

THE MEANING OF GRAVITY AT THE INTERNATIONAL CRIMINAL
COURT: A SURVEY OF ATTITUDES ABOUT THE SERIOUSNESS OF
MASS ATROCITIES

*Stuart Ford**

TABLE OF CONTENTS

I. INTRODUCTION	210
II. THE IMPORTANCE OF GRAVITY	212
III. CHOOSING A GRAVITY DEFINITION.....	215
IV. METHODOLOGY	217
V. THE PARTICIPANTS.....	221
VI. SURVEY RESULTS.....	222
A. Scale of the Crimes.....	223
1. Number of Direct Victims	223
2. Number of Indirect Victims.....	225
3. Extent of the Harm	226
4. Geographic Spread of the Crimes.....	226
5. Temporal Spread of the Crimes.....	227
B. Nature of the crimes	228
1. The Type of Crimes	228
2. The Identity of the Victims.....	231
3. Discriminatory Intent.....	231
C. Manner of Commission	233
1. The Means Employed.....	233
2. Plan or Policy	235
3. Abuse of Power or Official Capacity.....	237
4. Particular Cruelty.....	238
D. The Impact of the Crimes	239
1. Suffering of the Victims	239
2. The Increased Vulnerability of the Victims.....	240
3. The Terror Subsequently Instilled	241
4. The Social, Economic or Environmental Damage.....	241
VII. DISCUSSION	241

* Copyright © 2018 Stuart Ford, Associate 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. This article was improved by the comments of Professor William Ford. Any mistakes that remain are mine.

A. Strong Indicators of Crime Gravity	242
B. Moderate Indicators of Crime Gravity	243
C. Weak Indicators of Crime Gravity	243
D. Other Observations	245
1. Reliability of the Results	245
2. A Minority Rejects All Gravity Components?	245
3. Individuals Have Different Baselines	246
VIII. CONCLUSIONS	247
IX. NEXT STEPS.....	249

I. INTRODUCTION

Gravity is an enormously important concept at the International Criminal Court (ICC). The word appears nine times in the Rome Statute and is crucial at every stage of the proceedings.¹ It is an important factor in decisions about which situations to investigate, which individuals the court will try, and what sentences to impose on those convicted of violating international criminal law.²

Gravity may also be important for the long-term success of the court. The Rome Statute states that the ICC will exercise jurisdiction over “the most serious crimes” that “deeply shock the conscience of humanity.”³ It also claims that the prosecution of these “grave crimes” will “contribute to [their] prevention.”⁴ And it may be true that if the court is perceived as investigating and prosecuting the gravest crimes that this will help prevent such crimes.⁵ But the court must not only *claim* to prosecute the most serious offenses, it must also *be seen* to prosecute the most serious offenses. If people perceive that the ICC is prosecuting the most serious crimes, then it is more likely that they will view the court as legitimate⁶ and comply with the norms contained in the Rome Statute.⁷

¹ See *infra* Section III

² *Id.*

³ See Rome Statute, Preamble.

⁴ *Id.*

⁵ See *infra* Section II.

⁶ The legitimacy discussed throughout this article is perceived legitimacy, which is also sometimes called sociological legitimacy. It refers to how audiences subjectively perceive the legitimacy of the ICC. See Stuart Ford, *A Social Psychology Model of the Perceived Legitimacy of International Criminal Courts: Implications for the Success of Transitional Justice Mechanisms*, 45 VANDERBILT J. INT’L L. 405, n. 1 (2012) (discussing various types of legitimacy).

⁷ See *infra* Section II.

This dynamic places a great deal of weight on the meaning of gravity within the Statute. If the ICC's gravity definition does not accord with most people's expectations, then they are less likely to perceive the court as prosecuting the gravest offenses. If people do not perceive the court to be focusing on crimes of the greatest gravity, they are less likely to believe the ICC is legitimate, and less likely to conform to its norms.⁸ Thus, it is very important that the court's gravity decisions match people's expectations about those decisions. Unfortunately, there is very little scholarship on how the population as a whole views the gravity of mass atrocities.

This article aims to fill that void by surveying people about their understanding of crime gravity. The survey asked participants to rate the seriousness of different mass atrocities.⁹ The participants' scores were then used to determine which factors increase the perceived gravity of crimes. The results indicate that there are some areas where there is relatively broad agreement. Factors like the extent of the harm suffered by the victim and the nature of the crime committed were strong indicators of crime gravity.¹⁰ At the same time, factors like the number of indirect victims or the temporal scope of the crimes turned out to be weak indicators of crime gravity.¹¹

Overall, the results are good news for the ICC.¹² They need to be replicated in other cultures,¹³ but there does appear to be relatively broad agreement about some factors that most people associate with crime gravity. These factors could be used to construct a definition of crime gravity that most people would agree with and that they are more likely to view as legitimate.¹⁴ At the same time, some of the factors currently used by the ICC are weak indicators of crime gravity.

The survey results suggest that the gravity definitions currently in use by the ICC should be modified to focus on the strong factors and remove the weak factors. This way, the ICC's gravity definition will conform to what most people understand gravity to mean.¹⁵ Ultimately, having a definition of gravity that most people agree with will improve the court's legitimacy and help it accomplish its long-term goal of preventing violations of international criminal law.

This article will proceed as follows. Section II discusses the importance of the concept of crime gravity to the success of the ICC, while Section III

⁸ *Id.*

⁹ *See infra* Section IV (describing the methodology of the article).

¹⁰ *See infra* Section VII.

¹¹ *Id.*

¹² *See infra* Section VIII.

¹³ *See infra* Section IX.

¹⁴ *See infra* Section VIII.

¹⁵ *Id.*

describes the gravity definition that was tested in the survey. The survey's methodology is explained in Section IV and the demographics of the participants are presented in Section V. Section VI contains the results of the survey and a discussion of those results is found in Section VII. Finally, Section VIII reports the article's key conclusions, while some ideas for future research on crime gravity are presented in Section IX.

II. THE IMPORTANCE OF GRAVITY

Gravity is a subjective rather than an objective component of crimes. Unlike physical characteristics like mass, circumference or height, crime gravity cannot easily be measured except by looking at how people view the crime.¹⁶ Because it is a subjective characteristic, one possibility is that it does not have the same meaning for most people. Perhaps every person's view of gravity is idiosyncratic and there is no generalized meaning of gravity that most people hold.¹⁷ If true, this would have serious implications for the ICC. The Rome Statute uses the concept of gravity extensively and it is central to the work of the court.¹⁸ It is critical in deciding which cases to investigate, which individuals to charge, and what sentences to impose on those found guilty.¹⁹ But, if the term has no generally agreed meaning, it would be almost impossible to apply the Rome Statute in ways that most people consistently agreed with. This would decrease the ICC's legitimacy and effectiveness.²⁰

For example, if the meaning of gravity is idiosyncratic, then most people will not be able to understand or agree with the ICC's sentencing decisions. No matter what definition of gravity the court used, most people would not agree with it or the sentencing decisions that resulted from its application. Similar problems would arise with the court's decisions about which situations to investigate and who to prosecute. If gravity is idiosyncratic, then most people will not be able to understand or agree with

¹⁶ Cf. Stuart Ford, *Complexity and Efficiency at International Criminal Courts*, 29 EMORY INT'L L. REV. 1, 19 (2014) (discussing subjective variables).

¹⁷ Cf. Margaret M. deGuzman, *Choosing to Prosecute: Expressive Selection at the International Criminal Court*, 33 MICH. J. INT'L L. 265, 288 (2012) (arguing that there is not widespread agreement about how to identify the gravest crimes and that different people may use different criteria); see also Paul H. Robinson and Robert Kurzban, *Concordance and Conflict in Intuitions of Justice*, 91 MINN. L. REV. 1829, 1831-36 (2007) (noting that many theorists have questioned whether people have a common understanding of what affects the seriousness of a crime).

¹⁸ See Margaret M. deGuzman, *Gravity and the Legitimacy of the International Criminal Court*, 32 FORDHAM INT'L L.J. 1400 (2009).

¹⁹ See *infra* Section III.

²⁰ See deGuzman, *supra* note 18, at 1435-45 (arguing that the ICC's gravity decisions have a significant effect on its perceived legitimacy).

those decisions either. This would have dire consequences for the court's ability to affect change in the world.²¹ According to procedural justice theory, the ability of legal institutions to persuade individuals to comply with the law is tied to whether individuals perceive those institutions as legitimate.²² But people who cannot understand or agree with the ICC's decisions about where to investigate, who to charge and how much to punish them, are unlikely to perceive the ICC as legitimate.²³ This would make it less likely that the court will be able to persuade people to comply with international criminal law.²⁴

On the other hand, if there is a broadly accepted meaning of gravity and the ICC uses that concept of gravity in deciding where to investigate, who to charge, and how to sentence people, then people are more likely to view the ICC's decisions as legitimate. People will view the court as not just saying that it prosecutes the gravest offenses, but will see it as actually prosecuting the worst crimes. This should increase the court's perceived legitimacy.²⁵ And if people view the court as legitimate, they are more likely to comply with the requirements of international criminal law.²⁶ Thus, the question of whether there is a generally agreed meaning of gravity and what that

²¹ See Yuval Shany, *Assessing the Effectiveness of International Courts: A Goal-Based Approach*, 106 AMERICAN J. INT'L L. 225, 230, 248-49 (2012) (arguing that the success of international courts should be measured by their ability to affect change in the external world).

²² See generally Tom R. Tyler, *Psychological Perspectives on Legitimacy and Legitimation*, 57 ANNUAL REVIEW OF PSYCHOLOGY 375-79 (2006) (providing an overview of psychological research relating to legitimacy and 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.")

²³ Cf. Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME AND JUSTICE 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 and treat members of the public."); see also deGuzman, *supra* note 18, 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.")

²⁴ See Stuart Ford, *A Hierarchy of the Goals of International Criminal Courts*, 27 MINN. J. INT'L L. 179, 221-28, 234-42 (2018) (arguing that the most important goal of international criminal courts is to prevent violations of international criminal law).

²⁵ Cf. deGuzman, *supra* note 18, at 1439, 1444 (noting that the "Statute's preamble leaves no doubt that a key component of [the Court's] integrity is the adjudication of *serious* crimes" and also arguing that "'global society' will consider the ICC's actions legitimate when they conform to the common understanding of the Court's purpose"). That common understanding is the ICC will prosecute the most serious crimes. That is, after all, what the court promises to do in the Preamble to the Rome Statute.

²⁶ See deGuzman, *supra* note 18, at 1438 (arguing that the Court's perceived legitimacy "is a critical component of" its success). A similar argument is made in the Preamble of the Rome Statute, where the drafters of the Statute argued that if the ICC focused on prosecuting the most grave crimes, this would help end impunity and prevent future violations. See *infra* note 35.

generally agreed meaning is has important implications for the success of the ICC.

It seems plausible, given the subjective nature of the concept, that there is no generally agreed meaning of gravity, but it is far from certain. Crime seriousness (which effectively means the same thing as crime gravity) has been studied empirically in domestic settings.²⁷ It turns out that while crime gravity is a subjective characteristic, there is fairly broad agreement about the relative gravity of domestic crimes.²⁸ This has been demonstrated through surveys that ask people to rank different crime scenarios as more or less grave. The results show that most people have fairly consistent views about the gravity of the kinds of crimes that are most common in domestic criminal justice systems.²⁹ Thus, it seems possible that there will be similarly broad agreement about the relative gravity of the kinds of crimes that are typically investigated and prosecuted by international criminal courts.

On the other hand, there are significant differences between the crimes that are typically prosecuted in domestic settings and the crimes that are typically prosecuted at international criminal tribunals.³⁰ The vast majority of domestic crimes are non-violent drug crimes and property crimes.³¹ Murders are very rare and the kind of widespread and systematic violence directed against civilians that is the hallmark of international crimes is virtually unknown in domestic settings.³² As a result, it is not certain that the results of the domestic crime severity studies will translate into broad agreement about the gravity of international crimes. The only way to know for sure is to test it.

This article is the first attempt to understand how people view the gravity of international crimes. It uses a survey to test the Office of the Prosecutor's (OTP) gravity definition to see whether there is broad agreement about the gravity of international crimes and (if so) whether the OTP's definition accurately captures the factors people believe affect crime

²⁷ See generally Robinson and Kurzban, *supra* note 17.

²⁸ *Id.* at 1848, 1854 (“[A] substantial body of research indicates a broad consensus regarding the relative seriousness of different wrongdoings and the appropriate relative amount of punishment... Using a range of techniques, previous empirical studies confirm a nearly universal human intuition that serious wrongdoing deserves punishment.”).

²⁹ *Id.* at 1855-60 (describing the results of domestic crime severity studies and noting that “the studies consistently show a significant level of agreement on intuitions of justice, even across demographics”).

³⁰ See Stuart Ford, *What Investigative Resources Does the International Criminal Court Need to Succeed?: A Gravity-Based Approach*, 16 WASH. U. GLOBAL STUDIES L. REV. 1, 37-38 (2017).

³¹ *Id.*

³² *Id.*

gravity.³³ If there is a broadly held understanding of the gravity of mass atrocities, that should be visible in the survey results. Similarly, if gravity is an idiosyncratic concept, that too should be visible. Either way, we will know much more about how people understand the gravity of mass atrocities.

If the survey results show that there is broad agreement about the relative gravity of international crimes, this should have a significant impact on the ICC. First, unless the OTP has perfectly captured the meaning of gravity, its definition should be amended to match what people generally understand gravity to mean. This could profoundly affect how the ICC does its job. Any changes in the ICC's gravity definition would affect questions like which situations should be investigated, which cases should be prosecuted, and what sentences should be imposed. Second, and more importantly, if the ICC modifies its definition of gravity so that it is consistent with how most people view the seriousness of crimes, it should improve the legitimacy and the effectiveness of the court. This could improve the court's ability to prevent violations of international criminal law.

If the survey shows that the concept of gravity is idiosyncratic, this too would have enormous implications for the ICC. It would suggest that the ICC will probably never be able to convince a majority of its audience that it is acting legitimately because most people will not be able to understand the decisions it makes about where to investigate, who to charge, and how to punish those found guilty. For a majority of its audience, these decisions will differ from how the individual understands the gravity of mass atrocities and may cause the individual to doubt whether the ICC is really investigating and prosecuting the gravest offenses. This will limit the ability of the ICC to be perceived as legitimate and, as a result, limit the ICC's ability to prevent violations of international criminal law.

III. CHOOSING A GRAVITY DEFINITION

The concept of gravity is extremely important at the ICC.³⁴ It features prominently in the Preamble to the Rome Statute,³⁵ and the term appears

³³ The reasons for testing the OTP's gravity definition are explained below in Section III.

³⁴ See Stuart Ford, *What Investigative Resources Does the International Criminal Court Need to Succeed?: A Gravity-Based Approach*, 16 WASH. U. GLOBAL STUDIES L. REV. 1, 6-7 (2017); see also deGuzman, *supra* note 18, at 1400 ("The concept of gravity or seriousness resides at the epicenter of the legal regime of the International Criminal Court.").

³⁵ For example, the Preamble references "unimaginable atrocities that shock the conscience of humanity" and says that "such grave crimes threaten the peace, security and well-being of the world." Rome Statute, preamble. 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

nine times in the body of the Statute.³⁶ It is crucial at every stage of the proceedings. The Office of the Prosecutor (OTP) is required to consider “the gravity of the crime” when deciding whether to open a formal investigation.³⁷ Once the Prosecutor has requested charges against an accused, the Court must determine whether the case is “of sufficient gravity to justify” being tried at the ICC.³⁸ And, once a person has been found guilty, the gravity of their crimes is a key factor in determining their sentence.³⁹

Yet, despite its importance, the Rome Statute never defines gravity.⁴⁰ In a general sense, gravity refers to the seriousness of the crimes,⁴¹ but the lack of a clear definition for such a key term creates a problem for the court. It must know what gravity means to be able to make decisions about when to investigate, who to charge, and what sentences to impose on those found guilty. As a result, the court must have a definition of the term.⁴² And indeed, various gravity definitions have been proposed, but there is no single gravity definition that everyone agrees on.⁴³

This article needs a gravity definition that it can test. Thus, it must either select a definition from amongst those that already exist or create one from scratch. Largely for three reasons, this article will use the definition of gravity adopted by the OTP. First, the OTP’s definition is representative of the various gravity definitions that have been proposed. A number of scholars have suggested definitions of gravity.⁴⁴ So have different organs

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.

³⁶ Rome Statute, 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).

³⁷ Rome Statute, Art. 53(1)(c); *see also* Art. 53(2)(c).

³⁸ *Id.*, Art. 17(1)(d).

³⁹ *Id.*, Art. 78(1); *see also*, Art. 77(1)(b) (noting that a sentence of life imprisonment must be justified by the “extreme gravity” of the convicted person’s crimes).

⁴⁰ Ford, *supra* note 34, at 7.

⁴¹ *Id.* at 6.

⁴² deGuzman, *supra* note 18, at 1402.

⁴³ Ford, *supra* note 34, at 5-14.

⁴⁴ *Id.* at 12-14.

within the court.⁴⁵ While there are some minor differences, most of the proposed definitions are quite similar to the definition that has been adopted by the OTP.⁴⁶ Second, the OTP has given more thought to how to define gravity than any other organ of the court.⁴⁷ In short, the OTP's definition is both thoughtful and representative of the way gravity is defined at the ICC.

Third, this article requires a detailed gravity definition and the OTP's definition is the most comprehensive and detailed definition that has been provided by any organ of the court. The OTP defines gravity using four factors: (1) the scale of the crimes; (2) their nature; (3) the manner of their commission; and (4) their impact.⁴⁸ The OTP has also explained in some detail what it considers each of these factors to mean.⁴⁹ This detailed explanation is very important because it allows the concept of gravity to be broken down into a number of smaller components that can be individually tested.⁵⁰ In contrast, the definitions used by the other organs of the court are much less detailed.⁵¹ Thus, the OTP's definition is more suitable for testing because it has a large number of components that can be separately tested to see which are associated with an increase in the gravity of the crimes.

The goal of this article is to understand how people understand the gravity of mass atrocities. To accomplish this, it requires a gravity definition. That definition must be both representative of the gravity definitions in use by the Court and detailed enough that it can be broken into components that can be individually tested. This article will test the OTP's gravity definition because it meets these criteria.

IV. METHODOLOGY

The OTP's gravity definition was tested using a survey. The components of the definition were identified and then questions were

⁴⁵ *Id.* at 7-9 (describing the definition adopted by the OTP); *id.* at 11-12 (describing the definition adopted by the Pre-Trial Chambers).

⁴⁶ *Id.* at 18.

⁴⁷ *Id.* at 11-12 (noting that the Pre-Trial Chambers have largely adopted the OTP's definition without discussion).

⁴⁸ See Office of the Prosecutor, Regulations of the Office of the Prosecutor, Doc. No. ICC-BD/05-01-09, dated April 23, 2009, at Reg. 29(2); see also Office of the Prosecutor, Report on Preliminary Examination Activities 2015, dated Nov. 12, 2015, at para. 7; Office of the Prosecutor, Policy Paper on Preliminary Examinations, dated Nov. 2013, at para. 9 (same).

⁴⁹ See Office of the Prosecutor, Policy Paper on Preliminary Examinations, dated Nov. 2013, at paras. 59-66 (describing how the OTP evaluates the gravity of crimes it is considering investigating).

⁵⁰ To put it another way, each component of the OTP's definition can be seen as a hypothesis that that component is a factor in determining the gravity of mass atrocities. This article can then test these hypotheses to see whether they are valid.

⁵¹ Ford, *supra* note 34, at 11-12.

designed to test the effect of variations in the components. Each question was designed to test a single gravity component. For each question on the survey, participants were shown between two and four scenarios.⁵² 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 claims that the seriousness of crimes increase as “the extent of the damage caused by the crimes, in particular the bodily . . . harm caused to the victims” increases.⁵³ This hypothesis was tested using three scenarios where the only change between scenarios was the extent of the physical harm to the victim caused by the crimes:

Scenario 1

During a peaceful political protest, government soldiers attack the protesters. They beat the protesters with clubs and rifles. 100 of the protesters suffer minor injuries such as cuts and bruises.

Scenario 2

During a peaceful political protest, government soldiers attack the protesters. They beat the protesters with clubs and rifles. 100 of the protesters suffer injuries which require hospitalization, including broken bones and concussions.

Scenario 3

During a peaceful political protest, government soldiers attack the protesters. They beat the protesters with clubs and rifles. 100 of the protesters are permanently disabled as a result of their injuries.⁵⁴

Participants were asked to assign a numerical score to each scenario that represented the seriousness of the events described in the scenario.⁵⁵ To give participants a baseline against which to score the various scenarios, they were told that “The baseline scenario is: ‘A group of soldiers deliberately shoots and kills 10 unarmed villagers.’ This has been assigned

⁵² The ordering of the questions in the survey and the order of the scenarios within the individual questions were randomized so as to avoid as much as possible subconsciously influencing the participants’ answers.

⁵³ Office of the Prosecutor, Policy Paper on Preliminary Examinations, dated Nov. 2013, at para. 62.

⁵⁴ See Stuart Ford, Meaning of Gravity Survey – Version 1.1, Question 8, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3021041.

⁵⁵ *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.”).

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.⁵⁶ In effect, the survey used a magnitude estimation approach similar to that used in the National Survey of Crime Severity.⁵⁷

The result is that variations in the scores given by the participants should be related to changes in the gravity component being studied.⁵⁸ For example, the scenarios listed above are identical except for the harm that the victims suffer. In Scenario 1, the victims suffer “minor injuries such as cuts and bruises.” In Scenario 2, the victims suffer “injuries which require hospitalization, including broken bones and concussions.” In Scenario 3, the victims “are permanently disabled as a result of their injuries.” Therefore, if there are changes in the scores that participants give to these scenarios, those changes ought to indicate how they feel that changing the harm suffered by the victims changes the seriousness of the crime.

While the survey did not ask people to rank the scenarios in order of their seriousness, the scores that participants gave can be used to infer a rank ordering. If a participant gives scenario 3 a score of 200, while assigning scenario 2 a score of 100, this implies that scenario 3 is more grave than scenario 2. Thus, the scores can be used to create a rank ordering of the scenarios within each question.

Generating a rank ordering allows this article to test the hypotheses that underlie the OTP’s 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 OTP’s 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 would indicate that they do not believe that gravity increases as the harm to the direct victims increases.

The relative importance of the various gravity components can then be inferred from the percentage of the participants who agree that each factor

⁵⁶ *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.”).

⁵⁷ See MARVIN E. WOLFGANG ET AL., THE NATIONAL SURVEY OF CRIME SEVERITY (June 1985); see also Robinson and Kurzban, *supra* note 17, at 1839-40 (describing magnitude estimation studies).

⁵⁸ See Robinson and Kurzban, *supra* note 17, at 1842-43.

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. This can be used to create a hierarchy of the components of crime gravity.⁵⁹

There is another way that the scores can be used to assess the different components of the gravity definition. This involves looking at the ratios of the scores given to different scenarios related to the same gravity factor. For example, imagine that the gravity of component A doubles as the number of incidences of that component double but the gravity of component B only increases by 20% as incidences of that component double. This implies that component A has a greater relative importance to overall gravity because a doubling of that component will result in a greater increase in overall gravity. This approach makes sense when attempting to compare two components that can vary in their incidence, like the number of direct or indirect victims, or the number of crime sites.

This approach can also be used to compare components that either exist or do not exist. So, for example, components like the presence of a discriminatory motive either exist or they do not exist. Such components can be compared to other similar components, like the existence of a plan, to assess their relative importance. For example, if the presence of component X increases gravity by 50% but the presence of component Y only increases gravity by 25%, this implies that component X is a stronger indicator of gravity. This approach cannot be used to compare components that are qualitative in nature, like the nature of the harm suffered.

Ultimately, this second approach is more limited than looking at the percentage of participants that agree with the OTPs' hypotheses for two reasons. First, it cannot be applied to each gravity component, and second it only produces results for those participants who agree with the hypothesis being tested.⁶⁰ On the other hand, it measures something different from the percentage of participants who agree with a hypothesis, which tells us how broadly the population agrees that a particular factor affects gravity. For those who agree that a factor affects gravity, the ratios act something like a measure of the magnitude of that factor's effect on gravity. The ratios of the scores will be discussed where appropriate.

⁵⁹ For the results of this process, *see infra* Table 1.

⁶⁰ If a participant scores two scenarios the same (implying they do not agree with the hypothesis being tested) then the ratio of the scores for those scenarios will always be 1. As a result, it only makes sense to look at the ratios of scores for those participants that agree with the hypothesis being tested. *See infra* text accompanying notes 79-80.

V. THE PARTICIPANTS

The survey was originally written with 19 questions, but it proved too difficult and a significant number of the non-expert⁶¹ participants during initial testing did not complete the survey. After that, the questions were broken up into three different groups and participants were assigned one group of questions to answer. Shortening the survey improved the completion rate. As a result of breaking the survey up into three parts, there were a total of 331 participants, but each individual question was answered by about 100 participants.

The non-expert participants were identified using SurveyMonkey's Audience feature.⁶² They were all from the United States and were not directly paid for their participation.⁶³ The pool from which the participants were selected is described as "a diverse population of millions of people across the United States,"⁶⁴ but it is not representative of the general population.⁶⁵

Before any information was collected, some explanatory information about the survey and its purpose was provided and prospective participants were asked if they wanted to participate.⁶⁶ 85% of the prospective participants opted to take the survey. Those who agreed to take the survey were asked some demographic questions.⁶⁷ The ages of the participants varied, with the largest group (29%) being between 36 and 50 years of age. The gender of the group was 66% female and 34% male. A plurality of the group (48%) had completed a college degree. Only 2% of the participants

⁶¹ The survey was also administered to a group of self-identified experts in international criminal law. This article focuses on the responses of the non-expert participants. The results of the administration of the survey to the expert group may be discussed in a later article. Unless specified otherwise, all references to participants in this article means the non-expert participants.

⁶² See Audience feature, available at <https://www.surveymonkey.com/mp/audience/>.

⁶³ See "Who's Taking Your Survey" available at https://help.surveymonkey.com/articles/en_US/kb/SurveyMonkey-Audience?bc=Buying_Responses#Panel ("Contribute members take surveys for charity and a chance to win a sweepstakes prize. We believe that by offering these non-cash incentives, we limit problems such as satisficing and encourage respondents to provide honest, thoughtful opinions."). The participants were, however, compensated in a sense, as SurveyMonkey commits to making a small donation to a charity on behalf of the participants.

⁶⁴ *Id.*

⁶⁵ See "Balancing and Regional Representation", available at https://help.surveymonkey.com/articles/en_US/kb/SurveyMonkey-Audience?bc=Buying_Responses#Panel. ("We automatically balance Contribute panels according to census data of age and gender, while location tends to balance out naturally."). That said, the pool is balanced for age, gender and geographic location.

⁶⁶ See Ford, *supra* note 54, at 1 (describing informed consent procedure).

⁶⁷ See Ford, *supra* note 54, at 2.

were lawyers, and the vast majority (70%) knew either “nothing” or “very little” about the ICC.

The participants varied from the general population in the United States. According to the Census Bureau, the US population is 51% female.⁶⁸ Thus, the sample in the survey contains a higher percentage of women than in the population at large.⁶⁹ The sample is also more educated than the general population. Only 30% of the US population has a bachelor’s degree or higher.⁷⁰ As a result of these differences, it may eventually be desirable to replicate the results of this survey in a sample that more accurately reflects the US population. On the other hand, there is some evidence that surveys administered online produce similar results to surveys administered in more traditional ways, despite differences in the sample composition.⁷¹

VI. SURVEY RESULTS

This section discusses the results of the survey. The survey consisted of 19 questions designed to test the components of the OTP’s gravity definition. 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.⁷² The OTP has elaborated on the meaning of each of these factors in its Policy Paper on Preliminary Examinations:

62. The scale of the crimes may be assessed in light of, *inter alia*, the number of direct and indirect victims, the extent of the damage caused by the crimes, in particular the bodily or psychological harm caused to the victims and their families, or their geographical or temporal spread (high intensity of the crimes over a brief period or low intensity of crimes over an extended period).

63. The nature of the crimes refers to the specific elements of each offence such as killings, rapes and other crimes involving sexual or gender violence and crimes committed against

⁶⁸ See United States Census Bureau, Quick Facts, available at <https://www.census.gov/quickfacts/fact/table/US/PST045217> (showing that as of July 2016, the US population was 50.8% female).

⁶⁹ Given that the pool of prospective participants was gender-balanced, one explanation for the larger number of women in the data could be that men were more likely than women to decline to participate in the survey. This is speculative, however, because no data was collected on the individuals who declined to participate in the survey.

⁷⁰ *Id.*

⁷¹ See, e.g., Michael Buhrmester et al., *Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?*, 6 PERSPECTIVES ON PSYCHOLOGICAL SCIENCE 3 (2011).

⁷² See *supra* text accompanying note 48.

children, persecution, or the imposition of conditions of life on a group calculated to bring about its destruction.

64. The manner of commission of the crimes may be assessed in light of, *inter alia*, the means employed to execute the crime, the degree of participation and intent of the perpetrator (if discernible at this stage), the extent to which the crimes were systematic or result from a plan or organized policy or otherwise resulted from the abuse of power or official capacity, and elements of particular cruelty, including the vulnerability of the victims, any motives involving discrimination, or the use of rape and sexual violence as a means of destroying groups.

65. The impact of crimes may be assessed in light of, *inter alia*, the sufferings endured by the victims and their increased vulnerability; the terror subsequently instilled, or the social, economic and environmental damage inflicted on the affected communities.⁷³

Essentially, each of these paragraphs contains a number of hypotheses about what contributes to the overall gravity of the crimes. These hypotheses were used as the basis for the questions the participants answered.

A. *Scale of the Crimes*

In paragraph 62 of its Policy Paper on Preliminary Examinations, the OTP argues that the scale of the crimes depends on the: (1) the number of direct victims; (2) the number of indirect victims; (3) the extent of the damage, particularly bodily or psychological harm; (4) the geographic spread of the crimes; and (5) the temporal spread of the crimes.⁷⁴ Each of these hypotheses formed the basis for one or more questions designed to test whether they affected people's perceptions of the gravity of mass atrocities. The questions will be grouped below according to the hypothesis they test.

1. Number of Direct Victims

The OTP hypothesizes that the gravity of the crime increases as the number of direct victims⁷⁵ of the crime increases. This hypothesis was tested using two questions (question 13 and question 19). In question 13, a

⁷³ See Policy Paper on Preliminary Examinations, *supra* note 49, at paras. 62-65.

⁷⁴ *Id.* at para. 62.

⁷⁵ See Valentina Spiga, *Indirect Victims' Participation in the Lubanga Trial*, 8 J. INT'L CRIM. JUST. 183, 186 (2010). For purposes of this article, a direct victim is someone who suffers harm as a direct result of the commission of a crime.

rebel group attacks an undefended village and deliberately kills either 100, 200, or 400 unarmed villagers.⁷⁶ In question 19, government soldiers attack peaceful protesters and beat either 100, 150, or 250 of them with clubs and rifles inflicting serious injuries.⁷⁷ 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.

First, the individual scores participants gave to the different scenarios were used to assign a rank order to those scenarios.⁷⁸ In response to both questions, the vast majority of participants fell into one of two categories. Either they thought that the gravity of the crime increased as the number of victims increased or they thought that all of the crimes were equally serious irrespective of the number of victims. Interestingly, more people thought that the number of victims did not affect gravity than thought it did. In response to question 13, 53% of the participants scored all of the scenarios as equally grave, compared to 42% who said the scenario involving the largest number of victims was the most grave. In response to question 19, 57% of the participants felt that all three scenarios were equally grave, while 34% thought that the scenario involving the largest number of victims was most grave. Thus, it appears that for a slim majority of respondents, gravity does not depend on the number of direct victims. On the other hand, for a sizable minority, gravity increases as the number of victims increases.

The use of a magnitude estimation approach also allows for the study of the relative seriousness of the different scenarios.⁷⁹ Of course, the discussion of relative seriousness only applies to those participants who thought that the scenarios were of different gravity, so it excludes those participants who said that the number of victims did not affect seriousness. For those who thought the number of victims mattered, that translated into a doubling of the seriousness of the crimes as the number of victims doubled. For question 13, the median ratio of the gravity of 400 deaths to 200 deaths was 2.⁸⁰ Similarly, the median ratio of the gravity of 200 deaths to 100 deaths was 2. For question 19, the median ratio of the gravity of 150

⁷⁶ See Ford, *supra* note 54, at 16.

⁷⁷ *Id.* at 22.

⁷⁸ See *supra* Section IV.

⁷⁹ See *supra* Section IV.

⁸⁰ If one looks at the ratios of the scores given to the two scenarios (400 deaths vs 200 deaths), they span a range of ratios. On the low end, one participant ranked the scenario with 400 deaths as 1.125 times as serious as the scenarios with 200 deaths. On the high end, one participant ranked the scenario with more deaths as 10 times as serious as the scenarios with fewer deaths. But those results are both outliers. The median ratio (the ratio given by the participants in the middle of the distribution) was 2. The median is reported precisely to avoid the problem of outliers skewing the mean of the distribution. See JEFFERY T. WALKER & SEAN MADDAN, STATISTICS IN CRIMINOLOGY AND CRIMINAL JUSTICE (4th ed. 2013) at 100.

beatings to 100 beatings was 1.5. The median ratio of the gravity of 250 beatings to 100 beatings was 2. Essentially, it appears that for those participants who thought that gravity increased as the number of victims increased, there was a rough doubling of gravity for every doubling of the number of victims.

2. Number of Indirect Victims

The next hypothesis that was tested was whether the gravity of the crime increases as the number of indirect victims⁸¹ increases. This hypothesis was tested using a single question (question 4). In 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 200.⁸² 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.

When looking at the rank ordering of the crimes, the vast majority of participants (77%) thought that all three crimes were equally serious. Only 15% ranked the scenario with the largest number of indirect victims as the most serious scenario. This suggests that most people do not think that increasing the number of indirect victims of a crime significantly increases the gravity of the crime. This result seems meaningfully different from the results of the questions about direct victims, where nearly half of the participants felt that gravity increased as the number of victims increased.⁸³ This suggests that the number of indirect victims is a weaker gravity factor than the number of direct victims.

While it is true that some participants thought that gravity was affected by the number of indirect victims, even for those participants, the effect was not very large. If we look at the ratio of the score for the scenario with 200 orphans and compare it to the score for the scenarios with 50 orphans, the median ratio was only 1.5. Thus, even for those participants who thought that gravity increased as the number of indirect victims increased, the median participant thought that a quadrupling of the number of indirect victims resulted in only a 50% increase in the gravity of the crime. This is

⁸¹ See Spiga, *supra* note 75, at 186. For purposes of this article, an indirect victim is someone who suffers harm as a result of the harm suffered by a direct victim.

⁸² See Ford, *supra* note 54, at 7.

⁸³ See *supra* Section VI(A)(1).

significantly less than in the questions about direct victims.⁸⁴ This finding is consistent with the number of indirect victims being a less important component of overall crime gravity than the number of direct victims.

3. 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.⁸⁵ While the OTP’s definition 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 on the survey. 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.⁸⁶ 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.” In scenario 2, the victims suffer “injuries which require hospitalization, including broken bones and concussions.” In scenario 3, the victims “are permanently disabled as a result of their injuries.” 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.

When looking at a rank ordering of the crimes, the vast majority of participants (76%) said that the scenario resulting in permanent disability was the most serious one. A much smaller group (8%) said that the two scenarios involving permanent injuries and injuries which required hospitalization were equally serious and both more serious than the scenario involving minor injuries. Virtually nobody (less than 1%) said that the scenario involving minor injuries was the most serious. A further 15% of the participants said that all of the scenarios were equally serious. This suggests that most people agree that a crime increases in severity as the harm suffered by the victim increases.

4. Geographic Spread of the Crimes

This factor focuses on the geographic spread of the crimes,⁸⁷ which is

⁸⁴ For those participants who thought the number of direct victims affected the seriousness of the crime, a quadrupling of the number of direct victims would be expected to result in a quadrupling of the gravity of the offense. *See supra* Section VI(A)(1).

⁸⁵ *See* Policy Paper on Preliminary Examinations, *supra* note 49, at para. 62.

⁸⁶ *See* Ford, *supra* note 54, at 11.

⁸⁷ *See* Policy Paper on Preliminary Examinations, *supra* note 49, at para. 62.

defined in this article as the number of crime sites at which the crimes occur.⁸⁸ The OTP seems to hypothesize that the gravity of the crime will increase as the number of crime sites at which the crime occurs increases. 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.⁸⁹ 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. 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.

Contrary to the OTP's hypothesis, the vast majority of participants (80%) scored the two scenarios as equally grave. Only 18% of participants thought that the offense was more grave if the crimes were carried out at more locations. 2% of the participants said that the scenario where crimes occurred at fewer locations was more serious. These results suggest that most people do not believe that the gravity of an offense increases as the number of crime sites increases.

This conclusion is supported by the finding that, for those participants who said that gravity increased as the number of crime sites increased, the median ratio of the scores between the scenarios was only 1.33. In other words, even for those people who thought the offense was more serious if carried out at more locations, a tripling of the number of crime sites only increased the overall gravity of the offense by 33%. This is significantly smaller than the effect of doubling the number of direct victims,⁹⁰ which suggests that even for people who believe this factor matters it does not have a strong impact on overall gravity.

5. Temporal Spread 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.⁹¹ The OTP's reference to the "high intensity of the crimes over a brief period"⁹² suggests that the OTP believes that gravity increases as the period over which the crimes occurs decreases. This hypothesis was tested with a single question (question 11), which presents two scenarios where government

⁸⁸ See Stuart Ford, *The Complexity of International Criminal Trials Is Necessary*, 48 GEO. 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).

⁸⁹ See Ford, *supra* note 54, at 8.

⁹⁰ See *supra* Section VI(A)(1).

⁹¹ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 62.

⁹² *Id.*

soldiers set up a roadblock and execute rebel sympathizers.⁹³ 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 6 months. In scenario 2, the crimes occur over a period of 2 months. 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.

Similar to the question on geographic spread, the vast majority of participants (81%) felt that both scenarios were equally grave. The next largest group (at 15%) felt that the crimes that occurred over a shorter period were more grave. Only 4% of the participants thought that the crimes that occurred over a longer period were more serious. This provides some support for the idea that a minority of people believe that crimes become more serious as their intensity (i.e., the rate at which they are committed) increases. Nevertheless, the vast majority of participants did not believe that the temporal spread of the crimes affected their gravity.

For those who felt that gravity increased as intensity increased, the median participant scored scenario 2 as 39% more serious than scenario 1. Or, to put it another way, a threefold increase in intensity corresponded to about a 40% increase in gravity. This is significantly smaller than the expected effect of tripling the number of direct victims.⁹⁴ This finding is consistent with the results of the rank ordering and suggests that temporal spread is not a significant component of overall crime gravity.

B. Nature of the crimes

In paragraph 63 of its Policy Paper on Preliminary Examinations, the OTP expands on the meaning of "nature of the crimes." According to the OTP, the nature of the crimes depends on factors like: (1) the specific offense committed, "such as killings, rapes and other crimes"; (2) the identity of the victims; and (3) whether the crimes result from a discriminatory intent.⁹⁵ The questions related to those factors are discussed below.

1. The Type of Crimes

The OTP's description of the nature of the crimes implies that some crimes are worse than others and that the type of the underlying crime

⁹³ See Ford, *supra* note 54, at 14.

⁹⁴ See *supra* Section VI(A)(1). One would expect a tripling of the number of direct victims to result in a tripling of the gravity of the offense.

⁹⁵ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 63

affects its gravity.⁹⁶ This proposition was tested using two questions that asked participants to rank the severity of different scenarios that varied only by the type of crime committed. Question 2 describes how rebel sympathizers are arrested by the government and placed in detention, where they are abused by the guards.⁹⁷ 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 the more serious crimes will be scored higher than the less serious crimes.

A plurality of participants (40%) scored the scenario that involved murder as the most grave. The next largest group (22%) said that the scenarios involving rape and murder were equally serious, but that both were more serious than assault. 15% of the participants said that rape was the most serious offense. Almost nobody (2%) thought that assault was the most serious crime. Thus, a majority (79%) of participants believed that the type of crime committed affected the severity of the offense. In contrast, 20% of the participants said that all of the crimes were equally serious.

While the vast majority of participants agree that there is a ranking of the gravity of the crimes, they are split between three different approaches to ranking them: 1) murder is the most serious; 2) rape is the most serious; and 3) both are equally serious. Ultimately, the largest group of participants (40%) believed that murder was more grave than rape. The two other groups collectively, however, contain almost as many participants (37%).

There were also some gender differences in how participants answered this question. Among participants who indicated there was a hierarchy among the crimes, men were more likely than women (57% to 46%) to say that murder was the most serious. Women, in contrast, were more likely than men (32% to 20%) to say that rape and murder were equally serious. Men and women ranked rape as the most serious offense at essentially the same rate (20% vs 19%). Unfortunately, it is not possible to tell if these differences are meaningful.⁹⁸ If there are significant gender differences on

⁹⁶ *Id.* (stating that gravity depends on “the specific elements of each offense such as killings, rapes . . .” or other offenses).

⁹⁷ See Ford, *supra* note 54, at 5.

⁹⁸ A Pearson's chi-squared test was run on the results, but the differences did not achieve statistical significance. See WALKER & MADDAN, *supra* note 80, at 193-201 (describing the chi-squared test). This could be a result of the small sample size. It is possible that there is a significant difference between men and women and that a larger sample size would detect those differences. It is also possible that the differences in the male and female responses on this survey are the result of random chance and there is no meaningful difference between men

this issue, that result would be interesting as domestic surveys of crime severity have not shown large gender differences.⁹⁹

Even if there are gender differences in how people view rape and murder, comparisons of the ratios of the scores for those crimes suggest that the male and female participants are not that far apart in how they view them. While the median female participant ranked the rape and murder crimes as equally serious, the median male participant ranked murder as 20% more serious than rape. In contrast, the participants as a whole ranked rape and murder as almost twice as serious as assault. Ultimately, these figures suggest that while men may be slightly more likely than women to view murder as more serious than rape, both genders view rape and murder as quite similar in seriousness and significantly more serious than assault.

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.¹⁰⁰ 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.¹⁰¹ In the other scenarios, the villagers are either beaten, raped or killed.

The results for question 10 are similar to the results for question 2 above. A plurality of participants (44%) thought that murder was the most serious offense. Rape was seen as the most serious by 16% and rape and murder were viewed as equally serious by another 15%. Another 15% of the participants thought that all of the scenarios were equally serious. Almost nobody thought that assault or forced displacement were the most serious offenses (2% cumulatively). As with question 2, there were also some disparities by gender. Men were about 10% more likely to view murder as the most serious offense than women (58% vs. 49%).¹⁰² These results are consistent with the results of question 2, described above.

Collectively, the responses to these questions suggest that there is a discernable hierarchy in the seriousness of different crimes. Murder and rape are viewed by a clear majority of the participants (77% in question 2

and women on this issue. *Id.*

⁹⁹ See Robinson and Kurzban, *supra* note 17, at 1857-59 (describing domestic studies of crime seriousness and noting that none found a strong difference based on the gender of the participants).

¹⁰⁰ See Ford, *supra* note 54, at 13.

¹⁰¹ This serves as a proxy for the crime of forcible displacement of a civilian population. See Rome Statute, art. 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”).

¹⁰² See *supra* note 98. As with question 2, the differences were not statistically significant.

and 75% in question 10) as more serious than assault or forcible displacement. Very few participants (15% in question 2 and 15% in question 10) thought that there was no difference in the seriousness of the different underlying crimes. The participants are divided, however, about whether to view murder as the most serious offense, rape as the most serious offense or to view both as equally serious. Having said that, a plurality (40% in question 2 and 44% in question 10) believe that murder is more serious than rape.

2. The Identity of the Victims

The OTP claims 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.¹⁰³ This hypothesis was tested in question 3. In that question, rebel soldiers attack a village and kill 100 unarmed villagers.¹⁰⁴ The question presents four different variations on that scenario. The only difference between the scenarios is the identity of the victims. In one scenario, “most of the victims are women,” in another “most of the victims are men,” in the third scenario “most of the victims are elderly,” and in the final scenario “most of the victims are children.” This question tests whether the identity of the victims changes the gravity of the crime.

By far the largest group of participants (64%) felt that all four scenarios were equally serious. The next largest group (26%) felt that the scenario involving the children was the most serious. Collectively, all of the other possibilities account for the remaining 10%. The results are quite clear, most participants do not think that the identity of the victims matters in determining the gravity of the crime. A small but significant minority (26%) believes that killing children is worse than killing men, women or the elderly. For this group, killing children is significantly worse than killing other groups.¹⁰⁵ There is no significant group of participants who believe that crimes that affect women or the elderly are inherently more serious than crimes that affect men.

3. Discriminatory Intent

The OTP hypothesizes that crimes carried out with a discriminatory

¹⁰³ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 63. The OTP’s definition focuses on “crimes involving sexual or gender violence and crimes committed against children.”

¹⁰⁴ See Ford, *supra* note 54, at 6.

¹⁰⁵ The median participant in this group thought that killing children was 83% more serious than just killing men.

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.¹⁰⁶ These are references to crimes that are distinctive because they involve a discriminatory intent – genocide and persecution.¹⁰⁷ 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.¹⁰⁸ 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." 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.

Here, the participants were split into two groups. A majority (56%) thought that the two crimes were equally serious, while most of the remainder (41%) felt that the crime involving religious discrimination was more serious. Almost nobody (3%) thought the scenario that lacked discriminatory intent was the most serious. Among those participants that thought the religious violence was more grave, the median participant in that group indicated it was 75% more serious than the arbitrary violence.

A second question addressed ethnically-motivated violence. In question 18, government soldiers set up a roadblock and stop everybody that passes it.¹⁰⁹ 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."

¹⁰⁶ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 63.

¹⁰⁷ Persecution requires a discriminatory intent. See Rome Statute, 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").

¹⁰⁸ See Ford, *supra* note 54, at 4.

¹⁰⁹ See Ford, *supra* note 54, at 21.

The largest group of participants (66%) felt that the two crimes were equally serious. Only 29% of the participants thought that the ethnically-motivated murders were more serious. Similarly to question 3 above, very few people (5%) thought the scenario that lacked discriminatory intent was the most serious. For those participants that believed the ethnically-motivated murders were worse, the median participants assigned said it was 50% more serious.

Taken together, the results of these two questions demonstrate that a majority of people do not see crimes carried out because of religious or ethnic animus to be worse than that same crime carried out for an essentially arbitrary reason. For a sizable minority (41% in question 1 and 29% in question 18), crimes carried out with a discriminatory motive are more serious. For that minority, the presence of a discriminatory intent makes the crime significantly more serious. There is also a suggestion that the participants perceive religious discrimination to be more serious than discrimination based on ethnicity.¹¹⁰

C. *Manner of Commission*

According to the OTP, the manner of commission focuses on factors like: 1) the means employed to carry out the crime; 2) the degree to which the crimes are systematic or the result of an organized plan or policy; 3) whether the crimes resulted from the abuse of power or official capacity; and 4) elements of particular cruelty.¹¹¹ The OTP also suggests that the presence of discrimination is a factor in the manner of commission, but this factor was considered above when discussing the nature of the crimes and will not be discussed again here.¹¹² The perpetrator's degree of participation in the crime was not tested in the survey.¹¹³

1. The Means Employed

The OTP posits that the means employed to carry out a crime can affect its gravity. This hypothesis was tested with two questions. In question 12, the government forces all of the inhabitants of a village to flee.¹¹⁴ The only

¹¹⁰ The percentage of people who viewed the religious discrimination as more grave is greater than the percentage that viewed the ethnic discrimination as more grave (41% to 29%). In addition, those who thought that religious discrimination was a gravity factor thought that it made the offense 75% more serious, while those who thought that ethnic discrimination was a factor thought that it made the offense 50% more serious.

¹¹¹ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 64.

¹¹² See *supra* Section VI(B)(3).

¹¹³ Any subsequent testing of this concept should probably use questions based on the modes of liability outlined in the Rome Statute. See Rome Statute, art. 25.

¹¹⁴ See Ford, *supra* note 54, at 15.

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.

The largest group of participants (56%) thought that all of the scenarios had the same gravity. The rest of the group thought there was some sort of hierarchy, but there was widespread disagreement about what that hierarchy ought to be. 17% of the participants thought the scenario involving the soldiers threatening the villagers was most serious, while 14% thought the one involving cutting off power and water was the most serious. Another 5% thought the scenario involving the radio broadcasts was the most serious. The remainder of the participants (about 9%) were closely split among people who picked either scenarios 1 and 2, 2, and 3, or 1 and 3 as the most serious.

The results suggest that most people do not believe that the means of commission affected the gravity of the offense. Slightly less than half of the participants thought there was a hierarchy among the means of carrying out the crimes, but there was essentially no agreement amongst this group on what that hierarchy ought to be. People were split fairly evenly among the different scenarios. In short, there is little support for the OTP's hypothesis that the means of commission alters the gravity of the offense.

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.¹¹⁵ The only difference among the three scenarios is how the killings are 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 look very similar to the results of question 12 above. A majority (61%) viewed all of the crimes as equally serious. The remainder were split between the remaining options. 19% said that beating people to death was most serious. Another 10% said that beating people to death was equivalent to shooting people but that both were more serious than killing with bombs. Finally, 5% said that the shootings were the most serious while 3% said that the bombings were the most serious.

As with question 12, the clear majority position is that all of the scenarios are equally serious. There is no obvious second choice. Instead, while slightly less than 40% of the participants thought there was some sort

¹¹⁵ See Ford, *supra* note 54, at 18.

of hierarchy among the scenarios, there was no consensus about that hierarchy. Rather, the participants were pretty evenly split between the various possibilities. Ultimately, most participants do not believe that the means of commission affects the gravity of the offense.¹¹⁶

2. Plan or Policy

The OTP's gravity definition proposes that crimes are more serious if they are "systematic or result from a plan or organized policy."¹¹⁷ 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. 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." 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.

A plurality of the participants (42%) felt that all three crimes were of equal seriousness. For the rest, 19% said the premeditated killings were the worst, while another 19% said that the premeditated killings and the deliberate but spontaneous killings were joint worst. 12% of the participants said that the deliberate but spontaneous killings were the most serious. In contrast, only 2% said that the mistaken killings were the most serious.

A majority of the participants agree that the deliberate killings (both the spontaneous ones and the premeditated ones) were worse than the mistaken killings.¹¹⁸ Almost nobody thought that the mistaken killings were the worst. But within the group that thought the deliberate killings were the most serious, they were pretty evenly split between those that thought the premeditated killings were worse, those that thought the spontaneous

¹¹⁶ *But see infra* Section VI(C)(4) (showing that a majority of participants agreed that exceptional cruelty in the commission of a crime did increase its gravity).

¹¹⁷ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 64.

¹¹⁸ If you combine the participants who said that either the spontaneous or the premeditated killings were most serious with the participants who said that the spontaneous and premeditated killings were joint most serious, you end up with slightly more than 50% of the participants.

killings were worst, and those that thought both categories of deliberate killings were equally serious. In other words, there is support for the conclusion that deliberate killings are generally viewed as more serious than ones that result from mistakes. But there is little support for the OTP's assertion that planned or premeditated killings are more serious than ones that are deliberate but spontaneous. It should also be remembered that a sizable minority of participants thought that all of the scenarios were equally serious.

A similar issue was tested in question 7. In that question, rebels attack a village and kill 100 unarmed villagers.¹¹⁹ 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,¹²⁰ deliberate but not part of a larger policy,¹²¹ or deliberate and part of an organized policy.¹²²

The vast majority of the participants (80%) felt that all three scenarios were equally serious. The remaining 20% of the participants were fairly evenly split between thinking that scenario 1 was most serious (5%), scenario 2 was most serious (5%), and scenario 3 was most serious (6%). Most people did not feel that the existence of an organized policy that led to the violence made the resulting crime more serious. There is no obvious hierarchy amongst the remainder of the participants.

There is an apparent inconsistency between the results of question 14 and question 7. In question 14 only 42% thought that all the crimes were equally serious, while 80% of the participants who answered question 7 thought that all the crimes were equally serious. This is a large difference. While it is impossible to know with certainty why the two questions generate different answers, one possibility is that the inclusion of a scenario based on mistake in question 14 changed the way people answered it. A majority of the participants who answered question 14 thought the scenario

¹¹⁹ See Ford, *supra* note 54, at 10.

¹²⁰ 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.

¹²¹ 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.

¹²² 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.

based on a mistake was different from and less serious than the two scenarios based on a deliberate intent. It is possible that the inclusion of a scenario based on mistake caused more people to perceive a hierarchy in the crimes based on the level of culpability of the perpetrators.¹²³ Question 7, on the other hand, involves culpable conduct in every scenario. While in one of the scenarios, the rebel commander acts only negligently or recklessly, it is still a higher level of culpability than was present in the mistake scenario in question 14. This may explain the difference in the answers.¹²⁴

Even if the results of question 7 and question 14 look a little different, however, they both firmly reject the OTP's claim that the existence of a plan or organized policy increases the gravity of the resulting offense. In both questions, a crime that results from the implementation of a plan or policy was not viewed as significantly more serious than a deliberate but spontaneous crime.

3. 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.¹²⁵ This hypothesis was tested in question 9. In that question, government soldiers shoot and kill 100 unarmed villagers.¹²⁶ 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." 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 majority of the participants (63%) thought that both crimes were equally serious, while 34% thought that the scenario where the government

¹²³ Cf. Robinson and Kurzban, *supra* note 17, 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).

¹²⁴ If true, this would suggest that people do, at a high level, make a distinction between accidental and culpable conduct, even if they do not make a distinction between reckless, deliberate and premeditated conduct. It may make sense to explore this issue in more detail in future research.

¹²⁵ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 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").

¹²⁶ See Ford, *supra* note 54, at 12.

tricks the villagers into attending the meeting was the most serious. In effect, the majority of participants did not see the abuse of the government's position as changing the gravity of the offense. There was, however, a sizable minority that felt that abuse of power did increase the severity of the crime. For this group, the median score for the scenario involving abuse of power was 50% more severe than the one that did not involve it. So, for those that did think abuse of power mattered, it made a significant impact on the gravity of the crime.

4. 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.¹²⁷ In question 6, government soldiers kill 100 unarmed villagers.¹²⁸ 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." All of the villagers die in the fire. If particular cruelty matters, one would expect the scenario involving the fire would be scored as more serious.¹²⁹

While very few people (5%) thought that the shootings were the most serious, the bulk of the participants were almost evenly split (49% to 46%) between scoring the two scenarios as equally serious and scoring the scenario involving fire as more serious. For those that ranked the fire as more serious, the median score was 50% higher than for the shootings.

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.¹³⁰ 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." If particular cruelty is a gravity factor, one would expect the

¹²⁷ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 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).

¹²⁸ See Ford, *supra* note 54, at 9.

¹²⁹ Cf. *Prosecutor v. Milan Lukić & Sredoje Lukić*, Trial Judgment, Case No. IT-98-32/1-T, dated July 20, 2009, at para. 740 (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.").

¹³⁰ See Ford, *supra* note 54, at 19.

scenario involving the amputations to be scored as more serious.¹³¹

Similar to the results of question 6, the participants were fairly evenly split. 55% of the participants said that killing people by chopping of their hands and feet was more serious than simply shooting them. 44% of the participants said the two crimes were equally serious. Almost nobody (1%) said that the shootings were the most serious. Interestingly, for those people who said that the forcible amputations were more serious than the shootings, the median score was 90% higher. This is a higher ratio than in question 6 above.

Ultimately, it seems that people are pretty evenly split between those that think that killing someone with particular cruelty increases the gravity of the offense and those that do not. Within the group of people who believe that cruelty matters, there is some suggestion that there is a hierarchy of cruel means, with death by amputation more serious than death by burning, which is in turn more serious than shooting someone to death.¹³²

D. *The Impact of the Crimes*

The OTP hypothesizes that the impact of the crimes is dependent on factors like: (1) the suffering endured by the victims; (2) their increased vulnerability; (3) the terror subsequently instilled; and (4) the social, economic or environmental damage inflicted on affected communities.¹³³ Some of these factors have arguably already been tested in other questions.

1. Suffering of the Victims

The questions about particular cruelty can be seen as questions about the suffering of the victims.¹³⁴ About half the participants felt that it was worse to be burned to death or killed by amputation of one's limbs than by being shot to death. One plausible explanation for these results is that people perceive there to be greater suffering associated with death by burning and amputation and that is the reason they find those crimes to be more serious.¹³⁵ If so, then the questions about particular cruelty are, in

¹³¹ Cf. *Prosecutor v. Issa Hassan Sesay, Morris Kallon, & Augustine Gbao*, Sentencing Judgment, Case No. SCSL-04-15-T, dated April 8, 2009, at para. 109 (describing death by amputation as particularly cruel and brutal).

¹³² This is implied from the fact that people thought that death by amputation was 90% more serious than death by shooting, whereas participants thought that death by burning was 50% more serious than death by shooting. These numbers come from different questions which were answered by different participants, so the conclusion is somewhat tentative.

¹³³ See Policy Paper on Preliminary Examinations, *supra* note 49, at para. 65.

¹³⁴ See *supra* Section VI(C)(4).

¹³⁵ Cf. *Prosecutor v. Milan Lukić & Sredoje Lukić*, Trial Judgment, Case No. IT-98-32/1-T, dated July 20, 2009, at para. 740 (arguing that being burned to death is particularly cruel

effect, questions about whether the suffering endured by the victims is a factor in the gravity of the crimes. The results of those questions suggest that the suffering endured by the victims is a relevant component of gravity for about half the participants.

This could also explain why the questions about manner of commission did not find a hierarchy amongst the means employed to carry out the crimes.¹³⁶ In those questions, it was not obvious that any of the means employed imposed a greater degree of suffering on the victims.¹³⁷ If the factor that really matters is the suffering caused by the particular means employed, then it is not surprising that the means questions did not result in any discernable hierarchy.

2. The Increased Vulnerability of the Victims

The question about the identity of the victims can be seen as related to the increased vulnerability of the victims. One might expect that women, the elderly, and children are vulnerable victims and their vulnerability might impact the gravity of offenses directed at them. But the results of question 3 suggest that only when the victims are children does their vulnerability affect the gravity of the offense, and even then only for a minority of the participants.¹³⁸

The effect of the vulnerability of the victims was also 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 detention status 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,¹³⁹ while it is generally lawful to kill enemy soldiers during combat.¹⁴⁰ 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.

because of the “degree of pain and suffering inflicted on the victims as they are burnt alive”).

¹³⁶ See *supra* Section VI(C)(1).

¹³⁷ *Id.*

¹³⁸ See *supra* Section VI(B)(2).

¹³⁹ See JEAN-MARIE HENCKAERTS & LOUISE DOWSALD-BECK, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW VOLUME 1: RULES (2005 Cambridge University Press) at Rule 47 (noting that attacks on persons who are in the power of an adverse party are prohibited).

¹⁴⁰ *Id.* at Rule 1 (“Attacks may only be directed against combatants.”).

In reality, the participants were pretty equally divided. 51% of the participants thought that killing the captured soldiers was more serious, while 46% thought that both scenarios were equally serious. Only 3% of the participants thought that killing the soldiers during combat was worse than executing them afterwards. In essence, a bare majority of participants agrees that killing soldiers is worse when they have the status of detainees. Virtually everyone else thought there was no difference in crime gravity when the victims were detainees.

3. The Terror Subsequently Instilled

The effect of the terror subsequently instilled in the victims was not tested. It would be possible to devise questions to test this hypothesis, but there were limits on how many questions participants could realistically be expected to answer.¹⁴¹ For that reason, this issue was not tested on the survey.

4. The Social, Economic or Environmental Damage

The question about indirect victims tests a concept similar to the “social, economic, or environmental damage inflicted in affected communities.”¹⁴² The responses to question 4 suggest that the majority of participants did not believe that the indirect effects of a crime, like the effect on children who are orphaned by the death of their parents, was a significant factor in crime gravity. Of course, orphans are just one indirect effect of crime. It would be possible to devise questions that tested other aspects of the social, economic or environmental consequences of crimes. They were not, however, tested in this survey. Nevertheless, the responses to the question about indirect victims suggest that the consequential impacts of crimes will be less important to most people than the direct effects.

VII. DISCUSSION

The survey described in this article tested many of the components of the OTP’s gravity definition. The results suggest some of those components are strong indicators of crime gravity. At the same time, there are some components that are only moderate indicators of crime gravity and some that are weak indicators. This section will discuss those components and their effect on the gravity of crimes.

But first, it is necessary to define what this article means by strong,

¹⁴¹ See *supra* text accompanying note 61 (noting that the survey as originally constructed had 19 questions but that this proved too difficult for the participants).

¹⁴² See *supra* Section VI(A)(2).

moderate and weak indicators. A component will be designated a strong indicator of crime gravity if a majority of the participants in this survey associated it with an increase in the seriousness of the crime. A component will be designated as a moderate indicator if between 25% and 50% of the participants associate it with an increase in crime gravity. A component will be designated a weak indicator if less than 25% of participants associate it with increased crime gravity. The cut-offs between the categories are, to some extent arbitrary, but they help illustrate the level of agreement about these indicators among the survey participants.

A. Strong Indicators of Crime Gravity

One of the strongest indicators of crime gravity is the extent of the physical harm suffered by the direct victims. A large majority of participants (76%) felt that the crimes became more serious as the harm suffered by the victim increased. Only 15% of participants said that the extent of the harm did not affect the gravity of the crime.

The questions related to the type of crime demonstrate that this, too, is a very strong indicator of crime gravity. In the two questions related to this issue, the majority of participants (79% and 75% respectively) believed that rape and murder were significantly more grave than either assault or forced displacement of a civilian population. A minority (20% and 15% respectively) felt that all of the offenses were equally grave.

Within the majority that ranked rape and murder as the most serious offenses, however, there were splits between those who believed murder was the most serious, those who believed rape was the most serious, and those who believed that both were equally serious. Those who believed murder was the most serious were the largest group (at 40% and 44% respectively) but there were sizable minorities who believed that rape was the most serious and that rape and murder were equally serious. Even for those who said that murder was most serious however, it was not ranked as significantly more serious than rape. The best way to view this data is that there is a clear hierarchy between rape and murder on the one hand and assault and forcible displacement on the other, but that views are more mixed about the relative gravity of rape and murder.

Particular cruelty also turned out to be a strong indicator of gravity, although not as strong as the extent of the harm or the type of crime. In two questions that related to particular cruelty, a slight majority of participants (55% and 49% respectively¹⁴³) agreed that a crime that was carried out in a particularly cruel manner was worse than the same crime carried out in a less

¹⁴³ Of course, 49% is not a majority, but on balance the results across both questions suggest that a slight majority of participants agree that particular cruelty does matter.

cruel manner. The two questions on this issue compared death by fire and death by amputation to death by gunshot wound. Participants concluded that causing death by fire and death by amputation were both significantly graver than causing death by gunshot. For both questions, however, a sizable minority (44% and 46%) said that particular cruelty did not affect the gravity of the offense.

B. Moderate Indicators of Crime Gravity

A number of components of the OTP's definition turned out to be moderate indicators of crime gravity, with between 25% and 50% of the participants saying that they affected the gravity of the offense. The number of direct victims of the crime was a moderate indicator. The effect of varying the number of direct victims was tested using two questions. In both questions, a majority of participants (53% and 57% respectively) indicated that the number of victims did not affect the gravity of the offense. A sizable minority (43% and 34% respectively) believed that the gravity of the offense did increase as the number of direct victims increased. For those who believed that the number of victims matters, the gravity of the offense appears to double as the number of victims doubles.

The presence of a discriminatory intent was also a moderate indicator of gravity. This was tested with two questions, one related to religious discrimination and one related to ethnic discrimination. In both questions a majority of participants (56% and 66%) said that a discriminatory motive did not change the gravity. Sizable minorities (41% and 29%) indicated that it did.

The abuse of power or official capacity was a moderate indicator of crime gravity. A majority of participants (63%) said that the government's abuse of its power or official capacity did not affect the gravity of the offense. On the other hand, 34% of the participants felt that abusing power or official capacity did increase the gravity of the offense.

Finally, the weakest of the moderate indicators was the identity of the victim. Four different possible victim identities were tested: men, women, children and the elderly. The majority of participants (64%) did not think that the identity of the victims mattered to the gravity of the offense. A sizable minority of participants (26%) believed that the crime was worse when the victims were children. Very few people believed that the crimes were worse because the victims were men, women or the elderly.

C. Weak Indicators of Crime Gravity

There are a number of components of the OTP's definition that were weak indicators of gravity. For these components, less than a quarter of the participants indicated that they affected the seriousness of the resulting

offense. For example, the number of indirect victims is a weak indicator of gravity. Only 15% of the participants ranked the scenario with more indirect victims as the more serious, while 77% of the participants ranked the two scenarios as equally grave.

Both the geographic and the temporal spread of the crimes were also weak indicators. 80% of the participants said that crime gravity did not increase as the number of crime sites increased, while 81% believed that crime gravity did not change as the length of time over which the crimes were committed changed. Some participants did think these mattered (18% and 15% respectively), but they were a distinct minority.

The existence of a plan or policy that leads to the commission of a crime was also a weak indicator of gravity. In the question that tested the importance of a plan, only 19% of the participants said that killings that resulted from a plan were more serious than killings that resulted from a spontaneous decision to commit the crime. In contrast, 61% indicated that the spontaneous killings and the premeditated killings were equally serious, while 12% said that the spontaneous killings were the worst. Having a plan to commit an offense does not make it more serious for most participants.

The results of the question testing the salience of a policy were similar. The vast majority (80%) felt that the existence of a policy of committing the violations did not affect their seriousness. Only 6% of the participants felt that the killings that resulted from a specific policy were the most grave.

The means employed to carry out the crimes was also a weak indicator of crime gravity. This was tested using two questions and the results are quite similar. In response to both questions, a majority of participants (56% and 61% respectively) said the means employed did not affect the gravity of the offense. A minority of participants did say that the means mattered, but there was no agreement within this minority about the relative hierarchy of means. So, for example, in the question related to killing, the group that said means mattered were split between believing that beating people to death was most serious, shooting people was most serious, and that killing people with bombs was the most serious. The largest group within this minority was those who said that beating people to death was the most serious at 19%. This makes the means employed to carry out the offense a weak indicator of gravity.¹⁴⁴

¹⁴⁴ Having said that, the use of particular cruelty was a strong indicator of gravity. *See supra* Section VII(A). One way to reconcile these findings is that the means employed does not matter unless it results in particular suffering for the victims. *See supra* Section VI(D)(1).

D. Other Observations

1. Reliability of the Results

There is consistency in the survey results. There were several gravity components that were tested using two questions. This was done, in part, to permit the answers to the questions to be compared so as to test the overall reliability of the responses.¹⁴⁵ Thus, the number of direct victims, the type of crime, the means employed, plan or policy, and particular cruelty were all tested using two questions. It is striking that the results for the two questions are very similar for most of those factors. For example, in question 13 (number of direct victims), 53% of the participants scored all of the scenarios as equally grave, while in question 19 (number of direct victims) 57% of the participants scored all of the scenarios as equally grave. Responses were similarly consistent across the questions related to type of crime, the means employed, and particular cruelty. There was an apparent difference between the two questions related to plan or policy, but this may be explained by the fact that one of the two questions had a scenario that contained accidental deaths while the other question contained only deaths associated with a culpable mental state.¹⁴⁶

Even when two distinct but related concepts were tested, for example in the questions related to ethnic and religious discrimination or the questions related to geographic or temporal scope of the crimes, the responses were quite similar. The consistency of the responses across questions that test either the same concept or similar concepts suggests that the responses are not random and do represent the relative gravity of the different components. In other words, the consistency of the responses suggests that those responses are reliable indicators of people's perceptions of the gravity of the offenses.

2. A Minority Rejects All Gravity Components?

There appears to be a small minority of participants that ranked all the scenarios the same. Thus, even for the strongest indicators of gravity (the extent of the harm and the type of crime) there was a small group of participants (15-20%) who said that all of the scenarios were equally serious. These responses do not appear to be random. More than 80% of the participants who gave all of the scenarios in question 2 (type of crime) the same score also gave all of the scenarios in question 10 (also type of crime) the same score. About 50% of the participants who said that all of the

¹⁴⁵ See Robinson and Kurzban, *supra* note 17, at 1840-41 (noting that consistency across answers that test the same idea in different ways is an indicator of reliability).

¹⁴⁶ See *supra* Section VI(C)(2).

scenarios in the type of crime questions were equally serious also said that all of the scenarios in the extent of harm question (question 8) were equally serious. In other words, there appears to be a distinct group of participants who did not believe that any of the factors, even the ones on which there was broad agreement, had any impact on the gravity of the offense.

It would be useful to know more about this group. Why did they rank all of the scenarios the same? One possibility might be a philosophical belief that crime gravity is unquantifiable and cannot be reduced to a number.¹⁴⁷ Another possibility might be that some people have an upper threshold for gravity. For these people, perhaps any crimes more serious than the threshold will be viewed as equally serious. Mass atrocities may simply exceed the upper threshold for these participants. A third possibility is that the people in this group do believe that these factors affect gravity but either misunderstood the instructions or were not sufficiently motivated to answer the questions carefully.¹⁴⁸ It would be worth following up on this group in subsequent research.¹⁴⁹

3. Individuals Have Different Baselines

Another feature of the data is that people appear to have different baselines for crime gravity even when they agree on the relative importance of different factors. For example, 77% of the participants agreed that the number of indirect victims of a crime did not affect its gravity. But, while there was broad agreement that the number of indirect victims did not change the gravity of the offense, the absolute scores that people gave in response to that question varied significantly. While the median response by people who said all of the scenarios were equally serious was 200, there were several people who gave scores in the tens of thousands and one person who gave a score of 100,000.¹⁵⁰ The answers to the question about the geographic spread of the crimes are similar. 80% of the participants scored all of the scenarios equally, but that agreement about the relative seriousness

¹⁴⁷ See, e.g., deGuzman, *supra* note 17, at 288 (arguing the various gravity factors suggested in the literature are “incommensurable” and provide “an inadequate justification” for decisions that are based on gravity). See also *id.* at 289 (arguing that “[g]ravity simply does not have enough agreed content to provide convincing justifications” for decision-making by the ICC).

¹⁴⁸ See Robinson and Kurzban, *supra* note 17, at 1875 (noting that surveys carried out over the internet run the risk that participants will be confused or lack sufficient motivation to complete the tasks as instructed).

¹⁴⁹ For example, in future iterations of the survey it might be possible to ask people who mark all of the scenarios as equally serious to explain why they answered that way.

¹⁵⁰ To put it in statistical terms, while the median score was 200, the 75th percentile score was 750, the mean was 2,917 and the standard deviation was more than 12,000. There was a lot of variability in the results.

of the offense was belied by the absolute scores. The median score (among those who scored all of the scenarios the same) was 300, but there were many scores in the thousands and tens of thousands and one score of 300,000.¹⁵¹

If we look, instead, at questions where most people agreed there was a hierarchy of the scenarios, the results are similar. In response to question 8 (extent of the harm suffered), 76% of the participants said that the harm suffered was a factor in the gravity of the offenses. But, the absolute scores showed a large amount of variability even though there was broad agreement that this factor did affect gravity. The median score for scenario 3 (the most serious scenario) amongst people who agreed that the extent of the harm mattered was 100, but there were several people who gave scores in the thousands and one person who gave a score of nearly 85,000.¹⁵² In effect, it appears that even where there is broad agreement about whether a particular factor affects the gravity of an offenses, there are still significant differences about the absolute gravity of those offenses. This suggests that even when people agree about the relative importance of different factors, they are often operating from different baselines. This finding is consistent with the results of domestic studies of crime severity.¹⁵³

VIII. CONCLUSIONS

Gravity is central to how the ICC operates, but it is an inherently subjective concept and it seemed possible that there would be no widely-accepted meaning of crime gravity.¹⁵⁴ That does not turn out to be true. One of the most important results of this survey is that there are some areas of agreement about how to define the gravity of mass atrocities. Of course, not everyone agrees about all of the components of a gravity definition and some components are stronger indicators of gravity than others, but there does appear to be agreement about at least some components of gravity. Gravity is not an idiosyncratic concept. This is not entirely surprising given the results of research on domestic crime seriousness,¹⁵⁵ but it is still an important finding.

The various components tested in this article have been categorized as strong, moderate, or weak indicators of crime gravity depending on what percentage of the participants believed they affected the overall gravity of a

¹⁵¹ Again, to put it in statistical terms, while the median score was 300, the 75th percentile was 3,000, the mean was 7,270, and the standard deviation was more than 34,000.

¹⁵² The median score was 100, the 75th percentile score was 800, the mean was 2,300 and the standard deviation was more than 10,000.

¹⁵³ See Robinson and Kurzban, *supra* note 17, at 1854-55.

¹⁵⁴ See *supra* Section II.

¹⁵⁵ See *supra* text accompanying notes 27-29.

crime. The results are summarized below in Table 1. The number that follows each component in the table is the percentage of the participants that agreed that factor was associated with an increase in gravity. If there are two numbers following a component, that component was tested using two questions, and the numbers represent the results of each question.

Table 1 – Gravity Components
<p>Strong Indicators of Crime Gravity</p> <p>Extent of the physical harm suffered by victims (76%)</p> <p>Type of crime (79%, 75%)</p> <p>Particular cruelty (55%, 49%)</p>
<p>Moderate Indicators of Crime Gravity</p> <p>Number of direct victims (43%, 34%)</p> <p>Presence of a discriminatory intent (41%, 29%)</p> <p>Abuse of power or official capacity (34%)</p> <p>Presence of child victims¹⁵⁶ (26%)</p>
<p>Weak Indicators of Crime Gravity</p> <p>Number of indirect victims (15%)</p> <p>Geographic spread of the crimes (18%)</p> <p>Temporal spread of the crimes (15%)</p> <p>Existence of a plan or policy (19%, 6%)</p> <p>Means employed to carry out the crimes (19%)</p>

Within the strong and moderate categories, there has been an attempt to list the components in order of their relative importance. For example, among the strong indicators, the extent of the physical harm and the type of crime appear to be nearly equal in strength, although both are stronger indicators than particular cruelty. The number of direct victims and the presence of child victims are both moderate indicators of crime gravity, but the number of victims appears to be a stronger indicator than the presence of child victims and appears higher in the list. There is no obvious hierarchy

¹⁵⁶ This component has been renamed here. While the OTP appears to make a broader claim that the identity of the victim matters, the results of the survey suggest that is true when the victims are children but not when the victims are men, women or the elderly. Thus, this component has been renamed here to reflect that result.

among the weak indicators.

The existence of broad agreement about the components of the gravity of mass atrocities means that it should be possible for the ICC to adopt a definition of gravity that most people will find acceptable. Of course, there are variations in how people see gravity, so not everyone will accept every component of a gravity definition, but there does seem to be enough commonality that a widely-accepted definition is possible. This is good news for the ICC because people will be much more likely to view ICC as legitimate if they can understand and agree with the decisions it makes about which situations to investigate, which people to try, and what sentences to give to those it finds guilty. In turn, this will make it more likely that it can persuade people to comply with international law. Ultimately, this makes it more likely that the ICC can achieve its primary goal of preventing violations of international criminal law.¹⁵⁷

The results of this survey indicate that the existing definition of gravity used by the OTP is not perfect. While it does contain a number of strong indicators of crime gravity, it also contains a number of weak indicators of gravity. Assuming the results of this survey can be replicated elsewhere, then the OTP should modify its gravity definition. The weak indicators should be dropped from the definition. If the vast majority of people do not believe that the geographic or temporal spread of the crimes affects their gravity then it does not make sense for the OTP to use these factors in assessing the gravity of offenses. The same holds true for the number of indirect victims, the means employed to carry out the crimes, and the existence of a plan or policy. Removing these factors from the definition will produce results that are more likely to be viewed by the public as legitimate.

The OTP's definition should also be amended to recognize that not all indicators have an equal effect on crime gravity. The OTP should formally recognize that the strong indicators from Table 1 should be weighted more than the moderate indicators. This will also help ensure that the court's gravity decisions are perceived as legitimate by the public. Finally, the other organs of the court should also adopt a gravity definition that matches how most people perceive the gravity of mass atrocities. Adopting a definition of gravity that accords with most people's perceptions of gravity will make it more likely that the ICC will be successful in the long run.

IX. NEXT STEPS

While this article shows that there is broad agreement about some components of crime gravity in the United States, further research is needed.

¹⁵⁷ See *supra* Section II.

Most important would be replicating the results of the survey in other countries, because it is possible that other populations have different understandings of gravity. For example, while Professors Robinson and Kurzban argue that “that human intuitions of justice about core wrongdoing—both the sense that serious wrongdoing should be punished and the sense of the relative seriousness of wrongdoing—are deep, predictable, and widely shared,”¹⁵⁸ it seems plausible that some aspects of the gravity of mass atrocities are cultural. In other words, there is a distinct possibility that different cultures define gravity slightly (or perhaps even significantly) differently.¹⁵⁹ The only way to know is to test the definition of gravity in different cultures.

If the core components of gravity remain the same across cultures this would make the success of the ICC more likely in the long run as it would make it easier for the ICC to achieve legitimacy across different cultures. If, on the other hand, gravity varies significantly across cultures, this will make it harder for the ICC to achieve legitimacy because no matter what it does different cultural groups will view its work differently. Given that much of the ICC’s focus has been on crimes committed in Africa,¹⁶⁰ it probably makes sense to test the meaning of gravity somewhere in Africa next.

Further research is also needed on other potential gravity components. Not all of the gravity components proposed in the OTP’s definition were tested in this survey. Additional work should be done to test all of the OTP’s gravity components.¹⁶¹ In addition, there is no reason to believe that the OTP has identified all of the factors that could potentially affect crime gravity.¹⁶² It would make sense to test gravity components beyond those contained in the OTP’s definition. Additional factors may be found in the legal scholarship on crime gravity or by looking in more depth at domestic crime severity research. It probably also makes sense to follow up on the tentative finding that there may be gender differences with regard to some of

¹⁵⁸ See Robinson and Kurzban, *supra* note 17, at 1892.

¹⁵⁹ *But see id.* at 1852 (noting that “cross-cultural data suggest that questionnaire studies yield similar results in all of the cultures that have been studied”); *id.* at 1862-65 (noting that studies in a wide range of locations, including Ireland, Taiwan, Canada, Puerto Rico, India, Indonesia, Iran, Italy, and Yugoslavia, all demonstrated broad agreement about the relative seriousness of many domestic offenses).

¹⁶⁰ See Ford, *supra* note 34, at 25 (noting that as of July 2015, the ICC had begun or completed investigations in nine different situations, all of them in Africa).

¹⁶¹ For example, this survey did not test factors like the effect of varying the psychological harm of a crime or the environmental or economic consequences of crimes.

¹⁶² Indeed, the OTP acknowledges as much. Its own description of the components of crime gravity suggests that the factors it has identified are not exhaustive. See, e.g., Policy Paper on Preliminary Examinations, *supra* note 49, at para. 25 (noting that the factors that affect the scale of the crimes include “inter alia” the number of direct and indirect victims, the extent of the harm, and the geographical and temporal spread of the crimes).

the gravity components.¹⁶³

Finally, it makes sense to test different theories of criminal responsibility to see how the degree of participation in the crime affects perceptions of the criminal responsibility of individuals.¹⁶⁴ This question is less important when deciding whether to begin a formal investigation of a situation, but is very important when deciding what sentence to impose.¹⁶⁵ Having a better understanding of how a person's participation affects their criminal responsibility could help the ICC make better decisions about who to try and what sentences to impose on those found guilty. In this regard, it is worth noting that domestic studies of crime severity have found that people do make nuanced decisions about the appropriate punishment for a wrongdoer based on the way in which they participate in a crime.¹⁶⁶ So, there is reason to believe that the way in which a person participates in a serious violation of international criminal law affects what punishment people think that person deserves.

¹⁶³ See *supra* Section VI(B)(1).

¹⁶⁴ Degree of participation is suggested as a factor related to gravity in the OTP's definition but was not tested in this survey. See *supra* text accompanying note 113.

¹⁶⁵ See Rome Statute, art. 78(1) (noting that the court shall consider both the gravity of the crime and "the individual circumstances of the convicted person"). The Rules of Procedure and Evidence elaborate on this and make it clear that the court should consider the degree of the convicted person's participation in the crime when deciding on a sentence. See International Criminal Court, Rules of Procedure and Evidence, art. 145(1) (stating that the court should consider "the degree of participation of the convicted person" when determining the sentence").

¹⁶⁶ See Robinson and Kurzban, *supra* note 17, at 1842-45 (describing the results of a number of studies that showed that people thought perpetrators deserved different punishments depending on what specific role they played in the offense and what they knew at the time they contributed to the offense).