

A Novel Measure of the Need for Moral Cognition

Matthew Vanaman, Mary-Page Leggett, Laura Crysel & Robert Askew

To cite this article: Matthew Vanaman, Mary-Page Leggett, Laura Crysel & Robert Askew (2019): A Novel Measure of the Need for Moral Cognition, Basic and Applied Social Psychology, DOI: [10.1080/01973533.2018.1531000](https://doi.org/10.1080/01973533.2018.1531000)

To link to this article: <https://doi.org/10.1080/01973533.2018.1531000>



Published online: 12 Jan 2019.



Submit your article to this journal [↗](#)



Article views: 16



View Crossmark data [↗](#)



A Novel Measure of the Need for Moral Cognition

Matthew Vanaman^a, Mary-Page Leggett^b, Laura Crysel^c, and Robert Askew^c 

^aCUNY; ^bSeton Hall University College of Education and Human Services; ^cStetson University

ABSTRACT

This research developed the first measure of the *need for moral cognition*, or the tendency to seek out, talk about, reflect on, or otherwise engage with ethical or moral issues. Interpretative phenomenological analysis was conducted on transcripts from interviews with members of the target population and focus groups with content experts. Pilot items were administered to a large online sample; 2 latent factors were identified. Scores were assessed for test-retest reliability, and convergent-, discriminant-, and criterion-validity. Collectively, these studies demonstrate that the Need for Moral Cognition scale is a valid and reliable measure suitable for use in psychological research.

What is moral? Some individuals are more interested in finding the answer to this question than others. Although many prefer to follow the adage “never talk about politics or religion,” others are quite prone to engaging with such morally charged topics, particularly around the Thanksgiving table. What is different about those who prefer avoiding such topics and those who prefer steering family dinner conversation toward contentious moral subject matter? It could be that some enjoying thinking deeply for its own sake, predisposing them to spend more time thinking about moral issues than others. That is, some people have a high *need for cognition* (NFC), or “the tendency to engage in or enjoy thinking” (Cacioppo & Petty, 1982, p. 116), and have a pronounced tendency to “seek, acquire, think about, and reflect back on information to make sense of stimuli, relationships, and events in the world” (Cacioppo, Petty, Feinstein, & Jarvis, 1996, p. 198). Those high in the NFC may spend time seeking out, thinking about, and reflecting on moral issues, and there is indeed evidence to suggest that the NFC intersects with the moral lives of individuals. For example, those high in the NFC seem to derive higher cognitive enjoyment from morally ambiguous fictional characters (Krakowiak, 2015, p. 337) and come to counterintuitive utilitarian moral judgments more often than those low in the NFC (Wiech et al., 2013). Of interest, those high in the NFC tend to also show a high level of *moral sensitivity*, or the tendency to “perceive the presence of an ethical issue” (Sparks & Hunt, 1998, p. 92).

However, that an individual may have a higher-than-average tendency to *recognize* an element as a moral one does not necessarily say anything about one’s propensity to *engage* that element. Moreover, not everyone who is expected to have a high NFC, such as university faculty, becomes a moral philosopher, or even a hobbyist on the matter. Philosophy professors, for example, seem to have self-selected into areas of study requiring almost constant moral deliberation, whereas other high-NFC faculty preferred fields agonistic to moral questions (e.g., math). Taken to extremes, morals are viewed by some as a set of unquestionable or uninteresting rules to live by; for those at the other extreme, morals are a set of propositions to be explored curiously, reflected upon, and assessed on their merit.

This tendency to seek out, talk about, reflect on, or otherwise engage with issues of ethics or morality may be best described as a *need for moral cognition* (NFMC). Although the present study is the first to explicitly identify the NFMC by name, other research has investigated the effect of individual differences in the NFMC on moral judgment and behavior. Previous studies have reified the NFMC construct by asking whether those who self-identify as ethicists (Rust & Schwitzgebel, 2013; Schwitzgebel, 2009, 2013; Schwitzgebel & Rust, 2010; Schwitzgebel, Rust, Huang, Moore, & Coates, 2012), professional philosophers specializing in ethics (Schwitzgebel & Cushman, 2012; Schwitzgebel & Rust, 2010), and philosophy professors who teach ethics courses (Schwitzgebel & Rust, 2009)

exhibit different moral behavior than laypersons. Research along such lines has shown that the moral behavior of moral experts does not differ from laypersons (Rust & Schwitzgebel, 2013; Schwitzgebel, 2013; Schwitzgebel & Cushman, 2012; Schwitzgebel & Rust, 2009, 2010, 2014), whereas some studies found that moral experts behave *more* immorally than laypersons (Schwitzgebel, 2009; Schwitzgebel et al., 2012; Schwitzgebel & Rust, 2014). Other research has shown that moral experts may have different moral intuitions than the general population (Tobia, Buckwalter, & Stich, 2013).

As can be seen, studies examining behavioral differences between those with training in ethics or moral issues and those without have effectively presupposed a trait difference between these populations, that trait being the NFMC. If the NFMC is already being studied, what value does a self-report measure of the NFMC provide? So far, research has operationalized differences among individuals in the NFMC by categorizing participants as moral experts and nonexperts. This is, however, an indirect measurement of the NFMC, as these individuals are assumed a priori to be either high or low in the trait and categorized as such. A measure of the NFMC presents an opportunity to help determine whether having a heightened tendency to engage with moral issues affects moral judgment and behavior without relying on discrete categorization of participants according to researcher-defined criteria. A continuous self-report measure of the NFMC would also provide researchers with greater resolution of trait-differences among participants. It may also help to clarify future findings by enabling investigators to better control for the NFMC in research studying similar constructs. Last, given the field of psychology's recent calls for larger, more diverse, and more inclusive participant samples that better approximate the general population (Kitayama, 2017), a measure of the NFMC validated in a such a sample would allow researchers to study the trait in nonexpert populations.

Given these considerations, the current research sought to develop a valid and reliable measure of the NFMC. This was accomplished through the course of three studies, each with unique goals. Study 1 used qualitative methods to ensure content validity, using feedback from members of the target population and content experts to inform item development. Study 2 provided evidence of construct and structural validity by identifying and validating a factor structure in addition to testing hypotheses concerning the relationships between the NFMC scale scores and scores from

scales measuring theoretically related and unrelated constructs (i.e., convergent and discriminant validity). Last, Study 3 tested whether NFMC scales scores successfully discriminate among subpopulations known to differ in the trait (i.e., criterion validity) while examining the stability of scale scores across repeated assessments of the same individuals (i.e., test-retest reliability).

Study 1: Qualitative study and content validity

Study 1 sought to identify the various ways in which the NFMC manifests itself in the everyday life of individuals using an interpretative phenomenological analysis (IPA) approach as outlined by Smith (2015). One-on-one semistructured interviews (SSIs) with participants from the target population of English-speaking and reading adults were used to identify the ways in which the NFMC is expressed in the thoughts, experiences, and behaviors as reported from a first-person perspective. Following this, a group interview (focus group [FG] cohort) with a sample comprised of content experts provided additional conceptual clarity while helping to identify behavioral indicators of the trait they have noticed in other people. In short, following the IPA framework, we sought to identify indicators of the NFMC construct from *both* the self-evaluative and observer-expert perspectives.

Method

SSI cohort

Participants and procedures. Participants ($N=14$) were recruited through convenience sampling. The age of the sample ranged from 18 to 57 years ($M=26.9$, $SD=13.2$) and were mostly White ($n=10$) with four participants self-identifying as Hispanic or Latino. All participants spoke and read English as a first language. In sessions ranging from 30 min to 2 hr, participants were asked to describe the various ways in which morally laden issues are experienced in their day-to-day lives, for example, "Do you think about, read about, or otherwise contemplate moral issues when other people are not around?" and "Is there someone you know who approaches moral issues differently than you? If so, how are they different?" Guiding questions for both SSIs and the FG were developed following the guidelines of Smith (2015; Eatough & Smith, 2017) in conjunction with a review of the research literature on moral identity and moral judgment. Questions were left open-ended to allow participants to elaborate on their initial

responses and to allow the interviewer to record additional cognitive and behavioral indicators of the construct. The full questionnaire for the SSIs is available in section A of the appendices.

FG cohort

Participants and procedures. The FG cohort consisted of university faculty members with substantial experience teaching courses with moral and ethical content (e.g., business ethics, philosophy, and religious studies courses). Given the content covered in their courses, they were in a unique position to comment on behavioral indicators of those in their classes who are particularly high or low in the trait. The final sample consisted of seven faculty members whose ages ranged from 29 to 71 years ($M = 44.7$, $SD = 16.6$). The sample consisted of six men and one woman; each person was White and held a doctorate degree.

Responses from the SSI cohort informed the development of guiding questions for the FGs (see section B of the appendices for orientation script and full FG questionnaire). As with the SSIs, the FGs followed a loose script, and open-ended questions allowed respondents to elaborate with limited constraints and to respond to one another's comments. For example, "How would you recognize someone who has a heightened tendency to engage in moral cognition?" and "What differences do you notice between students who put extra thought into, ask more questions about, or see more layers in moral or ethical subject material than other students?" These questions sought to both clarify the nature of the NFMC construct and identify additional behavioral indicators.

Content analysis

Thematic content analyses following the methods outlined by Smith (2015) were conducted on the SSI and FG transcripts. The SSI and FG analyses were carried out as two independent but methodologically consistent procedures. In each case, transcripts were reviewed independently by coauthors. After separately identifying themes and sorting comments by theme, we met to identify areas of consensus and discrepancy. Retaining the commonly identified themes and comments, any discrepancies in theme identification or comment categorization were resolved through group deliberation until consensus was achieved. Themes identified in both the SSIs (representing the self-evaluative perspective) and the FGs (representing the expert-observer perspective) were consolidated. See sections C and D of the appendices for the full

list of themes and exemplary statements derived from SSIs and FGs, respectively.

Pilot items

After themes within the SSI and FG data were identified and responses appropriately sorted, a preliminary item pool was generated directly from respondent comments. Items were intended to closely resemble the vernacular of everyday speech and to represent their respective themes as they pertain to the NFMC construct. The number of items derived from each theme was roughly proportionate to the number of comments representing each theme. Items were subsequently inspected for content and face validity.

Results

The content analysis yielded five overlapping themes (e.g., Risk and Curiosity), two themes unique to the SSI data set (Reflection and Draining/Stimulating), and two themes unique to the FG data set (Willingness to Question Assumptions and Folk Relativism). A summary of themes (along with exemplary participant comments) from SSIs and FGs can be found in appendices C and D, respectively. A preliminary item pool ($N = 50$) was derived from participant responses and retained as much of the participants' original verbiage as possible (see section E of the appendices for the full list of preliminary items).

Discussion

Using an interpretative-phenomenological approach, we sought to identify the various cognitive, emotional, and behavioral manifestations of the NFMC construct by soliciting feedback directly from the target population via SSIs as well as from content experts in a FG setting. These methods uncovered fundamental elements of the construct and guided the development of 50 pilot items reflecting the types of statements that those particularly high or low in the NFMC would be expected to endorse in specific ways. These 50 items served as the preliminary item pool for Study 2.

Study 2: Development and preliminary validation of NFMC scale

The objective of Study 2 was to establish construct validity by identifying and validating the factor structure of the NFMC scale and assessing the performance

Table 1. Participant characteristics.

N	Semistructured interviews		Focus group		Validity sample		Test–retest sample	
	14		7		737		234	
M age in years (SD)	26.9 (13.2)		44.7 (16.6)		38.7 (13.2)		20.5 (3.8)	
Mdn age in years, (range)	21.0 (18.0–57.0)		41.0 (29.0–71.0)		36.0 (19.0–80.0)		20.0 (18.0–56.0)	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Sex								
Female	10	71.4	1	0.1	442	60.0	136	58.1
Male	4	28.6	6	0.9	288	39.1	96	41.0
Other	—	—	—	—	6	0.8	1	0.4
Missing	—	—	—	—	1	0.1	1	0.4
Race								
Caucasian	10	71.4	7	100.0	605	82.1	171	73.1
Black/African American	—	—	—	—	55	7.5	24	10.3
Native American	—	—	—	—	6	0.8	2	0.9
Asian/Pacific Islander	—	—	—	—	50	6.8	12	5.1
Other	4	28.6	—	—	21	2.8	23	9.8
Missing	—	—	—	—	—	—	2	0.9
Ethnicity								
Non-Hispanic/Non-Latino	10	71.4	7	100.0	680	92.3	186	79.5
Hispanic/Latino	4	28.6	—	—	56	7.6	46	19.7
Missing	—	—	—	—	1	0.1	2	0.9
Education level								
GED or high school diploma	—	—	—	—	96	13.0	—	—
Some college courses	11	78.6	—	—	168	22.8	—	—
Associate's degree or equivalent	—	—	—	—	101	13.7	—	—
Bachelor's degree or equivalent	2	14.3	—	—	259	35.1	—	—
Master's degree and above	1	7.1	7	100.0	113	15.3	—	—

of scale scores alongside scores from other theoretically related and unrelated measures. The dimensionality of the NFMC scale was assessed using exploratory and confirmatory factor analyses. Pearson's correlation coefficients among the NFMC scale scores and scores from measures of theoretically related and unrelated constructs served as evidence of convergent and discriminant validity.

Method

Recruitment, procedures, and exclusions

The field of psychology has been criticized for its overreliance on undergraduate research participants (Gallander Wintre, North, & Sugar, 2001), and previous findings indicate that Amazon Mechanical Turk (MTurk) participants are more diverse than samples consisting of undergraduate college students (Buhrmester, Kwang, & Gosling, 2011). With this in mind, participants for this study were recruited from Amazon MTurk via convenience sampling. Participants were administered a brief demographics questionnaire, the 50 items reflecting the NFMC, and several additional measures to facilitate convergent and discriminant validity testing. Participation was voluntary, and those who completed the survey were nominally compensated for their time. The survey was made available to those 18 years of age or older

residing in the United States, Canada, United Kingdom, and Australia.

Exclusions

To ensure integrity of the data, attention checks were embedded into each page of the survey (six checks total; e.g., “Please select ‘strongly disagree’ for this question”). We excluded responses from individuals who did not answer all the attention checks correctly ($n = 273$). Furthermore, we determined via consensus before data collection that the survey materials could not be completed thoughtfully in fewer than 7 min; responses from participants who completed the survey in less than 7 min ($n = 70$) were also excluded from the final sample. The median completion time was 15 min 17 s, and 16 min after exclusions.

Participants

Demographic characteristics of the sample are presented in Table 1 ($N = 737$). To enable validation of the dimensional structure of item responses, one half of the participants were randomly assigned to the development data set, and the remainder were assigned to the validation data set. The convergent and discriminant validity of the NFMC Scale was examined in the combined sample to reduce sampling error and to maximize precision in the estimation of population parameters (Trafimow & MacDonald, 2017).

Measures

The NFMC pilot scale. All 50 pilot items developed in Study 1 were administered. Participants were instructed to indicate the degree to which they agree or disagree with the items, responding on a 5-point Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

The NFC Scale. *Need for cognition*, defined as “the tendency to engage in and enjoy thinking” was assessed with the NFC Scale (Cacioppo & Petty, 1982, p. 116). The original 34-item NFC Scale demonstrates adequate construct validity and strong criterion validity (Cacioppo & Petty, 1982). This study used the 18-item short form developed by Cacioppo, Petty, and Kao (1984), which demonstrates similar dimensionality and internal consistency ($\alpha = 0.90$; $\alpha = 0.94$ observed in the sample in this study) as the 34-item form ($\alpha = 0.91$). It follows that if an individual is more likely than others to engage in thinking for its own sake, he or she should be more likely than others to also think about a specific domain (e.g., morality) for its own sake. It was therefore expected that the NFC Scale would demonstrate a medium-to-large correlation with the NFMC Scale.

The Need to Evaluate Scale. The Need to Evaluate (NTE) Scale measures *evaluative responding*, or the chronic tendency to engage in “the assessment of the positive or negative qualities of an object” (Jarvis & Petty, 1996). The NTE Scale has shown strong internal consistency ($\alpha = 0.83$ – 0.87 across three samples; $\alpha = 0.85$ observed in the current study) as well as strong test–retest reliability ($r = .84$). The scale also demonstrates strong criterion validity: Jarvis and Petty (1996) found that those high in the NTE are more likely to have formed attitudes toward a variety of social and political issues and are more likely to spontaneously generate opinions toward subjects (such as art) without provocation. It therefore stands to reason that the NTE should be theoretically related to the NFMC, as morality is a domain that requires evaluation of its precepts on the basis of their good or bad qualities. Thus, a medium-to-large correlation between the NTE and the NFMC scales was expected.

Openness to Experience and Extraversion subscales of the IPIP-120. Those high in Openness to Experience are characterized as being “imaginative and sensitive to art and beauty and have a rich and complex emotional life; they are intellectually curious, behaviorally flexible, and nondogmatic in their attitudes and values” (Costa & McCrae, 1992, p. 5). Extraversion is the

“dimension underlying a broad group of traits, including sociability, activity, and the tendency to experience positive emotions such as joy and pleasure” (Costa & McCrae, 1992, p. 5). McCrae and Costa developed a standardized instrument to assess the Big Five dimensions as outlined in their five-factor model (the NEO Personality Inventory; McCrae & Costa, 1985), though the instrument is expensive and time-consuming to administer. Maples, Guan, Carter, and Miller (2014) developed a free and relatively brief Item Response Theory–based 120-item personality measure (the IPIP-120) of the Big Five using items from the International Personality Item Pool, a database of personality measure items that resides in the public domain. The Openness to Experience scale possesses strong internal consistency ($\alpha = 0.85$ – 0.88 across two samples) as well as construct and criterion validity, as does the Extraversion scale ($\alpha = 0.85$ – 0.90 across two samples; Maples et al., 2014). Due to its ease of availability, the Openness to Experience and Extraversion subscales of the IPIP-120 instrument were used to measure openness to experience and extraversion. Responses from the sample in this study indicated an $\alpha = 0.87$ for the Openness to Experience subscale and an $\alpha = 0.91$ for the Extraversion subscale. Given that morality can be said to be both an intellectual and emotional endeavor, and that the act of questioning the validity of one’s beliefs requires openness to possible alternatives, we expected that Openness to Experience would demonstrate a medium to large correlation with the NFMC Scale. Although there is no theoretical reason to expect Extraversion to be positively correlated with the NFMC, Extraversion has been shown to be positively correlated with Openness to Experience (Aluja, García, & García, 2002). Thus, we expected a small but positive association between Extraversion and the NFMC.

The Need for Closure Scale. The need for closure (NFCL) is the “desire for an answer—any answer—on a topic compared to confusion and ambiguity” (Kruglanski, 1990). Although both the 45-item full version and the abbreviated 15-item version of the NFCL Scale are freely available for research or education purposes, the short form demonstrates similar internal consistency to the full measure ($\alpha = 0.87$ for the short form, $\alpha = 0.90$ for the long form), as well as similar convergent, discriminant, and criterion validity evidence (Roets & Van Hiel, 2011). This study used the short form of the NFCL, which exhibited reliability consistent with previous findings ($\alpha = 0.86$). The NFCL scale scores have been shown to demonstrate a

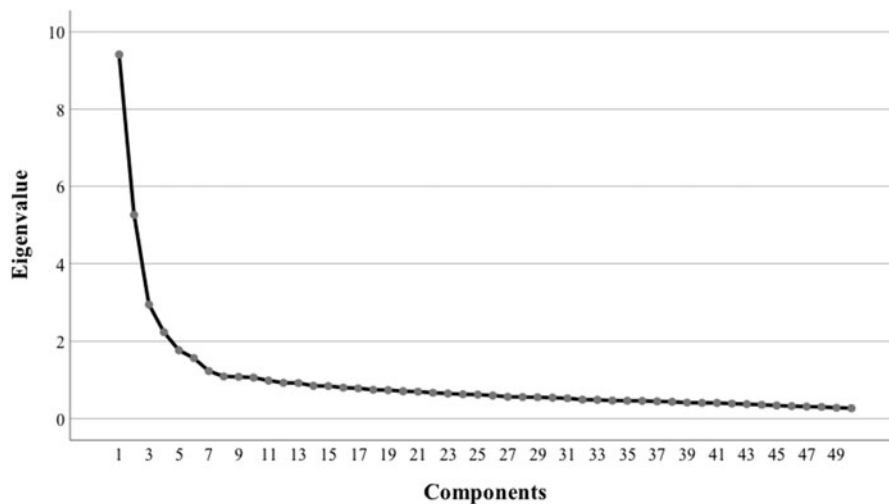


Figure 1. Scree plot from principal component analysis.

small, negative correlation with the NFC scale ($r_s = -.28$ to $-.29$; Roets & Van Hiel, 2011; Webster & Kruglanski, 1994). As noted by Webster and Kruglanski (1994), the NFCL may result in people attempting to avoid thinking in order to come to a *quick* resolution, or they may continue thinking in the hopes of an *eventual* resolution. Therefore, these two modes of thought—equally likely to manifest when one is confronted with any subject matter requiring effortful deliberation—should be also equally likely to occur when confronted with challenging moral propositions, which are often ambiguous or even inherently unsolvable. Thus, the NFCL was expected to demonstrate a similarly sized correlation with the NFMC as with the NFC, that is, a small-to-medium correlation.

Personal Need for Structure Scale. The Personal Need for Structure (PNS) Scale is the individual difference in the “tendency to reduce personal experiences into simplified representations” (Neuberg & Newsom, 1993). The PNS demonstrates adequate construct validity, internal consistency ($\alpha = 0.77$), and test-retest reliability ($r = .79$). Reliability observed in this study was consistent with previous observations ($\alpha = 0.83$). Given that the PNS was found to be independent of the NTE ($r = .03$; Jarvis & Petty, 1996, p. 172), and weakly correlated to the NFC ($r_s = -.23$ to 0.01 ; Neuberg & Newsom, 1993), the PNS should also be unrelated to the NFMC. However, because the PNS is also moderately negatively correlated with Openness to Experience ($r = -.42$), the PNS and NFMC scales were expected to demonstrate a small, negative correlation.

Emotional Intensity Scale. Emotional intensity (EI) is defined as “stable individual differences in the

strength with which individuals experience their emotions” (Geuens & De Pelsmacker, 2002). The EI Scale demonstrates strong internal consistency ($\alpha = 0.86$ – 0.91 across three samples; Geuens & De Pelsmacker, 2002), and strong convergent and discriminant validity. Reliability for the sample in this study was $\alpha = 0.86$ for the EI composite score and $\alpha = 0.84$ and 0.85 for the Positivity and Negativity subscales, respectively. It could be reasoned that those with especially high emotional intensity would be particularly motivated to engage with issues of morality (e.g., by moral outrage), or likewise particularly motivated to avoid them (e.g., by stress). In other words, the intensity (as opposed to frequency) with which one experiences emotions that could be positively related to moral engagement (e.g., moral outrage) or negatively related (e.g., stress) might be predictive of one’s willingness to engage with moral issues. On the other hand, previous research on the relationship between EI and constructs theoretically positively associated with the NFMC, such the NFC, have been shown to have weak relationships with EI (Geuens & De Pelsmacker, 2002). It was thus expected that the NFMC Scale would demonstrate at most a small but positive relationship with the EI Scale.

Results

Dimensionality

Before we conducted exploratory factor analysis (EFA), we used a principle component analysis (PCA) to determine the minimum and maximum number of factors to extract in the EFA. As shown in Figure 1, the PCA identified a minimum of two distinct components following Cattell’s scree test (elbow rule) and a maximum of 11 components following the Kaiser

criterion (eigenvalues ≥ 1). Informed by the results of the PCA, we next conducted a series of EFAs with solutions ranging from two to 11 factors, all with geomin rotations. It was determined a priori that items with factor loadings less than 0.50 and cross-loadings greater than 0.40 were to be excluded. We then compared solutions from the EFAs with the aim of identifying the solution that maximized model fit, factor interpretability, and simple structure. Following these a priori criteria, the two-factor solution was deemed optimal. Fit indices for this solution are presented in Table 2 along with factor loadings. Fit indices for EFA solutions 3 to 11 can be found in Appendix F.

Items composing Factor 1, Curiosity and Exploration, appeared to represent the degree to which people will search for and consider the rightness of moral positions other than their own (e.g., “I consider other people’s views on controversial topics before making up my mind”). Factor 2 appeared to capture the degree to which an individual is motivated to avoid negative consequences resulting from engaging with moral issues (e.g., “I am afraid of being wrong about what I believe”). As such, the items composing this factor represent Courage. A correlation of $-.26$ between the factors suggest that an individual could be both high in a willingness or drive to consider new moral frameworks but perhaps also feel intimidated by the prospect of discovering a perceived inadequacy in their personal virtues or at being perceived by others as morally lacking. The Curiosity and Exploration, Courage, and Composite (i.e., the grand mean of responses) scales each exhibited acceptable internal consistency reliability ($\alpha = 0.81, 0.73,$ and 0.74 , respectively).

The two-factor model was then fit to data from the validation sample using confirmatory factor analysis (CFA). Although chi-square suggested less than optimal model fit, multiple criteria indicated adequate fit, $\chi^2(26) = 133.88$, comparative fit index = 0.96 , Tucker–Lewis index = 0.95 , root mean square error of approximation = 0.11 (Hu & Bentler, 1999). All factor loadings were moderate to high in magnitude (i.e., ≥ 0.50), and only one residual correlation exceeded 0.20 (Items 5 & 9: $r = .24$) indicating limited local dependency. A weak correlation between factors was again found ($r = .33$), and all nine items identified in the EFA were retained after validation in the CFA. Internal consistency reliability of scales scores from the validation sample was consistent with the development sample ($\alpha = 0.82, 0.73,$ and 0.70 for the Curiosity and Exploration, Courage, and Composite scale scores, respectively). The final item set is

presented in Table 2 along with fit indices and factor loadings from the CFA.

Convergent and discriminant validity

Pearson’s correlation coefficients (r) were calculated to assess the relationships between NFMC scale scores and scores from measures of theoretically related and unrelated constructs (Table 3). In terms of directionality, all correlations were in line with a priori hypotheses regarding convergent and discriminant validity (as outlined in the Measures section). However, in some cases, effect sizes were smaller than anticipated (e.g., between scores from the NFMC and the NFC scales).

Post hoc exploratory analysis revealed several informative relationships between NFMC scores and scores from other measures (see Appendix H for the correlation matrix of all measures in this study). First, we found no correlation between NFMC scores and the Emotionality facet of Openness to Experience ($r = -.005$). This finding suggests that although the NFMC interacts positively with one’s intellectual habits, emotionality does not seem to correlate with one’s tendency to engage with moral and ethical subject matter. Second, NFMC scale scores demonstrated a very small correlation with the Assertiveness facet of Extraversion ($r = .08$), mitigating a possible concern that the NFMC might measure one’s tendency for debating or arguing—a common occurrence among those discussing morally charged topics. Third, these exploratory analyses revealed some important distinctions between the NFMC and the NFC, a measure of deep thinking. If the NFMC merely represents a general tendency for deep thinking, we would expect large correlations between NFMC composite and subscale scores and NFC scores. However, NFMC composite scores were only moderately correlated with NFC scores ($r = .46$), and the Curiosity and Exploration ($r = .34$) and Courage subscales ($r = .36$) were even less strongly correlated with NFC scores. Moreover, scores from the Curiosity and Exploration ($r = .34$) and Courage ($r = .35$) subscales of the NFMC were also more weakly correlated with the Intellect facet of Openness to Experience than the NFC ($r = .59$). This suggests that although there is shared variance among the NFC, Intellect, and NFMC subscales, the majority of the variance in the NFMC subscales is unique to deliberations of moral and ethical subject matter.

Discussion

Two distinct factors were identified and validated in Study 2: Curiosity and Exploration, and Courage.

Table 2. Model fit indices and items loadings.

		Exploratory factor analysis				Confirmatory factor analysis			
		Fit indices				Fit indices			
		χ^2 (df)				χ^2 (df)			
			3345.62 (1126)				133.88 (26)		
		CFI	0.793			CFI	0.962		
		TLI	0.775			TLI	0.948		
		RMSEA	0.074			RMSEA	0.105		
		<i>r</i>	-0.285			<i>r</i>	-0.329		
		C & E		Courage		C & E		Courage	
Item no.	Item	<i>L</i>	SE	<i>L</i>	SE	<i>L</i>	SE	<i>L</i>	SE
1	The beliefs of others are intriguing to me.	0.84	0.02	0.01	0.02	0.87		0.02	
2	I enjoy hearing other perspectives on social issues.	0.83	0.02	-0.03	0.03	0.91	0.02		
3	I often consider other peoples' views without necessarily agreeing with them.	0.67	0.03	-0.12	0.05	0.71	0.03		
4	Unusual world-views are often the most interesting.	0.67	0.03	-0.01	0.03	0.61	0.04		
5	I consider other peoples' views on controversial topics before making up my mind.	0.72	0.03	0.19	0.04	0.57	0.04		
6	I will only share my beliefs with certain people. (r)	0.00	0.03	0.67	0.03			0.71	0.03
7	I often just tell people what they want to hear when discussing a controversial topic. (r)	0.10	0.05	0.76	0.03			0.70	0.03
8	It makes me uncomfortable to share my views with others. (r)	-0.02	0.04	0.75	0.03			0.83	0.03
9	I am afraid of being wrong about what I believe. (r)	0.34	0.05	0.75	0.03			0.50	0.04

Note. Items were scored by taking the mean of the responses to Items 1, 3, 5, and 9 for the Courage subscale, Items 2, 4, 6, 7, and 8 for the Curiosity and Exploration (C & E) subscale, and all nine items for the Composite scale. (r) = reverse coded.

Curiosity and Exploration represents an individual's interest in comparing their own moral positions to those of others, whereas Courage appears to represent

one's tendency to engage with such issues despite the risk of negative consequences associated with doing so. Strong convergent and discriminant validity

Table 3. Correlations of the need for moral cognition scores composite and subscale scores with scores from other measures.

	Composite	Curiosity & exploration	Courage
Need for cognition	0.464	0.336	0.357
Need to evaluate	0.260	0.106	0.297
Openness to Experience	0.365	0.348	0.184
Imagination	0.207	0.300	-0.015
Artistic interests	0.255	0.243	0.129
Emotionality	-0.005	0.044	-0.059
Adventurousness	0.313	0.218	0.247
Intellect	0.457	0.337	0.347
Liberalism	0.084	0.115	0.007
Extraversion	0.273	0.199	0.210
Friendliness	0.238	0.124	0.243
Gregariousness	0.152	0.087	0.149
Assertiveness	0.078	0.067	0.055
Activity level	0.267	0.175	0.225
Excitement seeking	0.177	0.203	0.046
Cheerfulness	0.250	0.206	0.165
Need for closure	-0.300	-0.192	-0.261
Personal need for structure	-0.221	-0.158	-0.168
Emotional intensity	0.004	0.090	-0.093
Positivity	0.186	0.183	0.090
Negativity	-0.133	-0.009	-0.197

evidence supported the validity of NFMC composite and subscale scores through an assessment of the behavior of NFMC scale scores in relation to scores representing theoretically related and unrelated constructs, namely, the need for cognition, need to evaluate, openness to experience, the need for closure, personal need for structure, and emotional intensity.

Study 3: Longitudinal assessment and validation of NFMC scores

The purpose of Study 3 was to assess criterion validity and test-retest reliability of NFMC scale scores. Concurrent criterion validity evidence was assessed by examining differences in scores among subpopulations expected to differ in their level of the trait. That is, undergraduate students who choose to take courses in subjects ostensibly high in moral and ethical subject matter (philosophy and religious studies) were predicted to score higher than those who take classes with less moral content (music; science, technology, engineering, and mathematics [STEM]). A pre-post design was used to assess changes in NFMC scores over time and provide estimates of score precision (test-retest reliability).

Method

Participants and procedure

Undergraduate students ($N = 234$) ranging in age from 18 to 56 years ($M = 20.52$, $SD = 3.82$) were recruited from university courses in philosophy, religious studies, music, and STEM (for a list of the specific course

topics, see Appendix G). Participants were mostly White (73.1%), female (58.1%), and non-Hispanic/Latino (79.5%). Participants were administered the nine-item NFMC scale at both time points (T_0 , which occurred at the beginning of the school semester, and T_1 , which occurred 2 to 7 days after T_0).

Materials

NFMC scale. The measure consisted of the nine retained NFMC items developed in Study 1. Item responses were on a 5-point Likert scale, with 1 (*strongly disagree*) and 5 (*strongly agree*), and this scale was administered at both T_0 and T_1 .

Results

Criterion validity

To determine whether scores from the NFMC scale could distinguish among subpopulations expected to differ in their level of the trait, we compared the mean scores (at T_0) of students enrolled in philosophy ($n = 87$), religious studies ($n = 37$), music ($n = 70$), and STEM ($n = 38$) courses (see Table 4 for means and standard deviations). Hypotheses were supported: Students in philosophy and religious studies courses, which cover moral and ethical subject matter extensively, scored higher than students in music and STEM courses. Evidence was strongest for the Composite and Curiosity and Exploration subscale scores, though evidence was marginal (i.e., expected direction of effect but with small effect sizes) for the Courage subscale scores. Although our predictions were supported, effect sizes were smaller than anticipated (Cohen's $d_s = 0.37$ – 0.47 ; see Table 5 for effect sizes of pairwise comparisons among all course topics).

At the time of this study, the university where participants were recruited (a liberal arts institution) required its students to take a variety of courses in areas outside of their major. Therefore, one drawback to comparing students by course enrollment is that many students in these courses—particularly those in philosophy and religious studies courses—were not, in fact, majoring in the same content area as the course in which they were enrolled. A student's choice of major, which entails a commitment of several years of study on a particular topic, quite likely denotes far more interest in the topic than one's choice of enrolling in a single college course entailing a commitment of only one semester. We therefore followed our analysis of differences in NFMC scores by course enrollment with a post hoc evaluation of differences among majors. Table 4 reports the mean NFMC scores for

Table 4. Need for moral cognition scores by course topic and college major.

	<i>n</i>	Composite <i>M</i> (<i>SD</i>)	Curiosity & exploration <i>M</i> (<i>SD</i>)	Courage <i>M</i> (<i>SD</i>)
Course topic				
Philosophy	87	3.92 (0.43)	4.22 (0.47)	3.55 (0.82)
Religious studies	37	3.77 (0.38)	4.03 (0.55)	3.43 (0.88)
STEM	38	3.70 (0.51)	3.92 (0.65)	3.43 (0.79)
Music	70	3.71 (0.52)	4.03 (0.53)	3.32 (0.86)
College major				
Social sciences	32	3.98 (0.35)	4.28 (0.39)	3.60 (0.76)
Business/Pre-law	22	3.79 (0.46)	4.01 (0.60)	3.49 (1.04)
Philosophy/Religious studies	17	4.06 (0.29)	4.36 (0.37)	3.68 (0.69)
Arts, music, and humanities	79	3.73 (0.51)	4.04 (0.53)	3.34 (0.83)
STEM	56	3.70 (0.51)	3.95 (0.61)	3.39 (0.86)

Note. STEM = science, technology, engineering and mathematics.

Table 5. Cohen's *d* of pairwise comparisons of need for moral cognition scores by course topic.

	Philosophy	Religious studies	STEM	Music
Philosophy				
Composite	—			
Curiosity & exploration	—			
Courage	—			
Religious studies				
Composite	0.37	—		
Curiosity & exploration	0.37	—		
Courage	0.14	—		
STEM				
Composite	0.47	0.16	—	
Curiosity & exploration	0.53	0.18	—	
Courage	0.15	< 0.01	—	
Music				
Composite	0.45	0.13	-0.02	—
Curiosity & exploration	0.38	< 0.01	-0.19	—
Courage	0.27	0.13	0.13	—

Note. Effect sizes are between column groups and row groups, for example, $(\text{Philosophy}_{\text{Composite}} - \text{Music}_{\text{Composite}}) \div SD_{\text{pooled}} = 0.45$. STEM = science, technology, engineering and mathematics.

students from our sample with majors in STEM ($n = 56$); social sciences ($n = 32$); business and law ($n = 22$); arts, music, and humanities ($n = 79$); and philosophy and religious studies ($n = 17$). These categories were coded prior to our analysis, and those with majors in more than one of these categories (e.g., someone with a double major in business and philosophy) were excluded.

Our hypothesis that those studying philosophy or religious studies—content areas most concerned with moral and ethical subject matter—would score higher than those studying any other areas (e.g., STEM) was again supported. However, effect size differences (Table 6) were considerably larger when comparing students by major than by course enrollment: Cohen's *ds* between composite scores of those majoring in either philosophy or religious studies and any of business, arts, music, humanities, and STEM were all 0.7 or greater. Evidence was again strongest for the Curiosity and Exploration subscale but marginal for the Courage subscale scores. On the whole, these findings supported our ad hoc hypothesis that students would differ more by major in their

NFMC scores than they would course enrollment given the relative differences in commitment to a given topic.

Test-retest reliability. Although we have characterized the NFMC as a stable trait, it is possible that fresh exposures to moral and ethical subject matter (e.g., newly enrolled freshmen in morality-laden courses like philosophy and religious studies) might cause systematic changes in scale scores between T_0 and T_1 for some participants. All participants at T_1 were therefore prompted to answer whether their interest in moral and ethical subject matter had changed since the administration of the NFMC scale at T_0 . This item was placed at the end of the survey to ensure that participants would not feel prompted to be consistent in their responses. Only those who indicated no change between T_0 and T_1 ($n = 203$) were included in the test-retest sample. Interclass correlation coefficients (ICC) estimating absolute agreement in scores were selected because the data fit a two-way mixed-effects design—specifically *ICC(3,1)-Absolute* in the conventions of Shrout and Fleiss (1979), or *ICC(A,1)* in the conventions of McGraw and Wong (1996). ICCs indicated strong test-retest reliability across time points (ICC = 0.83, 0.83, and 0.83 for Composite, Curiosity and Exploration, and Courage scale scores, respectively). Changes in composite, Curiosity and Exploration, and Courage subscale scores between T_0 and T_1 were small (see Table 7 for effect size differences).

Because we excluded those whose interest in moral and ethical subject matter had changed, we may have artificially inflated stability estimates by excluding unsystematic changes in scale scores *along with* systematic changes. That is, we may have removed those whose interest in moral and ethical subject matter would have changed even without exposure to philosophy and religious studies courses. In this sense, the very people who might evidence inconsistency between time points were removed. To provide a more conservative estimate of test-retest reliability, we conducted an identical analysis on all participant data irrespective of interest change ($n = 230$). Results from this ICC did not differ (ICC = 0.83, 0.83, and 0.83 for Composite, Curiosity and Exploration, and Courage scale scores, respectively), and changes in effect size differences between administrations were negligible (see Table 7).

Discussion

The NFMC scale scores from Study 3 provided evidence of concurrent criterion validity by registering higher scores for subpopulations presumed to be

Table 6. Cohen's *d* of pairwise comparisons of need for moral cognition scores by major.

	Philosophy/Religious studies	Social sciences	Business/Pre-law	Arts, music, and humanities	STEM
Philosophy/Religious studies					
Composite	—				
Curiosity & exploration	—				
Courage	—				
Social sciences					
Composite	0.25	—			
Curiosity & exploration	0.21	—			
Courage	0.11	—			
Business/Prelaw					
Composite	0.70	0.46	—		
Curiosity & exploration	0.70	0.53	—		
Courage	0.22	0.12	—		
Arts, music, and humanities					
Composite	0.80	0.57	0.12	—	
Curiosity & exploration	0.70	0.52	-0.05	—	
Courage	0.45	0.33	0.16	—	
STEM					
Composite	0.87	0.64	0.19	0.06	—
Curiosity & exploration	0.81	0.64	0.10	0.16	—
Courage	0.37	0.26	0.10	-0.06	—

Note. Effect sizes are between column groups and row groups, for example, $(\text{Philosophy/Religious Studies}_{\text{Composite}} - \text{Social Sciences}_{\text{Composite}}) \div SD_{\text{pooled}} = 0.25$. STEM = science, technology, engineering and mathematics.

Table 7. Changes in need for moral cognition scores from T_0 to T_1 .

	<i>d</i>	Δ	ICC
Excluding interest change			
Composite	0.05	0.02	0.83
Curiosity and Exploration	0.09	0.05	0.83
Courage	-0.01	-0.01	0.83
Including interest change			
Composite	0.05	0.02	0.83
Curiosity and Exploration	0.10	0.05	0.83
Courage	-0.03	-0.02	0.83

Note. Effect sizes were calculated in accordance with recommendations of Dunlap, Cortina, Vaslow, and Burke (1996, S. 171, equation 3). ICC = intraclass correlation coefficient.

higher in the NFMC trait (those studying philosophy and religious studies) than those presumed to be lower in the trait (those studying STEM and music). Criterion validity was supported both when categorizing participants by course enrollment and when categorizing by college major. Last, data from repeated assessments indicated strong test-retest reliability and the relative stability of NFMC scores over time, and demonstrated that scale scores are robust to brief exposures to college courses with moral and ethical subject matter (viz., philosophy and religious studies courses).

General discussion

The goal of this series of studies was to develop a novel measure of the NFMC. Study 1 provided evidence of content validity by soliciting the perspectives of members of the target population in semistructured interviews and content experts in an FG session. A preliminary item pool was developed from written

transcripts of these sessions using Interpretative Phenomenological Analysis.

In Study 2, these items were tested for their structural and construct validity using a large sample of online participants. A two-factor solution was identified using EFA, and subsequently confirmed with CFA. One factor—*Curiosity and Exploration*—appeared to represent the degree to which one seeks out different moral perspectives. The other distinct factor—*Courage*—appeared to represent the degree to which an individual may be willing to engage moral issues despite the negative consequences associated with doing so. Correlations among scores from the final nine-item NFMC Scale were compared to those from scales measuring other theoretically related and unrelated constructs, meeting all a priori hypotheses.

Last, Study 3, which used responses from a sample of undergraduate students, showed that scores from the NFMC Scale differ among subpopulations expected to differ in their NFMC. More specifically, those enrolled in courses with relatively high amounts of morality-laden content (religious studies and philosophy courses) scored higher than those enrolled in courses that are not known to focus on issues of morality directly (STEM and music courses). These differences held true when categorizing participants by their college major, showing that philosophy and religious studies majors scored higher in the NFMC than those with majors in the social sciences, business and law, arts, humanities, and music. Moreover, the NFMC demonstrated strong test-retest reliability, showing that participants respond consistently to the items across repeated administrations.

Limitations and directions for future research

Study 1, although perhaps better than developing items without input from the target population and content experts, suffered from some disadvantages. Namely, the sample sizes were small, and we—rather than confederates naive to the study goals—conducted the interviews. Although there are advantages to using semistructured interviews over fully structured ones, particularly that participants have the opportunity to introduce new information into the conversation that would be constructive for item creation, there is also opportunity for the investigator to unconsciously influence the interviewee's responses. With respect to FGs, a prominent feature of this method is that participants interact and elaborate on one another's comments giving way to greater richness in responding (Kitzinger, 1994); however, it may also have the undesired effect of producing responses differing from those that would have been given privately. Additional steps could have been taken to ensure that items were validated to the desired level of content and face validity, such as seeking feedback from the target population about the preliminary item pool via cognitive interviewing before administering it in Study 2.

A key feature of the NFMC is the ability to measure the trait in the general population as opposed to relying on data from moral and ethical experts. As such, although the sample used to generate the preliminary item pool was assuredly more representative of the nonexpert population than, say, professional ethicists, it also could have been more diverse so as to achieve better demographics consistency with the general population at large. This same limitation could be applied to Study 2, as MTurk samples are slightly less diverse than the general population (Ross, Zaldivar, Irani, & Tomlinson, 2009). The NFMC Scale would therefore benefit from future validation in additional samples.

In Study 3, one issue of note is that although the NFMC Scale demonstrated evidence of concurrent criterion validity by distinguishing among subpopulations expected to differ in the trait, the mean differences were smaller than predicted. For example, the largest mean difference observed between any two groups was 0.36 (philosophy and religious studies majors and majors in STEM). These results may be explained by the fact that undergraduate students have had only brief exposure to their fields of study and that many students change majors or go on to pursue careers outside of their field of study. Although this small mean difference technically constitutes a large effect size ($d = 0.87$), future research

should nonetheless investigate whether scores between, for example, professional ethicists and laypersons elicit larger mean differences as these groups should differ more strongly. Furthermore, in regard to test-retest reliability, there was inconsistency in the amount of time between administrations at T_0 and T_1 such that the range (from 2 to 7 days) was wider than ideal. Future validation studies of this measure could involve additional assessment of test-retest reliability under stricter conditions.

More generally, the NFMC Scale requires additional evidence of its distinction from other similar constructs, such as moral sensitivity, or the tendency to recognize (as opposed to engage) moral content in one's environment (Sparks & Hunt, 1998). Furthermore, it might be argued that, given the NFMC items don't address moral issues directly (e.g., "The beliefs of others are intriguing to me"), our convergent and discriminant validity evidence may simply show that the NFMC represents a somewhat different form of deep or philosophical thinking than NFC, NTE, and Openness to Experience. More to this point, the NFMC Scale could have better distinguished itself from the NFC in terms of its correlations with other measures. For example, NFC and NFMC composite scale scores correlated with Openness to Experience about equally well ($r = .36$ for NFMC Composite and $r = .44$ for NFC). Moreover, given that philosophy and religious studies are topics requiring deep or philosophical thought, this alternative conceptualization may explain why those enrolled in philosophy and religious studies courses and majors scored higher on the NFMC than those enrolled in other courses and majors. As such, although the performance of NFMC scores were in line with a priori predictions, more construct validity evidence is needed to determine whether the scale successfully discriminates from other similar measures both in terms of convergent and discriminant validity and in terms of the ability to distinguish among subpopulations. Last, the NFMC Scale requires more evidence that it predicts relevant current (and future) attitudes and behaviors beyond student choices in courses and majors. One avenue would be to see whether the NFMC Scale predicts behaviors of professional ethicists, philosophers, and philosophy professors observed in previous research on moral engagement (e.g., Schwitzgebel & Rust, 2014). More specifically, showing that the NFMC Scale correlates with or predicts attitudes and behaviors associated with professional ethicists, philosophers, and philosophy professors—above and beyond

measures of deep or philosophical thinking—would serve as excellent criterion validity evidence.

Conclusion

The need for moral cognition—or the tendency to seek out, talk about, reflect on, or otherwise engage with issues of ethics or morality—is a trait already being studied by psychologists and philosophers (e.g., Schwitzgebel & Rust, 2014). A measure of the NFMC would be beneficial for advancing such research as well as research examining the effect of the moral deliberation on moral judgment and behavior and the effectiveness of interventions at increasing or decreasing the NFMC trait within individuals. It would help investigators determine whether being high in the NFMC makes one more likely to act in the service of others, whether leaders high in the trait are more concerned with the ethical implications of their decisions, or whether being high in the tendency to think about and otherwise engage with moral issues makes one less (or perhaps more) hypocritical than merely relying on one's moral intuitions. Despite its pertinence to these important questions, to date a measure of the NFMC has not been available for use in psychological studies. The three studies presented here sought to address this gap by developing and validating a sensitive measure of the NFMC capable of detecting differences among individuals as well as subpopulations expected to differ in their level of the trait. Taken together, the evidence provided here demonstrates that the NFMC scale is valid, reliable, sensitive to differences among individuals and subpopulations, and suitable for use in future research on moral judgment and behavior. This NFMC measure and scoring instructions are available in Appendix I and are free for all academic and educational purposes.

ORCID

Robert Askew  <http://orcid.org/0000-0002-9579-2279>

References

- Aluja, A., García, Ó., & García, L. F., (2002). A comparative study of Zuckerman's three structural models for personality through the NEO-PI-R, ZKPQ-III-R, EPQ-RS and Goldberg's 50-bipolar adjectives. *Personality and Individual Differences*, 33(5), 713–726. doi:10.1016/S0191-8869(01)00186-6
- 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. doi:10.1177/1745691610393980
- Cacioppo, J. T., & Petty, R. E., (1982). The need for cognition. *Journal of Personality and Social Psychology*, 41(1), 116–131. doi: 10.1037/0022-3514.42.1.116
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307. doi:10.1207/s15327752jpa4803_13
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253. <https://doi.org.ezproxy.gc.cuny.edu/10.1037/0033-2909.119.2.197>
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, 4(1), 5–13.
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1(2), 170.
- Eatough, V., & Smith, J. A. (2017). Interpretative Phenomenological Analysis. In C. Willig & W. Stainton-Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (2nd ed.). London, UK: Sage Publications.
- Gallander Wintre, M., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology/Psychologie Canadienne*, 42(3), 216–225. doi:10.1037/h0086893
- Geuens, M., & De Pelsmacker, P. (2002). Validity and reliability of scores on the reduced Emotional Intensity Scale. *Educational and Psychological Measurement*, 62(2), 299–315. doi:10.1177/0013164402062002007
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jarvis, W. B. G., & Petty, R. E. (1996). The need to evaluate. *Journal of Personality and Social Psychology*, 70(1), 172.
- Kitayama, S. (2017). Journal of personality and social psychology: Attitudes and social cognition. *Journal of Personality and Social Psychology*, 112(3), 357–360. <http://dx.doi.org/10.1037/pspa0000077>
- Kitzinger, J. (1994). The methodology of focus groups: the importance of interaction between research participants. *Sociology of Health and Illness*, 16(1), 103–121.
- Krakowiak, K. M. (2015). Some like it morally ambiguous: The effects of individual differences on the enjoyment of different character types. *Western Journal of Communication*, 79(4), 472–491. <https://doi-org.ezproxy.gc.cuny.edu/10.1080/10570314.2015.1066028>
- Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry*, 1, 181–197.
- Maples, J. L., Guan, L., Carter, N. T., & Miller, J. D. (2014). A test of the international personality item pool representation of the revised NEO personality inventory and development of a 120-item IPIP-based measure of the five-factor model. *Psychological Assessment*, 26(4), 1070–1084. doi:10.1037/pas0000004

- McCrae, R. R., & Costa, P. T. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources, Odessa.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. doi:10.1037/1082-989X.1.1.30
- Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology, 65*(1), 113–131. doi:10.1037/0022-3514.65.1.113
- Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences, 50*(1), 90–94. doi:10.1016/j.paid.2010.09.004
- Ross, J., Zaldivar, A., Irani, L., & Tomlinson, B. (2009). Who are the turkers? worker demographics in amazon mechanical turk. *Department of Informatics, University of California, Irvine, USA, Tech. Rep.*
- Rust, J., & Schwitzgebel, E. (2013). Ethicists' and Nonethicists' responsiveness to student Emails: Relationships among expressed normative attitude, self-described behavior, and empirically observed behavior. *Metaphilosophy, 44*(3), 350–371.
- Schwitzgebel, E. (2009). Do ethicists steal more books? *Philosophical Psychology, 22*(6), 711–725.
- Schwitzgebel, E. (2013). Are ethicists any more likely to pay their registration fees at professional meetings? *Economics and Philosophy, 29*(03), 371–380.
- Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language, 27*(2), 135–153.
- Schwitzgebel, E., & Rust, J. (2009). The moral behaviour of ethics professors: Peer opinion. *Mind, 118*(472), 1043–1059.
- Schwitzgebel, E., & Rust, J. (2010). Do ethicists and political philosophers vote more often than other professors? *Review of Philosophy and Psychology, 1*, 189–199.
- Schwitzgebel, E., & Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology, 27*(3), 293–327. doi:10.1080/09515089.2012.727135
- Schwitzgebel, E., Rust, J., Huang, L., Moore, A., & Coates, J. (2012). Ethicists' courtesy at philosophy conferences. *Philosophical Psychology, 25*(3), 331–340.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.
- Smith, J. A. (2015). *Qualitative psychology: A practical guide to research methods*. London, UK: Sage.
- Sparks, J. R., & Hunt, S. D. (1998). Marketing researcher ethical sensitivity: Conceptualization, measurement and exploratory investigation. *Journal of Marketing, 62*(2), 92–109.
- Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology, 26*(5), 629–638. doi:10.1080/09515089.2012.696327
- Trafimow, D. (2017). Using the coefficient of confidence to make the philosophical switch from a posteriori to a priori inferential statistics. *Educational and Psychological Measurement, 77*(5), 831–854.
- Trafimow, D., & MacDonald, J. A. (2017). Performing inferential statistics prior to data collection. *Educational and Psychological Measurement, 77*(2), 204–219.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049–1062. doi:10.1037/0022-3514.67.6.1049
- Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition, 126*(3), 364–372. <https://doi-org.ezproxy.gc.cuny.edu/10.1016/j.cognition.2012.11.002>