Articles

Patriarchy, Not Hierarchy: Rethinking the Effect of Cultural Attitudes in Acquaintance Rape Cases

Eric R. Carpenter*

Do certain people view acquaintance rape cases in ways that favor the man? The answer to that question is important. If certain people do, and those people form a disproportionately large percentage of the people in the institutions that process these cases, then those institutions may process these cases in ways that favor the man. In 2010, Dan Kahan published Culture, Cognition, and Consent, *a study on how people evaluate a dorm room rape scenario. He found that those who endorsed a stratified, hierarchical social order were more likely to find that the man should not be found guilty of rape.[1]*

If Kahan is right, radical change may be necessary. The institutions responsible for handling sexual assault complaints—law enforcement, the military, and university and college administrations—are stratified and hierarchical, and are likely over-populated by people who are attracted to hierarchical institutions and who hold hierarchical worldviews. These institutions may need to be overhauled—or even replaced. However, the study has a serious methodological flaw: It uses the Hierarchy-Egalitarianism Scale to measure those hierarchical worldviews, and as this Article demonstrates, this scale has reliability and validity issues.

This Article then applies a different methodology to the underlying data and shows that patriarchy, not hierarchy, explains the differences in guilt perceptions. This more accurate understanding of Kahan's data carries important policy implications. Rather than radical change, targeted training that addresses inaccurate rape beliefs may be enough to ensure accurate processing of these cases.

---

    1. Dan M. Kahan, *Culture, Cognition, and Consent: Who Perceives What, and Why, in Acquaintance-Rape Cases*, 158 U. PA. L. REV. 729 (2010) [hereinafter *Culture, Cognition, and Consent*].

[225]

TABLE OF CONTENTS

INTRODUCTION

Do certain people view acquaintance rape cases in ways that favor men? The answer to that question is important. If certain people do, and those people form a disproportionately large percentage of the people in the institutions responsible for handling sexual assault cases, then those institutions might process these cases in ways that favor men.

And if what characterizes this group is a deeply-held belief system, then radical change may be necessary. Trying to get someone to change deeply-held beliefs would be akin to trying to change someone from conservative to liberal, or from Christian to Jew. The institutions themselves may need to be overhauled—or even replaced.

In 2010, Dan Kahan published *Culture, Cognition, and Consent,*[2] which includes an important finding: Those who endorse a stratified, hierarchical social order were more likely to side with the man and find him not guilty in the context of a dorm room rape scenario.[3] This finding informs the problem described above. The institutions responsible for handling sexual assault complaints—law enforcement, the military, and university and college administrations—are stratified and hierarchical,

---

2. *Id.*
3. *Id.* at 733–35.

and are likely populated by people who are attracted to hierarchical institutions and who hold hierarchical worldviews. If Kahan is right, radical change may be necessary.

Unfortunately, his study may have a serious methodological problem. The study used the Hierarchy-Egalitarianism Scale ("H-ES"), which was designed by the Cultural Cognition Project ("CCP")[4] to measure those hierarchical worldviews. This Article shows that the scale has reliability issues (because of ambiguous and complex items) and validity issues (because it may not be measuring hierarchy). As a result of these problems, the link between hierarchical worldviews and the perception of guilt in sexual assault cases may not be accurate.

Regardless, Kahan's underlying data remains valuable. Most research on sexual assault is conducted on a small population: college students. Here, Kahan gathered data from a large sample that is representative of the general population. Further, the study uses a realistic hypothetical of a dorm room sexual assault, which is the type of case that is at the center of the sexual assault controversy.

This Article uses a different analysis of the items used in the CCP's scale to show that patriarchy, rather than hierarchy, predicts how people view sexual assault cases. This more accurate understanding of Kahan's data carries important policy implications. If beliefs about patriarchy account for the variation in guilt judgments in these cases, then it may not be necessary to radically restructure institutions. There may be nothing inherently wrong with hierarchical organizations handling sexual assault cases. Instead, targeted training related to these narrower belief systems about gender roles may be sufficient to ensure more accurate processing of these cases.

Part I of this Article reviews Kahan's 2010 study. Part II then analyzes the H-ES. In Part III, this Article offers Kahan's argument found in *Culture, Cognition, and Consent* that the H-ES is unidimensional and valid. Part IV conducts exploratory factor analysis of the H-ES that suggests that the scale is multidimensional. Part V then conducts structural equation modeling with a theoretical model that has gender and race facets and where the gender facet predicts the guilt variable while the race facet does not. This Part also shows that the CCP's single construct model has a poor fit to the data. In Part VI, this Article discusses the impact of

---

4. The CCP is "a group of scholars interested in studying how cultural values shape public risk perceptions and related policy beliefs." *Cultural Cognition Project*, YALE LAW SCH., http://www.culturalcognition.net (last visited Jan. 16, 2017). Dan Kahan is a member of this group. The scale that will be analyzed in this Article was developed by this group rather than by a single person. Various members of this group, separately and jointly, have published articles using this study. For clarity, the main text will attribute the development of the H-ES to the CCP and the footnotes will cite to the particular authors.

these findings on other CCP studies and suggests improvements to the scale. The Article concludes with a discussion of policy implications.

I. Kahan's Study and Findings

In 2009, Kahan conducted an experiment to see whether people who subscribe to certain worldviews perceive rape cases differently than people with contrasting worldviews. The respondents (n = 1500) were given a vignette of a sexual assault involving Lucy and Dave, two college students and casual acquaintances. One day, when Lucy was looking for her boyfriend in the dorms, she stopped by Dave's room to see his roommate. She had consumed an alcoholic drink prior to the visit. She went into the dorm room but the roommate was not there; however, Dave was. At this point, Lucy claims that she tried to leave but Dave blocked the door, pinned her down, and sexually assaulted her by inserting his penis into her vagina. Dave claims that she consented. Both agree that during the event, Lucy said "no" repeatedly—although Dave claims that she said it in a sexual way—and both agree that Lucy did not otherwise physically resist.[5]

Kahan randomly divided the subjects into five groups of 300 and gave each of them one of five legal conditions (either no law to apply, or four different versions of law to apply) which the respondents would use to evaluate the facts in the case.[6] Kahan measured thirteen dependent variables, including the outcome judgment of guilt ("Dave should be found guilty of rape").[7]

For use as independent variables, Kahan gathered demographic information from the respondents and also administered the H-ES.[8] He coded respondents as "hierarchs" if the respondent scored in the top third of the scale and as "egalitarians" if the respondent scored in the bottom third.[9] He then used "hierarch" and "egalitarian" as predictor variables.[10]

Kahan found that hierarchs were more likely than egalitarians to side with the man and find him not guilty of rape.[11] He also found that gender only had a meaningful effect when joined with the cultural worldview,[12]

---

5. *Culture, Cognition, and Consent, supra* note 1, at 737, 807–09. This fact pattern is based on *Commonwealth v. Berkowitz*, 641 A.2d 1161 (Pa. 1994).

6. *Culture, Cognition, and Consent, supra* note 1, at 767–68.

7. *Id.* at 769–71.

8. *Id.* at 765. The CCP also administered another scale, the Individualism-Communitarian Scale ("I-CS"). *See infra* Part II.B.

9. *Culture, Cognition, and Consent, supra* note 1, at 776.

10. *Id.* at 776 fig.3, 777 fig.4, 780 fig.5, 785 fig.7, 786 fig.8, 792 fig.10.

11. *Id.* at 793–94.

12. *Id.* at 782.

where hierarchical women were actually the most predisposed to side with the man.[13]

This study is not inconsistent with other research, but it does point to a more global variable as the explanatory variable. Other research has pointed to a more discrete variable: traditional, hierarchical gender role beliefs.[14] Those with traditional gender role beliefs tend to endorse certain inaccurate rape schemas more than those with nontraditional gender role beliefs. These beliefs are then associated with more discrete beliefs about rape, and ultimately with the outcome judgments in rape hypotheticals.[15] Studies have also found that acceptance of these inaccurate rape schemas is associated with siding with the man in the ultimate normative judgment about blame.[16]

However, Kahan's findings are only valid if the H-ES is itself valid and reliable. If the H-ES actually measures something else or does not measure anything at all, then those findings may be inaccurate and policy makers should not rely on them when making decisions. The next Part of this Article addresses the issue of whether the H-ES is a valid and reliable scale.

---

13. *Id.* at 787.

14. The traditional gender role construct has many potential facets, including the following beliefs: that the man should be in charge of the family unit; that women should remain at home rather than work outside the home; that men should pursue women while women should be passive; and that women should behave in sexually conservative ways. *See* Eric R. Carpenter, *The Military's Sexual Assault Blind Spot*, 21 WASH. & LEE J. C.R. & SOC. JUST. 383, 391–92 (2015). Of these facets, it is likely that expectations about sexual conservatism (particularly, that women should be lady-like) is the facet that plays the most central role in rape case processing. *Id.* at 394.

15. *See* Dominic Abrams et al., *Perceptions of Stranger and Acquaintance Rape: The Role of Benevolent and Hostile Sexism in Victim Blame and Rape Proclivity*, 84 J. PERSONALITY & SOC. PSYCHOL. 111 (2003); Kathryn B. Anderson et al., *Individual Differences and Attitudes Toward Rape: A Meta-Analytic Review*, 23 PERSONALITY & SOC. PSYCHOL. BULL. 295, 312 (1997); Gordon B. Forbes et al., *First- and Second-Generation Measures of Sexism, Rape Myths and Related Beliefs, and Hostility Toward Women*, 10 VIOLENCE AGAINST WOMEN 236, 250 (2004); Barbara E. Johnson et al., *Rape Myth Acceptance and Sociodemographic Characteristics: A Multidimensional Analysis*, 36 SEX ROLES 693, 704 (1997); Laura L. King & Jennifer J. Roberts, *Traditional Gender Role and Rape Myth Acceptance: From the Countryside to the Big City*, 21 WOMEN & CRIM. JUST. 1, 9, 12 (2011); Eliana Suarez & Tahany M. Gadalla, *Stop Blaming the Victim: A Meta-Analysis on Rape Myths*, 25 J. INTERPERSONAL VIOLENCE 2010, 2022 (2010); Lynda A. Szymanski et al., *Gender Role and Attitudes Toward Rape in Male and Female College Students*, 29 SEX ROLES 37 (1993); G. Tendayi Viki & Dominic Abrams, Brief Report, *But She Was Unfaithful: Benevolent Sexism and Reactions to Rape Victims Who Violate Traditional Gender Role Expectations*, 47 SEX ROLES 289 (2002); Rosanne Proite et al., *Gender, Sex-Role Stereotypes, and the Attribution of Responsibility for Date and Acquaintance Rape*, 34 J. C. STUDENT DEV. 411 (1993); Niwako Yamawaki, *Rape Perception and the Function of Ambivalent Sexism and Gender-Role Traditionality*, 22 J. INTERPERSONAL VIOLENCE 411, 413–18 (2007).

16. *See, e.g.*, Barbara Krahé, *Social Psychological Issues in the Study of Rape*, 2 EUR. REV. SOC. PSYCHOL. 279, 293 (1991); G. Tendayi Viki et al., *Evaluating Stranger and Acquaintance Rape: The Role of Benevolent Sexism in Perpetrator Blame and Recommended Sentence Length*, 28 L. & HUM. BEHAV. 295 (2004); Charlene L. Muehlenhard, *Misinterpreted Dating Behaviors and the Risk of Date Rape*, 6 J. SOC. & CLINICAL PSYCHOL. 20 (1988).

II. Analysis of the Hierarchy-Egalitarianism Scale

A. Overview of Scale Development

Researchers generally follow certain steps when developing scales and these steps provide a useful framework for evaluating the H-ES. The scale development steps are as follows: (1) clearly define the construct; (2) "generate the item pool"; (3) "determine the format for measurement"; (4) "have the item pool reviewed by a panel of experts"; (5) "consider inclusion of validation items"; (6) "administer items to a development sample"; (7) "evaluate the items"; and (8) decide on scale length.[17] When followed, these steps help to ensure that the resulting scale is reliable and valid.

"Reliability is the extent to which it is possible to replicate a measurement, reproducing the same value (regardless of whether it is the right one) on the same standard for the same subject at the same time."[18] As Lee Epstein and Gary King explain this concept, "[i]f any one of us stepped on the same bathroom scale one hundred times in a row, and if the scale were working reliably, it would give us the same weight one hundred times in a row—even if that weight is not accurate."[19]

Researchers who are studying a latent, unobservable variable often use a scale that consists of several questions that can measure discrete items that are related to that unobservable variable, and then use a score based on those items (called a factor score) to represent the value of that unobservable variable. There, "reliability concerns how much a variable influences a set of items"[20] and "[t]he more the score we obtain from a scale represents the true score of the [unobserved] variable and the less it reflects other extraneous factors, the more reliable our scale is."[21] Validity, on the other hand, "is the extent to which a reliable measure reflects the underlying concept being measured."[22] Returning to Epstein and King's scale example, "[i]f one's true weight is 150 and the scale, even one hundred times in a row, reports 125, we would not think much of that scale."[23] That scale would be reliable but not valid.

In this study, the specific issue is whether the items in the H-ES measure the CCP's targeted global construct, or whether instead the items measure another construct, like patriarchy, or even no construct at all. This

---

17. Robert F. DeVellis, Scale Development: Theory and Applications 73–110 (Leonard Bickman & Debra J. Rog eds., 3d ed. 2012); *see* Lee Anna Clark & David Watson, *Constructing Validity: Basic Issues in Objective Scale Development*, 7 Psychol. Assessment 309 (1995).

18. Lee Epstein & Gary King, *The Rules of Inference*, 69 U. Chi. L. Rev. 1, 83 (2002).

19. *Id.*

20. DeVellis, *supra* note 17, at 59.

21. *Id.* at 31.

22. Epstein & King, *supra* note 18, at 87.

23. *Id.*

is an issue of construct validity, which "is the degree to which an assessment instrument measures the targeted construct."[24] A component of construct validity is content validity. Content validity is "the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose."[25]

Steps one, two, and four address content validity: Scale developers clearly define the construct, generate an item pool, and then have that item pool reviewed by experts. The CCP has not published a detailed article on the H-ES's psychometric properties. Therefore, the following information comes from various sources.

B.   THE CCP'S CONSTRUCT DEFINITION: "GRID"

Because "[c]ontent validity is intimately linked to the definition of the construct being examined,"[26] we need to identify precisely what the CCP was trying to measure. The CCP's construct is derived from the cultural theory of risk developed by Mary Douglas, Michael Thompson, Aaron Wildavsky, and Karl Dake.[27] The cultural theory of risk proposes, "individuals select certain risks for attention and disregard others in a way that reflects and reinforces the particular worldviews to which they adhere."[28]

The theory advances a "grid/group" taxonomy. The "grid" dimension captures group regulation, social prescription, and structured social order.[29] The dimension "runs from minimum to maximum regulation,"[30] where "[i]n a high-grid environment, everything is classified and individual choice is heavily restricted."[31] There, structured social roles regulate the actions between individuals. Low-grid environments lack structure, and "individuals are increasingly expected to negotiate their own relationships with others."[32]

---

24. Stephen N. Haynes et al., *Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods*, 7 PSYCHOL. ASSESSMENT 238, 239 (1995).

25. *Id.* at 238, 239. *See* DEVELLIS, *supra* note 17, at 64 (construct validity is "the extent to which a measure 'behaves' the way that the construct it purports to measure should behave with regard to established measures of other constructs."). DeVellis contrasts this to criterion validity, which is the ability of a scale to predict relationships among variables. *Id.* at 61–62.

26. DEVELLIS, *supra* note 17, at 60; *see also* Clark & Watson, *supra* note 17, at 310.

27. Dan M. Kahan & Donald Braman, *Cultural Cognition and Public Policy*, 24 YALE L. & POL'Y REV. 149, 152–53 (2006) [hereinafter *Cultural Cognition and Public Policy*].

28. *Id.* at 154.

29. Mary Douglas, *Being Fair to Hierarchists*, 151 U. PA. L. REV. 1349, 1353 (2003).

30. *Id.* at 1352–55.

31. *Id.* at 1352.

32. MICHAEL THOMPSON ET AL., CULTURAL THEORY: POLITICAL CULTURE SERIES 6 (1990).

The "group" dimension captures group integration, identity, and the general boundary around that community:[33] it is "the amount of moral pressure to conform that a community puts on its members."[34] The higher the "group" value, "the tighter the control over admission into the group and the higher the boundaries separating members from nonmembers."[35]

FIGURE 1: DOUGLAS TYPOLOGY
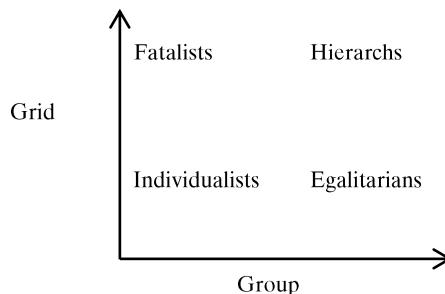
Four cultures coalesce at the extremes of those dimensions: hierarchs, egalitarians, individualists, and fatalists.[36] The "prime virtue" for hierarchs is order; for egalitarians, justice; and for individualists, liberty.[37] Fatalists are drop-outs or cast-outs, politically inactive and apathetic.[38] These four cultures are competitive and need their adversarial relationships to help define their own legitimacy.[39]

The basis for the taxonomy is "[t]he competing cultures of hierarchism and individualism."[40] They sit on the opposite, extreme parts of the taxonomy, where hierarchs are high-group and high-grid, and individualists are low-group and low-grid.[41] Hierarchs and egalitarians both have strong group boundaries, but the two cultures differ in that hierarchs also have strong internal boundaries and classifications while egalitarians

---

33. Karl Dake, *Myths of Nature: Culture and the Social Construction of Risk*, 48 J. Soc. Issues 21, 28 (1992); Douglas, *supra* note 29, at 1355.

34. Douglas, *supra* note 29, at 1352.

35. THOMPSON ET AL., *supra* note 32, at 6.

36. For a detailed discussion of these cultural categories, see *id.* at 5–11. *See generally* MARY DOUGLAS & AARON WILDAVSKY, RISK AND CULTURE: AN ESSAY ON THE SELECTION OF TECHNOLOGICAL AND ENVIRONMENTAL DANGERS (1982); Aaron Wildavsky & Karl Dake, *Theories of Risk Perception: Who Fears What and Why?*, 119 DAEDALUS 41, 43–44 (1990); Aaron Wildavsky, *Choosing Preferences by Constructing Institutions: A Cultural Theory of Preference Formation*, 81 AM. POL. SCI. REV. 3, 6 (1987); Dake, *supra* note 33, at 28–30.

37. Douglas, *supra* note 29, at 1367.

38. *Id.* at 1369.

39. Mary Douglas, *Four Cultures: The Evolution of a Parsimonious Model*, 47 GEOJOURNAL 411, 413 (1999).

40. Douglas, *supra* note 29, at 1353.

41. *Id.* at 1352.

are free to do what they like within the group with minimal regulation.[42] While hierarchs are set opposite to individualists, egalitarians are most strongly contrasted against hierarchs because they are the organized dissidents of that authoritative and regulated culture.[43]

While "grid" and "group" provide the taxonomy for the four worldviews, "grid" and "group" are not the actual constructs that these earlier scholars tried to measure. For them, the worldviews are the constructs. Dake and Wildavsky developed scales to measure them,[44] eventually focusing on just three (egalitarian, hierarch, and individualist, and ignoring fatalist).[45] Other researchers have followed this model and have used similar scales.[46]

The CCP scholars connect aspects of cultural theory with similar features found in social psychology, arguing "cultural commitments operate as a kind of heuristic in the rational processing of information on public policy."[47] They argue that individual citizens cannot personally investigate the facts that are used to support public policy arguments; therefore, citizens have to choose whom to trust among those who are advancing "facts"—and citizens trust those who share their worldviews.[48] This is congruent with social identity theory[49] and recognized psychological principles that explain motivated reasoning, like confirmation bias and the avoidance of cognitive dissonance.[50] Under social identity theory, those in

---

42. *Id.* at 1353.

43. *Id.* at 1368–69.

44. Dake, *supra* note 33, at 31.

45. Karl Dake & Aaron Wildavsky, *Individual Differences in Risk Perception and Risk-Taking Preferences, in* The Analysis, Communication, and Perception of Risk 15, 21 (B. J. Garrick & W. C. Gekler eds., 1991); Karl Dake, *Orienting Dispositions in the Perception of Risk: An Analysis of Contemporary Worldviews and Cultural Biases,* 22 J. Cross-Cultural Psychol. 61, 65 (1991); Wildavsky & Dake, *supra* note 36, at 47.

46. *See* Jean Brenot et al., *Testing the Cultural Theory of Risk in France,* 18 Risk Analysis 729, 732 (1998) (using all four constructs); *see also* Hank C. Jenkins-Smith & Walter K. Smith, *Ideology, Culture, and Risk Perception, in* Politics, Policy & Culture 17, 22 (Dennis J. Coyle & Richard J. Ellis eds., 1994) (not using fatalist); Ellen Peters & Paul Slovic, *The Role of Affect and Worldviews as Orienting Dispositions in the Perception and Acceptance of Nuclear Power,* 26 J. Applied Soc. Psychol. 1427, 1433–35 (1996) (combining fatalist and hierarchist); Richard J. Ellis & Fred Thompson, *Seeing Green: Cultural Biases and Environmental Preferences, in* Culture Matters: Essays in Honor of Aaron Wildavsky 169, 182 (Richard J. Ellis & Michael Thompson eds., 1997) (not using fatalist); Claire Marris et al., *A Quantitative Test of the Cultural Theory of Risk Perceptions: Comparison with the Psychometric Paradigm,* 18 Risk Analysis 635, 638 (1998) (using all four constructs). *But see* Douglas, *supra* note 29, at 1369–70 (providing that Mary Douglas thinks that ignoring fatalists is a mistake).

47. *Cultural Cognition and Public Policy, supra* note 27, at 151; *see also* Dan M. Kahan, *Cultural Cognition as a Conception of the Cultural Theory of Risk, in* Handbook of Risk Theory 725 (S. Roeser ed., 2012) [hereinafter Handbook of Risk Theory].

48. *Cultural Cognition and Public Policy, supra* note 27, at 151.

49. *See generally* Matthew J. Hornsey, *Social Identity Theory and Self-Categorization Theory: A Historical Review,* 2 Soc. & Personality Psychol. Compass 204 (2008).

50. *See Cultural Cognition and Public Policy, supra* note 27, at 165–66; Donald Braman et al., *Modeling Facts, Culture, and Cognition in the Gun Debate,* 18 Soc. Just. Res. 283, 289 (2005);

the "ingroup" trust others in the ingroup and disbelieve those in the "outgroup."[51] The CCP uses the cultural worldviews to define these ingroups and outgroups.

From this flows the CCP's definition of cultural cognition: "Cultural cognition refers to the tendency of individuals to conform their beliefs about disputed matters of fact (e.g., whether global warming is a serious threat; whether the death penalty deters murder; whether gun control makes society more safe or less) to values that define their cultural identities."[52] The CCP argues that these worldviews—rather than other predictive variables, like race and sex—best explain how people perceive risk.[53]

The CCP then modified the cultural theory taxonomy. For the purposes of this project, this move is important because it identifies the constructs that the CCP was trying to measure. The CCP continued to have "group" on the x-axis and "grid" on the y-axis and used three of the terms that other researchers continued to focus on—hierarchy, egalitarianism, and individualism.[54] However, the CCP added a new term, communitarianism,[55] and set individualism and communitarianism opposite each other on the x-axis to represent poles of the "group" dimension.[56] This move is generally consistent with Mary Douglas's "group" dimension, where the "individualism" label represents low group identity, and the "communitarianism" label represents high group identity.

The CCP kept "grid" on the y-axis, but then set hierarchy and egalitarianism opposite each other on that axis.[57] And importantly, this move is inconsistent with the Douglas model. Under the Douglas model, these two worldviews were the products of the two dimensions of "grid" and "group."[58] They were not the poles of the "grid" dimension. To be

---

HANDBOOK OF RISK THEORY, *supra* note 47, at 739. *See generally* Cass R. Sunstein, *Misfearing: A Reply*, 119 HARV. L. REV. 1110 (2006).

51. Hornsey, *supra* note 49, at 210–11.

52. *Cultural Cognition Project*, *supra* note 4.

53. *Cultural Cognition and Public Policy*, *supra* note 27, at 158–59. *See generally Culture, Cognition, and Consent*, *supra* note 2, at 733–34; Dan M. Kahan et al., Gender, Race, and Risk Perception: The Influence of Cultural Status Anxiety (Apr. 7, 2005) (unpublished manuscript), http:ssrn.com/abstract=723762 [hereinafter *Gender, Race, and Risk Perception*] (this article was later revised and published as Dan M. Kahan et al., *Culture and Identity-Protective Cognition: Explaining the White-Male Effect in Risk Perception*, 4 J. EMPIRICAL LEGAL STUD. 465 (2007) [hereinafter *Culture and Identity-Protective Cognition*]); HANDBOOK OF RISK THEORY, *supra* note 47, at 741.

54. Dan M. Kahan & Donald Braman, *More Statistics, Less Persuasion: A Cultural Theory of Gun-Risk Perceptions*, 151 U. PA. L. REV. 1291, 1297 (2003) [hereinafter *More Statistics, Less Persuasion*] (citing Wildavsky & Dake, *supra* note 36, at 11–13); *see* Wildavsky & Dake, *supra* note 36, at 44.

55. *More Statistics, Less Persuasion*, *supra* note 54, at 1303. The CCP initially used the term "solidarism" but later began using "communitarianism" instead. HANDBOOK OF RISK THEORY, *supra* note 47, at 730.

56. *Cultural Cognition and Public Policy*, *supra* note 27, at 153.

57. *See More Statistics, Less Persuasion*, *supra* note 54, at 1297.

58. Douglas, *supra* note 29, at 1353–54; *see* DOUGLAS & WILDAVSKY, *supra* note 36.

consistent with the Douglas model, the poles of the "grid" dimension would have been "structure" and "structureless."

The CCP then renamed the type of people that would fall within the resulting quadrants.[59] In the quadrant where Wildavsky and Dake set a worldview of "hierarchy," the CCP placed the label "hierarchical communitarian"; "egalitarianism" became "egalitarian communitarian"; "individualism" became "egalitarian individualist"; and "fatalist" became "hierarchical individualist." Note that while Wildavsky and Dake posited four separate constructs based on worldviews and created scales to measure those constructs, the CCP now posited just two constructs: grid and group.[60]

FIGURE 2: CCP TYPOLOGY
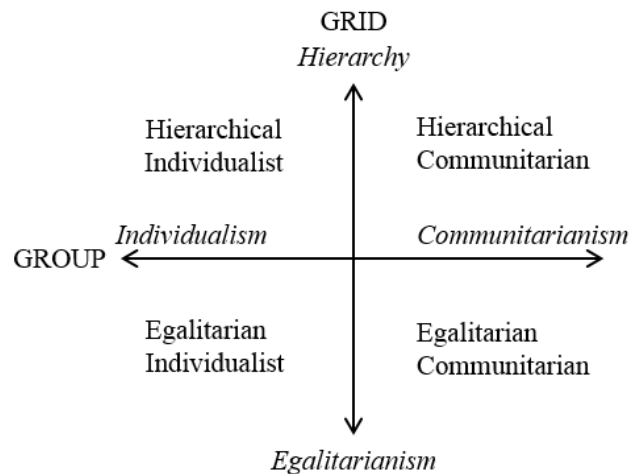
The CCP provides definitions of these two constructs. "Grid" represents a single, continuous measure of "how favorably or unfavorably disposed individuals are toward a social order that features differentiation and stratification of social roles based on observable and largely fixed

---

59. *See, e.g.,* Dan M. Kahan et al., *Cultural Cognition of Scientific Consensus,* 14 J. RISK RES. 147, 151 (2011) [hereinafter *Cultural Cognition of Scientific Consensus*]; Dan M. Kahan et al., *Affect, Values, and Nanotechnology Risk Perceptions: An Experimental Investigation* 11 (Cultural Cognition Project, Working Paper No. 22, 2007) [hereinafter *Affect, Values, and Nanotechnology Risk Perceptions*] (this study also formed the basis for Dan M. Kahan et al., *The Future of Nanotechnology Risk Perceptions: An Experimental Investigation of Two Hypotheses* (Cultural Cognition Project, Working Paper No. 46, 2008)); HANDBOOK OF RISK THEORY, *supra* note 47, at 739.

60. HANDBOOK OF RISK THEORY, *supra* note 47, at 730 (note that the CCP often refers to the "grid" construct as "hierarchy").

characteristics (including race, gender, sexual orientation, and class)."[61] At the poles, the CCP defines "hierarchy" as "favors deference to traditional forms of social and political authority and is protective of the roles and status claims they entail,"[62] and where "entitlements, obligations, opportunities and offices are all assigned on the basis of conspicuous and largely fixed attributes, such as gender, race, lineage, class, and the like."[63] The CCP defines "egalitarianism" as "abhors social stratification, distrusts the social and political authority structures that rest on such differentiation, and favors collective action to equalize wealth, status, and power."[64]

"Group" represents a single, continuous measure of "how favorably or unfavorably disposed individuals are toward a social order that treats individuals as responsible for securing the conditions of their own flourishing without collective assistance and that resists collective interference with individual strivings."[65] The poles of this attitude are individualism and collectivism. The CCP defines "individualist" as "prizes individual autonomy, celebrates free markets and other institutionalized forms of private ordering, and resents collective interference with the same,"[66] and where individuals "are expected to secure the conditions of their own flourishing without interference or assistance from the collective."[67] The CCP defined "communitarianism" (or "solidarism") as "logically opposed to individualism"[68] and where "collective needs trump individual initiative, and in which society is expected to secure the conditions of individual flourishing."[69]

---

61. *Culture, Cognition, and Consent, supra* note 1, at 770. The "observable and largely fixed characteristics" part of the definition represents a significant departure from the Douglas typology. *Id.* For a criticism of this modification, see Douglas, *supra* note 29, at 1364. Douglas believed that the CCP construct equated hierarchs to racists and sexists. *Id.* at 1362–63. For the CCP's response, see Dan M. Kahan & Donald Braman, Response, *Caught in the Crossfire: A Defense of the Cultural Theory of Gun-Risk Perceptions,* 151 U. Pa. L. Rev. 1395, 1408 (2003) [hereinafter, *Caught in the Crossfire*]; *see also* Handbook of Risk Theory, *supra* note 47. It may be that the CCP constructs have departed significantly from original cultural theory constructs and their constructs now more closely resemble those associated with the Social Dominance Orientation Scale. *See* Felicia Pratto et al., *Social Dominance Orientation: A Personality Variable Predicting Social and Political Attitudes,* 67 J. Personality and Soc. Psychol. 741, 754–55 (1994). For constructs more closely related with the Right-Wing Authoritarianism Scale, see Bob Altemeyer, *The Other "Authoritarian Personality," in* 30 Advances in Experimental Social Psychology 85, 86–87 (Mark Zanna ed., 1998); *see also* Bernard E. Whitley, Jr., *Right-Wing Authoritarianism, Social Dominance Orientation, and Prejudice,* 77 J. Personality and Soc. Psychol. 126 (1999). The items from these scales are available in the Online Appendix to this Article, which can be found at the following link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2881375 (last visited Jan. 16, 2017).

62. *More Statistics, Less Persuasion, supra* note 54, at 1297.

63. Dan M. Kahan & Donald Braman, *The Self-Defensive Cognition of Self-Defense,* 45 Am. Crim. L. Rev. 1, 28 (2008) [hereinafter *Self-Defensive Cognition*].

64. *More Statistics, Less Persuasion, supra* note 54, at 1297.

65. *Culture, Cognition, and Consent, supra* note 2, at 770.

66. *More Statistics, Less Persuasion, supra* note 54, at 1297.

67. *Self-Defensive Cognition, supra* note 63, at 28.

68. *More Statistics, Less Persuasion, supra* note 54, at 1303.

69. *Cultural Cognition and Public Policy, supra* note 27, at 153.

The CCP created two scales to measure those constructs, measuring the grid construct with the H-ES and the group construct with the Individualism-Communitarianism Scale ("I-CS").[70] Only the relative values derived from the H-ES were statistically or meaningfully significant in the CCP's acquaintance rape study,[71] so this Article will not focus on the I-CS.

A review of the definitions reveals a potential problem. The definition for "egalitarian," which is supposed to be a pole of the "grid" construct, includes part of the definition of the "group" construct. Again, according to the CCP, an egalitarian is someone who "abhors social stratification, distrusts the social and political authority structures that rest on such differentiation, *and favors collective action to equalize wealth, status, and power.*"[72] That last phrase deals with collective action or assistance, and that is a part of the "group" construct. To the extent that the CCP developed scale items to measure "egalitarian," it may have inadvertently targeted the wrong construct.

C. THE HIERARCHY-EGALITARIANISM SCALE ITEMS

The next step is to evaluate whether the H-ES items measure the targeted construct and whether the scale items are reliable. The CCP, likely facing resource constraints and looking for a creative way to test their theories, first turned to existing data from the General Social Survey ("GSS") to create a provision scale.[73] For the first study, the CCP hypothesized that a person's cultural worldviews could predict the person's risk attitude toward private gun ownership.[74]

To generate items that would measure the grid construct, the CCP looked for items in the GSS that tested attitudes about race, sexual orientation, the military, and capital punishment.[75] To generate items that would measure the group construct, the CCP looked for items that tested attitudes about public spending for regulatory and social welfare programs.[76] The CCP then used these scales to test its hypothesis that the measured latent variables would predict attitudes toward gun regulation,[77] and found that their measures did predict those attitudes.[78]

This particular study received some criticism related to its construct validity because the items did not measure constructs that were

---

70. *More Statistics, Less Persuasion, supra* note 54, at 1302–03.

71. *Culture, Cognition, and Consent, supra* note 2, at 763, 779 tbl.1.

72. *More Statistics, Less Persuasion, supra* note 54, at 1297 (emphasis added).

73. *Id.* at 1302.

74. *Id.* at 1295.

75. *Id.* at 1303. The scale items are located in the appendix to that article. *Id.* at 1326. The CCP did not use any of these items in their final scale.

76. *Id.* at 1303.

77. *Id.* at 1302.

78. *Id.* at 1307.

consistent with the earlier typology.[79] In response, the CCP stated that, "[w]e are currently developing more refined measures of cultural orientation for use in [future] survey[s],"[80] and recognizing the difficulty in finding items that would provide content validity, stated that, "[w]e're grateful to [our critics] for focusing our attention on the problem. We'd be even more grateful were [they able] to figure out a way to solve it."[81]

Soon after, the CCP ran one experiment that introduced the H-ES and I-ES. This experiment resulted in two unpublished studies: *Gender, Race, and Risk Perception*[82] and *The 'Wildavsky Heuristic.'*[83] The CCP generated items for these worldview scales by looking at previously used scale items and by creating new ones. In *Gender, Race, and Risk Perception*, the CCP tells us that "[i]tem development consisted of the adaptation of items used in previous studies . . . as well as the creation of new items . . . ."[84] In *The 'Wildavsky Heuristic,'* the CCP cites additional scales as possible sources of scale items, particularly from items in similar grid/group models.[85] Many of these items come from scales that have their own scale development and validity problems.[86] Recognizing these issues, the CCP states, "an effort was made to create a new set of items that would both exhibit better scale reliability and effectively distinguish the different dimensions of cultural orientation in factor analysis."[87]

For these first two studies, the CCP chose a four-point Likert scale[88] but later adopted a six-point scale.[89] The materials do not indicate that the CCP submitted this item pool to a panel of experts or included validation items, like items from a social desirability scale. The CCP does appear to have administered an item pool to a development sample, stating

79. *Caught in the Crossfire, supra* note 61, at 1407–10.

80. *Id.* at 1408 n.46. The CCP further responded to this concern in John Gastil et al., *The 'Wildavsky Heuristic': The Cultural Orientation of Mass Political Opinion* 28–30 (Yale Law Sch. Pub. Law & Legal Theory Research Paper Series, Research Paper No. 107, 2005).

81. *Caught in the Crossfire, supra* note 61, at 1411.

82. *Gender, Race, and Risk Perception, supra* note 53.

83. Gastil et al., *supra* note 80.

84. *Gender, Race, and Risk Perception, supra* note 53, at 12.

85. Gastil et al., *supra* note 80, at 28–31.

86. HANDBOOK OF RISK THEORY, *supra* note 47, at 729. To view a reproduced item pool, along with a brief discussion of criticisms of the previous scales, and items from similar scales which were not referenced by the CCP, see the Online Appendix, *supra* note 61. *See* Susanne Rippl, *Cultural Theory and Risk Perception: A Proposal for a Better Measurement*, 5 J. RISK RES. 147, 154 (2002).

87. Gastil et al., *supra* note 80, at 17. The CCP was aware of the criticisms of Karl Dake's scales. HANDBOOK OF RISK THEORY, *supra* note 47, at 730.

88. Gastil et al., *supra* note 80, at 39; *Gender, Race, and Risk Perception, supra* note 53, at 38. A Likert scale measures the respondent's agreement or disagreement with a particular statement in the survey instrument, generally from strongly agree to strongly disagree. *See generally* Rensis Likert, *A Technique for the Measurement of Attitudes*, 22 ARCHIVES PSYCHOL. 1 (1932) (describing the Likert scale measurement technique discussed in this Article).

89. Dan M. Kahan et al., *Cultural Cognition and Public Policy: The Case of Outpatient Commitment Laws*, 34 LAW & HUM. BEHAV. 118, 136 (2010) [hereinafter *Outpatient Commitment Laws*].

that they used "focus-group discussions and survey pretesting"[90] and conducted "extensive pretesting."[91] However, the CCP did not indicate what items were added or dropped during this process.

The CCP chose thirteen items:[92]

TABLE 1: H-ES ITEMS[93]

Code	Item
HEQUAL*	We have gone too far in pushing equal rights in this country.
HREVDIS1	Nowadays it seems like there is just as much discrimination against whites as there is against blacks.
EWEALTH*	Our society would be better off if the distribution of wealth was more equal.
ERADEQ*	We need to dramatically reduce inequalities between the rich and the poor, whites and people of color, and men and women.
EDISCRIM*	Discrimination against minorities is still a very serious problem in our society.
HREVDIS2*	It seems like blacks, women, homosexuals and other groups don't want equal rights, they want special rights just for them.
HCHEATS	It seems like the criminals and welfare cheats get all the breaks, while the average citizen picks up the tab.
EDIVERS	It's old-fashioned and wrong to think that one culture's set of values is better than any other culture's way of seeing the world.
HWMNRTS	The women's rights movement has gone too far.
ESEXIST	We live in a sexist society that is fundamentally set up to discriminate against women.
HTRADFAM	A lot of problems in our society today come from the decline in the traditional family, where the man works and the woman stays home.
HFEMININ*	Society as a whole has become too soft and feminine.
EROUGH	Parents should encourage young boys to be more sensitive and less "rough and tough."

Note: * items included in short form version of scale.

---

90. *Gender, Race, and Risk Perception, supra* note 53, at 12.

91. Gastil et al., *supra* note 80, at 17. The CCP indicates that it conducted pre-screening surveys with these scales in other studies, although these pre-screenings appear to have been used to measure worldviews and not to develop the scale. Dan M. Kahan et al., *Whose Eyes Are You Going to Believe? Scott v. Harris and the Perils of Cognitive Illiberalism,* 122 HARV. L. REV. 837, 859 (2009) [hereinafter *Whose Eyes Are You Going to Believe?*].

92. The CCP selected at least two items verbatim from two previous scales and adopted one with slight modification. The CCP appears to have written the remaining ten items. The scale found in *Gender, Race, and Risk Perception* contained the thirteen items found in the current scale plus one additional item: "EGAYMAR—A gay or lesbian couple should have just as much right to marry as any other couple." *Gender, Race, and Risk Perception, supra* note 53, at 38. For *The 'Wildavsky Heuristic,'* the CCP used the EGAYMAR item from the scale as a dependent variable and dropped it as an item in the scale. Gastil et al., *supra* note 80, at 39–40. This item (EGAYMAR) was not reintroduced to the scale in later studies conducted by the CCP.

93. *Gender, Race, and Risk Perception, supra* note 53, at 38 app.

Looking first at content validity, once the construct is clearly defined the items should cover the full domain of that construct: "In theory, a scale has content validity when its items are a randomly chosen subset of the universe of appropriate items."[94] Again, the CCP definition of "grid" is an attitude that favors or disfavors deference to traditional forms of social and political authority and is protective of the roles and status claims they entail, and where those roles and status claims are assigned on the basis of conspicuous and largely fixed attributes, such as gender, race, lineage, class, and the like. Here, no items measure lineage.

Next, scale items should be representative or proportional to the facets[95] of the construct.[96] Because "[t]he items in an assessment instrument should be distributed, or weighted, in a way that reflects the relative importance of the various facets of the targeted construct,"[97] the H-ES items should have been distributed (or weighted later during factor scoring) to reflect each facet's importance to the "grid" construct.

The distribution of items is not proportional, though. The H-ES has thirteen items that cover four of those facets: gender, race, sexual orientation, and class. Five of the thirteen scale items solely measure attitudes about gender roles (HWMNRTS, ESEXIST, HTRADFAM, HFEMININ, EROUGH), and two other items (ERADEQ, HREVDIS2) include gender among other dimensions, for a total of more than half of the items. One item (EDISCRIM) solely measures race, while two other items include a race dimension (ERADEQ, HREVDIS2), for a total of three. Two items appear to have been written to measure class (EWEALTH, HCHEATS), and class is included as a dimension in one other item (ERADEQ), for a total of three. One item (HREVDIS2) contains sexual orientation, among other facets.

Further, a single facet should not have overly redundant items.[98] If the aggregate score for a scale is disproportionately influenced by any one facet of the construct, the scale will lack content validity.[99] When a scale is designed to measure an entire global construct but has several items that measure a specific sub-construct, the result may be to "pull the item set as a whole away from the intended latent variable . . . to an alternative, more specific latent variable . . . ."[100]

---

94. DeVellis, *supra* note 17, at 60; *see* Haynes et al., *supra* note 24.

95. A facet is a "dimension of interest [that] is a potential source of variation . . . ." DeVellis, *supra* note 17, at 56.

96. Haynes et al., *supra* note 24, at 239; Clark & Watson, *supra* note 17, at 311.

97. Haynes et al., *supra* note 24, at 244–45; *see* Clark & Watson, *supra* note 17, at 311.

98. DeVellis, *supra* note 17, at 79.

99. Haynes et al., *supra* note 24, at 240.

100. DeVellis, *supra* note 17, at 79.

This scale has redundancy problems. In particular, the gender items stand out. With up to seven of the thirteen items drawing on gender hierarchy, the H-ES may really be measuring that sub-construct.

Continuing on with content validity, some of the items may not be measuring the intended construct. The "grid" items should only be measuring the respondent's agreement with social structure—a social order that features differentiation and stratification of social roles based on fixed characteristics—and should not be measuring agreement with collective interference or collective action. However, some items may also be tapping into the "group" dimension, likely because the definition of "egalitarian" included a "group" facet. Some items appear to tap into "group" fairly directly (EWEALTH, ERADEQ) while others (HEQUAL, HCHEATS, ERADEQ, HREVDIS2, ESEXIST, EROUGH, HWMNRTS) use words that seem to indicate agreement with collective action, collective interference, group identity, or group movements, such as "we need to" or "parents should."

Turning to item reliability, two of the items (ERADEQ, HREVDIS2) are complex or "double-barreled."[101] Complex items are ones that "convey two or more ideas so that an endorsement of the item might refer to either or both ideas."[102] Complex items have reliability problems: "[R]espondents will interpret complex items in different ways; accordingly, their responses will reflect the heterogeneity of their interpretations. . . ."[103] It is difficult to know whether people with the same belief systems will answer these two items in the same way, and it is difficult to understand what the items actually measure. For example, the item language for ERADEQ is as follows: "We need to dramatically reduce inequalities between the rich and the poor, whites and people of color, and men and women."[104] Some people might believe in race or income equality but still believe in traditional gender roles, or some may believe in gender equality (because they have more exposure to people of the opposite sex) but still have latent racist beliefs (because of a lack of exposure to other races).[105] A response to this item of "slightly agree" could mean that the respondent believes in race equality but not gender equality, or it could mean that the respondent believes in gender equality but not race equality. Further,

---

101. *Id.* at 82; Clark & Watson, *supra* note 17, at 312. This is also called construct-irrelevant difficulty and construct-irrelevant easiness. Samuel Messick, *Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning*, 50 Am. Psychol. 741, 742 (1995).

102. DeVellis, *supra* note 17, at 82.

103. Clark & Watson, *supra* note 17, at 312.

104. *Gender, Race, and Risk Perception, supra* note 53, at 26–27.

105. Racist and sexist beliefs are correlated and have similarities, but also have differences in their sub-constructs. Janet K. Swim et al., *Sexism and Racism: Old-Fashioned and Modern Prejudices*, 68 J. Personality & Soc. Psychol. 199, 209 (1995).

the next respondent with the same belief systems as the last respondent might answer the item in a different way.

Other items are ambiguous. In general, "a good item should be unambiguous."[106] For example, the item language for HCHEATS is as follows: "It seems like the criminals and welfare cheats get all the breaks, while the average citizen picks up the tab." This item appears to have been written to test class hierarchy, but it may actually test race hierarchy if many respondents assume that criminals and welfare cheats are minorities.[107] Likewise, the item language for EDIVERS is that it is "old-fashioned and wrong to think that one culture's set of values is better than any other culture's way of seeing the world." The word "culture" can mean many things to many people, and this item could be measuring beliefs about race, or national origin, or lineage.

Also, two of the items (HEQUAL, EDIVERS) are abstract or global and do not directly measure any facet, while other items list specific facets. Designing scales to measure abstract, global, or general constructs is perfectly acceptable: "Scales can be developed to assess constructs at each of many levels of abstraction."[108] But, the "scale's content should reflect the conceptual definition application *to that scale*"[109] and scale developers should "select item wordings that correspond[] with the intended level of variable specificity."[110]

Mixing abstract and specific items can cause problems. Sub-clusters can form on the abstract items based on their non-specificity.[111] The abstract items can also become ambiguous when surrounded by specific facets. For example, with the item HEQUAL (which states that "[w]e have gone too far in pushing equal rights in this country"), a respondent might agree or disagree with that item because of the influence of a particular facet that has already been listed in the surrounding items on that scale. The respondent might think that we have not gone far enough with race equality but agree with this item because he believes the women's rights movement has gone too far, and the surrounding items have been heavily weighted toward gender roles and the respondent now has gender roles on the mind.

In sum, the scale appears to have content validity issues[112] because the entire content domain is not represented; the items have proportionality

106. DeVellis, *supra* note 17, at 81.

107. Exploratory factor analysis suggests that this item measures race hierarchy. *See* discussion *infra* Part V.A.

108. Clark & Watson, *supra* note 17, at 310.

109. DeVellis, *supra* note 17, at 60 (emphasis in original).

110. *Id.* at 115.

111. *Id.* at 80.

112. *See* Eero Olli, Rejected Cultural Biases Shape Our Political Views: A Migrant Household Study and Two Large-Scale Surveys 290 (Mar. 29, 2012) (unpublished Ph.D. dissertation, University of Bergen), http://bora.uib.no/handle/1956/6103.

and redundancy issues; and some items may measure a different construct. Because of the redundancy problem, it may turn out that the gender items have narrowed the construct that the scale is measuring from global hierarchy to gender hierarchy. Further, several of the items have reliability issues because they are ambiguous, complex, or shift between specific and abstract forms.

D. THE CCP's ITEM EVALUATION

For item evaluation, the CCP initially reported a coefficient alpha of 0.80 for the H-ES items[113] and has consistently found a high coefficient alpha for the H-ES.[114] In *Culture, Cognition, and Consent*, Kahan reports an alpha of 0.89.[115] In *The 'Wildavsky Heuristic,'* the CCP indicated that it might have conducted factor analysis but does not report the results of that analysis.[116] The CCP also conducted structural equation modeling in one study, but the modeling was not related to item evaluation.[117] These discussions are the only reported discussions on the full H-ES item evaluation.[118]

Scale developers often use coefficient alpha to test for reliability: "Alpha is defined as the proportion of a scale's total variation that is attributable to the common source, presumably the true score of the latent variable underlying the items."[119] If a scale measures only one latent variable and the scale items are highly correlated with each other, then we attribute that correlation to the latent variable.

---

113. Gastil et al., *supra* note 80, at 17.

114. For the fourteen-item scale, the CCP reported a coefficient alpha for H-ES of 0.81. *Gender, Race, and Risk Perception, supra* note 53, at tbl.1. Later studies using the thirteen-item scale also reported coefficient alpha for the H-ES. *Affect, Values, and Nanotechnology Risk Perceptions, supra* note 59, at 11 (coefficient alpha of 0.81); Dan M. Kahan et al., *Cultural Cognition of the Risks and Benefits of Nanotechnology*, 4 NATURE NANOTECHNOLOGY 87, 89 (2009) (coefficient alpha of 0.81) [hereinafter *Risks and Benefits of Nanotechnology*]; *Self-Defensive Cognition, supra* note 63, at 35 (coefficient alpha of 0.82); *Whose Eyes Are You Going to Believe?, supra* note 91, at 860 (coefficient alpha of 0.85); Donald Braman et al., *Risk and Culture: Is Synthetic Biology Different?* 9 (Cultural Cognition Project, Working Paper No. 29, 2009) (coefficient alpha of 0.86) [hereinafter *Is Synthetic Biology Different?*]; Kahan et al., *Outpatient Commitment Laws, supra* note 89, at 125; Dan M. Kahan et al., *Who Fears the HPV Vaccine, Who Doesn't, and Why? An Experimental Study on the Mechanisms of Cultural Cognition*, 34 LAW & HUM. BEHAV. 501, 505 (2010) (coefficient alpha of 0.80) [hereinafter *Who Fears the HPV Vaccine*].

115. *Culture, Cognition, and Consent, supra* note 2, at 770.

116. Gastil et al., *supra* note 80, at 17. In a later article that summarized that article, the CCP states that "[b]ecause the cultural orientations were not conceptualized as uncorrelated, mean item scores (rather than factor scores) were used to generate reliable measures of egalitarian-hierarchy (alpha = .82) and individualism-solidarism (alpha = .79)." John Gastil et al., *The Cultural Orientation of Mass Political Opinion*, 44 POL. SCI. & POL. 711, 712 (2011).

117. *Self-Defensive Cognition, supra* note 63, at 40.

118. The CCP later modified this scale to form a short form scale and discuss the reliability and validity of that modified scale. *See* discussion *infra* note 130.

119. DEVELLIS, *supra* note 17, at 37.

As a measure of reliability, however, coefficient alpha has a fundamental assumption: the scale must be unidimensional.[120] And, as previously discussed, the H-ES may not be. Rather, the scale may be multi-dimensional, with gender, class, and race hierarchy all forming facets, and the scale may also measure some part of the "group" construct, too.

A scale can capture more than one dimension—say, racial, gender, and class attitudes—and still return a high coefficient alpha.[121] While the CCP reports high coefficient alpha values, a "high coefficient alpha does not indicate unidimensionality"[122] and "a relatively high alpha is no guarantee that all the items reflect the influence of a single latent variable."[123] Research has shown that "acceptable alpha levels can be obtained by aggregating distinct but correlated subscales."[124] Further,

> [P]sychometricians long have disavowed the practice of using reliability indices to establish the homogeneity of a scale. To understand why this is so, it is necessary to distinguish between internal consistency on the one hand and homogeneity or unidimensionality on the other. *Internal consistency* refers to the overall degree to which the items that make up a scale are intercorrelated, whereas *homogeneity* and *unidimensionality* indicate whether the scale items assess a single underlying factor or construct. As such, internal consistency is a necessary but not sufficient condition for homogeneity or unidimensionality. In other words, a scale cannot be homogeneous unless all of its items are interrelated, but . . . a scale can contain many interrelated items and still not be unidimensional.[125]

The gender item redundancy issue within the H-ES can also affect unidimensionality. When a scale has been designed to measure a general construct but has several items that are related to a sub-construct, the correlations among the sub-construct items "are likely to be greater than correlations between those items and others not [related to the sub-construct]."[126] This can then "undermine the unidimensionality of the item . . . ."[127]

A typical failure in many manuscripts submitted to one journal "was one in which investigators reported a coefficient alpha and apparently

---

120. Gregory T. Smith & Dennis M. McCarthy, *Methodological Considerations in the Refinement of Clinical Assessment Instruments*, 7 Psychol. Assessment 300, 302 (1995) ("For a scale to measure a unidimensional construct, its items must be parallel, alternative indicators of the same, underlying construct.").

121. John Hattie, *Methodology Review: Assessing Unidimensionality of Tests and Items*, 9 Applied Psychol. Measurement 139, 144 (1985); DeVellis, *supra* note 17, at 116.

122. Smith & McCarthy, *supra* note 120, at 303.

123. DeVellis, *supra* note 17, at 116.

124. Smith & McCarthy, *supra* note 120, at 301. "Because coefficient alpha is influenced by both internal consistency and scale length, it can be high when two internally consistent subscales, themselves only modestly interrelated, are combined." *Id.* at 303 (internal citations omitted).

125. Clark & Watson, *supra* note 17, at 315 (emphasis in original) (internal citations omitted).

126. DeVellis, *supra* note 17, at 79.

127. *Id.*

presumed the value indicated unidimensionality without testing that assumption . . . ."[128] This failure "could lead to inaccurate specifications of theory as well as misleading correlational and experimental findings."[129] That failure appears to have happened here.[130] The CCP has not published the scale's psychometrics so we do not know if the scale is unidimensional or multidimensional.

Determining whether a scale measures a global construct or instead returns an aggregate of related but distinct factors is usually done with factor analysis.[131] The exploratory factor analysis in Part V of this Article will help expose the scale's factor structure, particularly whether it is unidimensional or multidimensional.

III. KAHAN'S VALIDITY ARGUMENT IN *CULTURE, COGNITION, AND CONSENT*

Within the sexual assault context, Kahan recognized that the "grid" variable differed from the gender role constructs used by other researchers who study sexual assault, and he offered a lengthy reconciliation. Kahan stated:

---

128. Smith & McCarthy, *supra* note 120, at 301.

129. *Id.* at 300.

130. The CCP ran some studies of a short form version of the scale, using six items from the H-ES and six items from the I-CS. *Cultural Cognition of Scientific Consensus, supra* note 59, at 151, 173–74 app.1; Dan M. Kahan et al., *"They Saw a Protest": Cognitive Illiberalism and the Speech-Conduct Distinction,* 64 STAN. L. REV. 851, 869 (2012) [hereinafter *"They Saw a Protest"*]; HANDBOOK OF RISK THEORY, *supra* note 47, at 9. However, these results do not resolve these issues. The CCP reported an alpha for the short version of the H-ES of 0.87, and stated that the items "loaded appropriately on two separate factors, which were used as predictors for the study." *Cultural Cognition of Scientific Consensus, supra* note 59, at 151; *see "They Saw a Protest," supra,* at 869–70; HANDBOOK OF RISK THEORY, *supra* note 47. However, on the orthogonally rotated factor matrix, the items actually plotted on four separate factors, not two, with the H and E items plotting separately. *Id.* at 730 fig.28.5. In a recent study, the CCP reduced the H-ES to just two items (HEQUAL and EWEALTH) and reported that factor analysis showed that they loaded on the same factor with a coefficient alpha of 0.73. Dan M. Kahan et al., *"Ideology" or "Situational Sense?" An Experimental Investigation of Motivated Reasoning and Professional Judgment,* 64 U. PA. L. REV. 349, 379 (2016). In general, scale developers should run complete scale development on short form scales: "We advise against [reducing scales to short forms]: the psychometric properties of a measure cannot be imputed to a short form without empirical testing. Often, use of abbreviated measures attenuates reliability." Smith & McCarthy, *supra* note 120, at 306. Reducing scales to short forms can also cause issues with content validity: "Even more frequently, internal consistency is preserved but validity is attenuated because of reduced coverage of the target construct." *Id.* Global constructs have larger universes and covering the full domain can be difficult. Not covering the entire domain leads to construct underrepresentation. Messick, *supra* note 101, at 742.

131. Smith & McCarthy, *supra* note 120, at 303 ("[T]here are direct means of assessing the degree of subscale covariance (e.g., examination of correlation matrices and use of confirmatory factor analysis to test the degree of loss of model fit when combining scales).").

> Hierarchy is comparable but not identical to the attitudinal measures used
> to characterize subjects' gender-norm attitudes in [other studies of sexual
> assault]. Like those measures, Hierarchy includes items that relate to
> traditional gender roles and sexual equality.... However, it also includes
> items that relate to other dimensions of social stratification... unrelated to
> the gender-norm measures used in those studies.[132]

Kahan argued that this construct was valid because of the high
coefficient alpha he found for these scale items: "Because the reliability
of Hierarchy as a latent attitudinal measure indicates a high degree of
affinity between hierarchical gender attitudes and hierarchical attitudes
generally, there is no conceptual difficulty in using Hierarchy to test
hypotheses related to the former."[133] However, this statement is not
supported by the previous discussion about coefficient alpha.

Kahan then argued that its global construct would better explain the
variance than just a gender role facet: "Indeed, positive results obtained
by the use of Hierarchy are arguably *stronger* than ones based on gender-
role attitudinal scales. Hierarchy measures a disposition more general
than those measured by gender-role scales and is conceptually more
remote from the study's dependent variables, which themselves relate to
perceptions of sexual behavior."[134]

Kahan cites Paul Slovic and Ellen Peters for the proposition that the
"influence of distal [or global] variables is ordinarily smaller but more
important than the influence of proximal variables."[135] Slovic and Peters
were themselves responding to criticism about their choice of using a
global variable, but there, the authors were careful to point out that they
"purposely selected worldview items to minimize semantic overlap with
the risk attitudes and perception being explained."[136] They chose global
variables because they wanted to prevent their independent variables
from looking too much like their dependent variables, which is a
different issue than Kahan faced in his study.

Choosing to investigate a global variable versus a specific variable
(called construct hierarchy)[137] is not itself a problem, and specific

---

132. *Culture, Cognition, and Consent, supra* note 2, at 770.

133. *Id.*

134. *Id.* (emphasis in the original).

135. *Id.* at 770 n.157 (citing Paul Slovic & Ellen Peters, *The Importance of Worldviews in Risk
Perception*, 3 RISK DECISION & POL'Y 165, 169 (1998)).

136. Slovic & Peters, *supra* note 135, at 169.

137. Construct hierarchy refers to the generality or specificity of the construct that the scale developer
is trying to measure. *See* Andrew L. Comrey, *Factor-Analytic Methods of Scale Development in
Personality and Clinical Psychology*, 56 J. CONSULTING & CLINICAL PSYCHOL. 754, 755–56 (1988); Clark &
Watson, *supra* note 17, at 310–11; Smith & McCarthy, *supra* note 120, at 301. For our purposes, this is an
unfortunate term because the CCP sometimes calls the construct that we are exploring "hierarchy." I will
instead use the terms "global construct" and "specific construct" throughout this Article to capture the
concept rather than using "construct hierarchy." *See* DEVELLIS, *supra* note 17, at 79.

variables are not better or worse than global variables.[138] Further, specific facets can be combined and analyzed at an aggregate, global level. However, researchers can do that only if the correlations between the facets and the dependent variable do not differ from the main effect of the aggregated score.[139] If the facets perform differently than the aggregated score, the researchers need to report that result.[140]

The structural equation modeling in Part VI will explore whether the scale has different facets and whether those facets perform differently than the hypothesized global variable.

<div align="center">

IV. EXPLORATORY FACTOR ANALYSIS OF THE
HIERARCHY-EGALITARIANISM SCALE

</div>

I conducted exploratory factor analysis to better understand how the items related to each other and the scale's possible factor structure, which would then inform decisions about the theoretical model for the structural equation modeling.

A. SURVEY ADMINISTRATION AND DATA SCREENING

The data comes from a survey that was administered in 2009. Kahan used a private firm to administer the survey to 1500 people in the United States. The survey was conducted online using a pool of over one million people who are paid to participate in these surveys. The firm used a demographic-matching methodology that ensured that the sample was representative of the general population, so weighting is not necessary.[141]

Using Excel, I screened the data to see if any observations were missing data over ten percent and did not find any. I screened for unengaged respondents by running the standard deviation for each respondent's data and looking closely at those with low standard deviations. I removed those that I found. I screened the variables for outliers and found two in the variable EROUGH. I deleted those data points but retained the observations. The variable "PID7" (a seven-point Likert scale that measured political identification from liberal through conservative) was missing data at four percent. No other variable had missing data over one percent. After screening, n = 1487. I then randomly split the sample in half so that one half could be used for exploratory factor analysis (n = 770) and the other half for structural equation modeling (n = 717).[142]

---

138. The level of specificity of the scale should generally match the research question, however. DEVELLIS, *supra* note 17, at 75.

139. Smith & McCarthy, *supra* note 120, at 303.

140. *Id.*

141. *Culture, Cognition, and Consent, supra* note 2, at 765.

142. Leandre R. Fabrigar et al., *Evaluating the Use of Exploratory Factor Analysis in Psychological Research*, 4 PSYCHOL. METHODS 272, 277 (1999).

B.   Exploratory Factor Analysis

For the exploratory factor analysis ("EFA"), I used Statistical Package for the Social Sciences ("SPSS") (version 21) as the statistical software. I excluded missing data pair wise. The data is not normally distributed[143] and normality is a strict assumption for maximum likelihood factor solutions,[144] which is the common way of conducting EFA. Therefore, I used principal factor extraction methods (in SPSS, principal axis factors)[145] because principal factor methods do not have a distribution assumption.[146] I expected that any factors that were extracted would be correlated so I chose an oblique rotation (Promax).[147] I did not reverse code the items beginning with "E" so that it would be easier to spot criterion validity.

The correlation matrix (provided in Table 2) contained many items with correlations of r = 0.3 or greater.[148]

---

143. I conducted a visual inspection of the histograms for all of the H-ES items. Only ESEXIST and EROUGH looked somewhat normal. A Shapiro-Wilk's test (p > 0.05) showed that all of the variables were non-normal (all had p-values < 0.001). Many of the items were severely skewed (z-value > 2). All items except for HCHEATS showed negative kurtosis with many showing severe negative kurtosis (z-value > 7), meaning the distribution departed from a bell curve toward a uniform or flat distribution. *Id.* at 283. To view normality values, see the Online Appendix, *supra* note 61.

144. Frank J. Floyd & Keith F. Widaman, *Factor Analysis in the Development and Refinement of Clinical Assessment Instruments*, 7 Pscyhol. Assessment 286, 289 (1995).

145. *Id.*; Timothy A. Brown, Confirmatory Factor Analysis for Applied Research, 379, 387 (1st ed. 2006); Anna B. Costello & Jason W. Osborne, *Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis*, 10 Prac. Assessment, Res. & Evaluation 1, 2 (2005). I also ran exploratory factor analysis using maximum likelihood and nonweighted mean squares. The results were substantially the same: "[W]hen the common factor model holds reasonably well in the population and severe violations of distributional assumptions are not present, solutions provided by these methods are usually very similar." Fabrigar et al., *supra* note 142, at 277.

146. Fabrigar et al., *supra* note 142, at 277.

147. Promax is an oblique rotation used when the factors are assumed to be correlated. *See* Holmes Finch, *Comparison of the Performance of Varimax and Promax Rotations: Factor Structure Recovery for Dichotomous Items*, 43 J. Educ. Measurement 39, 42 (2006).

148. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was above 0.6 (0.92), indicating a high degree of common variance. Bartlett's Test of Sphericity showed significance (p < 0.001), meaning the null hypothesis that the correlation matrix is an identity matrix (where the variables are noncollinear) and that any non-zero correlations are due to sampling error is rejected. These indices support the factorability of the correlation matrix.

TABLE 2: CORRELATION MATRIX FOR H-ES EXPLORATORY FACTOR ANALYSIS

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. HEQUAL	-												
2. HREVDIS1	.56	-											
3. EWEALTH	-.40	-.27	-										
4. ERADEQ	-.43	-.31	.69	-									
5. EDISCRIM	-.48	-.41	.44	.57	-								
6. HREVDIS2	.62	.59	-.41	-.44	-.51	-							
7. HCHEATS	.46	.50	-.24	-.26	-.30	.56	-						
8. EDIVERS	-.26	-.13	.35	.40	.29	-.29	-.12	-					
9. HWMNRTS	.57	.44	-.41	-.44	-.44	.61	.41	-.32	-				
10. ESEXIST	-.30	-.23	.41	.46	.46	-.41	-.20	.36	-.42	-			
11.HTRADFAM	.40	.36	-.29	-.33	-.32	.49	.39	-.27	.50	-.24	-		
12. HFEMININ	.44	.40	-.36	-.40	-.40	.52	.39	-.30	.54	-.40	.47	-	
13. EROUGH	-.20	-.18	.40	.38	.37	-.29	-.14	.27	-.28	.40	-.23	-.38	-

Two factors were extracted. Both of the eigenvalues for the two factors exceeded the value set by parallel analysis,[149] and the screen plot was consistent with two factors. All of the items loaded on two factors, with all of the "H" items loading on Factor 1 and all of the "E" items loading on Factor 2 as shown in TABLE 3. All of the items loaded above 0.4.[150] The pattern matrix and communalities are provided in TABLE 3.

TABLE 3: EXPLORATORY FACTOR ANALYSIS PATTERN MATRIX AND COMMUNALITIES

Variable	Pattern Matrix Factor		Communalities	
	1	2	Initial	Extraction
HREVDIS1	0.80		0.45	0.51
HCHEATS	0.79		0.39	0.46
HREVDIS2	0.79		0.62	0.71
HEQUAL	0.70		0.52	0.55
HWMNRTS	0.56		0.52	0.54
HTRADFAM	0.53		0.36	0.35
HFEMININ	0.45		0.44	0.45
ERADEQ		0.83	0.59	0.64
EWEALTH		0.76	0.52	0.53
ESEXIST		0.63	0.37	0.41
EROUGH		0.59	0.28	0.30
EDIVERS		0.54	0.24	0.26
EDISCRIM		0.51	0.47	0.47

---

149. I ran Monte Carlo PCA for Parallel Analysis software, with variables set at thirteen, number of subjects at 750, and 1000 iterations. The criterion eigenvalue for the first extracted factor was 1.22, SD = 0.03, and the actual eigenvalue was 5.70; the criterion for the second was 1.17, SD = 0.02, and the actual eigenvalue was 1.50; the criterion for the third was 1.13, SD = 0.02, and the actual eigenvalue was 0.85.

150. Floyd & Widaman, *supra* note 144, at 294 (items should load greater than 0.3 or 0.4); Jacques Raubenheimer, *An Item Selection Procedure to Maximise Scale Reliability and Validity*, 30 S. AFR. J. INDUS. PSYCHOL. 59, 61 (2004) (noting that items should load greater than 0.4). Factor loading is a measure of an item's association with the underlying latent variable. *See generally* Floyd & Widaman, *supra* note 144, at 287.

The factors were strongly, negatively correlated (r = -0.66). Coefficient alpha for Factor 1 is 0.87 and for Factor 2 is 0.82. Two items, EDIVERS and EROUGH, had low communalities which indicates that they do not relate well to the other items or are unreliable.[151] The model explained less than fifty percent of variance (forty-eight percent),[152] which indicates that there is a lot of unique variance and random error, possibly because of the issues related to the scale validity and reliability that were previously discussed.[153]

The EFA provides a two-factor solution. Those factors do not appear to be related to particular facets of "grid," like race, gender, or class because those items did not load together. That might be because there are not enough items in the scale to measure each sub-factor: "[F]actor analysis might establish that the items can be subdivided into several subscales but that the initial pool does not contain enough items to assess each of these content domains reliably."[154] Generally, to detect facets or individual factors, researchers need three to six items per facet or factor.[155] Here, we only have one item in the scale that solely measures race and two that are designed to solely measure class. Sexual orientation is mentioned in only one item and there are no items on lineage.

Importantly, the EFA does not support the CCP's unidimensional scale design. If the H-ES was unidimensional, all of the items should load on one factor, with "grid" serving as the latent variable. A single dimension can sometimes load on two separate poles if the items are reverse coded.[156] Here, however, the loading reported by the EFA was not due to reverse coding because I ran the EFA both with and without reverse coding, and the results were the same. A single dimension can also sometimes load on two separate poles if the items are written in opposite directions—for example, some items but not others use "not" language.[157] That is not an issue here either. Rather, the EFA suggests that the H-ES might roughly measure two separate constructs.

---

151. Fabrigar et al., *supra* note 142, at 274 ("There are a number of reasons why communalities for measured variables might be low. One obvious reason is low reliability. As explained later, variance due to random error cannot, by definition, be explained by common factors. Because of this, variables with low reliability will have low communalities and thus should be avoided.").

152. Floyd & Widaman, *supra* note 144, at 295 (providing for a minimum of fifty percent and recommending above eighty percent).

153. *Id.* ("[A]ccounting for relatively little common variance challenges the relative importance of common factors as opposed to the specific factor variance associated with individual variables.").

154. Clark & Watson, *supra* note 17, at 311.

155. Floyd & Widaman, *supra* note 144, at 292 ("In general, three variables per factor are needed to identify common factors." (internal citations omitted)); Fabrigar et al., *supra* note 142, at 273 ("Research suggests that EFA procedures provide more accurate results when each common factor is represented by multiple measured variables in the analysis.... Methodologists have recommended that at least three to five measured variables representing each common factor be included in a study." (internal citations omitted)); Fabrigar and colleagues recommend four to six measured variables. *Id.* at 282.

156. Comrey, *supra* note 137, at 758.

157. *Id.*

V. STRUCTURAL EQUATION MODELING

The EFA suggests that the items do not measure a unidimensional construct and that there may be reliability issues with the items. I used structural equation modeling to see if the scale items could still deliver useful information.

Structural equation modeling has two components: a measurement model and a structural model.[158] The measurement model is usually a theory-driven confirmatory factor analysis, which is contrasted to the data-driven exploratory factor analysis that was run previously.[159] The structural model "displays the interrelations among latent constructs and observable variables in the proposed model as a succession of structural equations—akin to running several regression equations."[160] Statistical software then returns various model fit indices, which indicate the degree to which the specifications of the theory-driven model are "consistent with the pattern of variances and co-variances from a set of observed data."[161] Most indices "reflect the improvement in fit of a specified model . . . over the independence model, in which all structural parameters are fixed at zero."[162]

I organized the items into sub-constructs based on an analysis of item wording and the correlation matrix from the EFA. For the gender facet, I directed the model to load the gender items HWMNRTS, ESEXIST, HTRADFAM, FFMININ, and EROUGH onto a single factor. For the race facet, I directed the following to load on a single factor: EDISCRIM (purely a race item), HREVDIS1 (purely a race item), HREVDIS2 (the item starts with a race facet and is highly correlated with EDISCRIM and HREVDIS1), HEQUAL (highly correlated with EDISCRIM, HREVDIS1, HREVDIS2, and HCHEATS, although it is also highly correlated with HWMNRTS), HCHEATS (low correlation with EWEALTH and a strong correlation with the race items), and EDIVERS (low correlations with most items, and I theorized that many respondents would think of race when they think of "culture"). I directed EWEALTH and ERADEQ (the item starts with an income facet and is highly correlated with EWEALTH) to load on a class factor.

---

158. James B. Schreiber et al., *Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review*, 99 J. EDUC. RES. 323, 325 (2006).

159. *Id.* at 323.

160. *Id.* at 325.

161. Rick H. Hoyle, *The Structural Equation Modeling Approach: Basic Concepts and Fundamental Issues, in* STRUCTURAL EQUATION MODELING: CONCEPTS, ISSUES, AND APPLICATIONS 1, 3 (Rick H. Hoyle ed., 1995).

162. *Id.* at 7.

A. Confirmatory Factor Analysis

Starting with the theory-derived measurement model, I used Mplus (version 6.12) statistical software. Because of multivariate nonnormality (the Mardia's coefficient was significant), I ran a bootstrap with 2000 replicates. I used Full Information Maximum Likelihood ("FIML") to deal with missing data. As previously noted, n = 717.

The initial model fit was not good. The chi-square value for the overall model fit was significant (chi-square (62) = 494.72, p < 0.001).[163] However, for large samples sizes like the one studied here, "it can happen that the chi-square test is failed even though differences between observed and predicted covariances are slight."[164] Further, chi-square "is viewed by most as overly strict given its power to detect even trivial deviations of data from the proposed model,"[165] and so researchers should also look to other indices to evaluate model fit. Here, examination of other indices also showed an unacceptable model fit (RMSEA = 0.10, p(close) < 0.001; CFI = 0.89; TLI = 0.86; SRMR = 0.06).[166]

Modification indices indicated that ERADEQ had cross loading issues, meaning that it has substantial shared variance with all three facets, probably because it is a complex item that asks about race, gender, and class hierarchy, so I removed it. This meant that the class facet only had one observed variable (EWEALTH), so I removed this item and facet from the model. I then removed EDIVERS for cross loading issues which are probably the result of the ambiguous term "culture." I correlated the error terms for EROUGH and ESEXIST, and for EROUGH and HFEMININ, because the modification indices indicated that these items had a lot of shared error variance. This indicates the gender facet has possible subscales. The modification indices also indicated shared error variance between HCHEATS and HREVDIS1, so I correlated the error terms. This indicates that the race facet has possible subscales. I then removed EDISCRIM because of cross loading issues with two of the gender items.

The resulting model has two factors with the items loading appropriately on their respective factors (0.43–0.79 for gender and 0.58–0.88

---

163. In structural equation modeling, the chi-square statistic is really a "badness of fit" index. *Id.* A large chi-square and a small p-value (here, less than 0.01) means that there is a statistically significant difference between the estimated variance-covariance matrix and the actual variance-covariance matrix. Here, this statistic means that we reject the exact-fit hypothesis at the 0.01 level and suggests a lack of fit between the hypothesized model and the data. *See* Rex B. Kline, Principles and Practice of Structural Equation Modeling 199–201 (Todd D. Little ed., 3d ed. 2011).

164. Kline, *supra* note 163, at 201.

165. Ralph O. Mueller & Gregory R. Hancock, *Structural Equation Modeling, in* The Reviewer's Guide to Quantitative Methods in the Social Sciences 371, 379 (Gregory R. Hancock & Ralph O. Mueller eds., 2d ed. 2010).

166. Cutoff criteria: RMSEA < 0.06; CFI ≥ 0.95; TLI ≥ 0.95; SRMR ≤ 0.08. Schreiber et al., *supra* note 158, at 330 tbl.2.

for race). The chi-square value for the overall model fit was significant (chi-square (23) = 48.24, p < 0.001), suggesting a lack of fit between the hypothesized model and the data. However, examination of other indices also showed an acceptable model fit (RMSEA = 0.04, p(close) = 0.87; CFI = 0.99; TLI = 0.98; SRMR = 0.02).[167] The gender and race constructs were highly correlated (r = 0.87), which indicates that they may be sub-constructs of a more global construct.[168]

B.   THE STRUCTURAL MODEL

I used the results from the confirmatory factor analysis to establish the structural model. In addition to the nine observed variables (the scale items) loading on two latent variables (gender and race) with one exogenous variable (GUILTY), I included one control variable—political identification. The CCP reported that political identification as a liberal or conservative did not meaningfully influence the responses; however, other research has found that this variable predicts rape myth acceptance,[169] so I included it as a control variable in the model.[170]

As previously noted, Kahan applied a treatment condition by randomly dividing the subjects into five groups of 300 each, and giving each of them one of five legal conditions (either no law to apply, or four different versions of law to apply).[171] With the GUILTY variable converted to binary, Kahan reported that only one of these conditions was statistically significant.[172] I ran a univariate general linear model (ANOVA) with the treatment conditions as the independent variable and GUILTY as the dependent variable. The main effect was not significant (F(4709) = 0.40, p = 0.81, eta squared < 0.01), nor were any of

---

167. Keith F. Widaman & Jane S. Thompson, *On Specifying the Null Model for Incremental Fit Indices in Structural Equation Modeling*, 8 PSYCHOL. METHODS 16 (2003).

168. To contrast with this structure, I also ran confirmatory factor analysis on the structure revealed by the exploratory factor analysis, where the "H" items loaded on one factor and the "E" items loaded on another. The model fit was not good. The chi-square value for the overall model fit was significant (chi-square (64) = 355.54, p < 0.001), suggesting a lack of fit between the hypothesized model and the data. Examination of other indices also showed an unacceptable model fit (RMSEA = 0.08, p(close) < 0.001; CFI = 0.93; TLI = 0.91; SRMR = 0.05).

169. *See* Anderson et al., *supra* note 15, at 312; William D. Walker et al., *Authoritarianism and Sexual Aggression*, 65 J. PERSONALITY & SOC. PSYCHOL. 1036, 1038 (1993) (using the Right Wing Authoritarianism scale).

170. The CCP reported that race, age, and education did not meaningfully influence responses to the guilt variable. *Culture, Cognition, and Consent, supra* note 2, at 782. This is consistent with other research, so I did not include these as control variables. *See* Kimberly A. Lonsway & Louise F. Fitzgerald, *Rape Myths: In Review*, 18 PSYCHOL. WOMEN Q. 133, 142–45 (1994). The CCP reported that sex did not meaningfully influence responses. This is inconsistent with other research. *See id.* However, I did not include it in this model because sex does not itself involve a competing belief system relative to the two other latent variables in the model. After I ran the model, I later included sex, and it did not improve model fit and it was a statistically insignificant predictor of the outcome judgment.

171. *Culture, Cognition, and Consent, supra* note 1, at 767–68.

172. *Id.* at 778, 779 tbl.1.

the pair wise comparisons. Therefore, I did not include the treatment condition variable as a control variable in the model.

The chi-square value for the overall model fit was significant (chi-square (36) = 84.02, p < 0.001). However, examination of other indices showed an acceptable model fit (RMSEA = 0.04, p(close) = 0.82; CFI = 0.98; TLI = 0.98; SRMR = 0.03). TABLE 4 provides a correlation table, and FIGURE 3 provides the theoretical model.

TABLE 4: CORRELATION MATRIX FOR STRUCTURAL MODEL

Variable	1	2	3	4	5	6	7	8	9
1. HFEMININ	-								
2. HWMNRTS	.57	-							
3. HTRADFAM	.45	.49	-						
4. ESEXIST	-.36	-.40	-.24	-					
5. EROUGH	-.47	-.33	-.19	.43	-				
6. HREVDIS1	.41	.46	.33	-.29	-.24	-			
7. HREVDIS2	.56	.60	.47	-.38	-.36	.66	-		
8. HEQUAL	.53	.60	.41	-.38	-.33	.57	.69	-	
9. HCHEATS	.36	.37	.31	-.23	-.22	.50	.52	.43	-

Note: Spearman's rank correlation coefficient, all significant (two-tailed) at p < 0.01.

FIGURE 3: THEORETICAL MODEL
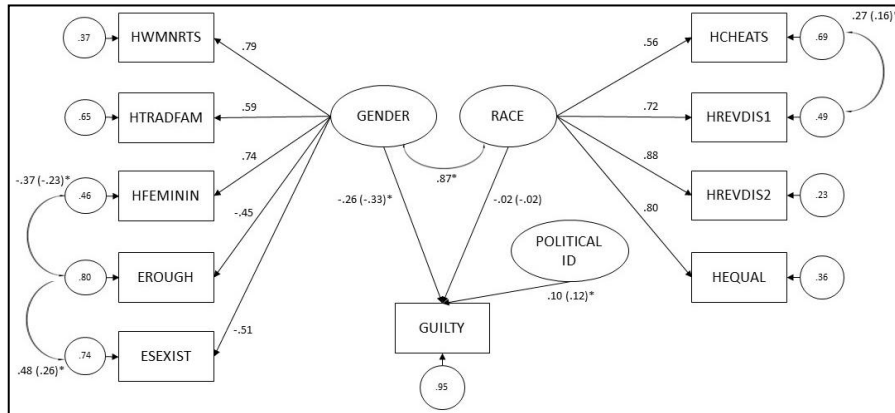
Note: Standardized path coefficients are in parentheses. *p < 0.01.

The data suggests that the gender items account for the variance in the outcome judgment and not the race items. The gender facet was a significant—both statistically and meaningfully—predictor of agreement with the outcome judgment of guilty. For every increase in one standard deviation from the mean of "Gender" (toward being more traditional), the mean of "Guilty" would be expected to decrease by 0.33 in its own standard deviations from its mean, while holding all other relevant

connections constant. The "Race" facet was not statistically significant. The "Political ID" facet was statistically significant but had a smaller effect. For every increase in one standard deviation from the mean of "Political ID" (toward being more liberal), the mean of "Guilty" would be expected to increase by 0.12 in its own standard deviations from its mean, while holding all other relevant connections constant.

This two-facet model had good model fit, and the results are consistent with other research in this area. While the gender facet in my model has predictive validity, it lacks construct validity. I did not develop these items to measure a clearly defined construct. However, the gender items appear to measure two dimensions of patriarchy: (1) beliefs about traditional, largely patriarchal gender roles, which tend to be benign or benevolent; and (2) also more hostile, sexist beliefs about the subordination of women in society.[173] Two items measured beliefs about the proper—but equally valued—roles that men and women should follow in the home and interpersonal relationships (HTRADFAM, EROUGH). Two measured more sexist beliefs about the subordinated position of women in politics and employment (HWMNRTS, ESEXIST). One measured a more hostile belief about the value of femininity (HFEMININ).[174] That said, a fair label to place on these collected gender items and the construct that they measure is *patriarchy*.

To contrast with this model, I also ran Kahan's model, with the thirteen items loading on one latent variable (grid), using PID7 as a control variable, and using GUILTY as the outcome variable. The model fit was not good. The chi-square value for the overall model fit was significant (chi-square (88) = 800.65, p < 0.001), suggesting a lack of fit between the hypothesized model and the data. Examination of other indices showed an unacceptable model fit (RMSEA = 0.11, p(close) < 0.001; CFI = 0.84; TLI = 0.81; SRMR = 0.06). The latent variable (grid) did predict the endogenous variable GUILTY (path coefficient = -0.37, p < 0.001), as did PID7 (path coefficient = 0.14, p = 0.001).[175]

The structural equation modeling of the CCP's model is consistent with the EFA, in that the data does not support the CCP's one-factor model. While Kahan's model had predictive validity (the grid latent variable was a significant predictor of agreement with the outcome judgment), the poor model fit suggested the grid latent variable did not have construct validity. Further, the path coefficient for the gender facet in my two-facet model (-0.33) was about the same as the path coefficient for Kahan's grid factor

---

173. *See generally* Peter Glick & Susan T. Fiske, *The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism*, 70 J. PERSONALITY & SOC. PSYCHOL. 491 (1996).

174. The race facet appears to measure race hierarchy, with three items that appear to measure agreement with programs like affirmative action and one that appears to measure hostile beliefs about the stereotyped black underclass.

175. To view the correlation matrix and theoretical model, see the Online Appendix, *supra* note 61.

(-0.37), which also suggests that in this rape hypothetical, the gender items are doing the work. While "grid" has predictive validity, the predictive validity appears to be generated by one of the sub-facets alone.

VI. Discussion

The data suggests that patriarchy, not hierarchy, predicts outcome judgments in acquaintance rape cases. This is inconsistent with Kahan's findings in *Culture, Cognition, and Consent.* The data also suggests that the H-ES does not measure a single construct of "grid," but rather has sub-constructs related to gender and race hierarchy.[176] This has important implications for other CCP studies. In particular, for the other CCP studies that involve racial issues and that used the H-ES,[177] a sub-factor related to race may explain the variation in outcome judgments rather than a global construct of "grid." Structural equation modeling of those studies could reveal the answer.

That said, the global construct of "grid" may exist. The high correlation between the two facets in my model suggests there might be a higher-order construct, and an improved scale with reliable items might be able to measure that construct. One option for improving the CCP's scale to measure this global construct would be to have all abstract or global questions,[178] like those used by John Jost and Erik Thompson when they measured a similar construct.[179] This would require the CCP to modify the "grid" definition somewhat to drop the facets.

Another option for improving the scale would be to measure the facets of this global "grid" construct. The "grid" items could consist of four

---

176. The exploratory factor analysis suggested two constructs, loading all of the "H" items onto one factor and all of the "E" items on another factor. Recognizing that the EFA solution was unsatisfactory (explaining less than fifty percent of the variance) and recognizing the issues with content validity and item reliability, the "H" items might roughly measure the CCP's worldview of "hierarchical communitarian" and the "E" items might roughly measure "egalitarian communitarian." As written, these might be tapping into the intersections of the grid and group dimensions, which are Douglas' worldviews of hierarchy and egalitarianism. With scale refinement, the Hierarchy and Egalitarianism model might work, and researchers would then have to decide if it is appropriate to combine the scores from those two worldview constructs into one aggregated score.

177. *See Culture and Identity-Protective Cognition, supra* note 53 (including gun risks that implicate race hierarchy and abortion risks that implicate gender hierarchy); *Self-Defensive Cognition, supra* note 63, at 22 (including a vignette where a white man shoots a black man in self-defense); *Whose Eyes Are You Going to Believe?, supra* note 91 (involving the use of police force on a fleeing suspect that implicates race hierarchy).

178. *See* DeVellis, *supra* note 17, at 74 ("There is general agreement in the social sciences that variables will relate most strongly to one another when they match with respect to level of specificity. Sometimes, a scale is intended to relate to very specific behaviors or constructs, while at other times, a more general and global measure is sought." (internal citations omitted)).

179. *See* John T. Jost & Erik P. Thompson, *Group-Based Dominance and Opposition to Equality as Independent Predictors of Self-Esteem, Ethnocentrism, and Social Policy Attitudes Among African Americans and European Americans*, 36 J. Experimental Soc. Psychol. 209 (2000). The items are listed in the Online Appendix, *supra* note 61.

items each on gender, race, class (if that is a facet rather than a byproduct of hierarchical beliefs), lineage, and sexual orientation. The importance of each sub-factor could be accounted for when calculating factor scores.[180] When the scale is used in projects that relate to a particular facet, like race or gender, researchers could then use structural equation modeling to see if a sub-factor behaves differently than the aggregated factor, as my model revealed here. The sub-factor would then be reported separately from the aggregated construct.

The CCP may also need to revise its construct definitions.[181] For the H-ES, the two-factor loading revealed by the EFA and the poor model fit revealed by the confirmatory factor analysis may be the result of the CCP's item development, where the CCP looked to items that were designed to measure features of Douglas' hierarch and egalitarian cultures.[182] Under Douglas' model, the constructs of hierarchy and egalitarianism represent cultures that form at the intersection of two dimensions—grid and group. As previously discussed, while the CCP set "hierarchy" and "egalitarianism" as poles on the "grid" construct, the CCP definition of "egalitarian" includes a "group" component and many of the H-ES items appear to also tap into the "group" construct. These items may actually be measuring features of the cultures that form at the intersection of the CCP's revised dimensions of the "grid" and "group," which the CCP calls "hierarchical communitarian" and "egalitarian communitarian,"[183] rather than the dimensions themselves. Here, the CCP may need to replace "egalitarian" on the y-axis with "structureless," and then return "egalitarian" to a culture formed by the intersection of the two dimensions. The H-ES items would then need to be revised to clean out any aspects of the "group" dimension.

While I have suggested that the H-ES has certain reliability and validity issues, the general trends that the CCP spotted in these other studies probably still exist. In this particular study, the scale still had predictive validity and it likely has predictive validity in others. But, there is a loss in accuracy and we cannot be certain what the factor scores actually represent. It is difficult to understand the research domain and interpret the score measures in a meaningful way.[184]

180. Haynes et al., *supra* note 24, at 244–45.

181. When the analysis suggests "that conceptualization of the target construct as, for example, a single bipolar dimension is countermanded by evidence that the two poles actually represent separate and distinct entities. ... revision of one's theoretical model may be in order." Clark & Watson, *supra* note 17, at 311–12.

182. Analysis of items used to measure the Douglas cultures reveals two dimensions. *See* Rippl, *supra* note 86, at 152.

183. *See* discussion *supra* note 61.

184. Floyd & Widaman, *supra* note 144, at 287.

Conclusion

The data suggests that beliefs related to patriarchy rather than worldviews related to hierarchy explain the guilt judgments in acquaintance rape cases. This finding suggests that reform efforts do not need to be targeted at creating separate organizations that are independent of existing, hierarchical law enforcement, military, university, or religious organizations. Rather, targeted training on certain gender role beliefs—particularly beliefs about the ways that women should behave sexually—may create bias-free organizations. The findings of my study are consistent with other findings in the field, and those findings show that those who hold traditional gender role or patriarchal beliefs subscribe to rape myths to a higher degree than those who do not hold those beliefs. When people use rape myths to resolve the discrete rape problem found in the legal file on their desk, they tend to find in favor of the man. Targeting those belief systems, rather than overhauling the organizations, may be enough to ensure bias-free rape case processing.