This postprint was originally published by Elsevier as:
Feltz, A., Cokely, E. T., & Tanner, B. (2021). **The free will and punishment scale: Efficient measurement and predictive validity across diverse and nationally representative adult samples.** *Consciousness and Cognition, 95*, Article 103215. https://doi.org/10.1016/j.concog.2021.103215.

The free will and punishment scale: Efficient measurement and predictive validity across diverse and nationally representative adult samples

Adam Feltz [a,b,*], Edward Cokely [a,c,d], Braden Tanner [a]

[a] University of Oklahoma, United States
[b] Center for Applied Social Research, United States
[c] National Center for Risk and Resiliency, United States
[d] Max Planck Institute for Human Development, Berlin, Germany

* Corresponding author.
  E-mail address: afeltz@ou.edu (A. Feltz).

Abstract

Theoretically, attitudes about freedom and punishment can shape people's decisions and cause pernicious disagreements (e.g., political policies). Several scales measure free will beliefs, partially to help understand disagreements about theoretical and practical issues. We contribute to these efforts by directly comparing existing measures and by introducing a short measure of free will related attitudes. Studies 1, 2, and 3 (Ns = 221, 225, 244) factor analyzed all items in existing scales of free will and moral responsibility, resulting in two prominent factors: Beliefs in Free Will and Beliefs in Punishment. Study 4 (N = 269) presents evidence for the 2-factor structure from a nationally representative sample. Study 5 (N = 108) gives evidence of the utility of the Free Will and Punishment scale in predicting free will relevant beliefs and attitudes. As such, the Free Will and Punishment scale is likely useful when longer instruments are not practically possible.

Keywords: Free will, Moral responsibility, Free will scales

1. Introduction

Beliefs about free will and moral responsibility form a cornerstone of how we understand ourselves and our place in the world. Theorists have argued that belief in free will is critical for things fundamental to human experience like love, friendship, respect, conceptions of uniqueness, desert, and merit (Kane, 1996). For reasons such as these, researchers have been interested in measuring how free and morally responsible people think that they are. There are a number of different instruments that have been developed to measure free will related beliefs, but to date no direct comparisons among these instruments have been conducted. These comparisons are necessary in order to determine whether instruments measure similar factors, to determine the relative predictive ability of instruments, and to contribute to a cumulative science about free will beliefs. In five studies, we offer evidence about the relations and predictive power of different measures of free will related beliefs. We also provide a 5-item measure of free will beliefs and punishment motivation with data from a nationally representative sample that offers substantial predictive power for judgments, attitudes, and behaviors related to free will. We close by discussing some practical and theoretical issues concerning measuring free will beliefs.

2. Free will measures

Historically, philosophers have been concerned about the nature of free will and under what conditions one could be free or morally responsible. For example, one prominent contemporary debate is whether free will and moral responsibility are compatible with determinism.[1] That is, regardless of whether determinism is *actually* true or false (something that is fairly difficult to prove with any degree of certainty), *if* determinism were true, *would* we be free and morally responsible? A large amount of empirical work has been devoted to help answering this question (Feltz & Cova, 2014; Sommers, 2010).

In a somewhat different vein, there is growing body of research suggesting that people's judgments about whether they are actually free and morally responsible can be different from judgments about hypothetical cases asking whether one is free and morally responsible if certain conditions were true (e.g., determinism; Feltz, 2015). Theorists working in this vein have been interested not just in what people think about some hypothetical situations concerning whether one *would* be free and morally responsible if something *were* the case. Rather, these theorists have been interested in whether people think that they actually have free will. One common way to attempt to determine aspects of people's free will related beliefs is by developing a set of questions or prompts and having individuals respond to them (e.g., "I have complete free will"). By compiling responses to these prompts, scales can be developed that measure free-will related beliefs (e.g., free will, moral responsibility, punishment). Many researchers have developed scales to measure people's free will relevant beliefs (Nadelhoffer et al., 2014; Paulhus & Carey, 2011; Rakos et al., 2008; Stroessner & Green, 1990; Viney et al., 1984; Viney et al., 1982).

There are significant advantages to having instruments to measure people's free will relevant beliefs. For example, a free will instrument can provide a standard way to assess free will beliefs. If different instruments are used in studies, there will remain some question whether the results reflect something about the people (e.g., different demographics or experimental manipulations) or something about the instruments themselves. These lingering questions about instruments can thereby hinder a cumulative science by making comparisons among studies using different instruments more difficult. Currently there are a plurality of instruments that could be used to measure free will related beliefs. These instruments have substantial differences in language and identify different free will relevant factors. Currently, little is known about the relations among these scales. It could be that most free will instruments measure basically the same underlying constructs, in which case proliferation of instruments is unnecessary. Or, the instruments could measure quite distinct constructs or measure the same constructs with different levels of fidelity. This lack of comparison threatens one of the primary reasons for measurement instruments in the first place—the ability to compare the results from different studies and contribute to cumulative science.

Our primary prediction was similar free will subscales across instruments would largely measure the same underlying factors. Consequently, a secondary goal was to attempt to provide a short measure of prominent free will relevant factors. Many researchers might avail themselves of this short instrument to help measure and predict behavior that would be prohibited by larger instruments. Results from five studies supported both of these goals.

3. Study 1

The purpose of Study 1 was to provide correlations among different free will instruments. A secondary goal was to conduct an exploratory factor analysis including all of the items from the major free will related scales. The exploratory factor analysis was the first step in a series of anticipated studies to identify relations among the items in currently existing free will scales.

3.1. Participants. Two hundred and sixty-nine participants were recruited from Amazon's Mechanical Turk. We required that the participants take the survey from a United States IP address and that they had an approval rate from Mturk that was greater than 85%. We used these same restrictions for all studies conducted with samples taken from Amazon's Mechanical Turk. Forty-eight participants were excluded for not completing the survey. The mean age was 36.31, $SD = 14.38$, range 18–83. Fifty-nine percent ($N = 121$) were male.

3.2. Materials. Participants received and responded to each of these scales in random order, followed by basic demographic information.
   The Free Will Inventory (Nadelhoffer et al., 2014). The Free Will Inventory is composed of five subscales. The *Free Will* subscale (5 items) measures whether one thinks one has free will. The *Determinism* subscale (5 items) measures the degree to which one thinks that determinism is true. The *Dualism* subscale (5 items) measures the degree to which one thinks that the mind is distinct from the body. The *Free Will Beliefs* subscale (7 items) measures one's beliefs about free will. The *Moral Responsibility* subscale (7 items) measures one's beliefs about moral responsibility.
   The Attitude Toward Punishment Scale (Viney et al., 1982). The Attitudes Toward Punishment Scale (ATP) is a 15-item measure of one's general acceptance of types of punishment. Viney et al. (1982) also developed a Free Will and Determinism scale that measured beliefs about free will and determinism, but we did not include that scale in these analyses. The primary reason this scale was

---

[1] Determinism is the thesis that "at any instant exactly one future is compatible with the state of the universe at that instant and the laws of nature" (Mele, 2006, p. 3)

Table 1
Correlations among free will subscales from Study 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Free Will Inventory																			
1. Free will	1																		
2. Determinism	0.09	1																	
3. Dualism	0.20**	0.11	1																
4. Free Will Belief	0.43**	0.03		1															
5. Moral responsibility Belief	0.44**	0.14*	0.43**	0.61**	1														
Attitudes Toward Punishment																			
6. Attitudes toward punishment	0.33**	0.19**	0.40**	0.42**	0.60**	1													
Free Will and Determinism Scale																			
7. Moral responsibility	0.51**	-0.05	0.29**	0.51**	0.60**	0.49**	1												
8. Free will	0.64**	-0.06	0.37**	0.51**	0.48**	0.35**	0.61**	1											
9. Personal agency	0.61**	-0.01	0.31**	0.56**	0.49**	0.33**	0.69**	0.76**	1										
10. Higher power	-0.05	0.39**	0.30**	0	0.08	0.31**	-0.13	-0.10	-0.17*	1									
11. Personal responsibility	0.48**	-0.02	0.22**	0.5**	0.57**	0.41**	0.73**	0.56**	0.66**	-0.17*	1								
12. Personal limitations	0.07	0.34**	0.45**	0.19**	0.27**	0.39**	0.07	0.15*	0.05	0.77**	0.12	1							
FAD-Plus																			
13. Free will	0.66**	0.16*	0.29**	0.47**	0.46**	0.50**	0.64**	0.58**	0.63**	0.01	0.56**	0.19**	1						
14. Scientific determinism	-0.02	0.58**	-0.12	0.06	0.13	0.09	-0.04	-0.11	-0.05	0.24**	-0.03	0.17*	0.10	1					
15. Fatalistic determinism	-0.03	0.60**	0.31**	0.1	0.18**	0.37**	-0.07	-0.03	-0.08	0.61**	-0.05	0.60**	0.13	0.51**	1				
16. Unpredictability	0.14*	0.22**	0.17*	0.28**	0.28**	0.27**	0.18*	0.20**	0.22**	0.06	0.16*	0.06	0.25**	0.35**	0.34**	1			
Free Will Determinism Scale																			
17. Religious-philosophical determinism	0.06	0.44**	0.39**	0.15*	0.26**	0.41**	0.03	0.08	0.05	0.70**	0.06	0.76**	0.19**	0.30**	0.65**	0.22**	1		
18. Libertarianism	0.64**	0.02	0.24**	0.51**	0.37**	0.26**	0.51**	0.72**	0.70**	-0.16*	0.49**	0.04	0.62**	-0.06	-0.07	0.22**	0.03	1	
19. Psychological determinism	-0.20**	0.49**	-0.04	-0.07	0.05	0.11	-0.24**	-0.27**	-0.23**	0.28**	-0.17*	0.18**	-0.11	0.69**	0.49**	0.31**	0.39**	-0.27**	1
Mean	5.01	3.01	4.93	5.05	4.96	3.83	4.85	4.81	5.13	2.62	4.9	2.99	3.64	2.87	2.38	3.39	2.64	4.69	3.16
SD	1.34	1.51	1.71	0.86	0.9	0.96	0.86	1.01	0.91	1.11	1.04	1.59	0.67	0.73	1.03	0.58	1.32	1.17	1.12

Note. * $p < .05$, ** $p < .01$.

Table 2
Item analysis from Study 1.

Factor 1: Free Will

Item	Low M	Low SD	High M	High SD	F	η_p^2
Libertarianism 4	3.04	1.13	5.71	0.49	259.28	0.71
Free Will 4	3.43	1.27	6.43	0.78	220.41	0.68
FAD-Plus 20	2.67	0.93	4.70	0.46	209.28	0.67
Rakos Free Will 4	3.41	1.13	5.79	0.49	203.06	0.66
Personal Agency 3	4.00	1.00	5.95	0.23	200.95	0.66
Libertarianism 2	3.31	1.24	5.75	0.44	190.15	0.64
Rakos Free Will 5	3.39	1.23	5.79	0.46	183.32	0.64
Personal Agency 1	4.20	0.92	5.93	0.26	183.86	0.64
Personal Agency 2	4.25	0.94	5.95	0.23	174.36	0.62
Free Will 3	2.08	1.25	6.09	1.18	241.72	0.61
Libertarianism 3	3.57	1.22	5.75	0.44	156.88	0.60
Personal Agency 4	4.18	1.09	5.98	0.13	151.30	0.59
Rakos Free Will 3	3.71	1.24	5.80	0.48	137.90	0.57
Libertarianism 1	3.71	1.29	5.77	0.43	128.72	0.55
FAD-Plus 4	2.75	0.85	4.41	0.68	126.96	0.55
Free Will 5	3.51	1.55	6.36	1.10	121.12	0.54
Rakos Free Will 2	3.94	1.29	5.88	0.33	117.83	0.53
Free Will Belief 1	4.59	1.17	6.63	0.80	112.36	0.52
FAD-Plus 12	2.71	0.94	4.32	0.86	86.26	0.45
Higher Power 3	4.08	1.04	5.73	0.84	82.71	0.44
FAD-Plus 26F	2.88	1.03	4.39	0.73	77.37	0.42
FAD-Plus 8	3.59	0.90	4.73	0.59	61.81	0.37
Rakos Free Will 1	4.02	1.36	5.71	0.89	59.01	0.36
Free Will Beliefs 4	5.00	1.34	6.18	1.32	20.91	0.17

Factor 2: Religious Determinism

Item	Low M	Low SD	High M	High SD	F	η_p^2
Religious Philosophical Determinism 6	1.00	0.00	4.76	1.11	579.07	0.85
Personal Limitations 1	1.18	0.72	5.06	0.93	551.38	0.85
Higher Power 1	1.04	0.20	4.71	1.12	520.71	0.84
Religious Philosophical Determinism 4	1.04	0.20	4.61	1.12	496.55	0.83
Personal Limitations 2	1.06	0.24	4.59	1.12	477.42	0.83
Higher Power 2	1.08	0.34	4.75	1.15	470.47	0.83
Religious Philosophical Determinism 5	1.00	0.00	4.55	2.03	339.88	0.77
Religious Philosophical Determinism 1	1.00	0.00	4.43	1.38	311.40	0.76
Religious Philosophical Determinism 3	1.00	0.00	4.18	1.47	234.80	0.70
FAD-Plus 13	1.32	0.65	3.57	0.99	182.14	0.65
FAD-Plus 1	1.14	0.45	3.14	1.31	103.80	0.51
FAD-Plus 17	1.70	0.81	3.27	1.22	58.10	0.37

Factor 3: Punishment

Item	Low M	Low SD	High M	High SD	F	η_p^2
Attitudes Toward Punishment 4	1.57	1.00	5.44	0.75	487.10	0.83
Attitudes Toward Punishment 3	1.84	0.90	5.21	0.98	325.58	0.77
Attitudes Toward Punishment 8	1.90	1.07	5.25	0.91	291.59	0.75
Attitudes Toward Punishment 12	1.69	1.21	4.96	1.30	170.59	0.63
Moral Responsibility Beliefs 3	3.41	1.57	6.37	0.66	156.13	0.61
Attitudes Toward Punishment 2	2.94	1.41	5.52	0.61	146.14	0.60
Attitudes Toward Punishment 6	2.27	1.20	5.04	1.17	137.71	0.58
Attitudes Toward Punishment 11	2.51	1.31	5.13	0.93	136.20	0.58
Attitudes Toward Punishment 10	2.16	1.42	4.96	1.15	118.80	0.55
Attitudes Toward Punishment 9	1.96	1.31	4.75	1.37	109.49	0.53
Attitudes Toward Punishment 13	2.65	1.15	4.98	1.11	107.33	0.52
Moral Responsibility Beliefs 5	3.88	1.73	6.42	0.94	86.10	0.47
FAD-Plus 16	3.20	1.17	4.56	0.61	53.97	0.35
Attitudes Toward Punishment 15	1.88	1.17	3.75	1.71	40.73	0.29
Attitudes Toward Punishment 1	2.18	1.45	4.21	1.79	38.90	0.28

Factor 4: Moral Responsibility

Item	Low M	Low SD	High M	High SD	F	η_p^2
Psychological Determinism 3	1.66	0.75	4.54	0.95	316.32	0.74
Psychological Determinism 5	1.57	0.81	4.61	1.04	298.55	0.73
Psychological Determinism 2	2.04	1.08	4.82	0.90	221.02	0.67
Psychological Determinism 6	1.86	1.07	4.37	1.29	126.81	0.54
Psychological Determinism 1	2.13	1.19	4.46	1.06	120.34	0.52
Psychological Determinism 7	2.39	1.33	4.64	0.90	109.64	0.50
Psychological Determinism 4	1.18	0.39	3.30	1.51	103.75	0.49
FAD-Plus 10	1.59	0.76	3.34	1.10	96.15	0.47

(continued on next page)

Table 2 (*continued*)

Factor 1: Free Will						
Item	**Low *M***	**Low *SD***	**High *M***	**High *SD***	***F***	η_p^2
FAD-Plus 24	2.23	1.01	3.80	0.84	80.21	0.42
FAD-Plus 14	2.46	1.14	4.04	0.74	74.67	0.40
FAD-Plus 6	1.61	1.09	3.30	1.26	57.86	0.35
FAD-Plus 21	2.48	1.11	3.80	0.92	46.84	0.30
Factor 5: Environmental Influences						
Item	**Low *M***	**Low *SD***	**High *M***	**High *SD***	***F***	η_p^2
Moral Responsibility 3	**3.54**	**0.68**	**5.96**	**0.20**	**589.49**	**0.86**
Moral Responsibility 5	**3.90**	**0.84**	**5.94**	**0.24**	**273.20**	**0.74**
Moral Responsibility 4	**3.56**	**1.05**	**5.90**	**0.30**	**228.07**	**0.70**
Moral Responsibility 2	4.26	0.80	5.98	0.14	22.33	0.69
Personal Responsibility 1	3.34	1.17	5.80	0.45	192.00	0.66
Moral Responsibility 1	3.48	1.07	5.76	0.52	182.99	0.65
Moral Responsibility 6	3.82	1.08	5.88	0.39	160.79	0.62
Attitudes Toward Punishment 5	4.14	1.05	5.50	0.79	53.62	0.35

Note. Item numbers refer to the item numbers in the original scales. Bold items were retained for Study 2.

not included was because the response options available to question were categorical (i.e., participants selected a sentence that best characterized their view) whereas all of the other instruments in our studies used Likert scale responses. To include Viney et al's Free Will and Determinism scale would require substantially altering that scale and thereby could introduce unknown measurement properties, making comparisons with other scales less reliable.

   The Free Will and Determinism Scale (Rakos et al., 2008). The Free Will and Determinism Scale measures six different free will relevant factors. The *Moral Responsibility* subscale (5 items) measures how morally responsible one thinks one is for actions. The *Rakos Free Will* subscale (5 items) measures how free one thinks one is. The *Personal Agency* subscale (4 items) measures how in control one thinks one is of one's actions. The *Higher Power* subscale (3 items) measures how much one thinks one is influenced by a higher power. The *Personal Responsibility* subscale (2 items) measures how blameworthy one thinks one is for one's action. The *Personal Limitations* scale (2 items) is a measure of how limited one is in acting, mostly because of the influence of some higher power.

   The FAD-Plus (Paulhus & Carey, 2011). The FAD-Plus measures four factors. The *FAD Free Will* subscale (7 items) measures whether one thinks that people have free will. The *Scientific Determinism sub*scale (7 items) measures whether people think that the world operates deterministically. The *Fatalistic Determinism* subscale (5 items) measure whether one thinks that the world operates fatalistically. The *Unpredictability* subscale (8 items) measures the extent to which people think that the world operates unpredictably.

   The Free-Will Determinism Scale (Stroessner & Green, 1990). The Free-Will Determinism Scale measures three factors. The *Religious Determinism* subscale (6 items) measures whether one thinks that God has a plan for one's life. The *Libertarianism* subscale (4 items) measures how much free will one thinks one has. The *Psychological Determinism* subscale (7 items) measures how much the environment influences one's choices.

   Demographic Information. We collected basic demographic information in all of the studies we conducted. The demographic information included gender, age, political orientation on a 7-point scale (1 = strongly liberal, 7 = strongly conservative), race, marital status, and income.

3.3. Results A sensitivity analysis suggested that the current study could reliably detect a minimum correlation of 0.19 ($power =$ 0.8, $alpha = 0.05$, two-tailed). There were strong correlations between similar subscales among the instruments (see Table 1). These relations suggest that many of the instruments measured the same underlying factors. To further explore these relations, all of the items from all of the scales were entered into an exploratory factor analysis. Twenty-four factors had eigenvalues greater than 1. The relatively large number of factors with eigenvalues greater than 1 is not surprising given the large number of items entered into the exploratory factor analysis. However, some factors explained more of the variance than others. Factor 1, the Free Will factor, explained by far the largest amount of the overall variance—about 20%. Factor 2, Religious Determinism, explained about 13% of the variance. Factor 3, Punishment, explained an additional 7% of the variance. Moral responsibility explained about 4% of the variance, and Environmental Influence explained an additional 3%. All other factors explained less than 3% of additional variance.

   Given that there were a handful of factors that explained the majority of the variance (while the remaining factors explained a small, trivial percent of the variance) and that there were many items that loaded on the major factors (greater than 12 for each), a smaller set of items would likely do as good a job as measuring the underlying factor as a larger set of items. To determine the most discriminating items for each factor, we conducted an item analysis. The logic of the item analysis was as follows. For each factor, rough quartiles were constructed based on the average response to the items in that factor (i.e., extreme groups analysis; Preacher, Rucker, MacCallum, & Nicewander, 2005). Then, each individual item was entered into an analysis of variance as the dependent variable and the high and low quartile for the factor as the independent variable. Those who are higher on the underlying factor should score the highest on the indicator item, and those who are lower on the underlying factor should score the lowest. In this way, we can identify items that provide the maximal discrimination between those low and high on the factor—items that are better than ones that

Table 3
Factor loadings from Study 2.

Item	Free Will	Moral Responsibility	Punishment	Environment	Higher Power
1	0.71				
4	0.41				
8	0.87				
2		0.73			
10		0.78			
15		0.68			
5			0.85		
6			0.65		
13			0.73		
7				0.63	
12				0.69	
14				0.46	
3					0.84
9					0.73
11					0.83

Note. All estimates are standardized.

only provide weak (or no) discrimination. The item analysis is reported in Table 2.

4. Study 2

Study 2 attempted to refine further the number of items based on the item analysis from Study 1. We selected the three best indicator items for the five factors identified in Study 1. We selected only 3 items for two reasons. First, we wanted to have a small set of items in an attempt to measure free will related beliefs. Second, previous research suggests that normally the minimum number of indicator items per factors should be three (MacCallum et al., 1999), even if there are some conditions where having fewer items is warranted (e.g., Gosling et al., 2003). So, as a starting point, we selected three items per factor. Study 2 was designed to submit these 15 items to a confirmatory factor analysis to provide additional evidence for the factor structure of free will relevant beliefs.

4.1. Participants. Two hundred and sixty-one participants were recruited from Amazon's Mechanical Turk. Thirty-six participants were excluded for not completing the survey. The mean age was 35.68, $SD = 13.24$ ranging from 18 to 72. Fifty-six percent were women ($N = 127$).

4.2. Materials. Participants responded to the 15 items identified from the item analysis in Study 1 (see Appendix). Then, basic demographic information was gathered.

4.3. Results. The confirmatory factor analysis of the 5-factor model failed all conventional fit criteria, but in informative ways, $\chi 2$ ($df = 80$) = 211.91, $p < .01$, $RMSEA = 0.09$, $90\% \ CI$ [0.07, 0.1], $p < .01$, $BIC = 10{,}871$, $AIC = 10{,}735$, $CFI = 0.9$, $TLI = 0.87$ (see Table 3 for factor loadings). The Free Will and Moral Responsibility factors were strongly correlated ($r = 0.9$), suggesting that the model could trim one of those factors without much loss of information. The Free Will and Environment factors had two items with relatively low factor loadings (standardized estimates = 0.63, 0.69, and 0.46), also suggesting that factor could be trimmed from the model. The Higher Power subscale was also eliminated from future studies, partially for empirical, but mainly for theoretical reasons. Theoretically, the items in the scale do not have much to do with free will and are not well represented in more recent free will scales. The Higher Power subscale represents choices being influenced and limited by higher powers, but those limitations are perfectly consistent with a person exercising free will (e.g., my choices are influenced by God's commands, but not determined or fated by those commands). Empirically, the higher power subscale is likely to be a predictor of one's general religious commitment (see Study 5 for evidence supporting this contention).

5. Study 3

Study 3 was designed to test the modified free will instrument based on the results of Study 2. Specifically, we eliminated a substantial number of items and subscales. Only the Free Will and Punishment subscales were retained (for justification, see above). One item was removed from the Free Will subscale because it had poor factor loading ("Luck plays a big role in people's lives," loading = 0.41). The remaining five items and two factors were then submitted to a confirmatory factor analysis.

Table 4

Free will and punishment scale items and subscales.

Item	Subscale
1. I have total free will.	Free Will
2. We could avert further offenses, if first offenders were not given such light sentences.	Punishment
3. Our courts and judges have been too lenient, that is one reason we have so much crime in our country.	Punishment
4. People ultimately have complete control over their decisions and their actions.	Free Will
5. The death penalty should be mandatory in cases of premeditated murder.	Punishment

Note. All items are measured on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree).
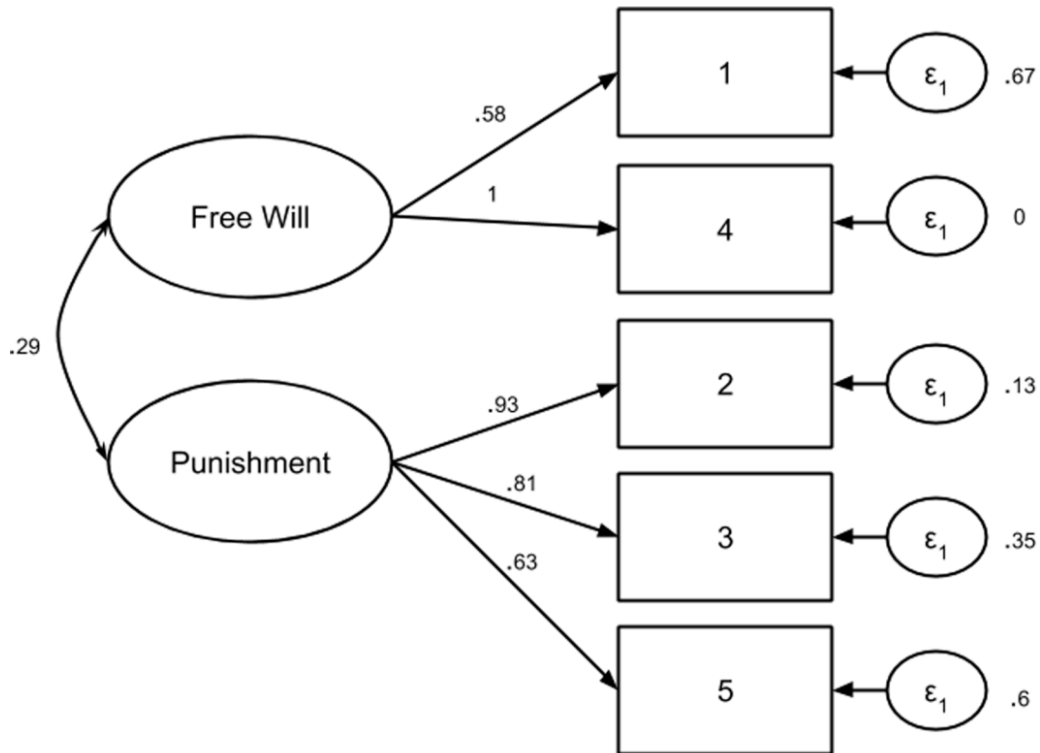

Fig. 1. Confirmatory factor analysis from Study 3 with standardized values.

5.1. Participants. Two hundred and seventy-six participants were recruited from Amazon's Mechanical Turk. Thirty-two were excluded for not completing the survey. The mean age was 35.17, $SD = 12.59$, $range = 19–83$, and 50% ($N = 115$) were male.

5.2. Materials. Participants received the 2-Item Free Will subscale and the 3-item Punishment subscale (1 = strongly disagree, 6 = strongly agree). We call this the Free Will and Punishment Scale (see Table 4 for item wording). Finally, basic demographic information was collected.

5.3. Results. A confirmatory factor analysis was conducted on the 2-factor model (see Fig. 1). The 2-factor model passed all conventional fit criteria: $\chi 2\ (df = 5) = 7.34$, $p = .2$, $RMSEA = 0.05$, $90\%\ CI$ [0, 0.11], $p = .47$, $BIC = 3642.54$, $AIC = 3694.98$, $CFI = 0.99$, $TLI = 0.99$.

Table 5
Descriptive statistics for free will and punishment scale subscales using the nationally representative sample from Study 4.

	N	Free Will Subscale			Punishment Subscale		
		Mean	SD	95% CI	Mean	SD	95% CI
Gender							
Male	138	4.26	1.15	[4.06, 4.45]	3.77	1.24	[3.57, 3.98]
Female	156	4.19	1.21	[4.0, 4.38]	3.84	1.23	[3.64, 4.03]
Age							
18–24	32	4.23	1.09	[3.84, 4.63]	3.75	1.01	[3.38, 4.12]
25–34	46	4.48	1.12	[4.15, 4.81]	3.69	1.15	[3.35, 4.03]
35–44	56	4.32	1.27	[3.98, 4.66]	3.78	1.47	[3.39, 4.17]
45–54	63	4.16	1.24	[3.85, 4.47]	3.82	1.15	[3.52, 4.11]
55–64	55	4.04	1.11	[3.74, 4.34]	3.78	1.34	[3.41, 4.14]
65+	42	4.13	1.19	[3.76, 4.50]	4.06	1.13	[3.70, 4.41]
Numeracy Score							
0–1	97	4.23	1.27	[3.97, 4.48]	4.12	1.23	[3.87, 4.36]
2	64	4.30	1.11	[4.03, 4.58]	3.91	1.21	[3.61, 4.21]
3–4	65	4.12	1.12	[3.84, 4.39]	3.58	1.07	[3.31, 3.85]
5–7	68	4.24	1.18	[3.95, 4.52]	3.49	1.31	[3.17, 3.81]

[2]The 90% confidence interval of the $RMSEA$ on the upper bound was slightly above the conventionally desirable threshold of 0.1, which may indicate some model misfit (Kline, 2011). The 2-Item Free Will ($Cronbach's\ alpha = 0.73$) and Punishment ($Cronbach's\ alpha = 0.82$) subscales had acceptable internal reliability. Taking the fit criteria together, the specified model has acceptable fit to the observed data.

6. Study 4

Study 4 was designed to provide evidence for the factor structure of the Free Will and Punishment scale from a national, representative sample. Data from Amazon's Mechanical Turk are generally reliable for many tasks, but there are some known issues with those data including non-naiveté and inattentiveness (M. Buhrmester et al., 2011; M. D. Buhrmester et al., 2018; Chandler et al., 2014; Thomas & Clifford, 2017). Data from a national sample can help alleviate some of the worries associated with Amazon's Mechanical Turk data

6.1. Participants. Two hundred and ninety-four participants were recruited using the Knowledge Networks recruitment service. These participants were recruited using random-digit dialing, one of the best ways to ensure representative sampling. Fifty-three percent ($N = 156$) were women. The mean age was 46.44 ($SD = 16.2$).

6.2. Materials. Participants received the Free Will and Punishment scale. Participants also completed the Ten Item Personality Inventory. Previous research has shown that personality—in particular, extraversion—has been related to free will related judgments (Feltz & Cokely, 2009; Feltz & Cokely, in press-a; Paulhus & Carey, 2011; Schulz et al., 2011). Finally, participants completed the Berlin Numeracy Test (Cokely et al., 2012). Numeracy has been related to better decision making across a host of domains such as finance, medicine, and health (Cokely et al., 2018). We included personality and numeracy to provide evidence for these relations. Basic demographic information was then gathered.

6.3. Results and discussion. The first step in analysis was to confirm the factor structure of the Free Will and Punishment scale (means, standard deviations, and confidence intervals for age, gender, and numeracy can be found in Table 5). The factors of the scale were specified in the same way as in Study 3. The model suggested that two items had correlated error variances (Items 2 and 3) for the Punishment subscale, so we allowed those variances to be correlated in the confirmatory factor analysis. Given those correlated variances, the model passed all conventional fit criteria: $\chi2\ (df = 3) = 2.34$, $p = .51$, $RMSEA = 0$, 90% $CI\ [0, 0.09]$, $pclose = 0.75$, $BIC = 4907.75$, $AIC = 4863.54$, $CFI = 1$, $TLI = 1$; see Fig. 2 for standardized loadings). The fit was substantially better than an alternative one-factor model with the same

---

[2] The error variance for Item 4 of the Free Will and Punishment Scale was negative but very close to zero (1^{-09}). In a subsequent, unreported study to verify this, the error variance was again negative but close to zero. Because the error variance was so close to 0, the error variance for Item 4 was constrained to 0, a tested practice in instances where the error variance is near zero (Chen et al., 2001).
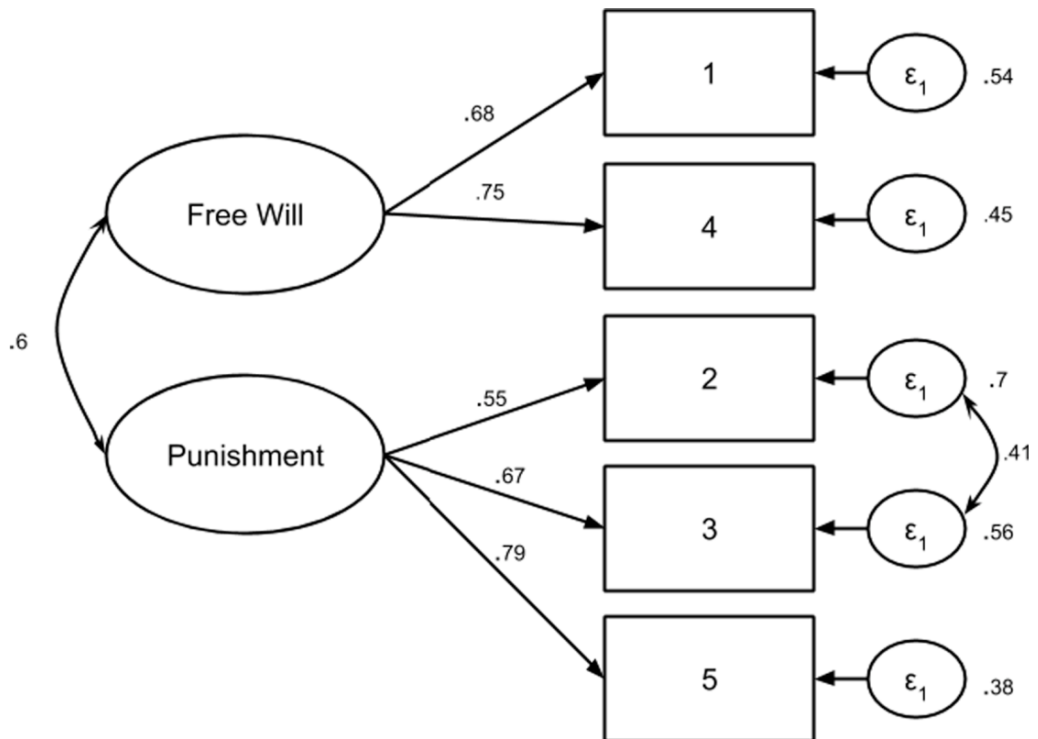
Fig. 2. Confirmatory factor analysis from Study 4 with standardized values.

Table 6
Correlations from Study 4.

	1	2	3	4	5	6	7	8	9
1. Free Will Subscale	1								
2. Punishment Subscale	0.41**	1							
3. Age	−0.07	0.09	1						
4. Numeracy	0.01	−0.18**	−0.04	1					
5. Extraversion	0.15**	−0.04	−0.02	0.03	1				
6. Agreeableness	0.06	0	−0.08	0.11	0.37**	1			
7. Conscientiousness	0.03	0.01	−0.02	0.14*	0.19**	0.21**	1		
8. Emotional Stability	0.21**	0.04	−0.09	0.02	0.32*	0.37**	0.27**	1	
9. Openness	0.09	0.01	0.04	0.09	0.18**	0.09	0.28**	0.12*	1
Mean	4.22	3.81	46.44	2.77	4.28	4.3	4.16	4.24	4.04
SD	1.18	1.24	16.2	2.02	0.94	0.96	0.77	0.85	0.87

Note. * p < .05, ** p < .01

correlated error variances: $\chi 2\,(df = 4) = 42.13$, $p < .001$, $RMSEA = 0.18$, $90\%\,CI$ 0.13 - 0.23, $pclose < 0.001$, $BIC = 4941.85.$, $AIC = 4901.33$, $CFI = 0.9$, $TLI = 0.75$. The punishment subscale had acceptable internal reliably ($Cronbach's\,alpha = 0.76$) while the 2-Item Free Will subscale had relatively weaker internal reliability ($Cronbach's\,alpha = 0.67$). We can conclude that the preferred 2-factor model has acceptable fit to the data.

   Correlations are reported in Table 6. A sensitivity analyses suggested that this study could reliably detect a minimum correlation of 0.16 ($power = 0.8$, $alpha = 0.05$, two-tailed). Replicating previous research, extraversion was correlated with the Free Will Subscale. Except for Emotional Stability, no other personality traits or demographic variables were strongly or significantly related to either of the two subscales. One intriguing, non-predicted result was that numeracy was related to lower punishment judgments. The relation suggests that those who are more numerate tend to have less retributivist tendencies toward punishment (see also Study 5).

7. Study 5

   Study 5 was designed to test whether the 2-factor Free Will and Punishment Scale fares well at predicting some judgments and beliefs compared to other scales.

Table 7
Correlations between the free will and punishment scale subscales and other free will scales in Study 5.

Scale	Free Will Subscale	Punishment Subscale	Mean	SD
Free Will and Punishment Scale				
Free Will Subscale	1		4.55	1.1
Punishment Subscale	0.21*		3.22	1.28
Free Will Inventory				
Free will	0.67**	0.31**	4.88	1.26
Determinism	−0.19*	0.17	2.97	1.42
Dualism	0.2*	0.29**	4.71	1.61
Free Will Beliefs	0.33**	0.14	4.84	0.93
Moral Responsibility Beliefs	0.31**	0.47**	4.64	0.82
Attitudes Toward Punishment	0.33**	0.81**	3.63	0.73
Free Will and Determinism Scale				
Moral responsibility	0.36**	0.41**	4.81	0.77
Free will	0.65**	0.3**	4.58	1.03
Personal agency	0.69**	0.18	5.17	0.86
Higher power control	0.17	0.27**	3.4	1.06
Personal responsibility	0.20*	0.36**	3.7	0.56
Personal limitations	−0.02	0.32**	2.6	1.37
FAD-Plus				
Free will	0.49**	0.51**	3.72	0.56
Scientific determinism	0.06	0.17	2.79	0.53
Fatalistic determinism	−0.18	0.34**	2.22	0.89
Unpredictability	0.11	0.01	3.28	0.59
Free Will-Determinism Scale				
Religious-philosophical determinism	−0.05	0.29**	2.39	1.15
Libertarianism	0.67**	0.2*	4.39	1.22
Psychological determinism	−0.23*	0.06	3.11	0.83

Note. * $p < .05$. ** $p < .01$.

7.1. Participants One hundred and eight participants were recruited from a mid-western university's psychology department's subject pool. The mean age was 20.1, $SD = 3.3$, range 18–52. Seventy-six percent were female ($N = 82$).

7.2. Materials. Participants were given the Free Will and Punishment Scale along with all the free will measures used in Study 1. Participants then responded to following two criteria sets: judgments (criminal punishment) and attitudes (limits on free will).

*Punishment Judgments. (*McCorkle, 1993*)*

A man approaches a couple in the parking lot of a casino and demands their cash. They refuse, so he pulls a knife stabbing and wounds them both. He then takes their money and runs away. Later he is arrested and brought to trial.

(1 = strongly disagree, 4 = strongly agree)

1. The primary concern with this criminal should be to make sure he is severely punished for his crime.
2. If the only way this criminal and others like him can be locked up is to build more prison, then so be it.
3. The only way to prevent this criminal from committing future crimes is to keep him locked up.
4. The courts are generally too easy on people who commit this sort of crime.

*Free Will Beliefs. (*items modified from* Viney et al. (1984))*
(7-point scale, 1 = strongly disagree, 7 = strongly agree).

1. All people, regardless of their level of intelligence, have free will.
2. Those are mentally disturbed still have free will.

Finally, participants completed the Berlin Numeracy Test (Cokely et al., 2012) and the Ten-Item Personality Inventory (Gosling et al., 2003). Basic demographic information was gathered.

Table 8
Correlations between the free will and punishment scale subscales, criterion variables, personality, numeracy, and politics from Study 5.

Outcome	Model	Variable	Adjusted R^2	df	F	p	R^2_{change}	F_{change}	$p\ F_{change}$
Punishment Judgments	1	Punishment	0.35	106	57.3	<0.001	0.35	57.29	<0.01
	2	2-Item Free Will	0.36	105	31.13	<0.001	0.02	2.57	0.06
	3	Attitudes Toward Punishment	0.41	104	26.08	<0.001	0.06	10.41	< 0.01
	4	Moral Responsibility Beliefs	0.43	103	21.45	<0.001	0.03	4.77	0.03
Free Will Beliefs	1	2-Item Free Will	0.14	106	18.27	<0.001	0.15	18.27	<0.01
	2	Libertarianism	0.18	105	12.35	<0.001	0.04	5.64	0.02
	3	FAD-Plus Free Will	0.20	104	9.92	<0.001	0.03	4.28	0.04

Note. * $p < .05$, ** $p < .01$.

Table 9
Regression analyses from Study 5.

Outcome	Model	Variable	Adjusted R^2	df	F	p	R^2_{change}	F_{change}	$p\ F_{change}$
Punishment Judgments	1	Punishment	0.35	106	57.3	<0.001	0.35	57.29	<0.01
	2	2-Item Free Will	0.36	105	31.13	<0.001	0.02	2.57	0.06
	3	Attitudes Toward Punishment	0.41	104	26.08	<0.001	0.06	10.41	< 0.01
	4	Moral Responsibility Beliefs	0.43	103	21.45	<0.001	0.03	4.77	0.03
Free Will Beliefs	1	2-Item Free Will	0.14	106	18.27	<0.001	0.15	18.27	<0.01
	2	Libertarianism	0.18	105	12.35	<0.001	0.04	5.64	0.02
	3	FAD-Plus Free Will	0.20	104	9.92	<0.001	0.03	4.28	0.04

7.3. Results and discussion. Consistent with previous studies, the Free Will and Punishment subscales had good internal reliability (Free Will subscale: Cronbach's alpha = 0.74, Punishment subscale: Cronbach's alpha = 0.83).[3] A sensitivity analysis suggested that the current study could reliably detect a correlation of 0.26 ($power = 0.8$, $alpha = 0.05$, two-tailed). The 2-Item Free Will and Punishment subscales from the Free Will and Punishment Scale predicted relevant subscales from other instruments (rs greater than 0.41, see Table 7). More conservative political orientation was related to more retributivist punishment judgments. The magnitude of the relation between the punishment subscale and numeracy was similar to what was found in Study 4, although not significant in this smaller sample. Entering extraversion, BNT, and political orientation in a multiple linear regression revealed that numeracy ($\beta = -0.19$, $t = 2.1$. $p = .04$) and political orientation ($\beta = 0.4$, $t = 4.49$. $p < .01$) were non-redundant, significant predictors of the Punishment subscale (Full Model $R^2 = 0.21$, $p < .01$). Extraversion trended toward being a significant predictor in this model ($\beta = 0.15$, $t = 1.7$. $p = .09$).

Next, composite scores were calculated for the punishment judgments (Cronbach's alpha = 0.73) and free will beliefs ($r = 0.66$). The composite score was the mean of responses to each of the criteria sets. The relevant subscale from the Free Will and Punishment Scale predicted the criteria items (see Table 8).[4]

To illustrate the predictive power of the Free Will and Punishment Scale compared to other scales, a series of hierarchical linear regression were conducted, one for each of the criteria sets. In the first block of each regression, the two Free Will and Punishment subscales were entered. In the second block, all of the remaining free will subscales were entered. Even when *all* of the other subscales were entered, at least one of the subscales from the Free Will and Punishment Scale was a significant predictor of the criterion set (see Table 9). More than that, the amount of variance explained by the Free Will and Punishment Scale was nearly as much as those predicted by even the best combination of longer scales. For Judgments, the Punishment subscale explained 76% of the variance that even the best model could predict. The 2-Item Free Will subscale similarly predicted 70% of the variance concerning Beliefs.

Finally, we conducted a series of direct, pairwise comparisons between the subscales of the Free Will and Punishment Scale and each of the other free will scales. In these analyses, we entered the two subscales from the Free Will and Punishment Scale and all of the subscales from one of the other instruments. The outcome variable in one set of regressions was Free Will Beliefs. In the other set of regression, the outcome variable was Punishment Judgments (full regressions are reported in Tables 10 and 11,). In nearly all the comparisons, the relevant subscale from the Free Will and Punishment Scale out-predicted any of the scales from the other instruments (see below for more discussion).

---

[3] We did not conduct a factor analyses on these data because the study was not powered to do so.
[4] As speculated, the Free Will and Determinism Higher Power subscale ($r = 0.66$) and the Free Will and Determinism Religious Determinism subscale ($r = 0.54$) were strongly correlated with general religious commitment "How religious are you?" on a 7=point scale, 1 = not at all 7 = extremely religious.

Table 10

Regressions Predicting the Free Will Beliefs Outcome Variable from Study 5 Note. The subscales from the Free Will and Punishment Scale (2-Item Free Will and Punishment) were entered along with the subscales from each of the other free will instruments. Italicized rows represent the best predictor

Free Will Beliefs	B	95% CI	SE	β	t	p
Free Will Inventory						
2-Item Free Will	*0.4*	*[0.1, 0.69]*	*0.15*	*0.33*	*2.67*	*0.01*
Punishment	0	[−0.23, 0.22]	0.11	0	0.03	0.97
Free will	0.16	[−0.11, 0.42]	0.13	0.15	1.19	0.24
Determinism	0.13	[−0.05, 0.31]	0.09	0.14	1.42	0.31
Dualism	0.09	[−0.08, 0.26]	0.09	0.11	1.05	0.3
Free Will Beliefs	−0.09	[−0.43, 0.26]	0.18	−0.06	0.48	0.63
Moral Responsibility Beliefs	−0.12	[−0.52, 0.29]	0.2	−0.07	0.57	0.57
Attitudes Towards Punishment						
2-Item Free Will	*0.41*	*[0.19, 0.64]*	*0.12*	*0.35*	*3.6*	*< 0.01*
Punishment	−0.08	[−0.4, 0.23]	0.16	−0.08	−0.52	0.61
ATP	0.02	[−0.02, 0.06]	0.02	0.16	1.01	0.31
Free Will and Determinism Scale						
2-Item Free Will	*0.3*	*[−0.02, 0.61]*	*0.16*	*0.25*	*1.85*	*0.07*
Punishment	−0.05	[−0.27, 0.18]	0.11	−0.04	0.41	0.69
Moral Responsibility	0.38	[−0.06, 0.81]	0.22	0.22	1.73	0.09
Free Will	0.25	[−0.11, 0.62]	0.19	0.2	1.27	0.17
Personal Agency	−0.12	[−0.6, 0.37]	0.24	−0.08	0.48	0.64
Higher Power Control	0.34	[−0.09, 0.76]	0.21	0.27	1.58	0.12
Personal Responsibility	−0.44	[−1.24, 0.35]	0.4	−0.27	1.1	0.27
Personal Limitations	0	[−0.48, 0.47]	0.24	0	0.02	0.98
FAD-Plus						
2-Item Free Will	*0.33*	*[0.07, 0.57]*	*13*	*0.27*	*2.57*	*0.01*
Punishment	−0.04	[−0.26, 0.19]	0.12	−0.04	0.31	0.76
Free Will	0.57	[0.02, 1.12]	0.28	0.24	2.07	0.04
Scientific Determinism	−0.27	[−0.7, 0.19]	0.23	−0.11	1.16	0.25
Fatalistic Determinism	−0.02	[−0.31, 0.28]	0.15	−0.01	0.11	0.92
Unpredictability	0.16	[−0.25, 0.57]	0.21	0.07	0.77	0.44
Free Will-Determinism Scale						
2-Item Free Will	0.22	[−0.07, 0.51]	0.14	18	1.52	0.13
Punishment	0.03	[−17, 0.22]	0.1	0.03	0.28	0.78
Religious-philosophical determinism	0.04	[−18, 0.26]	0.11	0.04	0.37	0.72
Libertarianism	*0.29*	*[0.03, 0.55]*	*0.13*	*0.27*	*2.23*	*0.03*
Psychological Determinism	−0.11	[−0.41, 0.19]	0.15	−0.07	0.73	0.47

8. General discussion

In five experiments, the utility of using a short measure of free will related attitudes was tested. In general, the Free Will and Punishment Scale had good internal statistical properties and offered good predictive ability for some beliefs and judgments. To help further establish the utility of the Free Will and Punishment Scale, it should be noted that the measure is quick. It takes an average participant about 2 min and 20 s to take each of the Attitudes Toward Punishment and the Free Will and Determinism Scale, nearly 5 min to take the Free Will Inventory, more than 3 min to take the FAD-Plus, and 2 min and 45 s to complete the Free Will and Determinism Scale. However, the Free Will and Punishment Scale only took on average about 45 s to complete. The relative speed in which the Free Will and Punishment Scale can be completed means that the measure can be included in a wide variety of studies where total test time is a consideration. The time (and other) savings of the short measure can be considerable, especially when using expensive samples like nationally representative or specialized samples (e.g., lawyers, doctors, or judges). In addition to time savings, the Free Will and Punishment Scale also did better on some of the criterion items than some of the other scales. While the Free Will and Punishment Scale did not out-predict the *optimal* set of *all* other instruments, the Free Will and Punishment Scale out-predicted most of the instruments in head-to-head comparisons (see Table 12 for a direct checklist for direct comparisons). In the head-to-head comparisons, the only subscale that did better than the 2-Item Free Will subscale predicting Free Will Beliefs was Stroessner and Green's (1990) Free Will-Determinism Scale. This might not be surprising, because the largest single predicting item was the item that the Free Will and Punishment Scale retained from their scale. Similarly the only scale that outperformed the Punishment subscale predicting Punishment Jusgments was Viney et al's. (1982) Attitudes Towards Punishment Scale. Again, this may not be surprising since all of the items from the Punishment subscale came from that instrument.

If the Free Will and Punishment Scale predicts beliefs and judgments, then measuring attitudes about free will and moral responsibility can play important roles in appropriately informing people's decisions. People naturally vary on a number of fundamental philosophical values and beliefs (Feltz & Cokely, in press-a; in press-b). Our focus here was on beliefs surrounding free will. But people do not always make good decisions, or even decisions that are based on their own values. For example, some people may value deterrence rather than retribution when a person is punished. However, in particular contexts, they may decide that retribution is more

Table 11
Regressions Predicting the Punishment Judgments Outcome Variable from Study 5.

Punishment Judgments	B	95% CI	SE	β	t	p
Free Will Inventory						
2-Item Free Will	0.04	[−0.08, 0.15]	0.06	0.07	0.67	0.67
Punishment	*0.25*	*[0.17, 0.34]*	*0.04*	*0.53*	*5.75*	*< 0.01*
Free will	−0.02	[−0.12, 0.09]	0.05	−0.04	0.33	0.74
Determinism	−0.06	[−0.14, 0.01]	0.04	−0.15	1.81	0.07
Dualism	0.02	[−0.04, 0.09]	0.03	0.006	0.68	0.5
Free Will Beliefs	0.07	[−0.06, 0.21]	0.07	0.11	1.06	0.29
Moral Responsibility Beliefs	0.09	[−0.07, 0.25]	0.08	0.11	1.08	0.28
Attitudes Towards Punishment						
2-Item Free Will	0.04	[−0.05, 0.13]	0.04	0.08	0.97	0.34
Punishment	0.11	[−0.01, 0.23]	0.06	0.23	1.82	0.07
ATP	*0.02*	*[0.01, 0.04]*	*0.01*	*0.43*	*3.23*	*< 0.01*
Free Will and Determinism Scale						
2-Item Free Will	0.07	[−0.06, 0.2]	0.07	0.13	1.12	0.27
Punishment	*0.25*	*[0.15, 0.34]*	*0.05*	*0.51*	*5.33*	*< 0.01*
Moral Responsibility	0.07	[−0.11, 0.25]	0.09	0.09	0.77	0.44
Free Will	−0.05	[−0.2, 0.1]	0.08	−0.08	0.66	0.51
Personal Agency	0.02	[−0.18, 0.22]	0.1	0.02	0.16	0.87
Higher Power Control	0.05	[−0.13, 0.22]	0.09	0.08	0,51	0.61
Personal Responsibility	0.08	[−0.24, 0.41]	0.17	0.11	0.51	0.61
Personal Limitations	−0.02	[−0.22, 0.17]	0.1	−0.05	0.25	0.8
FAD-Plus						
2-Item Free Will	0.09	[−0.01, 0.19]	0.05	0.16	1.78	0.08
Punishment	*0.3*	*[0.21, 0.39]*	*0.05*	*0.62*	*6.37*	*< 0.01*
Free Will	−0.1	[−0.32, 0.12]	0.11	−0.1	0.99	0.37
Scientific Determinism	−0.06	[−0.24, 0.13]	0.09	−0.05	0.63	0.53
Fatalistic Determinism	−02	[−0.14, 0.1]	0.06	−0.03	−0.34	0.74
Unpredictability	0.16	[−0.01, 0.32]	0.08	0.15	1.91	0.06
Free Will-Determinism Scale						
2-Item Free Will	0.09	[−03, 0.2]	0.06	0.15	1.46	0.15
Punishment	*0.25*	*[0.17, 0.33]*	*0.04*	*0.51*	*6.14*	*< 0.01*
Religious-philosophical determinism	0.08	[−0.01, 0.17]	0.05	0.15	1.77	0.08
Libertarianism	0.01	[−0.1, 0.12]	0.05	0.02	0.22	0.83
Psychological Determinism	0.01	[−0.11, 0.13]	0.06	0.01	0.12	0.91

Note. The subscales from the Free Will and Punishment Scale (2-Item Free Will and Punishment) were entered along with the subscales from each of the other free will instruments. Italicized rows represent the best predictor.

Table 12
Best predictor of the free will beliefs and punishment judgments from stepwise regressions directly comparing the free will and punishment scale versus each of the other free will scales.

	Free will beliefs	Punishment judgments
Free Will Inventory	✓	✓
Attitudes Towards Punishment	✓	X
Free Will and Determinism Scale	✓	✓
FAD-Plus	✓	✓
Free Will-Determinism Scale	X	✓

Note. Criterion items are the composite scores of the Free Will Beliefs and Punishment Judgment questions. A '✓" indicates the Free Will and Punishment Scale was the best predictor. An 'X' indicates that some other scale was a better predictor than Free Will and Punishment scale.

important than deterrence. In those contexts, those people could be making a mistake since they are making decisions in accordance with less important values at the expense of more important values.

The impacts of interventions are speculative, but having a short instrument to measure free will beliefs could be included in longer studies involving debiasing and education. Debiasing and educational studies tend to be long, so additional time is valuable. Including longer, more detailed free will instruments may not be feasible for many of these studies because of time constraints. Hence, there can be some real value including the Free Will and Punishment Scale in these studies to estimate the effects of debiasing on free will related beliefs and corresponding associations of free will beliefs on behaviors. As a welcome side-effect, direct comparisons between studies would also be easier to make using a standard free will measure across studies.

We want to emphasize that the Free Will and Punishment scale is one efficient and practical way to measure free will beliefs.

However, if one is interested in finer-grained or specific free will beliefs, then one of the other free will instruments may be preferable. We hope that the data will help researchers with instrument selection (and provide results from a national sample (i.e., Study 4) to which searchers can compare their results). For example, regardless of theoretical approaches to measurement, most of conceptually similar subscales in different free will measures were strongly correlated with one another (see Table 1 for full correlations). Take the free will subscales for an example. The correlations among these subscales ranged from 0.58 to 0.72. These data suggest that the subscales from different instruments largely measure the same latent traits.

In summary, there appears to be several good ways to measure free will beliefs depending on one's goals. Perhaps ironically, the results from our studies represent one way that informing individuals can help promote freedom. In this case, it is the researchers' freedom that could be promoted. Now, researchers have direct comparisons among free will instruments to consult so that they can select the appropriate instrument given their goals. Researchers can be more intentional about the instruments they are selecting and why (e.g., general free will beliefs in relation to longer surveys; fine-grained instruments that measure specialized facets of freedom and moral responsibility). This information, then, seems to be freedom promoting because it allows researchers to do more of what they value doing: good science. In this light, we follow in the steps of the multitude of free will theorists who attempt to help people reason better to become freer.

CRediT authorship contribution statement

Adam Feltz: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Project administration, Supervision. **Edward Cokely:** Methodology, Formal analysis, Methodology, Writing – review & editing. **Braden Tanner:** Formal analysis, Data curation, Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Fifteen items from Study 2

Instructions: Please carefully read the following statements and indicate your level of agreement (*1 = strongly disagree, 6 = strongly agree)*

1. I have total free will. (Stroessner & Green, 1990).
2. A person always has choices and therefore should be punished for making choices to harm others (Rakos et al., 2008).
3. My decisions are influenced by a higher power (Rakos et al., 2008).
4. Luck plays a big role in people's lives (Paulhus & Carey, 2011). (Reverse scored)
5. We could avert further offenses, if first offenders were not given such light sentences (Viney et al., 1982).
6. Our courts and judges have been too lenient, that is one reason we have so much crime in our country (Viney et al., 1982).
7. My exercise of free will is limited by my upbringing. (Stroessner & Green, 1990)
8. People ultimately have complete control over their decisions and their actions. (Nadelhoffer et al., 2014).
9. A person's choices are limited by a higher power's plan for him or her. (Rakos et al., 2008).
10. Human beings actively choose their actions and are responsible for the consequences of their actions. (Rakos et al., 2008)
11. God has planned out my life. (Stroessner & Green, 1990).
12. My free will is limited by such social conditions as wealth, career, and class. (Stroessner & Green, 1990).
13. The death penalty should be mandatory in cases of premeditated murder. (Viney et al., 1982).
14. My present behavior is totally a result of my childhood experiences. (Stroessner & Green, 1990).

A person should receive appropriate punishment for choosing to engage in bad or harmful behaviors. (Rakos et al., 2008)

References

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13*(2), 149–154. https://doi.org/10.1177/1745691617706516

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaivete among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112–130. https://doi.org/10.3758/s13428-013-0365-7

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research, 29*(4), 468–508.

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J., Petrova, D., & Garcia-Retamero, R. (2018). Skilled Decision Theory: From intelligence to numeracy and expertise. In A. Ericsson, R. Hoffman, A. Kozbelt, & A. Williams (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 476–505). Cambridge: Cambridge University Press.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. Judgment and Decision Making, 7 (1), 25-47. Retrieved from https://search.proquest.com/docview/927684255?accountid=12964, http://libraries.ou.edu/eresources/resolver.aspx?rft.genre=unknown&rft.atitle=Measuring+risk+literacy%3A+The+Berlin+Numeracy+Test&rft.au=Cokely%2C+Edward+T.%3B+Galesic%2C+Mirta%3B+ Schulz%2C+Eric%3B+Ghazal%2C+Saima%3B+Garcia-Retamero%2C+Rocio&rft.volume=7&rft.issue=1&rft.spage=25&rft.date=2012&rft.btitle=&rft.jtitle=Judgment+and+Decision+Making&rft.issn=1930-2975&rft.isbn=&sid=ProQ%3Apsycinfo_2012-03055-003.

Feltz, A. (2015). Experimental philosophy of actual and counterfactual free will intuitions. *Consciousness and Cognition, 36*, 113–130. https://doi.org/10.1016/j.concog.2015.06.001

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition, 18*(1), 342–350. https://doi.org/10.1016/j.concog.2008.08.001

Feltz, A., & Cokely, E. T. (in press-a). Extraversion and compatibilist intuitions: A ten-year retrospective and meta-analysis. Philosophical Psychology.

Feltz, A., & Cokely, E. T. (in press-b). Personality and philosophical bias. In J. Sytsma & W. Buckwalter (Eds.), A Campanion to Experimental Philosophy. New York: Wiley-Blackwell.

Feltz, A., & Cova, F. (2014). Moral responsibility and free will: A meta-analysis. *Conscious Cogn, 30*, 234–246. https://doi.org/10.1016/j.concog.2014.08.012

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

MacCallum, R. C., Widaman, K. F., Zhang, S. B., & Hong, S. H. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989x.4.1.84

McCorkle, R. (1993). Research note: Punish and Rehabilitate? Public attitdues toward six common crimes. *Crime and Delinquency, 39*, 240-252.

Mele, A. (2006). Free will and luck. Oxford; New York: Oxford University Press.

Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The Free Will Inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition, 25*, 27–41. https://doi.org/10.1016/j.concog.2014.01.006

Paulhus, D. L., & Carey, J. M. (2011). The FAD-Plus: Measuring Lay Beliefs Regarding Free Will and Related Constructs. Journal of Personality Assessment, 93(1), 96- 104. Pii 931472849, Doi 10.1080/00223891.2010.528483.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*(2), 178–192.

Rakos, R. F., Laurene, K. R., Skala, S., & Slane, S. (2008). Belief in free will: Measurement and conceptualization innovations. *Behavior and Social Issues, 17*(1), 20–40.

Schulz, E., Cokely, E. T., & Feltz, A. (2011). Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition, 20*(4), 1722–1731. https://doi.org/10.1016/j.concog.2011.04.007

Sommers, T. (2010). Experimental philosophy and free will. Philosophy Compass, 5, 192-212.

Stroessner, S. J., & Green, C. W. (1990). Effects of belief in free will or determinism on attitudes toward punishment and locus of control. *Journal of Social Psychology, 130*(6), 789–799.

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77*, 184–197. https://doi.org/10.1016/j.chb.2017.08.038

Viney, W., McIntyre, R., & Viney, D. W. (1984). Validity of a scale designed to measure beliefs in free will and determinism. *Psychological Reports, 54(3)*, 867–872.

Viney, W., Waldman, D. A., & Barchilon, J. (1982). Attitudes toward punishment in relation to beliefs in free will and determinism. *Human Relations, 35*(11), 939–949. https://doi.org/10.1177/001872678203501101