the court may hinge, in part, on whether there is a widely accepted understanding of the gravity of mass atrocities. If there is a widely accepted meaning of gravity and the ICC uses that concept of gravity in deciding where to investigate, whom to charge, and how to sentence people, then the court’s gravity decisions will match people’s expectations about those decisions.¹⁴ If the court’s gravity decisions match people’s expectations (i.e., if the court prosecutes those crimes that most people perceive as the most grave), those decisions are more likely to be viewed as the result of a fair process.¹⁵ And if they are viewed as fair, then the court is more likely to be perceived as legitimate.¹⁶ This, in turn, makes it more likely that people

2187 U.N.T.S. 3, preamble [hereinafter Rome Statute]. It goes on to say that these “most serious crimes . . . must not go unpunished” and claims that the ICC will help put an end to impunity for the gravest offenses and thereby “contribute to the prevention of such crimes.” *Id.* In effect, the Preamble promises that the ICC will focus on offenses of the greatest gravity and that by investigating and prosecuting the gravest crimes, it will help end impunity and prevent future violations of international criminal law. *See id.*

⁸ *See* Rome Statute, *supra* note 7, art. 7(1)(g) (referring to the gravity of crimes of sexual violence); art. 17(1)(d) (referring to the gravity of individual cases); art. 53(1)(c) (referring to the gravity of the crimes being investigated); art. 53(2)(c) (referring to the gravity of the crimes being investigated); art. 59(4) (referring to the gravity of the crimes being charged); art. 77(1)(b) (referring to the gravity of the crimes for which an individual has been found guilty); art. 78(1) (referring to the gravity of the crimes for which an individual has been found guilty); art. 84(1)(c) (referring to the gravity of official misconduct); art. 90(7)(b) (referring to the gravity of the charged conduct).

⁹ *See* Stuart Ford, *The Meaning of Gravity at the International Criminal Court: A Survey of Attitudes About the Seriousness of Mass Atrocities*, 24 U.C. DAVIS J. INT’L L. & POL’Y 209, 210, 215–16 (2018) [hereinafter Ford, *Meaning of Gravity at the ICC*].

¹⁰ Rome Statute, *supra* note 7, art. 53(1)(c). *See also id.* art. 53(2)(c) (requiring the OTP to consider the gravity of the alleged crime when *declining* to prosecute).

¹¹ *Id.* art. 17(1)(d).

¹² *Id.* art. 78(1). *See also id.* art. 77(1)(b) (noting that a sentence of life imprisonment must be justified by the “extreme gravity” of the convicted person’s crimes).

¹³ *See* Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 212–14.

¹⁴ *Id.*

¹⁵ *Id.*

¹⁶ *See* deGuzman, *supra* note 5, at 1435 (“Gravity helps to legitimize one of the Court’s most important decisions—the decision to act—and thereby serves to legitimize the Court itself.”); Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283, 284 (2003) (“[P]eople’s reactions to legal authorities are based to a striking degree on their assessments of the fairness of the processes by which legal authorities make decisions. . . .”).

will comply with the requirements of international criminal law.¹⁷ Since fostering compliance with international criminal law is the primary purpose of the ICC,¹⁸ the question of whether there is a generally agreed upon meaning of gravity has important implications for the ability of the ICC to achieve its most important goal.

Yet despite its importance, the Rome Statute does not define what gravity means.¹⁹ This lack of a definition has resulted in multiple attempts to define gravity.²⁰ The most comprehensive gravity definition in use within the court is the one that was created by the OTP.²¹ An earlier article by this author tested the components of the OTP's gravity definition to see if they matched people's subjective understanding of crime gravity.²² The results indicated that people's perceptions of crime gravity are not idiosyncratic.²³ Rather, there was broad agreement about at least some of the gravity factors that were tested.²⁴ There were some factors that were only weak indicators of crime gravity,²⁵ but there were several factors that had broad support, including factors like the extent of the harm suffered by the victims, the type of crime committed, and the presence of particular cruelty.²⁶ The results suggested that it should be possible for the court to create a gravity definition that would be widely viewed as fair and legitimate.

At the same time that the survey described in *The Meaning of Gravity* was administered to non-experts, it was also administered to a group of self-identified experts in international criminal law (ICL).²⁷ This Article reports the results of the expert portion of the survey. It also compares the results of the experts to those of the non-experts. It seeks to answer three main questions: First, are there meaningful differences between how experts and non-experts view the gravity of mass atrocities? Second, if so, what are those differences? And finally, assuming there are differences, what do those differences mean?

At a high level, the data indicate that there is broad agreement between experts and non-experts about the relative importance of the various gravity factors that were tested. This is probably good news for the court as it would be more difficult to agree upon and implement a new gravity definition if the experts and the non-experts had serious disagreements about the contents of that definition. At the same time, there were also some differences between the experts and non-experts. First, the

¹⁷ See Tom R. Tyler, *Psychological Perspectives on Legitimacy and Legitimation*, 57 ANN. REV. OF PSYCHOL. 375–79 (2006) (arguing that “authorities and institutions are viewed as more legitimate and, therefore, their decisions and rules are more willingly accepted when they exercise their authority through procedures that people experience as being fair”).

¹⁸ See generally Stuart Ford, *A Hierarchy of the Goals of International Criminal Courts*, 27 MINN. J. INT'L L. 179 (2018) (arguing that the prevention of violations of international criminal law is the most important goal of the ICC).

¹⁹ See Ford, *Investigative Resources*, *supra* note 2, at 7, 9–10.

²⁰ *Id.* at 7–14 (describing various attempts to provide a definition of crime gravity).

²¹ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 216–17.

²² See generally *id.* (describing a survey of people's attitudes about various factors that have been hypothesized to affect the gravity of international crimes).

²³ *Id.* at 247.

²⁴ *Id.* at 248–49.

²⁵ *Id.* at 243–44.

²⁶ *Id.* at 242–43.

²⁷ *Id.* at 221 n.61.

experts were more supportive of almost all of the gravity factors than the non-experts. This is not necessarily a problem for the court, but both the court and the experts need to be aware that non-experts have lower levels of support for the gravity factors that were tested in the survey. Second, some of the results suggest that non-experts do not intuitively agree with all of the norms of international criminal law. If non-experts disagreed with key features of international criminal law this could make it harder to persuade people to comply with that law. Thus, the question of the extent to which non-expert views on mass atrocities diverge from the norms of international criminal law deserves further study.

This Article proceeds as follows: Part I describes the survey's methodology, while Part II describes the demographics of the participants. Part III reports the results of the survey, and Part IV discusses what those results mean. The final Part presents the Article's conclusions.

I. METHODOLOGY

The participants' understanding of the gravity of mass atrocities was tested using a survey.²⁸ As described in an earlier article, the gravity definition that was tested was based on the definition currently in use by the Office of the Prosecutor (OTP) at the ICC.²⁹ Each question on the survey was designed to test a single component of that gravity definition. For each question on the survey, participants were shown between two and four scenarios. Each scenario was the same except for a change related to the component of the gravity definition being tested in that question.

For example, the OTP hypothesizes that the seriousness of crimes increases as "the extent of the damage caused by the crimes, in particular the bodily . . . harm caused to the victims" increases.³⁰ This hypothesis was tested in Question 8 of the survey.³¹ In that question, participants were shown three scenarios where government soldiers attack peaceful protesters. The only change between the scenarios is the extent of the physical harm to the protesters caused by the attack. In Scenario 1, the victims suffer "minor injuries such as cuts and bruises."³² In Scenario 2, the victims suffer "injuries which require hospitalization, including broken bones and concussions."³³ In Scenario 3, the victims "are permanently disabled as a result of their injuries."³⁴

Participants were asked to assign a numerical score to each scenario that represented the seriousness of the events described in the scenario.³⁵ To give participants a baseline against which to score the various scenarios, they were told that "[t]he baseline scenario is: 'A group of soldiers deliberately shoots and kills 10 unarmed villagers.'

²⁸ A copy of the survey is available online. See Stuart Ford, *Meaning of Gravity Survey—Version 1.1*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3021041 [<https://perma.cc/2WRW-GR6A>] [hereinafter Ford Survey].

²⁹ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 216–17.

³⁰ Office of the Prosecutor, *Policy Paper on Preliminary Examinations*, ¶62 (Nov. 2013).

³¹ See Ford Survey, *supra* note 28, at Question 8.

³² *Id.*

³³ *Id.*

³⁴ *Id.*

³⁵ *Id.* at 3 ("This survey asks for your opinion about how serious YOU think certain crimes are."); *id.* ("USE ANY number so long as it shows how serious YOU think the situation is.").

This has been assigned a score of 100 to show its seriousness.” Participants were asked to give scores to the different scenarios in light of the score given to the baseline scenario.³⁶ In effect, the survey used a magnitude estimation approach similar to that used in the National Survey of Crime Severity.³⁷

Variations in the scores given by the participants should be related to changes in the gravity component being studied.³⁸ For example, the scenarios described above from Question 8 are identical except for the harm that the victims suffer as a result of the attack. If there are changes in the scores that participants give to these scenarios, those changes ought to reflect how changing the harm suffered by the victims changes the participants’ perceptions of the seriousness of the crime.

The survey did not ask people to rank the scenarios in order of their seriousness, but the scores that participants gave can be used to infer a rank ordering.³⁹ Generating a rank ordering allows this Article to test the hypotheses that underlie the gravity definition. For example, the OTP’s definition hypothesizes that gravity increases as the severity of the harm to the victims increases. This can be tested by looking at the rank ordering of the scenarios in the question related to that gravity component. If a participant ranks the scenarios that have the most serious harm as more grave than those that have less serious harms, this indicates that the participant agrees with the hypothesis.

The magnitude estimation approach used in this survey also permits ties between scenarios. The presence of ties indicates that a participant does not believe that the gravity component being tested meaningfully affects the overall gravity of the crime. For example, if a participant gives the same score to each of the scenarios described above, it indicates that they do not believe that gravity increases as the harm to the direct victims increases.

The strength of the various gravity components can then be inferred from the percentage of the participants who agree that each factor affects gravity. The higher the percentage of participants who believe that a particular factor affects gravity, the stronger that factor is as an indicator of overall gravity. For example, a factor that 75% of the participants score as increasing gravity is a stronger indicator than one that only 25% of the participants score as increasing gravity.

II. PARTICIPANTS

The survey was administered to both experts and non-experts. The non-expert participants were identified using SurveyMonkey’s Audience feature.⁴⁰ They were all from the United States and were not paid for their participation.⁴¹ The pool from which

³⁶ *Id.* (“Use this baseline scenario to judge all the others. For example, if you think a situation is 20 TIMES MORE serious than the baseline scenario, you should give it a score of 2,000 or if you think it is HALF AS SERIOUS, you should give it a score of 50 and so on.”).

³⁷ See MARVIN E. WOLFGANG ET AL., THE NATIONAL SURVEY OF CRIME SEVERITY at v–vi (June 1985); see also Robinson & Kurzban, *supra* note 3, at 1830–40 (describing magnitude estimation studies).

³⁸ See Robinson & Kurzban, *supra* note 3, at 1842–43.

³⁹ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 219.

⁴⁰ See *Audience Feature*, SURVEY MONKEY, <https://www.surveymonkey.com/mp/audience/> [<https://perma.cc/5DPP-HDS8>].

⁴¹ See *Who’s Taking Your Survey*, SURVEY MONKEY, <https://help.surveymonkey.com/arti>

these participants were selected is described as “a diverse population of millions of people across the United States,”⁴² but it is not representative of the general population.⁴³ That said, the pool is balanced for age, gender and geographic location.⁴⁴ Further information on the non-expert participants can be found in an earlier article.⁴⁵

In addition, the survey was also given to a group of self-identified experts in international criminal law. Expert participation was obtained by sending an email inviting participation in the survey to the membership of the American Society of International Law’s International Criminal Law Interest Group.⁴⁶ The membership of the interest group is a self-selected group of individuals interested in international law generally (hence their membership in the American Society of International Law⁴⁷) and international criminal law specifically (hence their membership in the International Criminal Law Interest Group⁴⁸). In early 2017, when the email inviting participation was sent, the interest group had a membership of 473 people.⁴⁹ Ultimately, 46 individuals used the link in the email to take the survey, indicating a response rate of almost 10%.

The experts do not represent a random sample. The experts who responded to the survey chose to do so. Thus, there is the risk of selection bias in the composition of the sample.⁵⁰ In other words, there may be something about the characteristics of those experts that chose to respond to the invitation to participate that makes them different

cles/en_US/kb/SurveyMonkey-Audience?bc=Buying_Responses#Panel [https://perma.cc/3RWL-MLKS] (“Contribute members take surveys for charity and a chance to win a sweepstakes prize.”). The participants were, however, compensated in a sense, as SurveyMonkey commits to making a small donation to a charity on behalf of the participants.

⁴² *Id.*

⁴³ See *Balancing and Regional Representation*, SURVEY MONKEY, https://help.surveymonkey.com/articles/en_US/kb/SurveyMonkey-Audience?bc=Buying_Responses#Panel [https://perma.cc/DFS2-UD39].

⁴⁴ See *id.* (“We automatically balance Contribute and Rewards panels according to census data of age and gender, while location tends to balance out naturally.”).

⁴⁵ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 221–22.

⁴⁶ This was made possible with the help of Professor Jenia Iontcheva Turner, the Co-Chair of the International Criminal Law Interest Group, who emailed a link to the survey to the interest group membership on the author’s behalf.

⁴⁷ The American Society of International Law is one of the largest membership organizations devoted to the study of international law in the world, with “nearly 4,000 members from more than 100 nations includ[ing] attorneys, academics, corporate counsel, judges, representatives of governments and non-governmental organizations, international civil servants, students, and others interested in international law.” See *About ASIL*, ASIL, <https://www.asil.org/about/about-asil> [https://perma.cc/9BFN-6DK2].

⁴⁸ The International Criminal Law Interest Group is one of a number of interest groups within the American Society of International Law. It “has a primary purpose of facilitating networking on research, studies, drafts and proposals of a private or governmental nature concerning international criminal law and institutions.” See *International Criminal Law*, ASIL, <https://www.asil.org/community/international-criminal-law> [https://perma.cc/W8UJ-Y78N].

⁴⁹ See Email from Prof. Jenia Iontcheva Turner, Co-Chair of the International Criminal Law Interest Group, to Stuart Ford (June 8, 2018) (on file with author).

⁵⁰ See SHARON L. LOHR, *SAMPLING: DESIGN AND ANALYSIS* 4–6 (1999).

from the nonresponders. If this is true, it would be problematic to make inferences about the beliefs of the population from the beliefs of the sample. It is not possible to know whether this is true, however, because we do not know what the total population of ICL experts looks like or how the nonresponders differ from those who chose to respond to the survey. Moreover, it would be difficult to obtain a truly random sample of ICL experts.⁵¹ This Article will analyze the sample that was obtained. But it is worth noting that it is not a random sample and that there is the possibility of selection bias.

All the participants were asked some demographic questions before they took the survey. The experts spanned a range of ages, but the largest group (59% of the total) was between 36 and 50. There were more male experts than female experts (59% vs. 41%). Unsurprisingly, 100% of the experts had completed a graduate degree and 89% were lawyers. Moreover, they were virtually all self-professed experts in international criminal law. When asked to assess their own understanding of ICL, 85% said they knew a lot about it, while 13% said they knew a moderate amount. One participant said they only had “some” knowledge of ICL.

Table 1: Demographics		
	Non-Experts	Experts
Gender	66% female	41% female
Age	29% are 36–50	59% are 36–50
Education	20% had a graduate degree	100% had a graduate degree
Lawyers	2%	89%
ICL knowledge	70% knew “nothing” or “very little”	85% knew “a lot”

The expert participants look quite a bit different from the non-expert participants. For one thing, they were more likely to be male⁵² and their ages were concentrated in the 36–50 range.⁵³ They were also much more educated than the non-expert

⁵¹ There are two main problems with obtaining a random sample. First, it would be necessary to know the entire population of ICL experts. But there is no comprehensive listing of who is an ICL expert. Membership in ASIL’s international criminal law interest group is a good indicator of ICL expertise, but there are surely other ICL experts who are not part of ASIL. So, there is no easy way to ensure that all ICL experts receive an invitation to the study. The second problem is even trickier. Even if it were possible to identify all ICL experts, there is no easy way to ensure that a random sample of experts take the survey. Response rates are likely to be low to any invitation to participate and the nonresponders might bias the results. *See* LOHR, *supra* note 50, at 6.

⁵² Fifty-nine percent of the expert participants were male, but only 35% of the non-expert participants were male.

⁵³ Fifty-nine percent of the experts were in the 36–50 range, compared to only 29% of the

group and they were overwhelmingly lawyers.⁵⁴ But the most important way in which they differ from the non-experts (at least for purposes of this Article) is in their self-reported knowledge of international criminal law. Whereas 85% of the expert participants said they knew “a lot” about ICL, less than 1% of the non-experts said the same thing. In contrast, 70% of the non-experts said they knew either “nothing” or “very little” about international criminal law. None of the “experts” said they knew little or nothing about ICL. While the information is obviously self-reported, the participants in the expert group do claim to be experts on international criminal law. Given the pool from which the expert participants were drawn (members of a group dedicated to the study of international criminal law) this is not surprising, but it is important because the answers of this group serve as a proxy for the way that experts in ICL as a whole understand the gravity of mass atrocities.

III. SURVEY RESULTS

This section discusses the results of the survey. The survey consisted of nineteen questions designed to test the components of the OTP’s gravity definition.⁵⁵ That definition is comprised of four broad factors: (1) the scale of the crimes; (2) their nature; (3) the manner of their commission; and (4) their impact. The OTP has elaborated on the meaning of each of these factors in its Policy Paper on Preliminary Examinations.⁵⁶

The OTP’s explanation of its gravity definition contains a number of hypotheses about what contributes to the overall gravity of international crimes.⁵⁷ These hypotheses were used as the basis for the questions the participants answered. Some components of the gravity definition were tested using more than one question.⁵⁸ The questions will be grouped together below according to the gravity component they test. The components are ordered roughly from those that elicit the broadest support to those that elicit the least support.

A. Type of Crimes

The OTP’s gravity definition asserts that some crimes are worse than others and that the type of the underlying crime affects its gravity.⁵⁹ This proposition was tested using two questions that asked participants to rank the severity of different scenarios

non-experts. There were no experts in the 18–25 range and only one expert in the 65+ range. In contrast, 11% of the non-experts were in the 18–25 range and 18% were in the 65+ range.

⁵⁴ One hundred percent of the experts had completed a postgraduate degree, compared to only 20% of the non-experts. Eighty-nine percent of the experts were lawyers, compared to just 2% of the non-expert group.

⁵⁵ See Ford Survey, *supra* note 28.

⁵⁶ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶¶ 62–65.

⁵⁷ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 217.

⁵⁸ It should be noted, however, that not every component of the OTP’s gravity definition was tested in the survey. See, e.g., *id.* at 239–41 (noting that some factors related to the impact of the crimes were not tested).

⁵⁹ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 63 (stating that gravity depends on “the specific elements of each offense such as killings, rapes . . .” or other offenses).

that varied only by the type of crime committed.⁶⁰ Question 2 describes a situation where rebel sympathizers are arrested by the government, placed in detention, and then abused by the guards.⁶¹ The only difference between the scenarios is that in one scenario the prisoners are beaten suffering bruises and broken bones, in one scenario they are raped, and in the final scenario they are killed. The number of victims is the same in each scenario. In effect, the question tests whether participants believe that there is a difference in gravity that stems from whether the perpetrator commits assault, rape, or murder. If the OTP's hypothesis is correct, then there should be an observable hierarchy among the crimes.

The responses of the experts are consistent with the OTP's hypothesis. Seventy-four percent of them scored the scenario involving the murder of the detainees as the most serious offense. Fifteen percent of the experts said that rape was the most serious offense, while 9% said that rape and murder were equally serious.⁶² Nobody said that assault was the most serious offense. One of the expert participants, however, did score all of the scenarios as equally serious. Nevertheless, a clear majority of the experts agreed with the OTP's hypothesis that the type of crime committed affects the gravity of the offense.

The effect of the type of crime on gravity was also tested using a second question. In Question 10, rebel soldiers attack a village and capture 100 villagers.⁶³ In one scenario, the villagers are so scared by the attack that they subsequently flee the village in fear and do not return for more than a year.⁶⁴ In the other scenarios, the villagers are either beaten, raped, or killed.

The results of this question are very similar to the results of Question 2. Seventy-seven percent of the experts said that the scenario involving murder was the most grave. Ten percent said that the scenario involving rape was the most serious, while 6% said that rape and murder were equally serious. Taken together, the results of Questions 2 and 10 show that a majority of experts agree that there is a hierarchy of

⁶⁰ Here, type of crime refers to the legal qualification of the offense. For example, murder is a different crime from rape because it is comprised of different elements. For example, murder requires the killing of one or more persons, while rape requires "penetration, however slight, of any part of the body of the victim or of the perpetrator with a sexual organ, or of the anal or genital opening of the victim with any object or any other part of the body." *See* International Criminal Court, Elements of Crimes (2011), arts. 7(1)(a), 7(1)(g)(-1).

⁶¹ *See* Ford Survey, *supra* note 28, at 5.

⁶² The percentages presented for each question may not always total 100%. For example, the percentages reported for this question total 98%. There are two reasons why this might occur. First, this Article does not always present all of the results. Rather, it focuses on the most important results and sometimes omits options that were picked by very small numbers of participants. Second, the results are rounded to the nearest percentage. This can sometimes result in totals other than 100%. *See id.*

⁶³ *See id.* at 13.

⁶⁴ This serves as a proxy for the crime of forcible displacement of a civilian population. *See* Rome Statute, *supra* note 7, arts. 7(1)(d), 7(2)(d) (defining the crime against humanity of forcible transfer of a population as "forced displacement of the persons concerned by expulsion or other coercive acts from the area in which they are lawfully present, without grounds permitted by international law").

crimes and that murder is at the apex of that hierarchy. Rape is second in the hierarchy, while assault and forcible displacement are at the bottom.

B. Extent of the Harm

The OTP hypothesizes that the gravity of crimes increases as the “extent of the damage caused by the crimes, in particular the bodily or psychological harm caused to the victims and their families” increases.⁶⁵ While the OTP suggests that the extent of the psychological harm caused by the crime is a component of the overall gravity of the crime that is separate from the physical harm of the crime, that hypothesis was not tested. The survey included a single question related to the harm suffered by the victims and that question focuses on physical harm. In Question 8, government soldiers attack peaceful protesters and beat them with clubs and rifles.⁶⁶ The only thing that varied between the scenarios was the extent of the physical harm suffered by the victims. In Scenario 1, the victims suffer “minor injuries such as cuts and bruises.”⁶⁷ In Scenario 2, the victims suffer “injuries which require hospitalization, including broken bones and concussions.”⁶⁸ In Scenario 3, the victims “are permanently disabled as a result of their injuries.”⁶⁹ If the OTP’s hypothesis is true, one would expect the participants to score Scenario 3 as the most serious because it results in the most serious physical harm to the victims.

And indeed, 90% of the experts gave Scenario 3 the highest score, indicating agreement with the OTP’s hypothesis. Three percent of the expert participants scored Scenario 2 (involving broken bones and concussions) the same as Scenario 3 (involving permanent disability). Six percent of the experts scored all of the scenarios the same indicating that they saw no difference in the gravity of the offense as the extent of the harm suffered by the victims changed. Nevertheless, the majority of the experts agreed with the OTP’s hypothesis.

C. Number of Direct Victims

The OTP hypothesizes that the gravity of the crime increases as the number of direct victims of the crime increases.⁷⁰ For purposes of this Article, a direct victim is someone who suffers harm as a direct result of the commission of a crime.⁷¹ This hypothesis was tested using two questions (Question 13 and Question 19). In Question 13, a rebel group attacks an undefended village and deliberately kills either 100, 200, or 400 unarmed villagers.⁷² In Question 19, government soldiers attack

⁶⁵ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 62.

⁶⁶ See Ford Survey, *supra* note 28, at 11.

⁶⁷ *Id.*

⁶⁸ *Id.*

⁶⁹ *Id.*

⁷⁰ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 62 (noting that the “scale of the crimes may be assessed in light of . . . the number of direct . . . victims”).

⁷¹ See Valentina Spiga, *Indirect Victims’ Participation in the Lubanga Trial*, 8 J. INT’L CRIM. JUST. 183, 186 (2010).

⁷² See Ford Survey, *supra* note 28, at 16.

peaceful protesters and beat either 100, 150, or 250 of them with clubs and rifles inflicting serious injuries.⁷³ If this hypothesis is correct, one would expect the participants to score the scenarios with more direct victims higher than the scenarios with fewer direct victims.

The results of Question 13 were unambiguous. Seventy-seven percent of the experts scored the scenario involving the largest number of deaths as more serious than the scenarios involving smaller number of deaths. The remainder of the expert group (23%) scored all of the scenarios the same. Thus, while a minority does not believe that the number of victims affects the gravity of the offense, a clear majority believes that it does.

The results of Question 19 were slightly more ambiguous. While a majority of the experts still scored the scenario with the largest number of victims as the most grave, it was only a bare majority (53%). At the same time, 40% of the expert participants scored all of the scenarios in Question 19 as equally serious. A majority of the participants still agreed with the hypothesis that the gravity of the offense increases as the number of victims increases, but there was a significant minority that disagreed.

The results of Questions 13 and 19 appear to be somewhat inconsistent. Why did so many more of the participants view the scenarios in Question 19 to be equally serious compared to the scenarios in Question 13 (40% vs. 23%)? In both questions, the only thing that varied between the scenarios in each question was the number of victims. So, one would expect the questions to elicit similar responses. But there was one obvious difference between the two questions—the extent of the harm suffered by the victims. In Question 13, the victims are killed, while in Question 19 the victims suffer “injuries which require treatment at a hospital.”⁷⁴ The other results show that almost all the experts agree that the gravity of the offense increases as the extent of the harm to the victims increases.⁷⁵ The experts also agree that killings are a more serious offense than inflicting non-fatal injuries.⁷⁶ Thus, it seems possible that the experts answered Question 13 differently from Question 19 because they were more likely to see the crime committed in Question 13 as inherently more serious than the crime committed in Question 19.

D. Number of Indirect Victims

The next hypothesis that the survey tested was whether the gravity of the crime increases as the number of indirect victims increases.⁷⁷ For purposes of this Article, an indirect victim is someone who suffers harm as a result of the harm suffered by a direct victim.⁷⁸ This hypothesis was tested using a single question (Question 4). In

⁷³ *Id.* at 22.

⁷⁴ *Id.* at 16, 22.

⁷⁵ See *supra* Section III.B (discussing the effect of varying the harm suffered by the victims).

⁷⁶ See *supra* Section III.A (discussing the effect of varying the type of crime committed).

⁷⁷ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 62 (noting that the “scale of the crimes may be assessed in light of . . . the number of . . . indirect victims”).

⁷⁸ See Spiga, *supra* note 71, at 186.

this question, rebel soldiers attack a village and 50 unarmed villagers are killed. The only thing that changes between the scenarios is the number of children that are orphaned as a result of those deaths, which varies between 50, 100, and 200.⁷⁹ In all three scenarios, the number of direct victims (the people actually killed by the rebel soldiers) stays the same; all that changes are the number of people indirectly affected by those deaths (the orphaned children). If this hypothesis is true, then one would expect the participants to score the scenarios with more indirect victims higher than the scenarios with fewer indirect victims.

Here, the majority of the experts (72%) scored all of the scenarios as equally serious. The rest of the participants (28%) scored Scenario 1 (the scenario involving the largest number of orphans) as the most serious. In effect, a majority of the experts did not believe that varying the number of indirect victims affected the gravity of the offense. This constitutes a rejection of the OTP's hypothesis.

E. The Means Employed

The OTP posits that the means employed to carry out a crime can affect its gravity.⁸⁰ This hypothesis was tested with two questions. In Question 12, the government forces all of the inhabitants of a village to flee.⁸¹ The only difference between the scenarios is the means the government uses to force the villages to leave. In one scenario, it shuts off the water and electricity, in another scenario, it uses the radio to threaten to bomb the village unless the people leave, and in the final scenario soldiers enter the village and threaten the villagers in person. In all three scenarios, 500 villagers flee as a result of the government's actions. If the OTP is right that the means of carrying out a crime matters, then there should be a hierarchy among the various scenarios.

A majority of the experts disagreed that the means employed changed the gravity of the offense in Question 12. Fifty-two percent of the experts scored all of the scenarios as equally serious. The rest of the expert participants perceived a hierarchy among the seriousness of the scenarios, but they had different opinions about which means was most serious. Nineteen percent of the experts said that having soldiers threaten the villagers was the most serious, while 16% said that shutting off the water and electricity to the village was the most serious. Six percent of the experts said that the second and third scenarios were worse than the first scenario, while 6% said that the first and third scenarios were worse than the second scenario. The results are pretty clear. A majority of experts thought the means employed did not matter. Even for those that did think it mattered, there was little agreement about which means was the most serious.

Question 15 took a slightly different approach to testing whether the means of commission matters. In this question, government soldiers attack a village and kill 100 villagers.⁸² The only difference among the three scenarios is how the killings are

⁷⁹ See Ford Survey, *supra* note 28, at 7.

⁸⁰ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 64 (arguing that the seriousness of the crimes depends in part on "the means employed to execute the crime").

⁸¹ See Ford Survey, *supra* note 28, at 15.

⁸² *Id.* at 18.

carried out. In one scenario, the villagers are beaten to death, in another they are killed by bombs dropped from aircraft, and in the final one, they are shot to death.

The results of this question were similar to the results of Question 12. Sixty-five percent of the experts scored all of the scenarios as equally serious. Twenty-six percent of them said that Scenario 1 (the scenario where the villagers were beaten to death) was the most serious. The remaining experts were split between saying that Scenario 3 (shooting the villagers to death) was the most serious (3%) and that Scenario 1 and Scenario 3 were equally serious (6%). But ultimately, a clear majority of the experts saw no difference between the various means of killing the villagers.

F. Particular Cruelty

The OTP's assertion that particular cruelty in the manner of commission of a crime increases its gravity was tested using two questions.⁸³ In Question 6, government soldiers kill 100 unarmed villagers.⁸⁴ The only thing that changes between the two scenarios is the way in which the killings are committed. In one scenario, the soldiers shoot and kill the victims. In the other scenario, the villagers "are forced inside the local church and then the church is set on fire."⁸⁵ All of the villagers die in the fire.⁸⁶

If particular cruelty matters, one would expect the scenario involving the fire would be scored as more serious.⁸⁷ And indeed, a clear majority of the experts (65%) agreed that the scenario that involved burning the villagers to death was more serious than simply shooting them to death. The remaining experts (35%) said that both scenarios were equally grave.

The role of cruelty was also tested in Question 16. In that question, government soldiers set up a roadblock and stop and kill anybody believed to be a rebel sympathizer.⁸⁸ The soldiers kill 100 people this way. The only difference between the two scenarios is in how the killings are committed. In one scenario, the people are shot to death. In the other scenario, the people are "killed by having their hands and feet chopped off with an axe."⁸⁹

If particular cruelty is a gravity factor, one would expect the scenario involving the amputations to be scored as more serious.⁹⁰ A majority of the experts (57%)

⁸³ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 64 (arguing that the seriousness of the crimes depends in part on whether there are "elements of particular cruelty" in the manner of their commission).

⁸⁴ See Ford Survey, *supra* note 28, at 9.

⁸⁵ *Id.*

⁸⁶ *Id.*

⁸⁷ *Cf.* Prosecutor v. Lukić, Case No. IT-98-32/1-T, Trial Judgment, ¶ 740 (July 29, 2009) (describing the defendants' crimes as "horrific events [that will] remain imprinted on the memory for the viciousness of the incendiary attack, for the obvious premeditation and calculation that defined it, for the sheer callousness, monstrosity and brutality of herding, trapping and locking the victims in the two houses, thereby rendering them helpless in the ensuing inferno and for the degree of pain and suffering inflicted on the victims as they were burnt alive").

⁸⁸ See Ford Survey, *supra* note 28, at 19.

⁸⁹ *Id.*

⁹⁰ *Cf.* Prosecutor v. Sesay, Case No. SCSL-04-15-T, Sentencing Judgment, ¶ 109 (Apr. 8, 2009) (describing death by amputation as particularly cruel and brutal).

agreed that the scenario involving death by amputation was more grave. The rest of the experts (43%) said that both scenarios were equally serious. The results from Questions 12 and 16 are consistent. In both questions, a majority of the expert participants agreed with the OTP's hypothesis that particular cruelty increased the gravity of the offense.

There is a potential inconsistency between the results of the questions about the means employed discussed above⁹¹ and the results of the questions about particular cruelty. After all, most of the experts seemed to believe that the means did not matter in the earlier questions but did think it mattered in the questions on particular cruelty. One potential explanation for these results is that the means employed does not matter unless it is perceived as resulting in significant additional suffering for the victims, in which case it does matter.⁹²

G. Discriminatory Intent

The OTP hypothesizes that crimes carried out with a discriminatory intent are worse than the same crime carried out without that intent. Thus, the OTP's definition refers to "persecution, or the imposition of conditions of life on a group calculated to bring about its destruction" as factors that affect gravity.⁹³ These are references to crimes that are distinctive because they involve a discriminatory intent—genocide and persecution.⁹⁴ For that reason, two questions were devised to test whether a discriminatory intent made the crime more grave. One question tested the effect of religious discrimination and the other tested the effect of ethnic discrimination.

In Question 1, hundreds of people are arrested by the government and placed in prison.⁹⁵ Once in prison, some of them are beaten daily. There are two scenarios associated with this question and the only difference between them is why the prisoners are being beaten. In Scenario 1 they are beaten for breaking a rule against speaking. In Scenario 2, they are described as having different religious beliefs from the guards and are beaten "because of their religious beliefs."⁹⁶ If crimes that are committed with discriminatory intent are more serious than the same crime committed without a discriminatory intent, then most participants will give Scenario 2 a higher score.

The experts were split into two groups. The largest group, comprising 68% of the sample, thought that the crime committed with discriminatory intent was the most serious. The remaining 32% of the sample scored the two scenarios as equally

⁹¹ See *supra* Section III.E.

⁹² See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 239–40.

⁹³ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 63.

⁹⁴ Persecution requires a discriminatory intent. See Rome Statute, *supra* note 7, art. 7(1)(h) (prohibiting persecution against any identifiable group based on political, racial, national, ethnic, cultural, or religious grounds); see also *id.* art. 7(2)(g) (defining persecution as "the intentional and severe deprivation of fundamental rights contrary to international law by reason of the identity of the group"). The reference to conditions of life calculated to bring about the destruction of a group is a reference to genocide. See *id.* art. 6(c). The hallmark of genocide is discriminatory intent. *Id.* art. 6 (defining genocide as certain prohibited acts carried out with the "intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such").

⁹⁵ See Ford Survey, *supra* note 28, at 4.

⁹⁶ *Id.*

serious. Thus, a clear majority of the experts agreed with the OTP's hypothesis that a crime that resulted from religious discrimination was more serious than the same crime that did not include religious discrimination.

A second question addressed ethnically motivated violence. In Question 18, government soldiers set up a roadblock and stop everybody that passes it.⁹⁷ In both scenarios 100 people are detained and then executed. In Scenario 1, they are executed because they cannot make a payment of \$20. In the second scenario, they are executed because they "are from a different ethnic group than the soldiers."⁹⁸

The experts' responses to Question 18 were quite similar to their responses to Question 1. Sixty-one percent of the experts said that Scenario 2, the scenario involving ethnic discrimination, was more serious than Scenario 1. The remaining 39% said that the two scenarios were equally serious. Thus, in response to both questions about discrimination, a clear majority of the experts agreed that the presence of a discriminatory intent increased the gravity of the resulting crimes.

H. Plan or Policy

The OTP's gravity definition proposes that crimes are more serious if they are "systematic or result from a plan or organised policy."⁹⁹ This hypothesis was tested using two questions. Question 14 focused on the question of whether having a plan to commit a crime made a difference and involved government soldiers surrounding a town and eventually killing 100 people. The main difference between the three scenarios relates to how the killings take place.¹⁰⁰ In one scenario, the killings are the result of a mistake by the soldiers who believe a crowd is hostile when in fact it is peaceful. Another scenario involves a spontaneous decision by an officer to order the soldiers to attack and kill the people because he was angry they were protesting the government. The final scenario involved a systematic plan to kill the people. In that scenario, the participants are told that the soldiers are "seen going house to house checking the names of the inhabitants against a list they carry. Anybody whose name is on the list is taken away and shot."¹⁰¹ In effect, this scenario tests the relative seriousness of a similar crime carried out by mistake, as the result of a deliberate but spontaneous decision, or as the result of a premeditated plan.

The results suggest that the experts generally agree with the OTP's hypothesis. Fifty-five percent of the participants ranked the scenario involving a systematic plan to carry out the killings as the most serious. The next largest group of experts (23%) said that all of the scenarios were equally serious. Sixteen percent of the experts said that the two scenarios involving the systematic plan and the spontaneous but deliberate attack were equally serious and more serious than the scenario involving the mistaken killings. Six percent of the participants said that the spontaneous but deliberate killings were the most serious. Nobody said that the mistaken killings were the most serious.

These results suggest that there is broad agreement with the OTP's hypothesis. A majority of the experts agreed that a crime committed as part of a systematic plan

⁹⁷ *Id.* at 21.

⁹⁸ *Id.*

⁹⁹ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 64.

¹⁰⁰ See Ford Survey, *supra* note 28, at 17.

¹⁰¹ *Id.*

was more serious than one committed spontaneously or as the result of a mistake. There was also a clear hierarchy within the scenarios, with the planned killings being worse than the spontaneous one that was, in turn, worse than the mistaken ones. Having said that, nearly a quarter of the experts did not see any difference in the seriousness of the various scenarios.

A similar issue was tested in Question 7. In that question, rebels attack a village and kill 100 unarmed villagers.¹⁰² It is composed of three scenarios. In one scenario the rebel commander is angry that the villagers support the government and orders his troops to open fire. In another, the rebel commander knows the village is undefended but orders his troops to attack anyway. In the final scenario, the attack takes place as part of a policy created by the overall leader of the rebel group to kill everyone who supports the government. This question essentially focused on whether the attack is negligent,¹⁰³ deliberate but not part of a larger policy,¹⁰⁴ or deliberate and part of an organized policy.¹⁰⁵

Interestingly, 65% of the experts gave all three scenarios the same score, suggesting that they saw no difference in their seriousness. The next largest group (19%) scored the scenario involving the plan to kill all government sympathizers as the most serious. Six percent said that the scenario involving deliberate but spontaneous killings was the most serious and another 6% said that the scenario involving the attack on the undefended village was the most serious.

The results of these two questions look a bit different. In Question 14, a clear majority said that the existence of a plan made the crime more serious, while in Question 7, a clear majority said that all of the scenarios were equally serious and apparently rejected the OTP's hypothesis that crimes committed as part of an organized policy were more serious than spontaneous crimes. It is not clear why the responses to these two questions appear to be different, but the results of the non-experts showed similar differences.¹⁰⁶ One possibility is that the inclusion of a scenario based on mistake in Question 14 caused more people to perceive a hierarchy in the crimes based on the level of culpability of the perpetrators compared to Question 7.¹⁰⁷

¹⁰² *Id.* at 10.

¹⁰³ The rebel Commander's decision to launch an attack on the village despite knowing that it is not defended looks something like negligence or recklessness. The facts in that scenario do not suggest that the Commander intended the villagers to die, although a reasonable person would probably have realized that would occur as a result of an armed attack on an undefended location. *Id.*

¹⁰⁴ This is the scenario where the rebel commander orders the attack because he is angry. The killings are deliberate, but there are no facts that suggest the commander is acting as part of larger plan or policy. *Id.*

¹⁰⁵ This is the scenario that states the attack is taken to carry out a decision by the overall rebel leader that all government supporters must be killed. *Id.*

¹⁰⁶ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 236–37 (noting that the non-experts were also more likely to say that all of the scenarios in Question 7 were equally serious compared to Question 14).

¹⁰⁷ *Id.* Cf. Robinson & Kurzban, *supra* note 3, at 1844 (noting that in domestic studies of crime seriousness, participants tended to assign less seriousness to acts that are done mistakenly compared to acts that are reckless or intentional).

Another possibility is that most people perceive there to be a difference between a high-level policy that results in the commission of crimes and a premeditated plan to commit a particular crime. If this is true then plan and policy are really two different gravity factors and, while a plan is an indicator of increased crime gravity, a policy that results in the commission of crimes is not. This issue deserves further study.

I. Abuse of Power or Official Capacity

The OTP also claims that crimes are more serious when they result from an abuse of power or official capacity.¹⁰⁸ This hypothesis was tested in question 9. In that question, government soldiers shoot and kill 100 unarmed villagers.¹⁰⁹ There are two different scenarios and the only difference between them is that in one scenario the soldiers simply attack the village and kill the people. In the other scenario, the government “announces a public meeting” in the village and “invites all the inhabitants to it.”¹¹⁰ When the villagers arrive for the meeting, they are attacked and killed by the soldiers. Essentially, the only difference between the two crimes is that in one of the scenarios the government abused its position to trick the villagers into coming to a meeting where they were killed. In the other scenario, they are killed without the abuse of power. A slim majority of the experts (52%) said that the scenario involving the government tricking villagers into attending a meeting where they will be attacked and killed was the most serious. Virtually all the rest of the experts (45%) said that both scenarios were equally serious.

J. The Identity of the Victims

The OTP hypothesizes that the identity of the victims can affect the severity of the crime. For example, the OTP suggests that crimes are more serious when they are directed against women or children.¹¹¹ This hypothesis was tested in Question 3. In that question, rebel soldiers attack a village and kill 100 unarmed villagers.¹¹² The question presents four different variations on that scenario. The only difference between the scenarios is the identity of the victims. In one scenario, “[m]ost of the victims are women,” in another “[m]ost of the victims are men,” in the third scenario “[m]ost of the victims are elderly,” and in the final scenario “[m]ost of the victims are children.”¹¹³ This question tests whether the identity of the victims changes the gravity of the crime.

The experts appeared to agree with only a part of the OTP’s hypothesis. While 42% of the experts said that all of the scenarios were equally serious, 52% said that

¹⁰⁸ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 64 (arguing that the seriousness of the crimes are affected by “the extent to which the crimes . . . resulted from the abuse of power or official capacity”).

¹⁰⁹ See Ford Survey, *supra* note 28, at 12.

¹¹⁰ *Id.*

¹¹¹ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 63. The OTP’s definition focuses on “crimes involving sexual or gender violence and crimes committed against children.”

¹¹² See Ford Survey, *supra* note 28, at 6.

¹¹³ *Id.*

Scenario 4 (the scenario involving child victims) was the most serious. There was virtually no support for the belief that crimes involving women or the elderly were the most serious. Ultimately, there was a slim majority supporting the belief that crimes are more serious when the victims are children, even as a significant minority indicated they do not believe that the identity of the victims makes a difference in the gravity of the offense.

K. *Combatants Hors de Combat*

The effect of being *hors de combat*¹¹⁴ was tested in Question 17. In that question, government soldiers attack a rebel village. In Scenario 1, the rebel soldiers defending the village are captured. The government subsequently kills 100 of the captured soldiers. In Scenario 2, government soldiers attack the village and kill 100 rebel soldiers during combat. This question was designed to test whether being *hors de combat* affects the gravity of the crime. Detainees are under the control of an enemy power and thus vulnerable to abuse by that power. For this reason, it is generally unlawful to execute captured enemy soldiers,¹¹⁵ while it is generally lawful to kill enemy soldiers during combat.¹¹⁶ Thus, one might expect participants to score the scenario where captured soldiers are killed as more serious than the scenario where the soldiers are killed during combat because of the vulnerability of the detainees.

Virtually all the experts (90%) agreed that the killing of the captured soldiers was more serious. The remaining 10% said that both scenarios were equally serious. None of the experts thought that killing the soldiers in combat was more serious. Moreover, 50% of the experts gave the scenario involving the death of the soldiers in combat a score of zero, indicating that they did not view it to be a crime at all.¹¹⁷ The results indicate that there is strong agreement among the experts that killing combatants who are *hors de combat* increases the gravity of the offense.

L. *Geographic Extent of the Crimes*

This factor focuses on the geographic spread of the crimes, which is defined in this Article as the number of crime sites at which the crimes occur.¹¹⁸ The OTP

¹¹⁴ “Hors de combat” is a French phrase used in international humanitarian law. It literally translates as “out of combat” and is used to mean a person who is no longer participating in hostilities. It applies to combatants who are detained by an adverse party, combatants who are defenseless because of injuries, and combatants who have clearly indicated an intent to surrender. See JEAN-MARIE HENCKAERTS & LOUISE DOWSALD-BECK, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW VOLUME 1: RULES 166–68 (2005).

¹¹⁵ *Id.* at 164 (noting in Rule 47 that attacks on persons who are in the power of an adverse party are prohibited).

¹¹⁶ *Id.* at 3 (explaining Rule 1 provides that “[a]ttacks may only be directed against combatants”).

¹¹⁷ This is the “correct” answer under international humanitarian law. Killing enemy combatants during an armed conflict is not generally a crime under international law. See *supra* note 114.

¹¹⁸ See Stuart Ford, *The Complexity of International Criminal Trials Is Necessary*, 48 GEO.

hypothesizes that the gravity of the crime will increase as the number of crime sites at which the crime occurs increases.¹¹⁹ This hypothesis was tested using a single question (Question 5), which presents two scenarios where a rebel group attacks a number of villages over a three week period and kills 300 unarmed villagers.¹²⁰ In Scenario 1, the attacks occur at 15 villages, while in Scenario 2 they occur at 45 villages, but the total number of victims is the same.¹²¹ If the OTP's hypothesis is correct, one would expect the participants to score the scenario involving more crime sites as more serious than the scenario involving fewer crime sites.

The results do not strongly support the OTP's hypothesis. While 39% of the experts agreed that the scenario involving a larger number of crime sites was more serious, 58% of them scored both scenarios the same.¹²² These results show that a majority of the experts did not agree that increasing the number of crime sites increased the gravity of the offense.

M. Temporal Extent of the Crimes

This factor tests whether the temporal spread of the crimes (i.e., the amount of time over which the crimes occurred) affects their gravity.¹²³ The OTP's reference to the "high intensity of the crimes over a brief period" suggests that the OTP believes that gravity increases as the period over which the crimes occurs decreases.¹²⁴ This hypothesis was tested with a single question (Question 11), which presents two scenarios where government soldiers set up a roadblock and execute rebel sympathizers.¹²⁵ In both scenarios, the soldiers set up 12 roadblocks and kill 150 people. The only thing that changes is the amount of time over which these crimes occur. In Scenario 1, the crimes occur over a period of six months. In Scenario 2, the crimes occur over a period of two months.¹²⁶ If the OTP's hypothesis is correct, participants will score Scenario 2 (where the crimes occur over a shorter period) to be more serious than Scenario 1.

The majority of the experts (81%) said that both scenarios were equally serious. Thirteen percent said the scenario with the greater intensity of crimes (Scenario 2) was the most serious, while 6% said that the scenario with the lower intensity of crimes (Scenario 1) was the most grave. Ultimately, these results provide little support for the OTP's hypothesis that crime gravity is affected by the temporal extent of the crimes. The majority of the experts did not believe that the temporal scope of the crimes had any effect on crime gravity.

WASH. INT'L L. REV. 151, 161 (2015) (defining a "crime site" as a location where a crime or part of a crime has allegedly taken place).

¹¹⁹ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 62.

¹²⁰ See Ford Survey, *supra* note 28, at 8.

¹²¹ *Id.*

¹²² One participant scored the scenario with fewer crime sites as more serious, but this is an outlier. See *infra* Table 2.

¹²³ See Policy Paper on Preliminary Examinations, *supra* note 30, ¶ 62.

¹²⁴ *Id.*

¹²⁵ See Ford Survey, *supra* note 28, at 14.

¹²⁶ *Id.*

IV. DISCUSSION

Having looked at the responses of the expert participants, the next step is to see how the views of the experts differ from the views of the non-experts. The first problem is how to know whether a difference between the responses of the experts and the non-experts is meaningful. If 50% of the experts agree that the presence of a particular gravity component increases crime gravity, but only 49% of the non-experts say the same thing, is there a meaningful difference between the view of the experts and the non-experts? What if the difference were 5% instead of 1%? Would that be meaningful?

Table 2: The Views of the Experts and Non-Experts				
Hypothesis	Expert Agreement	Non-Expert Agreement	Difference	Statistically Significant?
Extent of Harm (Q. 8)	90%	76%	+14%	No, p=.386
Hors de Combat (Q. 17)	90%	51%	+39%	Yes, p<.0001
Type of Crime ¹²⁷ (Q. 10)	77%	44%	+33%	No, p=.052
Number of Direct Victims (Q. 13)	77%	42%	+35%	Yes, p=.013
Type of Crime (Q. 2)	74%	40%	+34%	Yes, p=.006
Discriminatory Intent (Q. 1)	68%	41%	+17%	Yes, p=.016
Particular Cruelty (Q. 6)	65%	46%	+19%	No, p=.134
Discriminatory Intent (Q. 18)	61%	29%	+32%	Yes, p=.001
Particular Cruelty (Q. 16)	57%	55%	+2%	No, p=.842
Plan or Policy (Q. 14)	55%	12%	+43%	Yes, p<.0001
Number of Direct Victims (Q. 19)	53%	34%	+19%	No, p=.407
Abuse of Official Capacity (Q. 9)	52%	34%	+18%	No, p=.172
Child Victims (Q. 3) ¹²⁸	52%	26%	+26%	No, p=.103
Geographic Extent (Q. 5)	39%	18%	+21%	Yes, p=.042
Number of Indirect Victims (Q. 4)	28%	15%	+13%	No, p=.268
Means Employed ¹²⁹ (Q. 15)	26%	19%	+7%	No, p=.805
Plan or Policy (Q. 7)	19%	6%	+13%	No, p=.222

¹²⁷ For this question (and Question 2, which tests the same hypothesis), the percentages reported indicate the percentage of participants who scored murder as the most serious crime. *See supra* Section III.A.

¹²⁸ While the OTP hypothesized that the identity of the victims mattered to the gravity of the offense, several identities were tested (women, men, elderly, and children) and only the presence of child victims seemed to affect the gravity of the offense. *See supra* Section III.J. Consequently, this hypothesis has been reframed to focus on child victims. The percentages reported are the percentages that agreed that the presence of child victims increased the gravity of the offense.

¹²⁹ For this question, the percentages reported indicate the percentage of participants that scored Scenario 1 as the most serious. Scenario 1 was chosen because both the experts and the non-experts picked it as the most serious of the scenarios most frequently. *See supra* Section III.E.

Hypothesis	Expert Agreement	Non-Expert Agreement	Difference	Statistically Significant?
Means Employed ¹³⁰ (Q 12)	19%	17%	+2%	No, p=.802
Temporal Extent ¹³¹ (Q. 11)	13%	15%	-2%	No, p=.848

The problem is that only a sample of the population of experts and non-experts took the survey. If a different sample took the survey or even the same sample took the survey on a different day, we might expect to see some variation in the results just from random chance.¹³² Thus, to say that there is a meaningful difference between the groups, the differences between the experts and non-experts have to be large enough that they are not likely to be the result of chance variations in the sample.

There is a statistical test to determine whether the differences are likely to have occurred by random chance or whether they are more likely to represent real differences between the groups.¹³³ That test is the chi-squared test.¹³⁴ If the p-value of the chi-squared test is less than .05, then the difference between the experts and non-experts is unlikely to have occurred by chance.¹³⁵ Rather, the differences probably represent a real difference in the views of the experts and non-experts. In that case, we would say that the differences are statistically significant.¹³⁶

Table 2 above summarizes the differences between the views of the experts and non-experts. The first column indicates the gravity hypothesis being tested. The second column contains the percentage of the experts that agreed with that hypothesis. If a hypothesis was tested using more than one question, then the results for each question appear separately. The third column contains the percentage of non-experts that agreed with that hypothesis. The fourth column shows the difference between the experts and the non-experts.¹³⁷ A positive result indicates the experts showed a higher level of agreement with the hypothesis than the non-experts. A negative result indicates the reverse. Finally, the fifth column indicates whether that difference is statistically significant. The hypotheses that are statistically significant have been shaded gray to make it easier to identify them. The rows are ordered from the hypothesis with the greatest percentage of expert agreement to the one with the lowest percentage of expert agreement.

¹³⁰ For this question, the percentages reported indicate the percentage of participants that scored Scenario 3 as the most serious. Scenario 3 was chosen because both the experts and the non-experts picked it as the most serious of the scenarios most frequently. *Id.*

¹³¹ For this hypothesis, the percentages reported indicate the percentage of participants who agreed that gravity increased as the intensity of the crimes increased (i.e., as the crimes were committed over a shorter period of time). *See supra* Section III.M.

¹³² *See* ALAN C. ACOCK, A GENTLE INTRODUCTION TO STATA 123 (Stata Press 3d ed. 2012).

¹³³ *Id.*

¹³⁴ *Id.* JEFFERY T. WALKER & SEAN MADDAN, STATISTICS IN CRIMINOLOGY AND CRIMINAL JUSTICE 193–94 (4th ed. 2013).

¹³⁵ *See* ACOCK, *supra* note 132, at 124.

¹³⁶ *See* WALKER & MADDAN, *supra* note 134, at 196 (discussing statistical significance).

¹³⁷ It was calculated as: (expert agreement) – (non-expert agreement) = difference.

There are both similarities and differences between the expert and non-expert responses. The sections below will discuss those similarities and differences.

A. General Agreement About the Relative Importance of the Gravity Factors

The most obvious similarity between the responses of the experts and the non-experts is that there is broad agreement about the relative importance of the various gravity factors. The factors that have the broadest support among the experts also tend to have the broadest support among the non-experts. Thus, factors like the extent of the harm, the type of crime committed, the number of direct victims, discriminatory intent and the presence of particular cruelty were among the strongest indicators for both the experts and the non-experts. At the same time, there was also agreement that some of the components of the OTP's gravity definition are poor indicators of crime gravity. Thus, factors like the temporal extent of the crimes, the number of indirect victims, the means employed, and the presence of a plan or policy had weak support among both groups.

To put it in quantitative terms, there was a high degree of correlation between the order in which the experts and the non-experts listed the factors in Table 2. The degree of correlation between the experts' and the non-experts' ranking of the gravity factors was evaluated using Spearman's rho, which is a test of the correlation in ranks between two lists.¹³⁸ The rank ordering of the gravity factors by the experts and the non-experts was highly correlated ($\rho=0.83$)¹³⁹ and that correlation was statistically significant ($p<.0001$). It is highly unlikely that the similarities in the rankings are the result of random chance. The two groups really do exhibit a high level of agreement about the relative importance of the various factors.

Thus, while there are some differences between the experts' and non-experts' perceptions of the relative importance of the various gravity factors, they are not fundamental in nature. This is probably good news for the ICC. If the experts and the non-experts had deep disagreements about the relative importance of the gravity factors, this might have made the work of the ICC harder in the long run.¹⁴⁰

This general agreement between the experts and non-experts also reinforces the need to amend the OTP's gravity definition. If large percentages of both the experts and the non-experts agree that factors like the temporal extent of the crimes, the means employed, or the number of indirect victims do not affect the gravity of mass atrocities, then why include them as gravity factors? It would be desirable to eliminate the weak factors and focus the definition on those factors that both groups agree are strong indicators of crime gravity.¹⁴¹

¹³⁸ See GRAHAM UPTON & IAN COOK, OXFORD DICTIONARY OF STATISTICS 401–02 (Oxford University Press 3d ed. 2014); ACOCK, *supra* note 132, at 201–02.

¹³⁹ Like the more common Pearson's correlation coefficient, Spearman's rho can have values between -1 and 1, with higher values indicating a higher degree of similarity between the rank ordering of the lists. See UPTON & COOK, *supra* note 138, at 401.

¹⁴⁰ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 212–14, 249 (arguing that the definition of gravity used by the ICC should match the expectations of non-experts to maximize the likelihood that those non-experts will be persuaded to comply with the norms of international criminal law).

¹⁴¹ See *id.* at 249.

B. Experts Show Higher Levels of Overall Agreement with the Gravity Definition

While there is general agreement about the relative importance of the various factors, the results in Table 2 also exhibit one clear difference. On virtually every question—the only exception being Question 11 about the temporal extent of the crimes—the experts agreed that the factors tested increased crime gravity at higher rates than the non-experts.¹⁴² While not all of the differences were statistically significant, the experts were almost always more supportive of the components of the OTP's gravity definition than the non-experts.

On one level, this is not surprising. The Office of the Prosecutor at the International Criminal Court has numerous international criminal law experts on staff.¹⁴³ And it seems very likely that the OTP's gravity definition was written by those experts. So, it is not surprising that a group of self-described experts largely agrees with a gravity definition that was probably written by other experts. But the high degree of agreement among the experts, combined with the lower overall level of agreement among the non-experts suggests there could be a problem.

The problem is that most experts, because they are experts, probably do not realize that there are differences between their views on the gravity of international crimes and the views of non-experts. After all, the experts show high levels of agreement with the OTP's gravity definition. Thus, they are likely to view the gravity definition as satisfactory because it accords with their own views on crime gravity. But the non-experts consistently show lower levels of agreement.

For example, of the nineteen questions on the survey, there was only one question on which more than 75% of the non-experts agreed with the hypothesis—Question 8 about the extent of the harm suffered by the victims. There were two questions where between 51% and 75% of the non-experts agreed with the hypothesis. Those were the questions on particular cruelty (Question 16) and combatants *hors de combat* (Question 17). The majority of the questions (9 of 17) received support from between 25% and 50% of the non-experts. The remainder (7 questions) received support from less than 25% of them.

In contrast, the experts showed much higher levels of agreement with many more of the questions. Four questions received support from more than 75% of them. Nine questions received support from between 51% and 75% of the experts. Three questions received support from between 25% and 50% of the non-experts, while three questions received support from less than 25% of them. Table 3 summarizes these differences.

¹⁴² See *supra* Table 2.

¹⁴³ See *Who's Who, Office of the Prosecutor*, INT'L CRIM. COURT, <https://www.icc-cpi.int/about/otp/who-s-who> [<https://perma.cc/DP6F-G5XC>].

Degree of Agreement	Experts	Non-Experts
76%–100%	4 questions	1 question
51%–75%	9 questions	2 questions
26%–50%	3 questions	9 questions
1%–25%	3 questions	7 questions

These results suggest that ICL experts will tend to significantly overestimate how much support there is for the gravity factors in the OTP’s gravity definition among non-experts. This is not a huge problem as the non-experts generally agree with the experts’ rankings of the relative importance of the factors.¹⁴⁴ But experts (and the ICC itself) should be aware that non-experts are likely to be generally less supportive of the various gravity factors that have been advanced.

C. Exploring the Differences

Out of the 19 questions on the survey, the responses of the experts were meaningfully different from the non-experts on seven of those questions. These are the seven questions highlighted in gray in Table 2, where the responses of the experts were statistically significantly different from the responses of the non-experts.¹⁴⁵ Of course, there were differences between the experts and non-experts on all twelve of the other questions as well, but the differences were small enough that they could be the result of random chance. Only for the seven questions highlighted in gray, can we be reasonably certain that the differences are meaningful.

In a number of cases, the statistically significant differences appear to be the result of the experts’ training in ICL. This is most striking in response to Question 17 about the treatment of combatants who are *hors de combat*. While 90% of the experts said that killing combatants *hors de combat* was more serious than killing combatants, only 51% of the non-experts agreed. In contrast, while only 10% of the experts said that both scenarios were equally serious, 46% of the non-experts thought they were equally serious.¹⁴⁶ These differences were highly statistically significant.¹⁴⁷ In other words, there was a real and meaningful difference between how the experts and the non-experts saw the treatment of combatants vs. non-combatants. Experts were much more likely to view killing combatants *hors de combat* as more grave.

In answering Question 17, experts were also much more likely than non-experts to give the “correct” answer—the answer that best accords with international humanitarian

¹⁴⁴ See *supra* Section III.A.

¹⁴⁵ See *supra* Table 2.

¹⁴⁶ See *id.*

¹⁴⁷ With a p-value of less than .0001, there is less than a 1 in 10,000 chance that these differences are the result of random chance.

law. Under international law, killing combatants during combat is not generally unlawful whereas killing combatants who are *hors de combat* is unlawful.¹⁴⁸ If killing combatants during combat is not a crime then logically it should be given a score of zero. And indeed, 50% of the experts gave the scenario involving the killing of the combatants a score of zero. In contrast, only 16% of the non-experts gave the killing of combatants a score of zero.

The responses to Question 17 suggest that the experts' answers were influenced by their training in ICL. They were much more likely to answer the question in a way that accords with ICL. But it also suggests that this particular rule—the killing of combatants during combat is lawful—is not one that is intuitively or generally understood by non-experts. They were much more likely to see the killings during combat as unlawful. Something similar appears to be happening in the responses to the questions about discriminatory intent.

The salience of a discriminatory intent was tested in Questions 1 and 18. In response to both questions, experts were much more likely than non-experts to say that the presence of a discriminatory intent made the crime more serious. In both cases, the differences were statistically significant.¹⁴⁹ Moreover, the differences may be the result of the experts' training in ICL. International criminal law stresses the importance of discriminatory intent. For example, the crime against humanity of persecution is defined by the presence of a discriminatory intent.¹⁵⁰ Indeed, persecution is often seen as the most serious crime against humanity precisely because of the requirement of a discriminatory intent.¹⁵¹ Similarly, the hallmark of genocide is the presence of a discriminatory intent.¹⁵² It is this discriminatory intent requirement that has led many courts to claim that genocide is the most serious crime under international law—the “crime of crimes.”¹⁵³

A majority of the experts agreed in both questions that a crime committed with a discriminatory intent was more serious than the same crime committed without

¹⁴⁸ See *supra* notes 112–15 and accompanying text.

¹⁴⁹ See *supra* Table 2.

¹⁵⁰ See *supra* note 94 (outlining various sections of the Rome Statute describing criminal intent). See also *Prosecutor v. Kvočka*, Case No. IT-98-30/1-A, Appeals Chamber Judgement, ¶ 460 (Int'l Crim. Trib. For the Former Yugoslavia Feb. 28, 2005) (“With regard to the required *mens rea* [for persecution], the Appeals Chamber reiterates that persecution as a crime against humanity requires evidence of a specific intent to discriminate on political, racial or religious grounds.”); *Prosecutor v. Brđanin*, Case No. IT-99-36-T, Trial Chamber Judgement, ¶ 996 (Int'l Crim. Trib. For the Former Yugoslavia Sept. 1, 2004) (“The crime of persecution . . . derives its unique character from the requirement of a specific discriminatory intent.”).

¹⁵¹ See, e.g., *Prosecutor v. Obrenović*, Case No. IT-02-60/2-S, Sentencing Judgement, ¶ 65 (Int'l Crim. Trib. For the Former Yugoslavia Dec. 10, 2003) (“The Trial Chamber considers that the seriousness of the crime of persecutions cannot be emphasised enough: this is a crime that can be committed in different manners and incorporates manifold acts. It is the abhorrent discriminatory intent behind the commission of this crime against humanity that renders it particularly grave.”).

¹⁵² See Robert Petit et al., *Exploring Critical Issues in Religious Genocide: Case Studies of Violence in Tibet, Iraq and Gujarat*, 40 CASE W. RES. J. INT'L L. 163, 211–13 (2007–08).

¹⁵³ *Id.* at 213.

that intent (68% in response to Question 1 and 61% in response to Question 18). On the other hand, the non-experts were much less likely to see the presence of a discriminatory intent as an indicator of crime gravity (41% and 29% respectively).¹⁵⁴ These results suggest that the experts' views may have been influenced by ICL's emphasis on the importance of a discriminatory intent in assessing the gravity of offenses. They also suggest that non-experts do not place nearly as much importance on discriminatory intent as the experts or ICL.

Something similar may be occurring in Question 13. This question tested whether crime gravity increases as the number of direct victims increases. There were statistically significant differences between the responses of the experts and the non-experts, with the experts much more likely to say that increasing the number of victims increased the gravity of the offense (77% for the experts versus 42% for the non-experts).¹⁵⁵ Again, this represents a question where the experts gave answers that are consistent with the outcomes one would expect under international criminal law. A number of cases have held that crime gravity increases as the number of victims increases.¹⁵⁶ This seems to be another example of a situation where the experts responses were influenced by their training in ICL. It also highlights another example of a rule of ICL that may not be widely accepted by non-experts.

There also appears to be a difference in how experts and non-experts view the hierarchy of crimes. How the type of crime affected crime gravity was tested in two questions (Question 2 and Question 10). The differences in Question 2 were statistically significant ($p=.006$) while the differences in Question 10 just barely missed statistical significance ($p=.052$). In both questions, the experts were highly likely to view murder as more serious than other crimes (77% in Question 10 and 74% in Question 2). The non-expert responses were a little different. The non-experts were much more likely to view rape as equal to or greater than murder in severity.

So, for example, in response to Question 2, 40% of the non-experts said that murder was the most serious crime. But a combined 37% said either that rape and murder were equal in seriousness or that rape was more serious than murder. Something very similar happened in response to Question 10. Forty-four percent of the non-experts viewed murder as the most serious offence, while a combined 31% thought that rape was either equal to murder in its severity or more serious. The responses for the experts and the non-experts are consistent across both questions.

¹⁵⁴ See *supra* Table 2.

¹⁵⁵ See *id.*

¹⁵⁶ See *Ndindabahizi v. Prosecutor*, Case No. ICTR-01-71-A, Appeals Chamber Judgement, ¶ 135 (Jan. 16, 2007) (“[T]he Trial Chamber did not err in considering the large number of victims at Gitwa Hill [in the Biseseo Hills, Kibuye prefecture] as an aggravating circumstance relevant to the sentence.”); *Prosecutor v. Serugendo*, Case No. ICTR-2005-84-I, Judgement and Sentence, ¶ 90 (June 12, 2006) (“The number of victims which resulted from the incitement to genocide and persecutions is indeed an aggravating factor.”); *Prosecutor v. Babić*, Case No. IT-03-72-S, Sentencing Judgement, ¶ 47 (Int’l Crim. Trib. For the Former Yugoslavia June 29, 2004) (holding that in assessing gravity, “primary consideration” is to be given to “the number of victims”); *Prosecutor v. Češić*, Case No. IT-95-10/1-S, Sentencing Judgement, ¶ 32 (Int’l Crim. Trib. For the Former Yugoslavia Mar. 11, 2004) (holding that the number of victims is among the “particular circumstances to be considered”).

The experts are much more likely than the non-experts to see murder as the most serious offense. The non-experts are much more likely to see murder as roughly equivalent to rape in its severity.¹⁵⁷

It is not clear that this difference has much to do with the content of ICL. While there is some support in ICL for the proposition that killings are worse than crimes that do not result in death, it is far from a black-and-white rule.¹⁵⁸ Rather, several courts have stressed that the overall gravity of the crime is driven by the particular circumstances of its commission rather than rigid rules about the gravity of particular offenses.¹⁵⁹ Ultimately, it is not clear why the experts perceive a much more rigid hierarchy of the crimes than the non-experts.

One possibility is that the differences stem from the fact that the sample of experts contained a higher percentage of men than the sample of non-experts.¹⁶⁰ If women are more likely than men to view rape as equal in gravity to murder, then that could explain the differences between the expert and non-expert groups. And indeed, there were some differences between the responses of the men and women among the non-expert participants. Women were more likely than men to say that rape and murder were equally serious.¹⁶¹ Those differences were not, however, statistically significant.¹⁶² Thus, they could have been the result of random chance. As a result, it is difficult to draw any firm conclusions on this point. Nevertheless, the question of whether there are meaningful differences in how men and women perceive gravity deserves further study.

¹⁵⁷ For more discussion of this issue, see Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 229–31.

¹⁵⁸ See *Prosecutor v. Furundzija*, Case No. IT-95/17-1-A, Appeals Chamber Judgement, ¶ 246 (Int'l Crim. Trib. For the Former Yugoslavia July 21, 2000) (holding that “the view that crimes resulting in loss of life are to be punished more severely than those not leading to the loss of life . . . to be too rigid and mechanistic”). *But see* *Prosecutor v. Delalić*, Case No. IT-96-21-A, Appeals Chamber Judgement, ¶ 732 (Int'l Crim. Trib. For the Former Yugoslavia Feb. 20, 2001) (“A failure to prevent or punish murder or torture committed by a subordinate must be regarded as being of greater gravity than a failure to prevent or punish an act of plunder, for example.”).

¹⁵⁹ See *Prosecutor v. Kamuhanda*, Case No. ICTR-99-54A-A, Appeals Chamber Judgement, ¶ 357 (Sept. 19, 2005) (“The principle of individualization requires that each sentence be pronounced on the basis of the individual circumstances of the accused and the gravity of the crime.”); *Prosecutor v. Seromba*, Case No. ICTR-2001-66-I, Trial Chamber Judgement, ¶ 381 (Dec. 13, 2006) (“The Chamber recalls that an evaluation of the gravity of offences is based on the crimes charged against the accused, that is, the individual circumstances under which the offences were committed, and not on a hierarchy of crimes.”); *Prosecutor v. Blagojević*, Case No. IT-02-60-T, Trial Chamber Judgement, ¶ 832 (Int'l Crim. Trib. For the Former Yugoslavia Jan. 17, 2005) (“[T]he Appeals Chamber has stressed that the sentence should be individualised and that the particular circumstances of the case are therefore of primary importance.”).

¹⁶⁰ See *supra* Table 1 (noting that 41% of the experts were female compared to 66% of the non-experts).

¹⁶¹ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 230–31 (discussing differences in how men and women answered the questions about the type of crime).

¹⁶² *Id.*

Taken together, the experts' responses to the questions about the treatment of combatants *hors de combat*, discriminatory intent and the number of victims, suggest two conclusions. First, experts' views about ICL appear to be shaped by ICL. This hardly seems surprising. After all, to become an expert in ICL one must study ICL. It seems logical that the study of ICL will affect how people view the gravity of international crimes. But it also means that experts need to be aware that by becoming an expert in ICL, it is likely that their views have become different in some ways from those of non-experts. Experts cannot simply assume that everyone shares their views.

Second, and potentially more troubling, there appear to be several ICL rules that do not match up particularly well with non-experts' expectations about what constitutes a grave crime under international law. For example, the non-experts were much less likely to attach importance to the presence of discriminatory intent. In contrast, they were more likely to focus on what was done to the victims rather than why it was done. At the same time, the idea that combatants are lawful targets during an armed conflict is fundamental to international humanitarian law, but most non-experts did not make a distinction between killing a combatant during combat and killing a combatant who is *hors de combat*.

It is not necessarily a problem if non-experts' expectations about the law do not match the law itself. There may be valid reasons for the rules that exist even if they are not intuitive for non-experts.¹⁶³ But people may be less likely to comply with laws that do not match their expectations.¹⁶⁴ In that sense, it seems important to know which rules non-experts are more likely to agree with and which ones they are less likely to agree with.

CONCLUSION

This Article has explored the differences between how experts and non-experts perceive the gravity of mass atrocities. And, while there are some differences, the most important conclusion is that both groups largely agree about the relative importance of the various gravity factors that were tested. Those factors that the non-experts thought were the most important also tended to be the factors that the experts thought were most indicative of crime gravity and vice versa.¹⁶⁵ This indicates that there is broad agreement between experts and non-experts about which factors contribute to crime gravity. This agreement is good news as the court is most likely to succeed

¹⁶³ For example, the deliberate killing of enemy combatants during combat is one of the principle ways in which armed conflict occurs. Indeed, the St. Petersburg Declaration of 1868 stated, "the only legitimate object which States should endeavor to accomplish during war is to weaken the military forces of the enemy." St. Petersburg Declaration, Nov. 29, 1868, preamble, Int'l Comm. of the Red Cross. It goes on to say that it is therefore legitimate "to disable the greatest possible number [of the enemy's forces]." *Id.* Prohibiting the killing of enemy combatants would effectively prohibit armed conflict. Most states have been unwilling to accept a blanket prohibition on armed conflict. As a result, international humanitarian law generally permits the killing of enemy combatants. See HENCKAERTS & DOWSALD-BECK, *supra* note 114, at 3 (noting Rule 1 provides that "[a]ttacks may only be directed against combatants").

¹⁶⁴ See Ford, *Meaning of Gravity at the ICC*, *supra* note 9, at 212–13.

¹⁶⁵ See *supra* Section IV.A.

if it meets the expectations of non-experts. This is more likely to occur if both experts and non-experts agree about the meaning of gravity.

At the same time, there is one overall difference between the responses of the experts and the non-experts. The experts almost uniformly showed higher levels of support for the various gravity factors than the non-experts.¹⁶⁶ This finding is not necessarily problematic given that both groups largely agree about the relative ranking of the factors, but ICL experts need to be aware that non-experts show lower overall levels of agreement.

Finally, there are some gravity factors where the responses of the non-experts are significantly different from the responses of the experts. While this Article cannot explain all of these differences, it appears that several stem from the study of international criminal law itself. There were several questions—particularly the question related to the treatment of combatants who are *hors de combat*, the questions related to discriminatory intent, and the question related to the number of victims—where the experts' answers were consistent with the “correct” answer under international criminal law but were significantly different from the answers of the non-experts.¹⁶⁷ This suggests that the non-experts did not agree with the particular rule of ICL that underlaid the question.

The difference is most striking with regard to the treatment of active combatants versus combatants who are *hors de combat*. The non-experts were more likely to view the killing of active combatants as an indicator of crime gravity. In contrast, the experts were more likely to view the killings of active combatants as not a crime at all, which is consistent with what one would expect under ICL.¹⁶⁸ This difference suggests that the experts have internalized the rules of ICL, but that the non-experts do not intuitively recognize such a rule. Exploring areas where the expectations of non-experts differ from the rules of ICL deserves further study.

¹⁶⁶ See *supra* Section IV.B.

¹⁶⁷ See *supra* Section IV.C.

¹⁶⁸ See *supra* Section IV.C.