



The Depravity Standard III: Validating an evidence-based guide

Michael Welner^{a,*}, Kate Y. O'Malley^a, James Gonidakis^a, Alisha Saxena^{b,1}, Jada Stewart-Willis^a

^a The Forensic Panel, 224 West 30th Street, Suite 806, New York City, NY, 10001, United States

^b Teachers College, Department of Clinical Psychology, 428 Horace Mann, New York, NY 10027, United States

ARTICLE INFO

Keywords:

Murder sentencing
Evil
Heinous, atrocious, and cruel (HAC)
Depravity Standard
Sentencing reform

ABSTRACT

Purpose: The Depravity Standard is an evidence-based guide developed to operationalize an approach to distinguish the worst of crimes in a consistent manner that minimizes bias. This phase of the research was designed to validate the Depravity Standard items and develop a scoring mechanism.

Methods and results: Inter-rater reliability was performed by two groups of trained raters, with each of the 25 Depravity Standard items finding high agreement. To distinguish the relative severity of each item as they may occur in a murder case, an online public survey was devised. U.S. participants ($n = 1273$) rated each item on a scale of 1–100 (100 = most depraved). The items were then applied to 770 case files of adjudicated murder convictions to establish content validity. 582 cases were retained for further analysis, and merged with survey data to establish a percentile scoring system.

Conclusions: The Depravity Standard is validated for application to murder cases to inform the presence or absence of the 25 items of depravity. It enables assessment of relative depravity of a perpetrator's intent, victim choice, actions, and attitudes. Application of the Depravity Standard relies on evidence, minimizes bias and prejudice, and promotes fairness in sentencing and release decisions.

1. Introduction

Codes of criminal law in America have established terms such as “heinous,” “atrocious,” “cruel,” “depraved,” “vile,” and other analogues of evil to refer to the worst of crimes. The law uses these terms to distinguish those crimes that warrant more severe sentences. HAC (heinous, atrocious, cruel) aggravators, as they are commonly known in legal parlance, were specifically devised for application to murder cases.

Over the years, HAC aggravators have been consistently challenged in court for their vague terminology and resulting contribution to arbitrary sentencing. Higher court opinions, however, have upheld the use of HAC aggravators. At the same time, court decisions have struggled to bring clarity and consistency to these distinctions.

In *Gregg v. Georgia* (1976), the United States Supreme Court upheld the Georgia aggravator of “heinous, atrocious, and cruel” as constitutional but expressed concern that a jury would have difficulty deciding this issue. Writing for the majority, Justice Potter Stewart noted that the problem of jury inexperience could be “alleviated if the jury is given guidance regarding the factors about the crime and the defendant that the state, representing organized society, deems particularly relevant to the sentencing decision.” (at 192). In *Godfrey v Georgia* (1980), the

Court determined that the death penalty must not be imposed by “standardless” sentencing, and that jurors require instruction in the otherwise ambiguous HAC language. Then, in *Walton v. Arizona* (1990), the Court clarified that aggravating factors needed to be identified through objective circumstances.

Even reliance upon objective circumstances, however, does not ensure that the interpretation of what is a depraved murder is not arbitrary or vague. *Norris v. State* (1999) involved a case of three bar patrons who were shot, one after the other, while sitting in a booth. The Alabama Court of Appeals was divided over whether the victims suffered “psychological torture” because they were aware they were going to die. The majority opinion ruled that the murders happened quickly enough that they did not fit the “torture” criteria. What is quickly enough? Without direction, the trier of fact must rely on visceral judgments which are vulnerable to bias and may provide different interpretations of the law from one case to the next. Specificity safeguards against the consequences of vagueness.

Courts' emerging recognition of the importance of substantive and evidence-driven arguments on HAC was illustrated in *Dixon v. Ryan* (2016). The District Court in *Dixon* upheld a finding of “cruel, heinous and depraved” because the prosecutor argued specific history and evidentiary findings at trial in support of the aggravator. However,

* Corresponding author.

E-mail address: drwelner@forensicpanel.com (M. Welner).

¹ This work was carried out while the author was doing an internship at The Forensic Panel.

arguments highlighting particular evidence may still be specific to a given prosecutor; they are detailed, but not necessarily informative of any reliable definition of what a heinous murder is. In addition, an elective interpretation of “heinous,” even if it provides an example that other prosecutors follow, still falls short of the *Gregg* aspirations of relying on developed societal standards.

Jurors and judges in American courts currently have no guidance to assess the level of heinousness or depravity of a crime. Inconsistency in defining the worst of murders continues to bedevil criminal casework. Moreover, the subjective interpretations of “depraved” and other HAC aggravators continue to fall short of accounting for societal standards.

In recent years, court decisions mandating large scale early release to relieve prison overcrowding has highlighted a related dilemma. How can early release decisions be rendered fairly among the sizeable numbers of prisoners of the same class of crime? In California, the state most notably associated with the magnitude of court-mandated release (see; *Brown v. Plata*, 2011), parole decision-making in murder cases has been criticized for the non-specific and widespread designation of first and second degree murders as “heinous, atrocious, and cruel” to avoid early release (Ball, 2009). When a court forces the issue, and demands that prisons be liquidated, what mechanism gives clarity to a parole board's newly mandated discipline for designating those prisoners never to be released?

How, for example, does one make a fair decision about which murder, attempted murder, or manslaughter defendant warrants early release? If both represent the same level of risk to the community, what makes one offender more suitable for release? An evidence-driven distinction of the worst of murders would better inform such decisions and promote public confidence in sensitive decisions.

The Depravity Standard is a 25-item evidence-based guide for the appraisal of criminal depravity in accordance with the goals set out by the Supreme Court in *Gregg*, *Godfrey*, and *Walton*. It has been developed to assist triers of fact to objectively assess a crime's relative severity in order to inform fair sentencing and release decisions. In an effort to minimize arbitrary sentencing decisions that result from inadequate scrutiny of relevant evidence, and from implicit and explicit biases, the Depravity Standard focuses investigation of depravity upon elements of the crime itself as opposed to the defendant's personal background.

In Welner, O'Malley, Gonidakis, Saxena, and Burnes (2018), the authors detailed the methodology of developing and refining the Depravity Standard in Studies 1–3 of this research. In Study 1, items of depravity were formed from review and thematic analysis of the specific rationale of 110 appellate court cases upheld as “heinous, atrocious, or cruel.” Fifteen core elements of a perpetrators' intent, actions, and attitude were supplemented with data from 91 students and professionals in the industry, proposing aspects of crimes that rendered them “depraved.” The efforts from these exercises resulted in 26 items for closer study.

In Study 2, the public opinion of 25,096 U.S. participants was captured to determine whether the items derived from Study 1 warranted inclusion in a final Depravity Standard. No items were flagged for exclusion from further study. The final development phase, Study 3, applied these items to actual closed felony murder cases. Through this review process, extensive qualifying and disqualifying definitions for each item were developed and refined.

In the course of the study, one item was dropped because of concerns that its overlap with other items would create vagueness in its definition, allowing for discrepant application. Twenty-five items remained for further study of the Depravity Standard (Table 1).

This article details the validation of the Depravity Standard (Studies 4 and 5) and its proposed application in U.S. court sentencing, in early release decision-making by parole officials, and for pardon decisions by elected officials.

2. Study 4 Phase A: inter-rater reliability

Each of the 25 items of the Depravity Standard include extensive definitions with various qualifying and disqualifying examples of the items' potential application in murder cases. This high level of detail minimizes arbitrariness and promotes consistent application in case-work, however common or obscure the potential fact pattern. An important aspect of validating the tool, therefore, is ensuring inter-rater reliability. The data for this study is a subset of the data set reported in Study 3, described in depth in Welner et al. (2018).

2.1. Method

2.1.1. Sample

From the 770 cases reviewed as part of Study 3, a subsample of 250 cases were randomly selected for inclusion using a random sample generator (Haahr, 2006).

2.1.2. Procedure

Each of the 250 cases in Study 3 were rated twice by two independent groups of raters for the presence of the Depravity Standard items. As part of Study 3, each case was assigned an overall rating of present (Yes), absent (No) or insufficient data (ID) for each of the 25 items. The ID responses represented either a lack of information (i.e., in an autopsy report, the photo of the body was too overexposed to see any detail) or a rater's uncertainty about the information presented (i.e., the defendant's statement contradicts a co-defendant's statement and there is no way to determine which is the true account from the provided materials without speculation).

2.2. Data analysis

Data was entered into IBM SPSS version 22.0 for Windows (IBM Corp., 2013) for analysis. The data was screened for coding errors, and responses for the presence of each item were coded as 1 = Yes, 2 = No, and 3 = ID. Ratings of “No” and “ID” have the same functional significance in a criminal justice context – namely, that proof of guilt requires presence of evidence beyond a reasonable doubt. With the exception of prior convictions, a court may not use aggravating factors to impose a harsher sentence than usual unless the jury found those factors to be true beyond a reasonable doubt (i.e., *Cunningham v. California*, 2007). Therefore, absence of evidence equals evidence of absence. As such, the ratings of “No” and “ID” were combined for analysis and re-coded to 1 = Yes and 2 = No/ID.

Upon preliminary analysis of Cohen's Kappa (Cohen, 1960) for inter-rater reliability, high agreement between raters yielded a low or negative Kappa value. Manual review of the 250 cases demonstrated that ratings of “No/ID,” where the item is absent or cannot be determined with confidence, appeared significantly more frequently than ratings of “Yes,” where the item is present without doubt.

The lower occurrence of “Yes” ratings can be attributed to the extensive development and validation phases of the research, where items were refined to describe specific criteria of depraved intent, actions, attitudes, and victim choice that reflect depravity, or the “worst-of-the-worst” murders. The thorough development process was described in Study 2 (Welner et al., 2018), and provided participant raters with specific criteria for each of the items to ensure that, when scored, the items reflected exceptional qualities in a crime. Raters therefore had the necessary guidance to highlight only a small subset of offenses for which any of the items were present. This meant that the worst, most depraved offenses only appeared as a small percentage of overall cases, and this discriminant sensitivity was the hypothesized result.

The AC₁ statistic (Gwet, 2008) was applied to each of the 25 items across the 250 random cases in IBM SPSS version 22.0 to determine the level of agreement between independent raters. The AC₁ statistic is noted by Gwet (2008) to measure ‘true’ inter-rater reliability in that it

Table 1
List of Depravity Standard items.

Item number	Item description
1	Intent to emotionally traumatize the victim, maximizing terror, through humiliation, or intent to create an indelible emotional memory of the event – INTENT
2	Intent to maximize damage or destruction, by numbers or amount if more than one person is victimized, or by degree if only one person is victimized – INTENT
3	Intent to cause permanent physical disfigurement – INTENT
4	Intent to carry out a crime for excitement of the criminal act – INTENT
5	Targeting victims who are not merely vulnerable, but helpless – VICTIM CHOICE
6	Exploiting a necessarily trusting relationship to the victim – VICTIM CHOICE
7	Influencing depravity in others in order to destroy more – INTENT
8	Crime reflects intent of progressively increasing depravity – INTENT
9	Carrying out a crime in order to terrorize others – INTENT
10	Carrying out crime in order to gain social acceptance or attention, or to show off – INTENT
11	Influencing criminality in others to avoid prosecution or penalty – INTENT
12	Disregarding the known consequences to the victim – ACTION
13	Intentionally targeting victims based upon prejudice – VICTIM CHOICE
14	Prolonging the duration of a victim's physical suffering – ACTION
15	Unrelenting physical and emotional victimization; amount of victimization – ACTION
16	Exceptional degree of physical harm; amount of damage – ACTION
17	Unusual and extreme quality of suffering of the victim, including terror and helplessness – ACTION
18	Indulgence of actions, inconsistent with the social context – ACTION
19	Carrying out crime in unnecessarily close proximity to the victim – ACTION
20	Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation – ACTION
21	Pleasure in response to the actions and their impact – ATTITUDE
22	Falsely implicating others, knowingly exposing them to wrongful penalty and the stress of prosecution – ATTITUDE
23	Projecting responsibility onto the victim; feeling entitlement to carry out the action – ATTITUDE
24	Disrespect for the victim after the fact – ATTITUDE
25	Indifference to the actions and their impact – ATTITUDE

removes the evaluation of all agreements that occurred by chance. Additionally, Gwet points out that kappa and pi analyses both consider that every rating agreement could have occurred by chance, which in the present study would proffer misleading results as the presence of depravity is not random. With the available training having led to a distinguishable prevalence of “No/ID” ratings, the AC₁ recalibrates the chance-agreement yield to the appropriate lesser power. It is important to account for this readjustment when conducting agreement statistics since each measure is influenced by trait prevalence rates (Gwet, 2002), wherein this study involved a particular score (No/ID) that had an agreement chance of > 50%. Since Kappa statistics can lead to low or negative values in data sets with agreement yields above or below 50%, the AC₁ statistic (Gwet, 2008) was used.

2.3. Results

The AC₁ statistic was calculated for items 1 through 25 of the Depravity Standard for the 250 random cases. The AC₁ statistic may be interpreted using the cut-off ranges of Landis and Koch (1977), Altman (1991), or Fleiss, Levin, and Paik (2003) (Table 2). Depending on the cut-off scale used, seventeen Depravity Standard items demonstrated “almost perfect,” “very good,” or “excellent” agreement (0.81 < κ < 1.00), seven had “substantial,” “good,” or “intermediate to good” agreement (0.61 < κ < 0.80), and one (Item 20) demonstrated “fair” or “poor” agreement at κ = 0.363.

Item 20 is the action of responding excessively to a trivial irritant, in which one's actions are clearly a disproportionate reaction to the

Table 2
Cut-off scales of kappa value.

Landis and Koch (1977)	Altman (1991)	Fleiss et al. (2003)
< 0.0 Poor		
0.00 to 0.20; Slight	< 0.20; Poor	< 0.40; Poor
0.21 to 0.40; Fair	0.21 to 0.40; Fair	
0.41 to 0.60; Moderate	0.41 to 0.60; Moderate	0.40 to 0.74; Intermediate to good
0.61 to 0.80; Substantial	0.61 to 0.80; Good	
0.81 to 1.00; Almost perfect	0.81 to 1.00; Very good	> 0.75; Excellent

perceived provocation. It was determined that this item should be evaluated and further refined to ensure better consensus. The next section describes the steps taken to further refine the item definition and the re-rating of cases for a second interrater reliability analysis of this item.

3. Study 4 Phase B: inter-rater reliability follow-up

Interrater reliability for Item 20 “Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation” resulted in low agreement between raters. As such, this item was probed for discrepancies between the raters' documented rationale for their ratings for this item, described below. The qualifying and disqualifying examples for the item were revised accordingly, and the participant training protocol was amended to incorporate a clearer explanation of the item. The follow-up to Study 4 was undertaken and re-analysis was then conducted to ascertain if Item 20 could reach suitable inter-rater reliability.

3.1. Method

3.1.1. Participants

This next phase of the study included a set of 10 male and female participants, forming two rater groups of 5, who were selected and trained in the same process as in Study 3 and Study 4 Phase A, but with the revised definition of Item 20.

3.1.2. Procedure

Participants who were initially assigned the 250 cases for rating were engaged in discussion sessions with each other and the Research Assistants and Research Director to collectively distill the elements of a case that reflect an offender's excessive response to a trivial irritant. The group discussions also served to eliminate confusion as it pertained to the verbiage of the item's examples in the Evidence Guide, the manual that fully explains the definitions as well as qualifying and disqualifying examples for each Depravity Standard item. Confusing or frequently misunderstood sentences were updated. The new group of 10 participants were provided with the same training for items, but with the revised explanation of qualifying and disqualifying examples for Item 20.

This new group of participants rated the same 250 homicide cases for Item 20 using the updated description.

3.2. Data analysis and results

The AC₁ statistic was re-calculated on the case ratings for this group of participants for Item 20, along with Cohen's Kappa for reference. The results for both statistics for all items are presented in Table 3.

With the revised training, the inter-rater agreement for Item 20 increased significantly from 0.363 to 0.818. The final definitions for the Depravity Standard demonstrate high inter-rater reliability across all items when applied to felony murder cases. The Depravity Standard items can therefore be expected to be reliably used by independent raters, professionals, and jurors who are provided and trained in the guidelines for case assessment.

4. Study 5 Phase A: weighting the items

During Study 2 of the Depravity Standard's development, the researchers established that public respondents affirm the 25 Depravity Standard items as being somewhat or especially representative of depravity. Study 5 measured the relative amount of depravity the public attributes to each item. This study aimed to inform societal standards for how each Depravity Standard item should be weighed when present, to inform a valid scoring mechanism.

Weighting specific qualities of depravity assists both inexperienced and experienced triers of fact alike to make informed and evidence-based decisions about the severity of a given offense. Furthermore, in cases of multiple or collaborating perpetrators and conspirators, weighting items of the Depravity Standard informs appraisal of the relative culpability of co-defendants, especially because the guide explores intent and therefore accounts for those who are a crime's unseen hand. Study 5 also allows for assessment of inter-item correlations.

By focusing on the quality of depravity as opposed to quantity of depraved items alone, the features of criminal behavior that society deems most and least severe can impact sentencing in accordance with societal preferences. Data from survey participants is examined below, in addition to an analysis of construct validity. Finally, a scoring mechanism incorporating the public survey data is presented.

Table 3
Inter-rater reliability results.

	AC ₁	AC ₁ Re-rated	COHEN'S	COHEN'S Re-rated
Item 1	0.961		0.553	
Item 2	0.867		0.303	
Item 3	0.988		0.394	
Item 4	0.992		0.000	
Item 5	0.944		0.289	
Item 6	0.900		0.611	
Item 7	0.878		0.455	
Item 8	0.643		0.476	
Item 9	0.988		0.566	
Item 10	0.975		0.390	
Item 11	0.930		0.300	
Item 12	0.764		0.367	
Item 13	0.975		0.239	
Item 14	0.923		0.416	
Item 15	0.766		0.458	
Item 16	0.728		0.389	
Item 17	0.770		0.430	
Item 18	0.843		0.404	
Item 19	0.824		0.407	
Item 20	0.363	0.818	0.358	0.566
Item 21	0.920		0.585	
Item 22	0.866		0.389	
Item 23	0.794		0.476	
Item 24	0.889		0.282	
Item 25	0.715		0.449	

4.1. Method

4.1.1. Participants

Participants were 9492 respondents from around the world. The survey was online, and therefore accessible by anyone who elected to take part. However, recruitment efforts targeted United States residents, in order to inform about the collective attitudes of individuals who could potentially comprise a jury in a U.S. court. Participants were randomly assigned to one of four categories; Murder/Attempted Murder, Violent Crimes, Non-Violent Crimes, or Sex Crimes.

After removing incomplete entries, a total of 1663 participants entered into this study for Murder/Attempted Murder. Data was also removed if: 1) participants were from countries other than the U.S., 2) participants did not report their country of residence, and therefore could not be conclusively included in this study's sample, 3) participants were under 18 years of age and therefore ineligible to serve on a U.S. jury, and 4) participants did not consider themselves fluent in English. 1273 participants were retained for analysis.

Of the participants retained for analysis, 49.6% percent were male and 50.4% were female. Participants included residents of all fifty states and the District of Columbia. The most represented states were California (17%), Ohio (9%), Texas (7%), New York (6%), and Florida (5%). Additional demographic characteristics of participants can be seen in Table 4. Not all participants elected to answer all of the demographic questions.

4.1.2. Participant recruitment

Study 5 received approval from The Forensic Panel's Institutional Review Board for the Protection of Human Participants. Participation in the Depravity Standard research was voluntary. Participants self-selected to complete the survey by visiting the website www.depravitystandard.org. To raise awareness of the research and encourage participants to visit the website and complete the survey, several recruitment techniques were utilized. During media appearances on forensic and investigative topics, including this research, the principal investigator encouraged viewers and readers to participate in the online survey. Additionally, several prior articles had been written about the project (Welner, 2003, 2009, 2013). The researchers also used snowballing techniques by allowing participants, upon completion of the online survey, to enter the email addresses of friends and family who may be interested in participating. A Twitter account (@whatistdepraved) and Facebook account (www.facebook.com/DepravityStandard/) highlighted crimes with thought-provoking fact patterns to encourage readers to contribute their own input on what makes a crime depraved by participating in the surveys. Participants from earlier phases of the research were emailed and invited to complete the latest survey online. A number of interested parties posted a link to the research on their social media accounts. Interns at The Forensic Panel placed posters about the research (with permission from their schools) on their college research recruitment boards. A number of teachers and professors called attention to the research in their classrooms.

No compensation was offered for survey completion. Participants read the statement: "By entering the survey I agree that I have read and understood the informed consent information," and were provided with two hyperlinks to view the informed consent information. Participants were also sent a copy of the informed consent form to their email address when they started the survey. For minors who actively looked up the website and opted to complete the survey despite the notice that participation was open to U.S. participants aged 18 and older, the researchers felt it would be impractical to obtain parental consent. The content of the survey was concise and far less detailed and graphic than other readily viewable internet content such as news media sites that regularly describe crime, TV programs and movies that portray violence, and gaming sites (Yoon & Somers, 2003).

Table 4
Demographics.

General demographics						
Age of participants when survey completed (n = 1159)						
18–24	25–34	35–44	45–54	55–64	65–74	75 +
18.2%	19.5%	19.2%	18.0%	17.3%	7.0%	0.8%
Education (n = 1270)						
Doctoral	Graduate degree	Undergraduate degree	High school diploma/GED	Some high school		
6.8%	25.4%	44.5%	22.7%	0.7%		
Income (n = 1171)						
< \$25,000	\$25,000–\$50,000	\$50,000–\$75,000	\$75,000–\$100,000	> \$100,000		
11.6%	19.8%	18.0%	19.7%	30.9%		
How would you classify the influence of religion in your life and in your thinking? (n = 1270)						
Not traditional, but spiritual	Not traditional, not spiritual		Very traditional, very spiritual		Very traditional, not spiritual	
39.9%	21.6%		30.3%		8.1%	
Community type (n = 1272)						
Rural						Suburban/urban
22.9%						77.1%
Ethnicity (n = 1260)						
White	Black	Hispanic	Asian	American Indian/Alaska Native	Two or more	Native Hawaiian/Pacific Islander
87.5%	3.5%	4.0%	2.1%	1.1%	1.7%	0.2%
Profession (n = 1247)						
Administrative and clerical n = 77				Not working (unemployed/retired) n = 64		
Attorney – civil n = 23				Other n = 30		
Attorney – criminal defense n = 8				Professor – law/criminology n = 9		
Attorney – criminal prosecution n = 8				Professor – other n = 21		
Clergy n = 13				Sales n = 51		
Computers/IT/telecommunications n = 112				Scientist n = 38		
Corporate – professional n = 71				Small business owner n = 46		
Financial services/real estate n = 46				Student – post graduate n = 59		
Government/civil service n = 45				Student – high school n = 6		
Healthcare n = 57				Student – undergrad n = 141		
Homemaker n = 44				Teacher/childcare n = 49		
Hospitality n = 24				Trade worker/construction/artisan n = 22		
Law enforcement/military/corrections n = 75				Transportation n = 17		
Media n = 15				Writer/actor n = 11		
Mental health professional n = 65						
Legal demographics						
Do you oppose the death penalty, without exception? (n = 1273)						
Yes				No		
20.3%				79.7%		
Has a close family member ever been sentenced to prison? (n = 1272)						
Yes				No		
18.4%				81.5%		
Have you ever been sentenced to prison? (n = 1271)						
Yes				No		
1.9%				98.1%		

(continued on next page)

Table 4 (continued)

Legal demographics	
Have you/family member ever been a victim of violent crime? (<i>n</i> = 1271)	
Yes	No
34.3%	65.5%

Due to rounding, some percentages may not total one hundred.

4.1.3. Procedure

Participants either searched for the research online, were hyperlinked to the research through an article as noted above, or manually typed in the website address to arrive at the research project's landing page. Participants signed in using an email address. This form of entry enabled the researchers to forward informed consent information to them (along with contact information for several helplines should any feelings of discomfort emerge). The email address submitted by participants also reduced the likelihood that participants would take the survey twice, for the site rejected attempts to enter the survey with an email address already used to participate.

Participants then proceeded to an initial starting page that generated a random number that assigned them to one of four crime categories: 1 = Assault; 2 = Murder/Attempted Murder, 3 = Non-Violent Crimes, and 4 = Sex Crimes. The participants were unaware as to which category each number referred. Each was asked to choose their assigned number from a drop-down box and to read the participant instructions, including information about consent. Informed consent was established by proceeding and subsequently completing and submitting the survey.

This paper reports on participants of the Murder category. There were 25 survey questions, each representing one of the 25 items of the Depravity Standard. Participants were instructed to read each item, use a drop-down menu to view longer descriptions of each item, and to then rate the item on a scale of 1 = Depraved to the least degree to 100 = Depraved to the most extreme degree by dragging a bar until the desired number was reached on the scale of 1–100. The order of the 25 items was randomly generated for each participant to control for order effects.

Instructions encouraged participants to use the entire range of 1–100 as they believed the items should be scored, in order to discourage homogeneous response styles that would make distinctions among items more difficult. Participants could adjust earlier answers as they proceeded through rating each of the 25 items. Once participants were satisfied with their answers, they clicked a button to submit their results and proceeded to a demographics questionnaire.

Participants were queried about numerous demographics. Researchers collected this data in order to compare results across subgroups, to be presented in subsequent manuscripts, and to explore how representative the sample was in comparison to the general United States population. After answering the demographic questions, participants reviewed and submitted their results. Participants may then have referred the survey to friends if they so chose.

4.2. Data analysis & results

Participants rated the level of depravity for each item surveyed on a scale of 1–100 (1 is the least depraved and 100 is the most depraved an item could be) and completed a series of demographic questions. Data was screened to exclude respondents who only filled out the demographic information but did not complete the questionnaire items in the survey. Outliers were also removed (i.e., participants scoring “0” for all 25 items) as well as participants without variance in their scores (i.e. rating “100” for 24 items, and “99” for one item). Participants who completed fewer than 19 of the 25 (75%) questionnaire items were also excluded from analysis, since the weighting of items was relative to

how a participant perceived each item on a scale of 1–100. If fewer than 19 items were accounted for, or if there was little or no variance in scores, this did not represent a sufficient comparative assessment of depravity for the purpose of weighting. Any participants with clearly fabricated demographics were also removed.

Participants from outside the United States or who did not provide a response for their country or Zip Code were also removed from analysis. Data was also removed from participants under the age of 18 and therefore unable to serve on an American jury, and from participants who did not consider themselves fluent in English. A final sample of 1273 cases was entered into IBM SPSS version 22.0 (2013) for analysis.

Violations of the Sixth Amendment's guarantee of an “impartial jury” are analyzed using the *Duren* test (*Duren v. Missouri*, 1979; reaffirmed in *Berghuis v. Smith*, 2010). The first prima facie factor that establishes a violation is where the group of potential jurors alleged to be excluded is a “distinctive” group in the community. A group is typically considered “distinctive” if it is based on race/ethnicity or gender. Therefore, chi-square goodness-of-fit analyses were employed to compare our sample against national data for ethnicity (Kaiser Family Foundation, 2016a) and gender (Kaiser Family Foundation, 2016b). Review of the chi-square goodness-of-fit statistical results, revealed no statistical difference for gender, $X^2(1, N = 1273) = 0.164$, $p = 0.685$, however most ethnic minority groups were under-represented in the sample, $X^2(4, N = 1260) = 389.708$, $p < 0.001$. To account for this difference the data was weighted, ranging from 0.7 for White participants, to 4.5 for Hispanic participants

To analyze the overall ranking of the Depravity Standard items, and specifically what their weighted score was compared to other items in the measure, the 1–100 responses were averaged across participants to generate a percentage. Analyses were run on the weighted and unweighted data to determine the overall score for each of the 25 items. The average level of depravity was determined for each item in the form of a percent. Summary *t*-tests were run to compare the weighted and unweighted totals with no statistically significant differences between item values (equal variances assumed) (Table 5).

Inter-item correlations were additionally performed to allow for the assessment of reliability through examination of internal consistency of the items. Cronbach's alpha was determined for the weighted ($\alpha = 0.959$) and unweighted ($\alpha = 0.955$) data, with both falling above the recommended minimum standard of 0.90 for applied research (Nunnally, 1978). The inter-item correlations were 0.485 (weighted data) and 0.465 (unweighted data). Ideally, the average inter-item correlation for a set of items should be between 0.15 and 0.50 (Briggs & Cheek, 1986). For a valid measure of a narrow construct, like depravity in crime, a much higher mean intercorrelation of 0.40–0.50 is recommended (Clark & Watson, 1995), thereby suggesting that the Depravity Standard items are reasonably homogenous while containing sufficiently unique variance so as not to be isomorphic, or a different expression of the same item (Piedmont, 2014). The weighted data was kept for further analyses.

5. Study 5 Phase B: applying the definitions to case files

The Supreme Court has asserted 1) the need to guide jurors, 2) that decisions of “heinous, atrocious, and cruel” must be informed by

Table 5
Depravity rating by item.

Item	Description	Unweighted percentage	Weighted percentage	Summary t-test significance ^a
1	Intent to emotionally traumatize the victim, maximizing terror, or through humiliation, or intent to create an indelible emotional memory of the event	77.99 (n = 1272)	76.41 (n = 1272)	0.066
2	Intent to maximize damage or destruction, by numbers or amount if more than one person is victimized, or by degree if only one person is victimized	76.77 (n = 1273)	75.79 (n = 1273)	0.272
3	Intent to cause permanent physical disfigurement	79.91 (n = 1273)	78.78 (n = 1273)	0.161
4	Intent to carry out a crime for excitement of the criminal act	70.01 (n = 1273)	69.07 (n = 1273)	0.350
5	Targeting victims who are not merely vulnerable, but helpless	80.80 (n = 1273)	80.53 (n = 1273)	0.746
6	Exploiting a necessarily trusting relationship to the victim	70.90 (n = 1273)	70.72 (n = 1273)	0.854
7	Influencing depravity in others in order to destroy more	69.89 (n = 1273)	69.42 (n = 1273)	0.612
8	Crime reflects intent of progressively increasing depravity	79.55 (n = 1272)	79.00 (n = 1272)	0.512
9	Carrying out a crime in order to terrorize others	72.60 (n = 1273)	72.14 (n = 1273)	0.618
10	Carrying out crime in order to gain social acceptance or attention, or to show off	58.02 (n = 1272)	56.98 (n = 1272)	0.336
11	Influencing criminality in others to avoid prosecution or penalty	60.67 (n = 1272)	60.23 (n = 1272)	0.685
12	Disregarding the known consequences to the victim	66.12 (n = 1273)	65.80 (n = 1273)	0.750
13	Intentionally targeting victims based upon prejudice	64.58 (n = 1273)	64.40 (n = 1273)	0.885
14	Prolonging the duration of a victim's physical suffering	88.45 (n = 1273)	87.65 (n = 1273)	0.196
15	Unrelenting physical and emotional victimization; amount of victimization	80.06 (n = 1272)	79.52 (n = 1272)	0.527
16	Exceptional degree of physical harm; amount of damage	79.41 (n = 1272)	78.81 (n = 1272)	0.472
17	Unusual and extreme quality of suffering of the victim, including terror and helplessness	85.79 (n = 1272)	84.66 (n = 1272)	0.132
18	Indulgence of actions, inconsistent with the social context	64.24 (n = 1272)	63.14 (n = 1272)	0.285
19	Carrying out crime in unnecessarily close proximity to the victim	66.91 (n = 1273)	65.49 (n = 1273)	0.166
20	Excessive response to trivial irritant; actions clearly disproportionate to the perceived provocation	56.03 (n = 1271)	55.04 (n = 1272)	0.353
21	Pleasure in response to the actions and their impact	80.05 (n = 1272)	78.80 (n = 1272)	0.176
22	Falsely implicating others, knowingly exposing them to wrongful penalty and the stress of prosecution	61.77 (n = 1271)	60.99 (n = 1272)	0.464
23	Projecting responsibility onto the victim; feeling entitlement to carry out the action	66.54 (n = 1270)	66.25 (n = 1271)	0.789
24	Disrespect for the victim after the fact	71.61 (n = 1270)	70.31 (n = 1271)	0.219
25	Indifference to the actions and their impact	67.49 (n = 1269)	67.05 (n = 1270)	0.673

^a p < 0.001.

societal standards, and 3) that standards must be clear, objective and offer specific and detailed guidance (e.g., *Gregg v. Georgia, 1976; Walton v. Arizona, 1990*). To assess whether the 25 items developed in earlier studies through appellate court review and input from the general public reliably measure depravity in actual case files, Study 5 was undertaken.

Our Research Assistants and Research Director collected and reviewed hundreds of closed felony murder case files, randomly selected for inclusion by numerous jurisdictions from around the United States. The researchers' data mined the cases to discern the presence or absence of items of depravity as defined in the earlier studies.

It was hypothesized that Depravity Standard items would only appear in a sub-group of the overall group of cases, and that multiple items would appear in still smaller subsets. If items appeared more frequently, the Depravity Standard may not be narrowing enough to the worst of cases. If items never appeared, the definitions could be too restrictive, unless the rare presence of that item demonstrates it to be uncommon, but not impossible. Items that seldom presented were reviewed to account for whether low frequency was due to definitions that were unduly restrictive, or because of the genuine rarity of that crime feature. For example, intent to cause physical disfigurement does occur in certain violent attacks, although our findings from three cases outlined below suggest this is rare. This study was therefore designed to test the content validity of the Depravity Standard measure.

5.1. Method

5.1.1. Participants

Participants were male and female research volunteers accepted into the internship program at The Forensic Panel. The participants, aged approximately 20 to 40 years, were students or recent graduates in the health science disciplines. Participants heard about the internship by visiting The Forensic Panel and the Depravity Standard websites, on

www.internships.com, by recommendation from their academic program advisors, and through peer-referral. Participants applied to become involved in the research and were selected after an interview focusing on their qualifications, dedication, and aptitude, and references from academic recommendations. Approximately 3% of those applying for the internship were accepted. Participation in the program was voluntary, and some received school credit for the six-month internship.

5.1.2. Sample & data collection

More than 150 jurisdictions within all states in the U.S. were individually invited to participate in this research study. Cases were ultimately obtained from nine counties in Arkansas, Illinois, Mississippi, Missouri, Florida, New York, Louisiana, and Hawaii. Each jurisdiction provided a different total number of cases, depending on personnel constraints and availability of in-office files. Cases were collected by a researcher at the District/Prosecuting Attorney offices, and were retrieved by either scanning hardcopy files or downloading (if the file was digitized), and saved on a secure network accessible only by staff and research interns who had a signed Non-Disclosure Agreement.

The case files contained a range of data sources to inform the study, presented in [Table 6](#).

Different case files had varying source materials available. However, the multiplicity of data sources followed the process of forensic assessment, which seeks to corroborate to the degree possible. This contrasts to professional opinions that are otherwise impressionistic in the absence of sufficient evidence. Only cases adjudicated as guilty were included in the sample to enhance confidence that the defendant was indeed responsible for what was alleged.

Incomplete cases were removed from analysis (n = 27). Cases were defined as incomplete if information germane to the Depravity Standard was clearly missing due to a lack of documented information about the crime, or if the documents were so heavily redacted that it

Table 6
Types of case documents reviewed to mine data for study 5 Phase B.

Police and legal records	Medical records	Statements and confessions
Affidavits	Autopsy photos	Defendant statements
Arrest warrants & probably cause	Autopsy reports	Victim/victim impact statements
Booking information	Certificates of death	Witness statements
Case summary/case status reports	Death investigation reports	Character witness statements
Charge report	Defendant medical records	Trial/court transcripts/grand jury testimony
Complaints & indictments	EMT/EMS/paramedic/ambulance reports	
Counts	Forensic evaluation reports	
Crime scene photos	Hospital reports	
Crime scene reports	Psychiatric/psychological reports	
Criminal history transcripts/rap sheets	Victim medical records	
Custody reports		
Diagnostic reports		
Evidence reports		
Felony information		
Felony review screening sheets		
Investigative reports		
Judgements		
Memorandums		
Police reports/incident reports/field reports/arrest reports		
Polygraph reports		
Presentencing reports		
Property records		
Prosecutor letters		
Supplemental reports		
Surveillance footage		

was unclear as to what took place. Cases from these jurisdictions reflected diverse sections of the United States with different ethnic and socioeconomic composition. These jurisdictions included small towns in which homicide was rare and medium and large metropolitan areas (Table 7).

Forty of the 770 murder cases were identified during the review process for exclusion from further analysis. These cases included non-murder felony cases that had been accidentally included, duplicate cases involving co-defendants, or instances when an insanity defense was successfully employed.

Closer scrutiny revealed that capital cases were over-represented in the Arkansas sample. Further investigation highlighted a misunderstanding regarding the request for a random sample of cases covering all types of murder charges. Table 8 presents a breakdown of

Table 8
Breakdown of cases rated by type supplied and included per jurisdiction.

Jurisdiction	Capital murder	Murder 1st	Murder 2nd	Murder/homicide (level not provided)	Other crime type	Duplicate case
Little Rock, AR	102	42	2	2	2 ^a	
Lake County, IL		29		19		
St. Clair County, IL		67	3		1 ^a	
Clay/Lowndes/Oktibbeha Counties, MS	12	7		72	7 ^a	
Jackson County, MO		72	11	10	1 ^a	1 ^a
Jacksonville, FL		22	43	21	9 ^a	
Utica, NY		2	70	11	9 ^a	
Jefferson Parish, LA		44	58	5	1 ^a	
Kaua'i County, HI			4		9 ^a	
TOTAL	114	285	191	140	39 ^a	1 ^a

^a Cases included in error, or where a co-defendant was charged with a non-murder felony, and duplicate cases removed from further study.

Table 7
Murder cases supplied and included per jurisdiction.

Jurisdiction	Cases supplied	Incomplete cases	Total included
Little Rock, AR	150	0	150
Lake County, IL	50	2	48
St. Clair County, IL	71	0	71
Clay/Lowndes/Oktibbeha Counties, MS	100	2	98
Jackson County, MO	100	5	95
Jacksonville, FL	100	5	95
Utica, NY	100	8	92
Jefferson Parish, LA	112	4	108
Kaua'i County, HI	14	1	13
Total	797	27	770

cases.

Non-murder and duplicate cases were removed from review in this study. Results from Little Rock, AR were additionally separated in analyses as they did not reflect a random sample of murder cases.

5.1.3. Procedure

5.1.3.1. Rater training. Participants completed a two-hour training session upon entry into the program. Participants were trained in the descriptions of the items and their extensive qualifying and disqualifying examples, covering the breadth of criminal scenarios and contexts in which a murder or attempted murder may occur. Over a period of several weeks, all participants then completed up to ten training cases and were provided with individualized feedback by an experienced researcher to ensure accurate scoring of the items. When it was determined that participants had a keen understanding of the 25 Depravity Standard items, cases were randomly assigned for rating.

5.1.3.2. Case rating. Participants reviewed each of the randomly assigned cases, applying the definitions and qualifying and disqualifying examples of each of the 25 items under consideration. At least two independent participants rated each case. The process of rating involved completing a chart reviewing each of the 25 items against each of the documents contained within the case, one at a time. For each case document, a rating of present (Yes), absent (No), or insufficient data (ID) was recorded for each of the 25 items, along with an accounting of evidence identified by the rater as the basis for their decision. The ID responses represented either a lack of information such as blurry crime scene photos, or a rater's uncertainty about the case information presented, such as conflicting evidence that would make any determination subjective.

Then, an overall score for each of the items was completed for the case, with an inventory of all of the evidence for the score. See Table 9 for an example of the results for Item 13 "Intentionally targeting victims based upon prejudice" as completed for one of the cases.

Table 9
Example of case rating for Item 13.

	Police records	Medical records	Statements and interviews	Defendant Letter	Defendant Statement	Character Witnesses
Overall Rating: present (YES), absent (NO), or insufficient data (ID)	Assessment, Family Background	Presentence Diagnostic Report	Victim & Witness Statements	Defendant Letter	Defendant Statement	Character Witnesses
Item 13 YES: The perpetrator and his friend selected the campsite as the target because the victims were homosexuals, which the perpetrator determined based on seeing a gay pride (rainbow) flag flying outside the victims' campsite. During the crime, the perpetrator repeatedly shouted "faggots" at the victims, saying "God said to kill you and your kind."	ID, this assessment does not contain enough detail to determine.	Yes, these documents indicated that the defendant harbored prejudice towards homosexuals.	Yes, these documents indicated that the defendant intentionally targeted the victims based on his desire to terrorize and harm them for being homosexuals. Witnesses reported the defendant shouting "faggots" and "God said to kill your kind" to the victims.	ID, does not contain enough detail to determine as the perpetrator did not acknowledge the victims homosexuality in this document.	Yes, this statement indicated that the defendant harbored prejudice towards homosexuals, even though he felt remorseful for harming the victims.	ID, none of the character witnesses expressed that the defendant targeted the victims based on prejudice.

Table 10
Number of items present in cases.

Number of items in random cases (n = 582)	Total cases with the number of items present	Number of items in Arkansas cases	Total cases with the number of items present
0	145 (24.9%)	0	26 (17.6%)
1	124 (21.3%)	1	24 (16.2%)
2	110 (18.9%)	2	16 (10.8%)
3	72 (12.4%)	3	21 (14.2%)
4	42 (7.2%)	4	18 (12.2%)
5	39 (6.7%)	5	11 (7.4%)
6	23 (4.0%)	6	15 (10.1%)
7	7 (1.2%)	7	7 (4.7%)
8	10 (1.7%)	8	4 (2.7%)
9	4 (0.7%)	9	2 (1.4%)
10	2 (0.3%)	10	3 (2.0%)
11	1 (0.2%)	11	1 (0.7%)
12	2 (0.3%)		
14	1 (0.2%)		

Table 11
Frequency of the 25 Depravity Standard items.

Random sample (n = 582)	Arkansas (n = 148)
Item 16 (21.6%)	Item 15 (32.4%)
Item 15 (21.5%)	Item 17 (29.1%)
Item 20 (20.4%)	Item 20 (26.4%)
Item 12 (19.8%)	Item 12 (25.7%)
Item 17 (18.7%)	Item 16 (24.3%)
Item 23 (17.4%)	Item 25 (23.6%)
Item 25 (14.8%)	Item 23 (20.9%)
Item 8 (10.0%)	Item 6 (20.3%)
Item 6 (9.5%)	Item 8 (16.9%)
Item 18 (9.1%)	Item 2 (12.2%)
Item 19 (8.9%)	Item 14 (11.5%)
Item 24 (8.6%)	Item 18 (10.8%)
Item 22 (7.6%)	Item 19 (10.8%)
Item 21 (6.5%)	Item 21 (9.5%)
Item 7 (6.4%)	Item 7 (9.5%)
Item 2 (6.0%)	Item 22 (8.8%)
Item 5 (4.6%)	Item 5 (8.1%)
Item 14 (4.0%)	Item 1 (5.4%)
Item 1 (2.2%)	Item 24 (4.7%)
Item 10 (1.2%)	Item 11 (4.7%)
Item 9 (1.0%)	Item 9 (3.4%)
Item 11 (1.0%)	Item 13 (1.4%)
Item 13 (0.7%)	Item 10 (0.7%)
Item 3 (0.5%)	Item 3 (0.7%)
Item 4 (0.5%)	Item 4 (0.0%)

The full template contains a row for all 25 items, and each case would have two full templates completed by two individual raters. The overall results of Yes, No, or ID for each rater were entered into an excel spreadsheet for comparison.

These templates were reviewed by an experienced Research Assistant to form a consensus rating for each of the 25 items per case. Any discrepancies in ratings between participants were reviewed in depth. If a discrepancy was present due to a clear misunderstanding of an item, additional training was provided to the rater. If a discrepancy identified an opportunity to refine an item's qualifying and disqualifying examples to be clearer, the item was reviewed as part of Study 3.

5.1.3.3. Debriefings. Participants were debriefed regularly to ensure adequate supervision, to promote discussion about the cases internally, to confirm that participants were undertaking the exercise properly and with the requisite commitment, and to make sure the nuances of instruction were being followed in a uniform fashion. Debriefing sessions also ensured counseling was available to those for whom the murder case material was especially impactful. Participants were debriefed at least once per week, however they were also aware that

Table 12
Percentiles of depravity scores.

25th	30th	35th	40th	45th	50th	55th	60th	65th	70th	75th	80th	85th	90th	95th	99th +
41.28	65.26	66.25	78.80	84.66	125.76	136.56	149.87	165.14	209.28	224.73	280.96	330.14	370.90	464.86	716.15 +

they could request a meeting at any time with either a Research Assistant or the Research Director for any reason. Participants were also encouraged to switch cases at any time, with no questions asked; however, this only occurred twice.

5.2. Results

After removing non-murder cases and one duplicate case, analyses were performed on the random sample of 582 cases, and the 148 cases from Arkansas separately. The greatest number of cases in each group had no elements of depravity present (random sample of cases, 24.9%; Arkansas 17.6%). The case that presented with the highest number demonstrated 14 items of depravity. Table 10 lists the number of items present in each of the cases.

Table 11 lists the percentages of frequency with which each item occurs in the overall sample across all jurisdictions that provided a random sample of murder cases (excluding Arkansas), as well as the Arkansas cases separately. All items of the Depravity Standard were represented rarely to infrequently. Item frequencies for the random sample of cases range from 0.5% to 21.6% in the total sample of random cases, and 0% to 32.4% in the Arkansas sample. The results of this study further demonstrate the utility of the Depravity Standard in creating a narrowed class of the worst of homicide offenders. This narrowing holds true even among the predominantly capital cases from Arkansas.

6. Study 5 Phase C: scoring cases with the Depravity Standard

In order to afford a practical everyday application of the validated Depravity Standard – Murder Evidence Guide to cases in the justice system, we devised a percentile scoring system. This method ensured the variable weight of different items would directly reflect the input of the general public, as well as the frequency with which items appeared in the murder case sample.

6.1. Method

Cases rated in Study 5 Phase B, excluding Arkansas, allowed for analyses of 582 murder cases from jurisdictions across the U.S. The overall percentages scored for each of the 25 items by participants 18 years and older in the U.S. were merged with the 582 cases to determine overall scores for each of the cases. The public survey data weighted by ethnicity was the final sample used.

For example, Item 1 has a score of 77.41. If this was the only item present in a case, this would be the overall Depravity Score for that case. If a case had two items present, for example Item 7 (69.42) and Item 13 (64.40), that case would have an overall Depravity Score that is the sum of those two items, or 133.82.

Based on these Depravity Scores, we are able to ascertain overall percentiles, to mark individual Depravity Scores for a case against other murders. The percentiles are demonstrated below in Table 12.

Therefore, if a hypothetical case had three of the more depraved items present such as Items 5 (80.53), 14 (87.65), and 17 (84.66), the Depravity Score for that case would be 252.84. Based on the percentiles, this case would be more depraved than over 75% of murders in the U.S. If, however, a case was present with items that the survey participants scored for being less depraved, four or more items would need to be present to reflect a 75th percentile. For example, if Items 10 (56.98), 11 (60.23), 18 (63.14), and 20 (55.04) were present in a case, the Depravity Score would

be 235.39. These four items in combination would still total less than the three items present in the first case. The scoring system directly reflects the impact of societal input, embedded in the experience of murder cases. In this fashion, the Depravity Standard scoring guide assists triers of fact who are necessarily inexperienced.

This merging of survey data with case analysis to devise percentiles allows for a quick and easy comparison of one murder against others in the same category. Our data demonstrates that cases of nine items or more are so seldom found (even in a more selected group such as the Arkansas sample) as to demonstrate a truly unusually rare degree of depravity beyond that point. With a 99+ percentile demonstrated at 716, scores beyond this measure are so extreme to make comparison beside the point.

Nevertheless, scores in lower ranges are influenced by the specific items present. The specific nature of intent, action, attitude, and victim choice in these crimes may vividly illustrate significant differences in depravity between crimes, and within the same crime among co-defendants with different roles, different actions, and different reactions.

7. Discussion

The Depravity Standard is a guide to aid the trier of fact in assessing evidence pertinent to sentencing decisions for the worst of crimes. Apart from an expansive effort to incorporate public attitudes, and to account for demographic differences, the research has used a large number of highly resourced and adjudicated guilty murder cases to establish content validity. The researchers were fortunate enough to have been provided access to these hundreds of pertinent cases from diverse jurisdictions that have their own unique social contributors to the makeup of murder cases in their area. The Depravity Standard research has incorporated much larger numbers of public participants and case files, and source materials within those case files, compared to the volume of data historically relied upon in justice research. As an evidence-based inventory of specific components of criminal heinousness, the Depravity Standard enables court participants to identify evidence of whether a crime, or even a particular perpetrator among co-defendants in a crime, reflects depravity and to what degree.

The Depravity Standard promotes investigative diligence beyond the determination of guilt. Prosecuting authorities need no longer assert heinous charges without specific guidance as to whether a crime demonstrates evidence of heinousness. This reduces the likelihood that defendants are overcharged for bias or other forces that impact justice, such as political, personal, or media considerations that have little to do with what actually transpired in a case. Defense and prosecution are on equal footing to ascertain and to establish the presence or absence of evidence for 25 specific examples of intent, actions, attitudes, and victim choice. The definitions of each of these items are very detailed, with clear inclusion and exclusion criteria. Study 4 Phase A demonstrated that each of these items has high inter-rater reliability. The guidance provided by the Depravity Standard therefore encourages more complete investigation prior to trial.

The impact of bias is diminished by a process that studies evidence spanning the arc of the crime – before, during, and after. The current approach to determining depravity, with no guidance, may prompt the trier of fact into visceral reactions that cannot protect against bias. An operationalized system organized around well-demarked aspects of what occurred protects fact-finding from being distracted by who the offender is, or the personalities or theatrics of the trial.

Judges and juries with access to such information could be far better informed to assess the depravity of a murder relative to other murders,

and to better compensate for their lack of experience in formally assessing heinous acts. Qualitative and quantitative improvements in evidence gathering will prevent injustices that can arise from limited investigations that seek only to resolve the identity of a perpetrator. The literature on wrongful convictions critiques the consequences, for example, of investigations that go no further than merely achieving evidence thought sufficient to gain an indictment and conviction (Findley, 2002; Garrett, 2012). Closer scrutiny of the more granular elements of a crime, such as motivation and exact nature of actions, may prove to be the first indication that evidence for guilt is not so definitive as originally thought. Lowering the risk of injustice at the point of charging an offense is an additional byproduct of the increased scrutiny that the Depravity Standard promotes.

We acknowledge certain limitations of the research to date. First, minority participants in Study 5, were underrepresented. Weighting our responses to compensate for this underrepresentation did not demonstrate a significant difference between the weighted and unweighted data. Study 5 surveys of the general public continue, and we will continue to recruit more minority participants and to seek a data pool that is increasingly representative of each demographic in the national population.

Much higher participant numbers in our continued online surveys, as more people become aware of this research and its benefits to fairness in the justice system, may change the understandings within subgroups. The Depravity Standard has been devised as a living, breathing instrument that can evolve with societal attitudes over time. We will reassess the statistical values based on public survey responses at a later date. Should there be an evolution of societal attitudes relative to any of the items of the Depravity Standard, that will be reflected in the weighting of that item in future iterations of the Depravity Standard.

The studies presented here reflect a review of murder offenses. Cases that are more homogeneous within murder – such as gang-activity murders, or domestic relations murders – may indicate differences within their sub-niches to be incorporated in later versions of the Depravity Standard. We now have the methodological framework to drill down further on these subcategories be they revenge killing, terrorism cases, or other subtypes. The Depravity Standard is a crime magnification tool. We encourage the scientific and legal community to take this methodology and to scrutinize subtypes of crime to a higher informative power. The byproducts will enhance fairness in justice in the same manner that higher magnification enhances a range of different disciplines.

8. Applications

The qualitative and quantitative guidance about homicide from the Depravity Standard informs determinations of depravity prior to trial and at criminal sentencing. However, the Depravity Standard also informs early-release decision-making. Parole and corrections officials confront difficult choices in an era of prison overcrowding and mandates to release large numbers of inmates, including those with previous murder convictions. Elected officials considering pardons or commutations likewise deliberate unguided.

Murder convictions inspire tremendous emotion. For some cases, that emotion fades with the newsprint of years gone by. For others, deaths of surviving family and changing of the involved attorneys over time erode the intensity of passions held by those abruptly forced to grieve or who lived the early investigation and discovery of the crime and its facts. Advocacy for and against leniency in release is impacted by these realities that have nothing to do with the facts of the intent, actions, victim choice, or attitude.

In other cases, constituencies maintain a vested interest in a case as a symbol or cause. The tenacity of an advocate or the influence of one's counsel may have outsized impact on release decisions. The Depravity Standard is a guide that does not fade with time or changes in assigned counsel, or the politics of a case. The evidence before, during, and after a case is preserved for its depravity or for its unremarkable quality

relative to other murders through the use of the Depravity Standard. The case's players age, but its evidence does not, thus enhancing data integrity.

The Depravity Standard informs early release decisions with guidance about a homicide's degree of depravity relative to other murders, and aids the decision of whether and when leniency is warranted. Evidence-based release decisions limit bias, promote fairness, and enhance public trust in justice. Elected officials who utilize the Depravity Standard to inform pardon decisions demonstrate evidence-based decision-making rather than cronyism or other qualities that undermine public confidence in justice.

The specificity and evidence-focus of the Depravity Standard brings objectivity and fairness to a murky world of early-release determinations. This research, in demonstrating how cases vary in severity, informs a future provocative scenario: Two murderers are being considered for early release. One has been incarcerated for fifteen years. One has been incarcerated for ten years. Both are roughly the same age, and have identical risk profiles for re-offense in the community. Which murderer is more appropriate for early release? Conventional wisdom would suggest that the murderer who was incarcerated for fifteen years would be more appropriate for release.

Now consider that the person in prison for fifteen years committed a murder whose depravity was in the 75th percentile relative to other murders. The person in prison for ten years committed a murder whose depravity was in the 40th percentile. Who is more deserving of early release? We are satisfied that the injection of evidence on both what a person intended to do and what a person did will lead responsible parties to more fair and just consideration of such cases.

By identifying and refining items for inclusion in the Depravity Standard, we are delineating the meaning of depravity as a byproduct of court-decisions, public survey, a range of forensic scientist input, and case sampling. In this manner, the Depravity Standard will serve not only as a guide, but also as a reference point. However, sentencing decisions are made by judges and juries and not by the Depravity Standard. The Depravity Standard serves only as a guide to those deliberating sentences as to examples of depravity in a given crime, as compared to other similarly charged crimes, and how significant these aspects of depravity are when present. There is otherwise no validated inventory or methodology to inform the elements of depravity in murder.

The Depravity Standard assists decision-makers in the spirit of the United States Supreme Court decisions of *Gregg*, *Godfrey*, and *Walton*. It informs otherwise inexperienced jurors as well as judges who otherwise have no validated guidance. Furthermore, the Depravity Standard relies in part on public input to establish societal standards and narrows a class of the worst of homicide offenders. The Depravity Standard focuses its inquiry on intents, actions, victim choice, and attitudes of a crime in order to minimize reliance on otherwise subjective judgments and presumptions in reaching sentencing decisions.

Assessment of what depraved factors are present in a crime can be separated from potential biases relating to who perpetrated the crime or why it was carried out. Courts have avenues for admissibility of character evidence, and the Depravity Standard does not negate the importance of such evidence. What the Depravity Standard does, however, is enable the before, during, and aftermath of crime to be carefully scrutinized with the focus and magnification needed to distinguish the severity of a crime by its unique features.

The presence or absence of an item can be reliably scored by a trained legal, investigative, forensic science, or corrections professional. This enables multiple disciplines to utilize the Depravity Standard as a valuable guide, often without the potentially cumbersome procedures of requiring an expert witness to interpret fundamental and straightforward evidence and to undertake justice decision-making.

Scoring the Depravity Standard is straightforward and integrates data derived from public survey with data from adjudicated case files. Public input informs the weight assigned to each item. At trial and

sentencing, prosecutors would present evidence that a particular item is present, and defense attorneys would counter with evidence that this same item was not present or could not be proven. Judges and jurors would then determine what items were present. The numerical weights of each of the items present would be added together. The sum of the weights of these items would then be compared to data from all of the murder cases analyzed, to reflect a percentile figure. A jury would then be able to specifically appraise any murder, and the role of any defendant within that murder, in the broader context of other murders and be able to confidently determine whether it is worse than most other murders, or not. This process for scoring ensures the distinction of the worst of crimes by a fair and informed decision-maker, guided in the face of inexperience. The Depravity Standard is a mechanism that narrows a class of the worst of offenders, but does so without limiting the myriad of ways in which one can carry out an exceptional crime.

The Depravity Standard is also being validated for application to sex crimes like rape and sexual assaults, violent crimes such as battery and assault, and even non-violent crimes such as property offenses, arson, fraud, and embezzlement. The next phase of the research will involve development of the training package, and the release of the Depravity Standard - Murder Evidence Guide and scoring pack to the justice system.

This development of an evidence-driven guide to distinguishing murder by severity has potential implication for other distinctions in the law, apart from heretofore undefined “heinous,” “depraved” and related terminology. In a number of states, laws distinguish “Aggravated Murder” from “Murder.” In New York, a conviction of first degree Murder results in a sentence of 20 to 25 years in prison; a conviction for Aggravated Murder results in a sentence of life in prison without parole. In Washington state, conviction for First Degree Murder carries a sentence of life in prison, but conviction for Aggravated Murder carries a sentence of life without possibility of parole.

What makes a murder aggravated? For those states that have such distinctions, the features may vary. In Ohio, these include having intentionally killed someone under age 13, or targeted a police officer. In Oregon, committing or soliciting murder-for-hire elevates to Aggravated Murder, as well as murder that occurs in the course of intentionally maiming or torturing the victim. Utah's Aggravated Murder statute includes murder that is “especially heinous.” The variability in what different states denote to be Aggravated Murder may reflect sensitivities unique to that state's history. Our research, however, has demonstrated that survey respondents from very different states showed little difference in how they weighed each of the items of the Depravity Standard. The Depravity Standard provides evidence-driven guidance to the distinction of Aggravated Murder from First Degree Murder.

The methodology of the Depravity Standard is adaptable to justice systems outside of America and to war crimes tribunals. The approach of item development based on earlier higher court decisions, and online survey of public attitudes of host countries can be replicated anywhere – and for any class of crime. Researchers in this project have been encouraged by dialogue with colleagues from countries in South America, Asia, and the Middle East seeking to develop a Depravity Standard adaptable to their own cultural context. Our findings demonstrate the healthy coexistence between evidence-driven justice and a cultural context that incorporates the many faces of diversity inherent to every country.

War crimes investigations are complicated in many unique respects. Justice in war crimes investigations is impeded by numerous considerations. The Depravity Standard is a very detailed guide that emphasizes investigative rigor of case evidence and the intent and attitudes bookending it. In the context of war crimes, evidence may be readily available from witnesses that include children, leaders, planners, and those involved in logistics. Political entities and vested economic interests that sweep justice under the rug, as well as propagandists who instigate disingenuous outcry, are cut out of influence on

justice.

In that regard, the Depravity Standard holds promise for many countries whose citizens do not enjoy the protections of the American justice system nor its freedom to refine itself of its shortcomings. The Depravity Standard is therefore a significant advance in the potential for justice to aid in human rights globally and in the evidence-based application of justice in world tribunals.

9. Conclusion

The development of the Depravity Standard will meet a long-standing need within the American criminal justice and corrections systems. Juries bearing the burden of deliberating “heinous, atrocious, and cruel” elements of a crime will be able to specifically consider the presence or absence of relevant evidence informing specific items of the Depravity Standard, as opposed to reaching into one's gut with only the guidance of attorneys' ability to argue. Parole and corrections officials will also benefit from guidance that minimizes bias, undue external influence, or corruption. The reliability and validation studies published here demonstrate how this non-denominational guide promotes fairness and a reliance on facts.

The definitions were also formulated and refined in order that an item, when scored present, was uncommon or even rare relative to the overall sample of its crime cohort, and the degree of depravity for each item was informed by public opinion. This priority aimed at ensuring the Depravity Standard, even when denoting elements of depravity to be present, highlights a narrowed subset of crime that society at large would consider depraved. A public that is served by evidence-based measures that inform critical justice interests is one that engages its legal system with more investment in justice.

Acknowledgements

Thank you to Victor Weedn, M.D., J.D., Ph.D., Robert Cancro, M.D., Research Assistants Thomas Ludorf, Welyn Craig, and all interns who provided support on these Depravity Standard research phases.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.
- Ball, W. D. (2009). Heinous, atrocious, and cruel: Apprendi, indeterminate sentencing, and the meaning of punishment. *Columbia Law Review*, 109(5), 893–972.
- Berghuis v. Smith*, 559 U.S. 314 (2010).
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality and Individual Differences*, 54, 106–148.
- Brown v. Plata*, 563 US 492 (2011).
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- IBM Corp (2013). *IBM SPSS statistics for windows, version 22.0*. Armonk, NY: IBM Corp.
- Cunningham v. California*, 549 U.S. 270 (2007).
- Dixon v. Ryan* (2016 U.S. Dist. LEXIS 33999).
- Duren v. Missouri*, 439 U.S. 357 (1979).
- Findley, K. A. (2002). Learning from our mistakes: A criminal justice commission to study wrongful convictions. *California Western Law Review*, 38(2), 333–353.
- Fleiss, J., Levin, B., & Paik, M. (2003). *Statistical methods for rates & proportions* (3rd ed.). New York: Wiley & Sons.
- Garrett, B. L. (2012). Introduction: *New England Law Review* symposium on “convicting the innocent”. Paper symposium. *New England Law Review*, 46, 671–687.
- Godfrey v. Georgia*, 446 U.S. 420 (1980).
- Gregg v. Georgia*, 428 U.S. 153 (1976).
- Gwet, K. (2002). Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity. *Statistical Methods for Inter-Rater Reliability Assessment Series*, 2, 1–9.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1), 29–48.
- Haahr, M. (2006). Random.org: True random number service. Web Resource, available at <http://www.random.org>.
- Kaiser Family Foundation (2016a). Population distribution by race/ethnicity 2015.

- Downloaded from <http://kff.org/other/state-indicator/distribution-by-raceethnicity/>.
- Kaiser Family Foundation (2016b). Population distribution by gender 2015. Downloaded from <http://kff.org/other/state-indicator/distribution-by-gender/>.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Norris v. State*, 793 So. 2d 847 (Ala. Crim. App. 1999).
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Piedmont, R. L. (2014). Inter-item correlations. In A. C. Michalos (Ed.). *Encyclopedia of quality of life and well-being research* (pp. 3303–3304). Dordrecht: Springer.
- Walton v. Arizona*, 497 U.S. 639 (1990).
- Welner, M. (2003). Legal relevance demands that evil be defined and standardized. *Journal of the American Academy of Psychiatry and Law*, 31, 417–421.
- Welner, M. (2009). The justice and therapeutic promise of science-based research on criminal evil. *Journal of the American Academy of Psychiatry and Law*, 37(4), 442–449.
- Welner, M. (2013). Classifying crimes by severity: From aggravators to depravity. In J. E. Douglas, A. W. Burgess, A. G. Burgess, & R. K. Ressler (Eds.). *Crime classification manual* (pp. 55–72). (3rd ed.). San Francisco: John Wiley & Sons.
- Welner, M., O'Malley, K. Y., Gonidakis, J., Saxena, A., & Burnes, J. (2018). The Depravity Standard II: Developing a measure of the worst of crimes. *Journal of Criminal Justice* (in press).
- Yoon, J. S., & Somers, C. L. (2003). Aggressive content of high school students' TV viewing. *Psychological Reports*, 93, 949–953.