UNIVERSITY OF CALIFORNIA

Santa Barbara

Seeing What We Measure or Measuring What We See?
The Content-Analytic Deconstruction of Visual Moral Exemplification via Unidimensional and Multidimensional Codebooks

A Thesis submitted in partial satisfaction of the

requirements for the degree Master of Arts

in Communication

by

Musa Inayat Malik

Committee in charge:

Professor René Weber, Chair

Professor Ronald E. Rice

Professor Daniel Lane

December 2023

The thesis of Musa Inayat Malik is approved.

Ronald E. Rice

Daniel Lane

René Weber, Committee Chair

October 2023

Seeing What We Measure or Measuring What We See?

The Content-Analytic Deconstruction of Visual Moral Exemplification via

Unidimensional and Multidimensional Codebooks

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my mother, Farrah Inayat Malik, whose tireless determination to provide her children with an education has been a source of inspiration throughout my life. Her sacrifices have allowed me to dream beyond the confines of our lives and have instilled in me the values of hard work and perseverance. This achievement is a testament to her enduring love and belief in me.

I acknowledge my wife, Anna Maria Jaskiewicz, my best friend and unwavering pillar of strength. Her unconditional love, encouragement, and support have given me the courage to turn my dreams into reality. Anna's steadfast belief in me, even during the darkest moments of this journey, has served as my guiding light. She has enriched my life in ways words cannot fully convey.

I owe everything that I am to you both. Thank you.

ABSTRACT

Seeing What We Measure or Measuring What We See?

The Content-Analytic Deconstruction of Visual Moral Exemplification via Unidimensional and Multidimensional Codebooks

by

Musa Inayat Malik

Given the inherently subjective nature of moral content analysis (MCA), high intercoder reliability (ICR) statistics reported in extant MCA literature have been questioned. This paper argues that greater precision in identifying morally relevant content cues via multidimensional coding schemes can minimize the likelihood of subjective coder interpretations being applied during the coding task, thereby yielding high ICR statistics. Accordingly, and derived from Moral Foundations Questionnaire - 2 (MFQ-2), we offer the seminal version of the Moral Foundations Content Codebook (MFCC-1) which provides a rule-based framework for the identification of morally relevant cues, specifically in *visual content*. Our findings support the claim that a multidimensional MCA codebook, as opposed to a unidimensional one such as the coding manual for Model of Intuitive Morality and Exemplars (MIME), yields comparatively greater ICR statistics. However, these still remain below thresholds for minimum acceptable reliability. In our post-hoc analyses, we find that acceptable reliability for MFCC-1 can *also* be achieved but only after constraint adjustment

procedures that take into account the degree of coder disagreement (i.e., *slight* versus *severe*). Additionally, we observe enhancement of discriminant, predictive, and external validities when using ratings as derived from MFCC-1. We discuss these findings and provide explanations inspired from the cognitive bias literature that speculatively rationalize why coders using unidimensional MCA codebooks, in general, can be expected to perform worse than their counterparts.

I. Introduction

Moral Foundations Theory (MFT; Graham et al., 2013) suggests that there are five innate, universal moral foundations that exist among individuals across cultures, and that each foundation is associated with its virtue and vice dimensions that respectively indicate whether a particular moral foundation has been upheld or violated. These include Care / Harm (i.e., compassion, empathy, or sympathy toward others, particularly those in need or distress), Fairness / Cheating (i.e., equal and just distribution of rights, opportunities, and resources among individuals or groups), Loyalty / Betrayal (i.e., commitment or allegiance to one's in-group, expressed through actions that uphold and support the interests and wellbeing of the group), Authority / Subversion (i.e., recognition and respect for legitimate leadership, societal rules, and established hierarchies), and Purity / Degradation (i.e., respect for and adherence to principles cleanliness, sanctity, and moral integrity, often related to bodily, social, and spiritual concerns).

Drawing upon MFT and Exemplification Theory (Zillmann, 1999), the Model of Intuitive Morality and Exemplars (MIME; Tamborini, 2013; Tamborini & Weber, 2020) proposes a reciprocal association between media and audiences. According to the MIME, exposure to media content that exemplifies disparate innate motivations can heighten the significance of those motivations in the minds of audiences. As these intuitive motivations become more salient, viewers actively seek out additional media content that features these motivations. Consequently, media creators are motivated to produce content that aligns with the desired intuitive motivations of their audiences (Tamborini, 2013; Tamborini & Weber, 2020). For analytical purposes, the MIME operationalizes specific moral representations in media content along MFT dimensions. Indeed, several studies have utilized the pragmatic

utility of MIME variables across multiple domains of moral content analysis (MCA) research (e.g., Aley & Hahn, 2023; Hahn et al., 2017; Malik et al., 2021).

From a conceptualization perspective, while both the MFT and MIME have accelerated MCA research within the previous decade, an increasing number of alternative paradigms have been posited that undertake the dissection of specific MFT foundations, contending that individual moral foundations delineated by MFT potentially exhibit multidimensional characteristics. These include, *inter alia*, perceptions of fairness encompassing notions of equality or equity (Atari et al., 2023), recognition of multiple components within the construct of purity (Inbar & Pizarro, 2022), applications of group-oriented social justice concerns pertaining to the moral domains of Care and Fairness (Janoff-Bulman & Carnes, 2013), as well as examinations of interpersonal dynamics within a given context to comprehend the predominant moral motivations exhibited (e.g., unity, hierarchy, equality, and proportionality) (Rai & Fiske, 2011).

From an operationalization perspective, given the intensely subjective nature of the moral construct, high intercoder reliability (ICR) statistics reported in previous MCA literature have also been questioned (Hopp & Weber, 2021; Weber et al., 2018). While content analysis is formally regarded as a research methodology for establishing replicable and valid inferentials that situate extracted content features within context sensitive meanings (Krippendorff, 1989), recent critiques emphasize there exist fundamental challenges associated with the overcoming of psychological constraints within MCA coders. These constraints are attributed to individual differences in moral intuitions across different coders and have been argued to remain saliently active during traditional MCA coding tasks (Hopp & Weber, 2021; Weber et al., 2018). In contrast to previous studies reporting high

ICR statistics, and within the paradigm of traditional MCA, recent research (Tamborini et al., 2021) has also started to demonstrate that coders are indeed unable to reach acceptable intercoder agreement on multiple moral content categories despite leveraging the extensively utilized MIME coding manual (Tamborini et al., 2017). Therefore, this paper argues that there remains an opportunity for much needed research on further MCA methodological development.

In addition to acknowledging that latent constructs activate coder schema in a differential fashion (see Potter & Levine-Donnerstein, 1999 for a detailed discussion on projective latent constructs), this paper contends that greater precision in identifying morally relevant cues (and in the current study specifically within *visual* content) can indeed facilitate the direct measurement of deeply latent moral constructs in reliable and valid ways, especially given their potential for minimizing the likelihood of subjective coder interpretations being applied during the coding task (Neuendorf, 2011). More specifically, the central claim of this paper is that *multidimensional* MCA coding schemes, as opposed to *unidimensional* MCA coding schemes, can help achieve greater coding precision and, subsequently, higher ICRs. Accordingly, this paper seeks to make the following three important contributions.

First, this paper offers the seminal version of the Moral Foundations Content Codebook (MFCC-1) which provides a fine-grained, rule-based framework for the identification of moral content cues and facilitates the multidimensional, as opposed to unidimensional, evaluation of individual moral foundations derived from the recently introduced Moral Foundations Questionnaire - 2 (MFQ-2; Atari et al., 2023). MFQ-2 not only provides revised survey items that contribute towards increased robustness of MFT in a cross-cultural

context but also introduces the decomposition of the Fairness moral foundation into the separate components of Equality and Proportionality (see Atari et al., 2023 for a detailed discussion on the empirical significance of this decomposition). In this manner, MFCC-1 encourages MCA literature to keep up with theoretical innovations already being introduced in state-of-the-art MFT research. Additionally, MFCC-1 expands the provisioning of moral content codes via the introduction of variables that are also reflective of moral vices, going beyond the sole provisioning of moral virtues as outlined within MFQ-1 and MFQ-2 derived survey items.

Second, MFCC-1 captures substantial information about the context of unique moral acts and introduces the measurement of situational factors (e.g., family, military, education, etc.) for understanding how context-specific factors influence the saliency of moral exemplification in visual content. In addition to providing important insights that facilitate comparative post-hoc analyses for the MFCC-1 and MIME coding schemes, this paper argues that given the influence disparate contextual variables have upon human moral evaluations, the incorporation of these additional variables will further facilitate the contextualization of individual moralized acts via content analyzing them in interactional relationships, as opposed to in isolation as is common within extant MCA literature, and thereby result in more robust ecological validity.

Third, while traditional MCA research has remained largely focused on textual corpora as primary data sources, this paper prioritizes the relatively underexplored novelty of an extensive MCA on static images extracted from the Socio-Moral Image Database (SMID; Crone et al., 2018). While distinctions within the activation dynamics of moral intuitions and moral reasoning as initiated via visual versus verbal stimuli remain unarticulated within

the framework of MFT, researchers have argued that they are theoretically informed by information processing paradigms (Yang et al., 2018). Visual stimuli, owing to their direct and immediate nature, hold a heightened capacity to promptly capture attention, evoke emotional responses, and activate innate physiological reactions (Geise & Baden, 2015; Lang et al., 2015). Conversely, the sequential and cognitive nature of verbal information aligns it more with the fostering of deliberate moral reasoning. In light of potentially delineated pathways that could underscore cognitive divergence in the processing of text-based stimuli as compared to visual representations (Slovic et al., 2017) as well as our objective of expanding the scope of MCA research into multiple modalities, we believe that the choice of visual imagery for the current study remains an interesting as well as relatively understudied one.

A. Current Literature

MCA research has increasingly utilized MFT as a pragmatic conceptual framework for the identification of moral cues in textual and visual content. However, MCA literature remains almost entirely focused, and thereby limited, in its adoption of unidimensional coding schemes that operationalize higher order representations of individual moral foundations in media content whilst rendering scarce attention to the multidimensional nature of individual foundational constructs (e.g., how can individual subdimensions of Loyalty / Betrayal be operationalized) as well as their sensitivity to non-moral content factors (e.g., how can situational factors such as military action or marital conflict modulate the saliency of Loyalty / Betrayal). For instance, moral foundations have been routinely operationalized via the development of idiosyncratic coding manuals that researchers leverage for assessing the binary presence (or absence) of a given moral domain (Bowman

et al., 2014; Brown & Silver, 2022; Hoover et al., 2020; Wang & Liu, 2021). Alternatively, researchers have also remained reliant on the logic of the MIME (Tamborini, 2013; Tamborini & Weber, 2020) and in turn leveraged an extensively utilized, but limited in replicability (see Tamborini et al., 2021; Weber et al., 2018), coding manual that operationalizes and identifies moral intuitions (Grizzard et al., 2017; Klebig et al., 2021) or motivations (Aley et al., 2021; Hahn et al., 2019; Hahn et al., 2022; Lewis & Mitchell, 2014; Tamborini et al., 2017) salient across a range of narratives. Furthermore, some studies have also attempted to extend beyond simple moral prevalence focused paradigms via the integration of moral exemplification in content with character specific evaluations, albeit still in a relatively straightforward fashion (e.g., hero/villain status, affable/surly expressions) (Hahn 2021, Tamborini et al., 2017). Finally, Computer-aided Text Analysis (CATA) approaches have also drawn interest from MCA researchers who frequently utilize off-the-shelf, dictionary-based solutions, commonly implemented with a bag-of-words (BoW) model logic and constructed with the assistance of expert or crowd coders, for estimating the prevalence of individual moral foundations across large textual corpora (Bayrak & Alper, 2021; Hoewe et al., 2020; Long & Eveland, 2021; Malik et al., 2021; Wendell & Tatalovich, 2021; Zhou et al., 2022).

B. Optimality in Construct Measurement

The unidimensional evaluation of moral constructs in content analytic research (i.e., Care / Harm, Fairness / Cheating, Loyalty / Betrayal, Authority / Subversion, Purity / Degradation) can be argued to be suffering from similar problems as observed in the adoption of single-item (*SI*) measures in traditional questionnaire research designs. Indeed, *SI* measures in survey research are frequently considered as an expedient, scalable, and

cost-effective solution (Allen et al., 2022). But *SI* measures are only useful for unambiguous constructs that remain particularly narrow in scope (e.g., job satisfaction) (Diamantopoulos et al., 2012). Criterion deficiency and measurement unreliability concerns are rightfully raised when *SI* measures are utilized for the evaluation of latent constructs (DeVellis & Thorpe, 2021) suggesting that it remains implausible for *SI* measures to exhaustively represent the multidimensionality of complex psychological constructs (e.g., justice ethics). For instance, in MCA research, Loyalty / Betrayal codings almost never make distinctions between moral exemplifications that suggest, *inter alia*, nationalistic indoctrination, alliance with enemy forces, political apathy, tribalistic attitudes, patriotism, communal pride, and sports spectatorship, all of which could be differentially associated with other moral foundations. While such concepts, albeit not exhaustively, are indeed tapped into via MFQ-2 survey measures these remain entirely unaccounted for during traditional MCA *SI* coding tasks. In this regard, multiple-item (*MI*) indicators are regarded as more accurate and valid representations of the latent construct in question (Sarstedt & Wilczynski, 2009) as these result in adjusted measurement values via the averaging out of errors and particularities inherent to the evaluation of unique items (Sarstedt et al., 2016), allow for the application of statistical procedures that provide internal consistency measures (Bergkvist & Rossiter, 2007), and even enhance predictive validity of the measurement tool via the explanation of greater variance of specific external criterion (Graf et al., 2018). In fact, extending beyond survey research, the incorporation of multiple indicators in content codebooks for the investigation of individual latent constructs has also been successfully implemented in previous content analytic research. Indeed, and as expounded in Neuendorf (2011), reliable measurements via multiple indicators have certainly been achieved for disparate latent

constructs including, *inter alia*, defamation (Simon et al., 1989), journalistic framing (Jones & Himelboim, 2010; Van Gorp, 2007), trust in nurses (Radwin & Cabral, 2010), political humor (Morris, 2009), website interactivity (Ghose & Dou, 1998), website user satisfaction (Muylle et al., 2004), welfare policy centered public discourse (Kinney, 2006), and cultural collectivism versus individualism (Zhang, 2009).

Keeping the aforementioned arguments and explanations in mind, this paper first seeks to address the following hypothesis:

H1: *MI moral measures (as measured via MFCC-1 coding scheme) will exhibit higher ICR statistics as compared to SI moral measures (as measured via MIME coding scheme).*

C. Moral Foundations' Dissociability

A central proposition of MFT is that moral foundations are sufficiently distinct to be treated as separate cognitive modules (Haidt & Joseph, 2008). As the authors of MFT have previously explained (Graham et al., 2009; Graham et al., 2013; Haidt & Joseph, 2008), these five candidate foundations emerged from a meticulous examination of lists of virtues (or moral concerns) that exhibited widespread presence, though not necessarily universality, across cultures, and concurrently possessed plausible and established evolutionary explanations underpinning related psychological mechanisms. Within this analytical framework, it was found that the overarching human inclination towards caregiving, nurturance, and safeguarding vulnerable individuals resonated with discussions on the attachment system (Bowlby, 1969) while the widespread human preoccupation with principles of fairness, reciprocity, and justice were aligned with evolutionary literature on reciprocal altruism (Trivers, 1971). Consequently, the Care / Harm and Fairness / Cheating

foundations were collectively termed as the “individualizing foundations” which are essentially characterized by their emphasis on individual rights and welfare and are derived from the ethic of autonomy (Shweder et al., 1997).

Similarly, virtues of loyalty, patriotism, and self-sacrifice in service of the collective, coupled with a heightened vigilance against traitorous actions concurred with research into the evolutionary development of coalitional psychology (Kurzban et al., 2001), and virtues associated with subordinates (e.g., obedience and respect for authority) paired with virtues attributed to authorities (e.g., leadership and safeguarding) harmonized with theories concerning the evolution of hierarchical structures in primate societies (de Waal, 1982) and the evolution of consent-based human hierarchies (Boehm, 1999). Finally, virtues centering on purity and sanctity, integral to many religious doctrines, found alignment with theories regarding the evolution of disgust sensitivity and contamination aversion (Rozin et al., 2000). It is essential to recognize that practices associated with purity and pollution were argued to hold significance beyond mere hygienic considerations, serving vital social functions like demarcating cultural boundaries of a group (Soler, 1979) and curbing the inclination towards self-centered behaviors frequently linked to human instinctual tendencies (e.g., lust, hunger, material greed) via the fostering of a more elevated and spiritual mindset.

These three distinct foundations, Loyalty / Betrayal, Authority / Subversion, and Purity / Degradation, have been collectively referred to as the “binding foundations” which essentially underpin the emphasis on loyalty to groups, adherence to duties, and cultivation of self-control, characteristics that are integral to conservative and religious moralities. Importantly, however, these “binding foundations” have been found to further cluster in a

differential manner with Loyalty / Betrayal and Authority / Subversion being derived from the ethic of community (Shweder et al., 1997) while Purity / Degradation being derived from the ethic of divinity (Shweder et al., 1997).

The explanations outlined above, therefore, strongly imply that an effective, reliable, and ecologically valid MCA coding scheme should possess the capacity to identify the content representations of unique moral foundations in dissociable fashions and in line with the stated theoretical predictions of MFT. Accordingly, this paper seeks to address the following additional hypothesis:

H2: *MI moral measures (as measured via MFCC-1 coding scheme) will exhibit greater discriminant validity as compared to SI moral measures (as measured via MIME coding scheme).*

D. Situational Factors

Communication scholars have written on the importance of considering the context for decades since it is a fundamental part of decoding messages (Pettegrew, 1988). MCA researchers have also argued strongly in favor of adopting context-sensitive approaches for moral content extraction (Weber et al., 2018; Hopp & Weber, 2021; Kennedy et al., 2021; Cingel & Krcmar, 2020). Unsurprisingly, therefore, for a complete understanding of foundation-specific moral exemplification in visual content, it remains essential to content analyze them within contextually embedded paradigms. Indeed, disparate non-moral factors have been shown to influence human moral evaluations, including, *inter alia*, racial affiliation in the context of criminality (Barmaki, 2020; Jardina & Piston, 2021; Majavu, 2020), sex differences and their influence upon expected gender roles (Hoffmann & Musch, 2019; Roper, 2019; Stoet & Geary, 2018), ageism and biased social perceptions (Hehman et

al., 2014; Marques et al., 2020; Ng & Lim-Soh, 2021), asymmetric target evaluations as a function of national narcissism (Bertin et al., 2022; Kervyn et al., 2008; Wollast et al., 2022), and the role of religious identification in modulating ingroup-outgroup prejudice (Everett et al., 2016; Gervais et al., 2011; Strabac & Listhaug, 2008). It can therefore be argued that studies seriously compromise their ecological validity when content analyzing moral exemplification in naturalistic media content via their disregard of disparate situational factors which could potentially influence, in substantial ways, the insights generated from their research procedures (Hester & Gray, 2020).

Keeping the aforementioned arguments and explanations in mind, this paper further seeks to address the following research question:

RQ1: *What is the relationship between visual moral exemplification (as differentially measured via the MFCC-1 and MIME coding schemes) and situational factors?*

Additionally, insights generated from RQ1 are meant to further facilitate the *post-hoc* analysis of differentials in discriminant and predictive validities achieved via both the MFCC-1 and MIME coding schemes.

E. Research Bias in Selecting Content Domains

MCA research has remained focused on textual corpora or visual narratives as primary data sources. Static images, on the other hand, have remained relatively underexplored. This is surprising as contemporary morality is routinely exemplified within the visual image paradigm (Clark, 2020; Joo & Steinert-Threlkeld, 2022; Korkmazer et al., 2021; Mortensen & Trenz, 2016; Pradantyo et al., 2021). Indeed, while textual comprehension remains a consequentialist outcome of human learning processes, visual information processing has

been argued to be an evolutionarily antecedent capacity that contributes towards enhanced cue accessibility (Barry, 1997; Gazzaniga, 1998; Gibson & Zillmann, 2000), strengthened evocation of emotional reactions (Grabe & Bucy, 2009; Tukachinsky et al., 2011), as well as the provisioning of an important channel for message framing and persuasion effects (Geise & Baden, 2015; Joo et al., 2014; Soroka et al., 2016). Accordingly, as visual stimuli are inherently processed in an involuntary manner, their discernible impact on the formulation of moral judgments can be argued to primarily transpire through the elicitation of moral intuitions while the impact of verbal information can be argued to predominantly be emanating from its propensity to engage cognitive processes linked to moral reasoning (Yang et al., 2018). Furthermore, research has demonstrated that there indeed remains cognitive divergence in the processing of text-based stimuli as compared to visual representations (Slovic et al., 2017), that textual information is routinely associated with properties that elicit high-level construal within audiences (Carnevale et al., 2015), and that high-level construal is differentially associated with individual emphasis on moral values (Napier & Luguri, 2013).

Previous research on images, content analytic and otherwise, has also applied visual framing perspectives for studying morally-contentious issues surrounding conflict (Bouko et al., 2021; Brantner et al., 2011; Esfandiari et al., 2021; Ireri, 2022; Issa, 2022), racial, gender, and sexual prejudice (Askanius & Keller, 2021; Askanius, 2021; Batova, 2021; Braumüller et al., 2020; CohenMiller et al., 2020; Fahmy, 2004; Kapidzic & Herring, 2015; Průchová Hrůzová & Zápotocký, 2022; Rogers, 2020), political representation (Bast, 2021; Haßler et al., 2021; Mortensen, 2015; Schill, 2012), beauty standards (Akinro & Mbunyuza-Memani, 2019; Heuer et al., 2011; Varava, 2016; Yan & Bissell, 2014),

sustainability (Carolan, 2022; Krause & Bucy, 2018; Milanesi et al., 2022; Rebich-Hespanha & Rice, 2016; Rebich-Hespanha et al., 2015), and homelessness (Bowen & Capozziello, 2022; Goldfischer, 2018; Lundberg, 2021). In light of the above outlined evidence, therefore, this paper argues that there is substantial opportunity for MCA researchers to extend their efforts towards the static image content domain as well.

II. Methods

A. Data Collection

Static images for MCA were retrieved from the SMID (Crone et al., 2018). Extant moral psychology stimuli datasets have been critiqued for their lack of emphasis upon conducting repeated ecological validations, expanding the modality and context space within which moral content is represented, accounting for the prevalence of multiple moral classes contributing towards moral complexity within a unique content piece, as well as departing from a focus on moral violations and incorporating instances of moral upholding within content. In response to these concerns, Crone et al. (2018) crowdsourced the SMID moral stimulus dataset consisting of 2, 941 public images (including moral content not only limited to human-based moral actions but also animals and inanimate symbols as targets of moral evaluations), evaluated on affect, arousal, moral wrongness, and the intensity of five moral foundations as posited by the original MFT (Graham et al., 2009), via 820, 525 individual judgments across 2, 716 participants. While the current study focuses on an "expert" content analytic approach that does not utilize these crowd-sourced ratings in its primary analyses, *post-hoc* evaluations of the relationships between MFCC-1 & MIME variables with the available SMID crowd-ratings, stratified along SMID participants' gender

and political affiliations, were conducted for purposes of external validity comparisons. These *post-hoc* analyses are accordingly reported in the results section.

B. Coder Groups

Two content analyses were conducted using *separate* and *independent* human coder groups. Each coder group received equivalent involvement incentives, i.e., course credit, and were subject to the same amount of training time. Coder groups were small by design, i.e., $g1 = 2$, $g2 = 2$, a common procedure in traditional content analyses, and consisted of undergraduate research assistants who participated for a total of two academic quarters at the University of California Santa Barbara. Coders, *independently* of each other, received an initial training which lasted for about one hour. Subsequently, coders, again *independently* of each other, attended weekly one-hour research meetings where issues were discussed and questions clarified.

Differentiations between these coder groups pertain to the specific coding tasks assigned to them. For the first coder group (*g1*), coding instructions and conceptual definitions of moral foundations followed established protocols as outlined in traditional MIME content analyses (Tamborini et al., 2017). In particular, *g1* was responsible for generating content codings corresponding to the unidimensional evaluation of individual moral constructs as outlined in the MIME coding manual (see Table 1 in *SM*). On the other hand, for *g2* conceptual definitions of moral foundations were generated from MFQ-2 and operationalizations were provided as outlined in the MFCC-1 (see Table 2 in *SM*). In other words, *g2* was responsible for generating content codings corresponding to the multidimensional evaluation of individual moral constructs.

Additionally, annotation procedures on SMID were conducted using a *third* and *independent* human coder group, i.e., *g3* ($n = 2$). *g3* received coding instructions for identifying the contextual settings moralized acts are embedded within as outlined in the MFCC content codebook (see Table 3 in *SM*). Importantly, *g3* was responsible for generating content codings corresponding to the content categories SMID images are reflective of (e.g., family, military, education, etc.). In summary, keeping in mind their similarities as well as their differentiations, we conceptualize the organization of our coder groups in the following manner: *g1* = MIME coders, high-involvement, medium training; *g2* = MFCC-1 coders, high-involvement, medium-training; *g3* = General coders, high-involvement, medium-training.

C. MIME Moral Measures

g1 were instructed to code unique SMID images in a sequential fashion. For an individual and unique image *i*, and in line with Weber et al. (2018), *g1* was first required to specify which moral foundation MF_m is most salient, where MF_m corresponds to the relevant variable code as provided in Table 1 (*SM*). Additionally, *g1* was required to specify which moral foundation MF_m is second-most salient in the presented image, thereby providing a two-layered saliency level SL_n variable where n corresponds to primary and secondary saliency levels respectively. Subsequently, this procedure was formalized as below:

$$\forall\, MF_m \cdot \forall\, SL_n : m \in \{1, \ldots, 5\} \cdot n \in \{P, S\}$$
$$MM_{i,\, MF_m,\, SL_n} = f : f \in \{0, 1\} \Rightarrow 0 \vee 1$$

where f indicates the binary presence or absence of a unique moral foundation and $MM_{i,\,MF_m,\,SL_n}$ is reflective of the moral measure score for unique image i and at saliency level SL_n. Furthermore, and in line with Weber et al. (2018) and under the given condition that $MM_{i,\,MF_m,\,SL_n}$ equals 1, *g1* was required to indicate the valence of identified moral foundation MF_m for unique image i. Accordingly, this procedure was formalized as below:

$$\exists\ MM_{i,\,MF_m,\,SL_n} = 1$$
$$MMV_{i,\,MF_m,\,SL_n} = v \times 1 : v \in \{1, \ldots, 5\}$$
$$\Rightarrow 1 \geq MMV_{i,\,MF_m,\,SL_n} \leq 5$$

where v is a valence score indicated on a 5-point Likert scale (e.g., completely loyalty, mostly loyalty, both loyalty and betrayal, mostly betrayal, completely betrayal) and $MMV_{i,\,MF_m,\,SL_n}$ is reflective of the moral measure valence score for unique image i, moral foundation MF_m, and at saliency level SL_n. In summary, $1 \geq MMV_{i,\,MF_m,\,SL_n} \leq 2$ indicates the upholding of moral foundation MF_m, $4 \geq MMV_{i,\,MF_m,\,SL_n} \leq 5$ indicates the violation of moral foundation MF_m, and $MMV_{i,\,MF_m,\,SL_n} = 3$ suggests within-foundation moral ambivalence (Weber et al., 2018). Accordingly, and in line with this logic, for each of the 10 MIME dimensions (e.g., betrayal) MIME ratings for an individual SMID image ranged between 0 (i.e., completely absent) and 3 (i.e., completely present).

D. MFCC-1 Moral Measures

g2 was also instructed to code unique SMID images in a sequential fashion. For an individual moral dimension MD_m and unique image i, *g2* was required to specify whether

or not a unique moral dimension item MDI_n is present, where MDI_n corresponds to the relevant variable code as provided in Table 2 (*SM*). In this fashion, the identification of moral signals via multiple-item MFCC-1 procedures was formalized as below:

$$\forall\, MD_m \cdot \forall\, MDI_n : m \in \{1, \ldots, 12\} \cdot n \in \{1, \ldots, 6\}$$

$$MM_{i,\, MD_m} = \sum_{n=1}^{6} f : f \in \{0, 1\}$$

$$\Rightarrow 0 \geq MM_{i,\, MD_m} \leq 6$$

where f indicates the binary presence or absence of a unique moral dimension item and $MM_{i,\, MD_m}$ is simply an aggregate moral measure with a lower-bound of 0 and an upper-bound of 6. Accordingly, and in line with this logic, for each of the 12 MFCC-1 dimensions (e.g., betrayal) MFCC-1 ratings for an individual SMID image ranged between 0 (i.e., completely absent) and 6 (i.e., completely present).

E. Situational Factors

Situational factors (SF) were selectively identified via the compilation of important topical factors that are routinely discussed in public polling codebooks (Gallup, 2022) as well as academic research (Guo & Vargo, 2020; Vargo & Guo, 2017). Accordingly, a non-exhaustive yet significantly diverse set of 24 SF were operationally defined in MFCC-1 as outlined in Table 3 (*SM*). Accordingly, *g3* was instructed to identify relevant SF_m in SMID images and this procedure was formalized as below:

$$\forall\, SF_m : m \in \{1, \ldots, 24\}$$
$$SF_i = c : c \in \{0, 1\}$$

where c indicates the binary presence or absence of a unique SF. It is important to note here that *g3* coders *collaboratively* performed these annotations for each SMID image. Peer

discussion was encouraged so as to achieve resolution of coder inferential ambiguities as well as remain consistent with definitional guidelines as provided in the MFCC-1 codebook.

E. Coder Training

Coders (*independently*) were required to attend in-person training procedures with the lead author of this paper before they were considered qualified to evaluate SMID images. The training procedure included reading detailed descriptions of each moral foundation, a step-by-step guideline for SMID image coding (with examples), as well as the opportunity to practice coding SMID images that had already been selected and evaluated by the lead author. Coding instructions and the conceptual definitions followed protocols as per the MFCC-1 and MIME coding schemes. Importantly, coders were required to verbally complete several comprehension checks designed to assess their understanding of unique coding items (for MFCC-1) and unique moral foundations (for MIME) and the coding procedure. In-person feedback was given to each coder and training was not considered to be complete until conceptual and definitional expertise was deemed to have been reached according to the lead author.

III. Results

Analyses have been conducted using the *Python* programming language. All datasets, scripts, and Supplemental Material (*SM*) are being made publicly available via the Open Science Framework at https://osf.io/rfpgh/?view_only=f1dd79c02fa84e049e69c63d43899c4d.

A. Primary Reliability Analyses

In response to H1, we computed a range of reliability metrics based on an “exact agreement” logic. For example, agreement would be considered as exact if dimensional

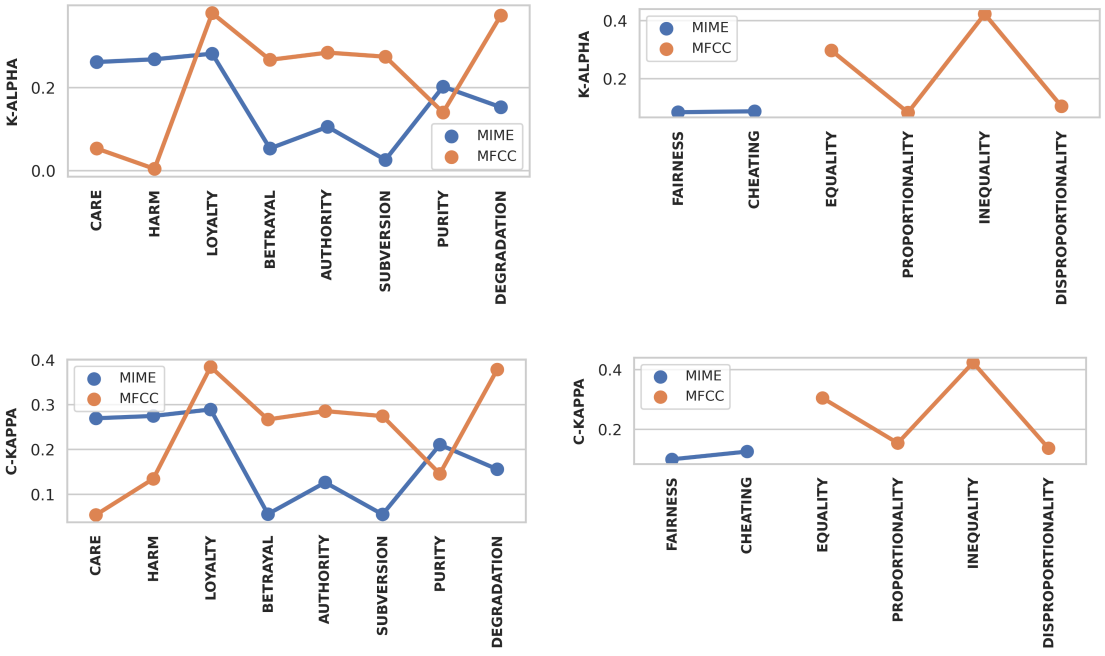
ratings across MIME coders were equivalent to "2" along the 4-point scale (i.e., 0 to 3). Similarly, agreement would be considered as exact if dimensional ratings across MFCC coders were equivalent to "2" along the 7-point scale (i.e., 0 to 6). Reliability metrics were calculated using the open-source *Python* library *agreement* 0.1.1[1] which provides a programmatic implementation of previously formulated reliability metrics for inter-rater agreement measurement (Gwet, 2014). Specifically, we report the following metrics: Cohen's kappa (Cohen, 1960), Krippendorff alpha (Krippendorff, 1970), Scott's pi (Scott, 1955), Bennett et al.'s S score (Bennett et al., 1954), and Gwet's gamma (Gwet, 2008). Additionally, for purposes of remaining consistent with reliability metrics as reported for SMID (Crone et al., 2018) and using the *Python* package *pingouin* 0.5.3[2], we computed intra-class correlation coefficients (ICC) (McGraw & Wong, 1996) where both target (SMID image) and rater (coder) were treated as random effects, coefficients were calculated based on absolute agreement (as opposed to consistency), and the coefficient reflected the reliability of a single rating.

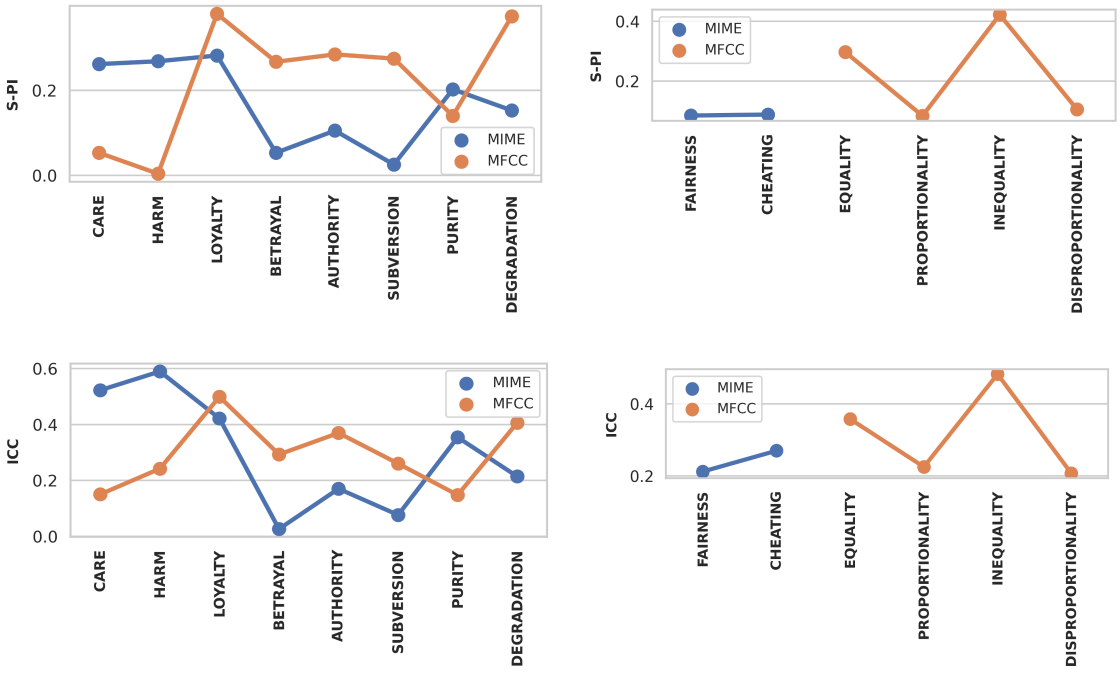
We find partial support for H1. As can be descriptively observed in Figure 1, MFCC-1 ratings exhibit *comparatively* higher reliability metrics across the majority of moral dimensions while MIME ratings outperform MFCC-1 specifically along the Care, Harm, and Purity dimensions. Overall magnitudes across reliability metrics remain low, generally ranging between 0.2 and 0.4 (see Tables 4 & 5 in *SM* for exact estimates), but also consistent with ICC-specific reliability metrics as observed across crowd coders in the original SMID paper (see Figure 1 in *SM*). Additionally, we find that these general trends remain consistent across traditional MCA measures as derived from Cohen's kappa,

[1] https://pypi.org/project/agreement/
[2] https://pypi.org/project/pingouin/

Krippendorff's alpha, and Scott's pi. Interestingly, our results from Bennett et al.'s S score and Gwet's gamma provide insights that deviate from the general inferentials we provide above (see Tables 4 & 5 in *SM* for exact estimates). While MFCC-1 ratings still outperform MIME ratings across a majority of dimensions, the specific nature of those dimensions and the absolute magnitudes in reliability that is achieved remains contentious. We refrain from engaging in further interpretation on these two specific reliability metrics but speculate that the variation in general trends as observed through the perspective of these two reliability metrics can potentially be attributed as to how expected chance-agreement is differentially conceptualized across reliability measures (Lovejoy et al., 2016). For purposes of maintaining consistency with reporting norms in the MCA literature, we defer to the interpretations as derived from Cohen's kappa, Krippendorff's alpha, and Scott's pi.

Figure 1. Reliability estimates as derived from MIME and MFCC-1 ratings.





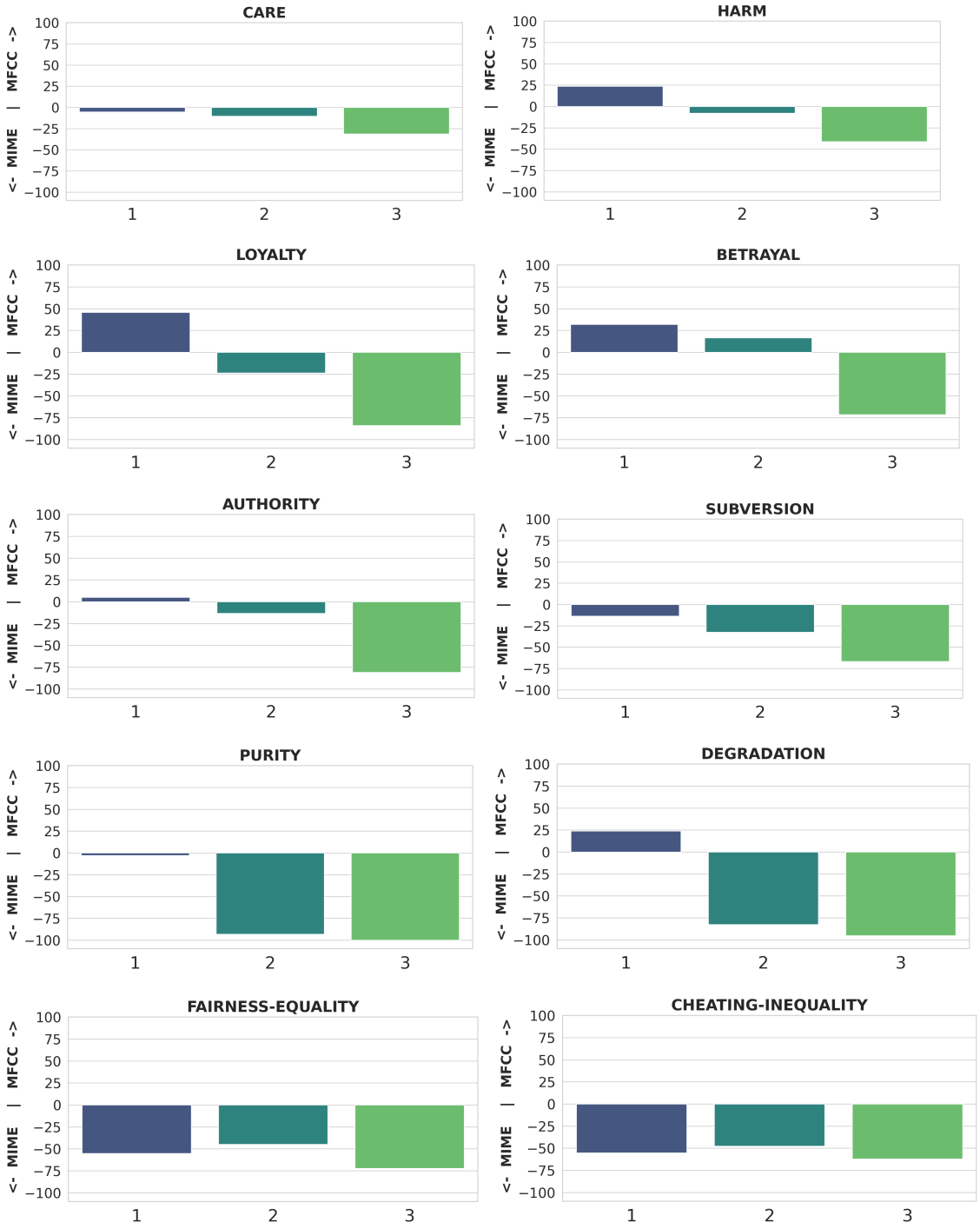
Unfortunately, however, our findings still provide no definitive evidence indicating whether or not the MFCC-1 rating protocol remains superior to the MIME rating protocol. At best, our findings suggest that the adoption of a multidimensional MCA codebook holds the potential to provide consistently *less worse* reliabilities as compared to a unidimensional one. Accordingly, we wanted to conduct *post-hoc* error analyses that specifically model the proportion of SMID images reflecting coder disagreement as a function of within-codebook coder differentials. In other words, rather than focusing on *maximizing* coder agreement, we were interested in evaluating whether either of the codebooks was successful in *minimizing* coder disagreement. For instance, coder disagreements can be slight in nature (e.g., single-point deviations) or more severe (e.g., three-point deviations). Therefore, from an alternative analytical perspective, we can expect that a reliable codebook would be more likely to be associated with slight, as opposed to severe, coder disagreements.

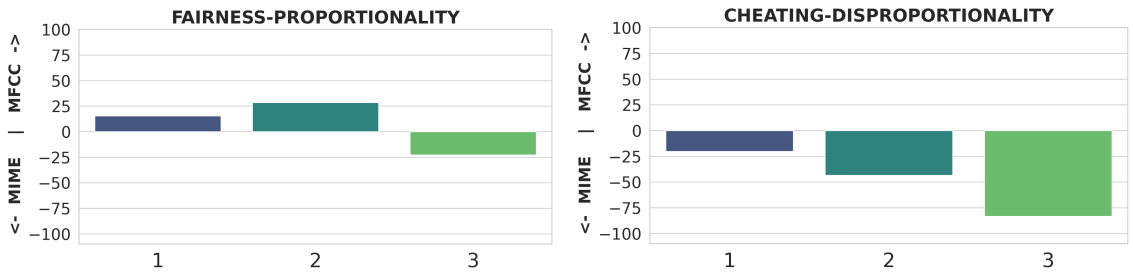
B. Secondary Reliability Analyses

We computed the absolute of differentials between MFCC-1 and MIME coder pair ratings (e.g., | 3 - 2 | = 1; | 0 - 3 | = 3) and for each differential category extracted the proportion of SMID images associated with it. Differential categories ranged from a minimum of 1 to a maximum of 6 for the MFCC-1 rating protocol and a minimum of 1 to a maximum of 3 for the MIME rating protocol. For purposes of standardization in visual comparisons, we conceptualized MFCC-1 associated absolute differentials greater than 3 (i.e., 4, 5, and 6) as being equivalent to the differential category of 3. In this way, the differential category of "1" represents slight disagreement while "3" represents severe disagreement across both codebooks.

As can be descriptively observed in Figure 2, MIME ratings exhibit more *severe* disagreement across each and every moral dimension as compared to MFCC-1 ratings. These effects also remained consistent for the contrasting of Fairness / Cheating disagreement with its MFCC-1 counterparts. Additionally, we observed that MFCC-1 ratings were consistently likely to be associated with *slight* disagreements as compared to MIME ratings. Interestingly, for the specific dimensions of Equality / Inequality, we observed that MFCC-1 ratings generally outperformed MIME ratings across every differential category.

Figure 2. Proportion of SMID images associated with a codebook as a function of degree of disagreement. Positive values on the y-axis indicate association with MFCC-1 while negative values indicate association with MIME.





In line with the above exploratory analyses, we collapsed MFCC-1 and MIME coder ratings that exhibited *slight* disagreements, i.e., thresholding coder agreement within one level above or below each other. We would like to emphasize here again that coder pairs in our study had remained completely independent, i.e., they were unable to communicate among each other to clarify confusion or resolve inferential ambiguities. As referenced in Weber et al. (2018), such designs are rare and that traditional content analysts instead report,

"spending months in training sessions with coders, during which time they refined categories, altered instructions, and revised data sheets until the coders felt comfortable with what was expected of them and the analysts were convinced they were getting the data they needed. It is typical for analysts to perform reliability tests during the development of coding instructions until the reliability requirement is met as well" (Krippendorff 2013, p. 130).

We motivate our *post-hoc* analytical experimentations with the logic that the resolution of slight differences in coder interpretations is a desirable outcome for MCA where marginal deviations between coder pairs can be indicative of measurement error that might be unrelated to differences in moral convictions and remains quite simply an artifact of the coding procedure at-hand (e.g., unrefined codebooks). Accordingly, in cases where *slight* disagreements emerged, we adjusted the rating of the first coder to match that of the second. For the sake of analytical robustness, we adjusted the rating of the second coder to match

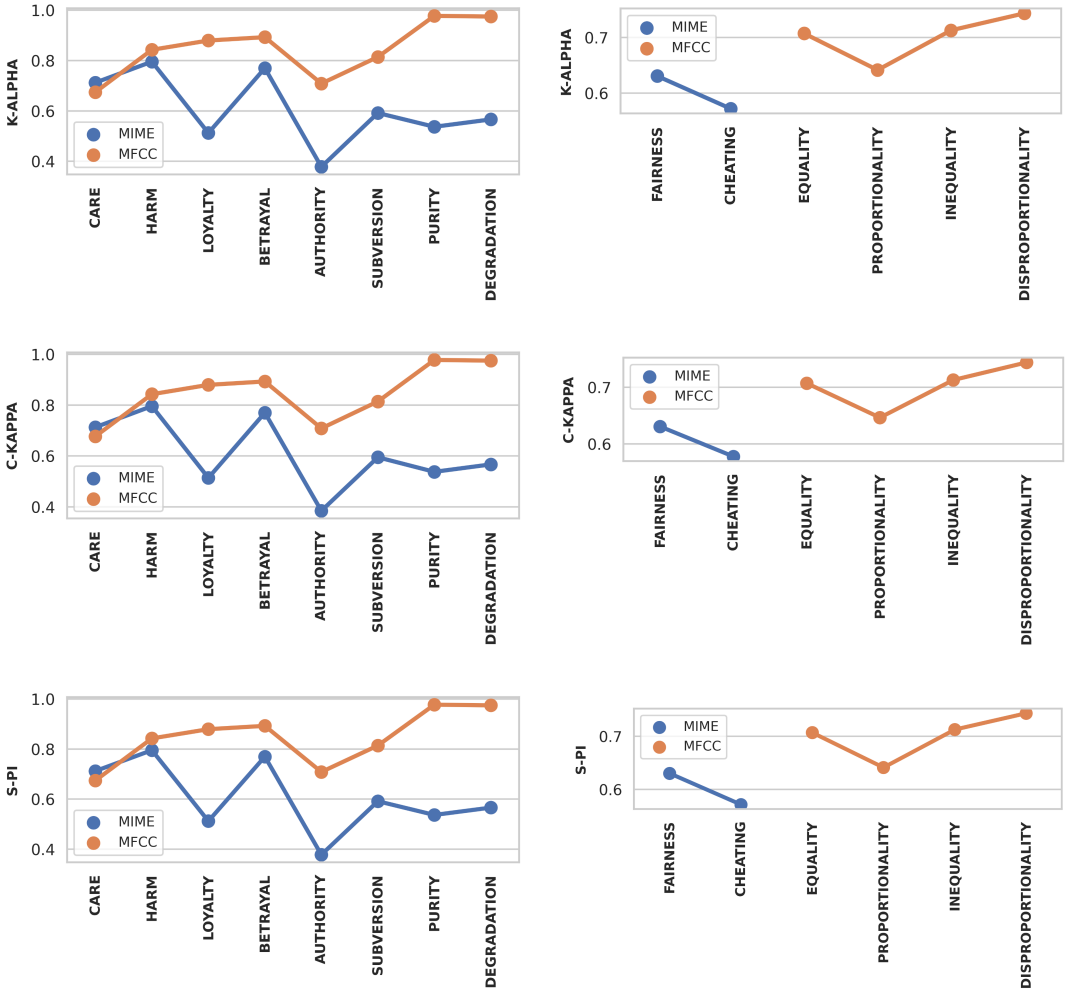
that of the first as well. In this way, we were able to compute an interval that reflected MFCC-1/MIME reliability ratings adjusted for *slight* disagreement.

As a reminder, ICR statistics within the context of content analyses are leveraged to correct for chance agreement, i.e., the level of agreement that would occur by chance alone, without any systematic agreement or shared understanding among coders, serving as a baseline against which the actual level of agreement can be compared. If the observed agreement between coders is significantly higher than the expected chance agreement, it suggests that there is some degree of consistency or pattern in the coding process beyond what would be expected due to random chance. Within our revised datasets adjusted for *slight* disagreements, higher ICA statistics would indicate that there is systematic agreement between coders that extends beyond random chance while lower ICR statistics would more strongly reflect chance agreement or systematic disagreement between coders.

Interestingly, and as observed in Figure 3, our adjusted reliability statistics provide strong evidence indicating that the MFCC-1 rating protocol indeed remains superior to the MIME rating protocol not only in a comparative fashion, as was observed in our previous analyses (see Figure 1), but also in an absolute manner, consistently yielding adjusted reliability metrics that exceed minimum acceptable thresholds for traditional content analytic purposes, i.e., every MFCC-1 dimension remains greater than 0.6. This substantial increase in ICR statistics is not observed for the majority of MIME dimensions (see Tables 6 through 9 in *SM* for exact estimates as derived across both codebooks) lending credence to our previous findings that the MIME codebook is more likely to be associated with severe disagreements that can't be accounted for with a straightforward *post-hoc* reliability adjustment. We refrain from further interpretations of these exploratory analyses (as these

extend beyond the scope of the current paper) but we strongly encourage future research to explore the important interactions we demonstrate herein between the degree of codebook complexity, adjustment in reliability parameters, and signal in moral content.

Figure 3. Disagreement adjusted reliability estimates as derived from MIME and MFCC-1 ratings.

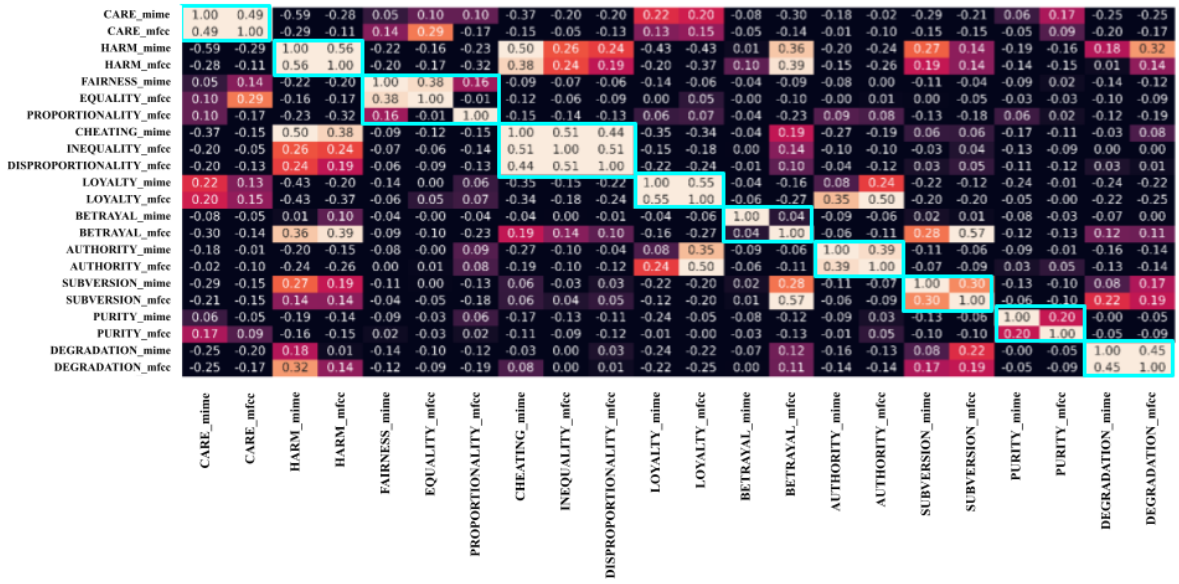


C. Convergent Validity Analyses

Before responding to H2, we first wanted to see *to what extent* MFCC-1 ratings exhibit convergent validity with MIME ratings. Provided that these two measurement codebooks evaluate similar constructs, albeit in distinct ways, we did expect to see a relatively high degree of within-dimension convergence. As demonstrated in Figure 4, this expectation was

indeed confirmed as MFCC-1 demonstrates highest within-dimension correlations (with the exception of betrayal) for each respective dimension as measured via the MIME (see green boxes on the diagonal which emphasize within and across codebook correlations). Importantly, however, we note that these codebook-specific dimensions are not perfectly correlated and each seems to be measuring aspects of an individual moral construct that remain uncaptured via the other. Thus, while both codebooks remain convergent with respect to their broader identification of overlapping moral dimensions, they provide unique perspectives that remain unaccounted for when evaluated separately. Interestingly, and with relevance for advancements in MFT research in general, we observe that Fairness / Cheating (as measured via the MIME) correlates meaningfully with notions of Equality / Inequality as well as Proportionality / Disproportionality lending further credence to the scholarly viewpoint that performing the decomposition of Fairness / Cheating remains a viable direction for future MCA research (Atari et al., 2023).

Figure 4. Correlation matrix demonstrating within and between dimension correlations across MIME and MFCC-1 ratings.



D. Discriminant Validity Analyses

In response to H2, we first implemented a Box-M test (Box, 1949) to check for equality of multiple variance-covariance matrices across the two datasets. The analysis rejected the null hypothesis, indicating that the variability of the data points (variances) and the relationships between variables (covariances) differed significantly between overlapping dimensions of the two datasets ($\chi^2(36, 1932) = 2951.5, p < .001$), suggesting that these datasets may not be directly comparable. Accordingly, and as demonstrated in Figures 5 (correlation matrices), 6 (hierarchically clustered heatmaps), and 7 (network representations of nomological structures) demonstrate, MFCC-1 reveals dimensional association patterns that group together in systematic and theoretically meaningful ways. For instance, we see MFCC-1 ratings discriminate remarkably well between moral virtues (i.e., Care / Equality / Proportionality / Loyalty / Authority / Purity) and moral vices (i.e., Harm / Inequality / Disproportionality / Betrayal / Subversion / Degradation). In other words, MFCC-1 ratings suggest that, on average, SMID images that exemplify the upholding of moral foundations are unlikely to simultaneously exemplify moral violations as well. While recent research has indeed purported the proposition that "within-foundation" and "between-foundation" vice-virtue exemplification should be considered valid phenomena within moral conflict research (Hopp et al., 2020), we believe that the theoretical feasibility of such exemplification patterns within SMID images, a database curated for maximizing moral signal in visual content (as opposed to emphasizing moral ambiguity) remains implausible. In line with this logic, we observe that MIME ratings exhibit a complex nomological structure with several "within-foundation" and "between-foundation" vice-virtue exemplification patterns emergent (e.g., Care / Harm, Purity / Harm, Loyalty / Subversion,

Care / Subversion) suggesting that MIME ratings are more likely to result in findings within which opposing moral dimensions manifest themselves simultaneously. Furthermore, and again as observed in Figures 5, 6, and 7, we find evidence that MFCC-1 ratings provide empirical discrimination between the "individualizing" upholding dimensions of Care / Equality and the "binding" upholding dimensions of Loyalty / Authority. We further observe that the moral violations "Inequality / Disproportionality", associated with Fairness, cluster together in meaningful ways while the "binding" violations of Betrayal / Subversion / Degradation also correlate together. Additionally, "Harm" considerations are meaningfully associated with almost every other moral violation which remains consistent with previous research, and alternative theoretical paradigms, highlighting the important moderating role of harm concerns in moral exemplification (Schein & Gray, 2018). MIME ratings provide meaningful clustering of the "individualizing" violations of Harm/Cheating as well as the "binding" violations of Subversion/Degradation. However, we also observe that MIME ratings suggest potential associations between the dimensions of Care / Loyalty as well as Authority / Fairness / Purity / Betrayal. While these findings are indeed intriguing, we would like to highlight that these remain inconsistent with MFT propositions that suggest distinctive discrimination between the "individualizing" and "binding" moral foundations.

Figure 5. Independent correlation matrices demonstrating within and between dimension correlations across MFCC-1 (top matrix) and MIME (bottom matrix) ratings.

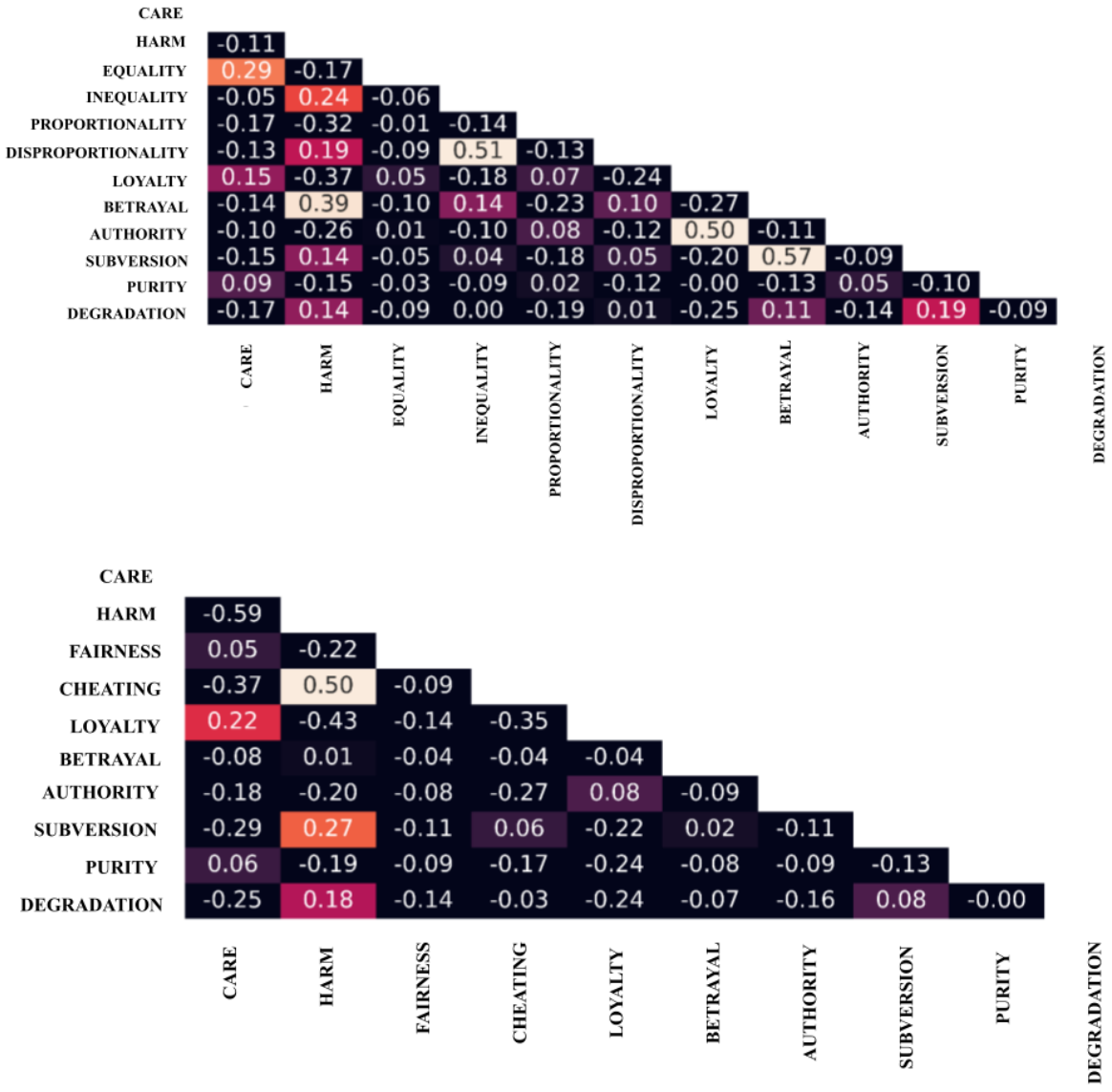


Figure 6. Independent hierarchically clustered heatmaps demonstrating dimensional clustering patterns across MIME and MFCC-1 ratings.

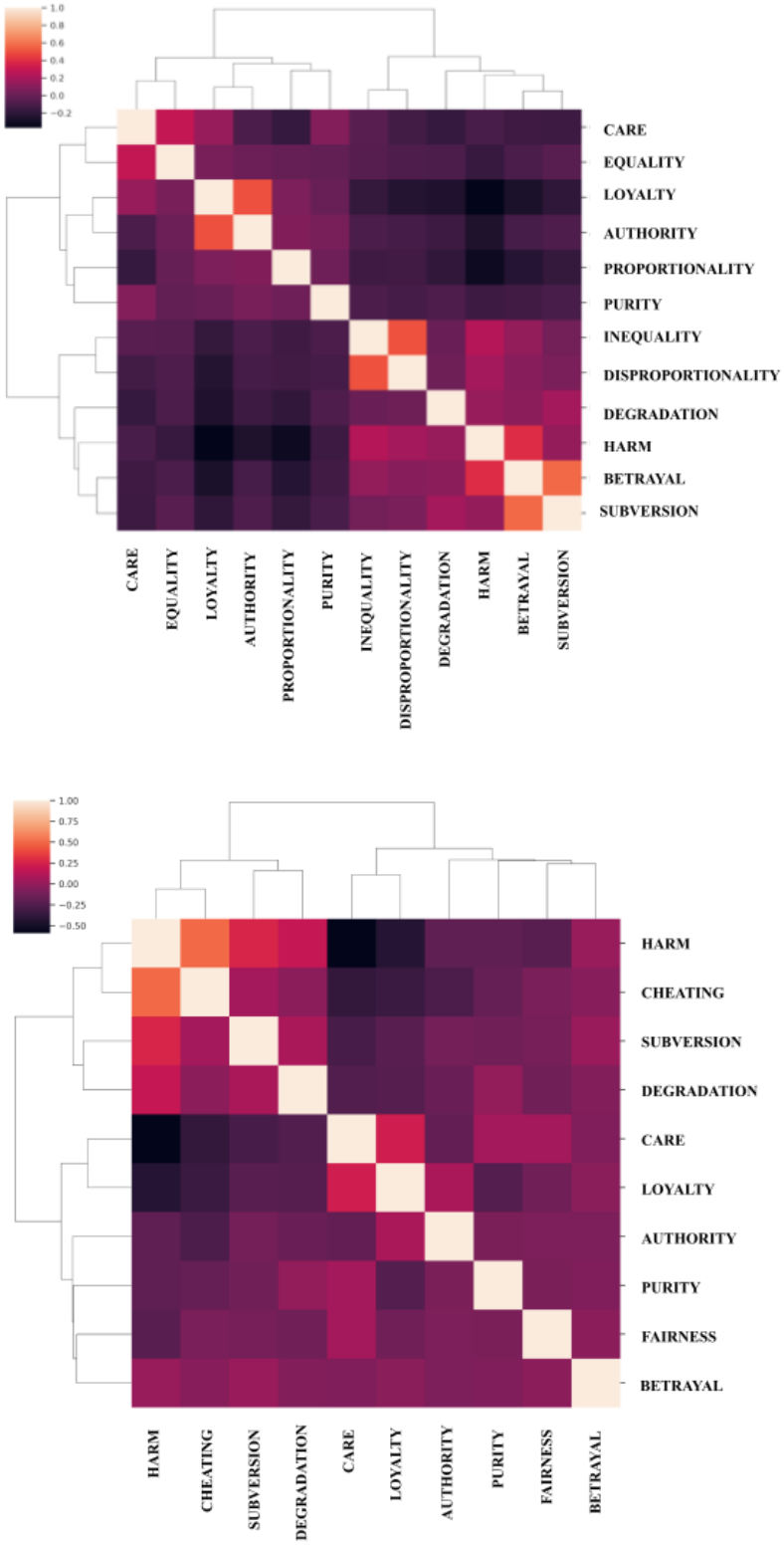
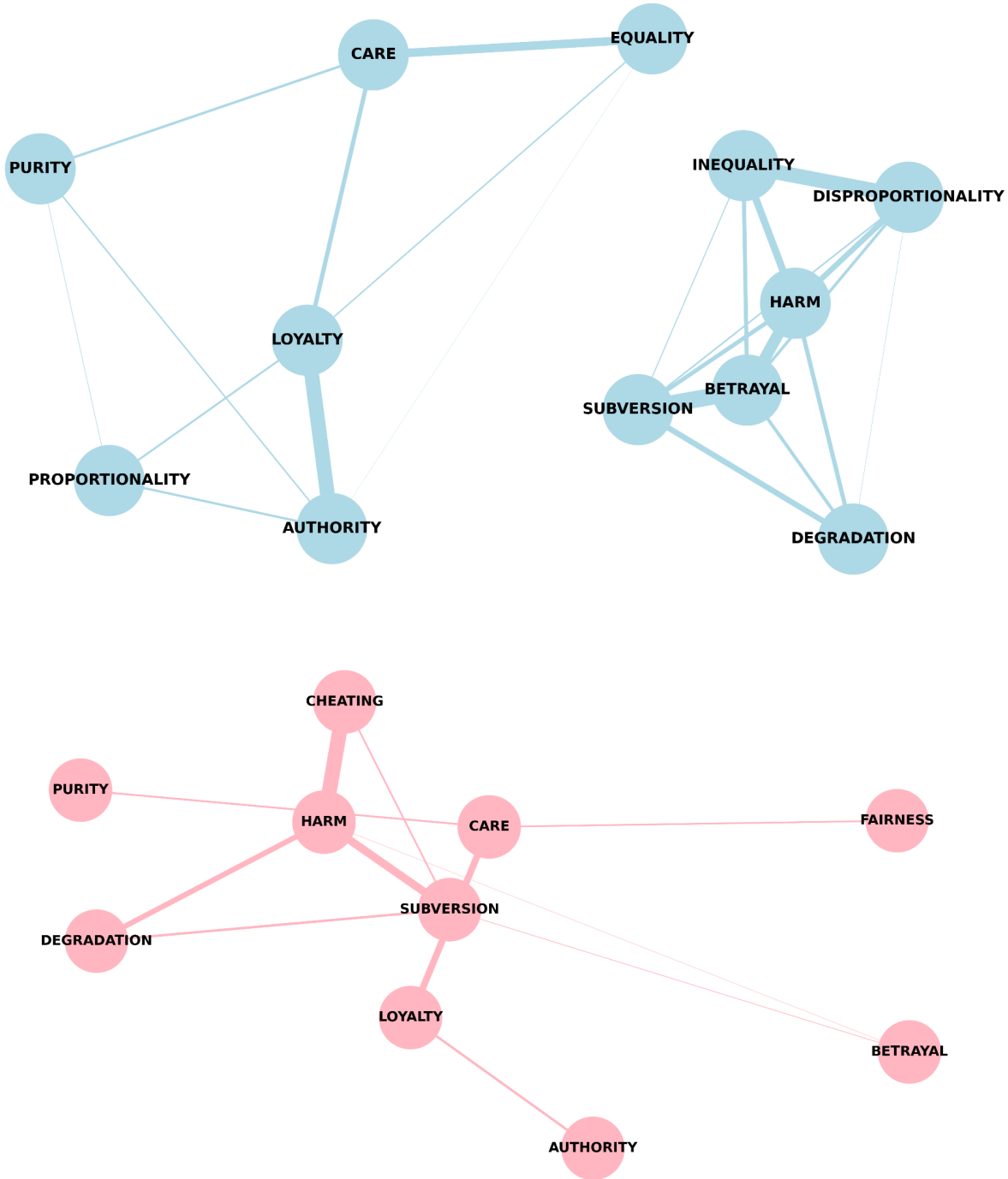


Figure 7. Independent network representations demonstrating correlational patterns across MIME and MFCC-1 ratings. Thicker edges and smaller inter-node distance indicates higher correlations.



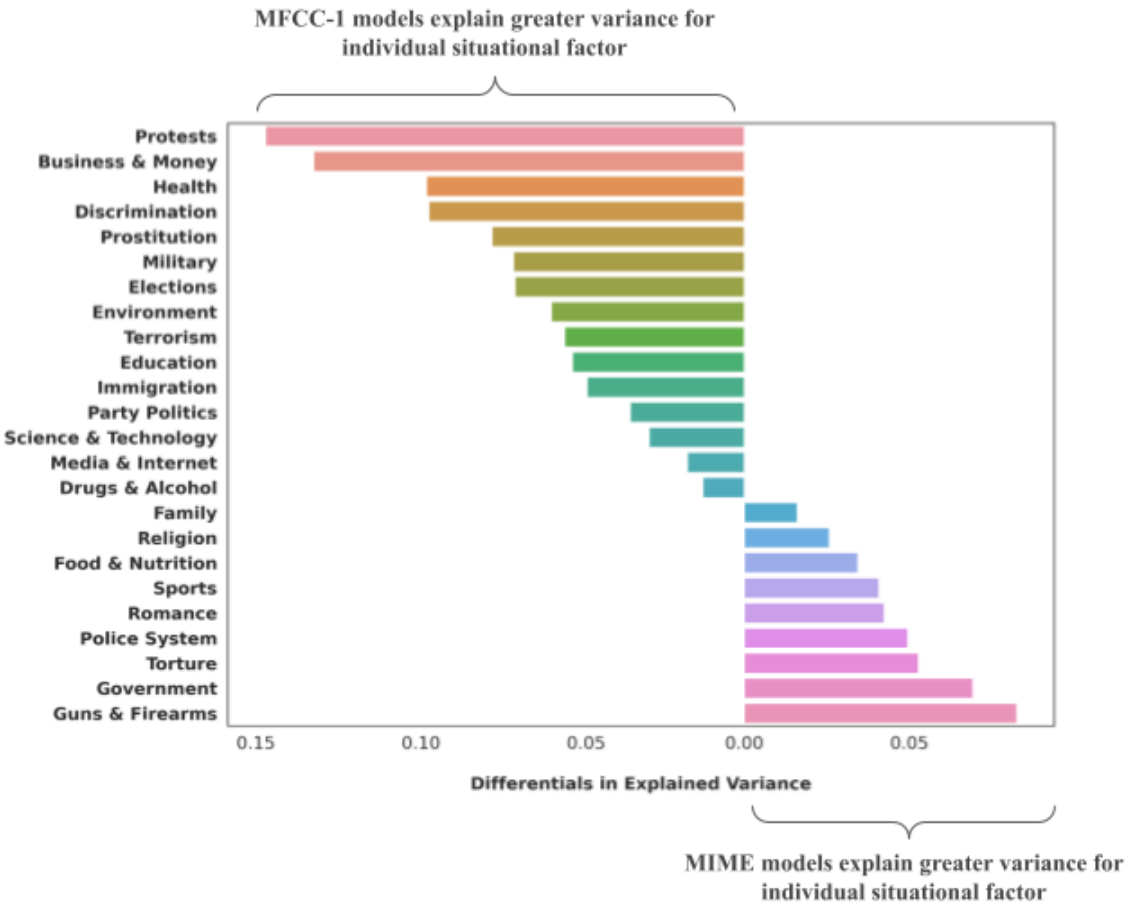
E. Predictive Validity Analyses

Building on these findings for establishing discriminant validity, and in response to RQ1, we further investigated the relationships between visual moral exemplification (as measured via MFCC-1 & MIME coding schemes) and our diverse set of 24 identified situational factors (e.g., prostitution, drugs & alcohol, police system). We believe these exploratory analyses hold the potential to further provide empirical evidence that emphasize the unique associations of moral variables with distinct contexts within SMID images. Accordingly, two sets of logistic regression models were implemented using the Python *statsmodel*[3] 0.14.0 package. The first model included MFCC-1 dimensional ratings (n = 12) as independent variables while the second model leveraged MIME dimensional ratings (n = 10). Both models predicted unique situational factors as their binary outcome variable (1 = present, 0 = absent). Model performance was evaluated by computing the differentials between their respective *pseudo R-squared* measures.

As demonstrated in Figure 8, logistic regression models leveraging MFCC-1 ratings exhibited superior predictive capabilities as compared to MIME models with 15/24 contextual factors being better predicted via MFCC-1 models. Interestingly, the magnitude of explained variance differentials was also substantially greater for MFCC-1 models than for MIME models. In particular, MFCC-1 ratings were most strongly associated with the prediction of situational factors, *inter alia*, Protest / Business & Money / Health / Discrimination / Prostitution / Military while MIME ratings were most strongly associated with the prediction of Guns & Firearms / Government / Torture / Police System.

[3] https://www.statsmodels.org

Figure 8. Pseudo R-squared measures as derived from independent MIME and MFCC-1 logistic regression models for predicting unique situational factors.



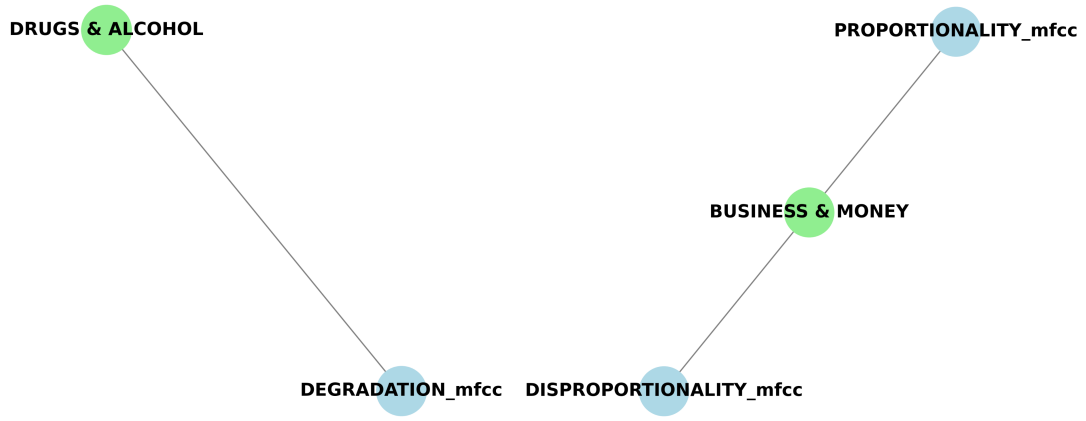
To evaluate the unique predictive capacity of individual MFCC-1 and MIME dimensional ratings in predicting specific situational factors, an additional series of independent logistic regression models were implemented. These models included the complete set of MFCC-1 and MIME dimensional ratings (n = 22) as independent variables and predicted unique situational factors as their binary outcome variable (1 = present, 0 = absent). False Discovery Rate (FDR) correction (Benjamini & Hochberg, 1995) was applied for multiple hypothesis testing to control the proportion of incorrect rejections of null hypotheses (i.e., significant false positives or Type 1 error). For reference, the least

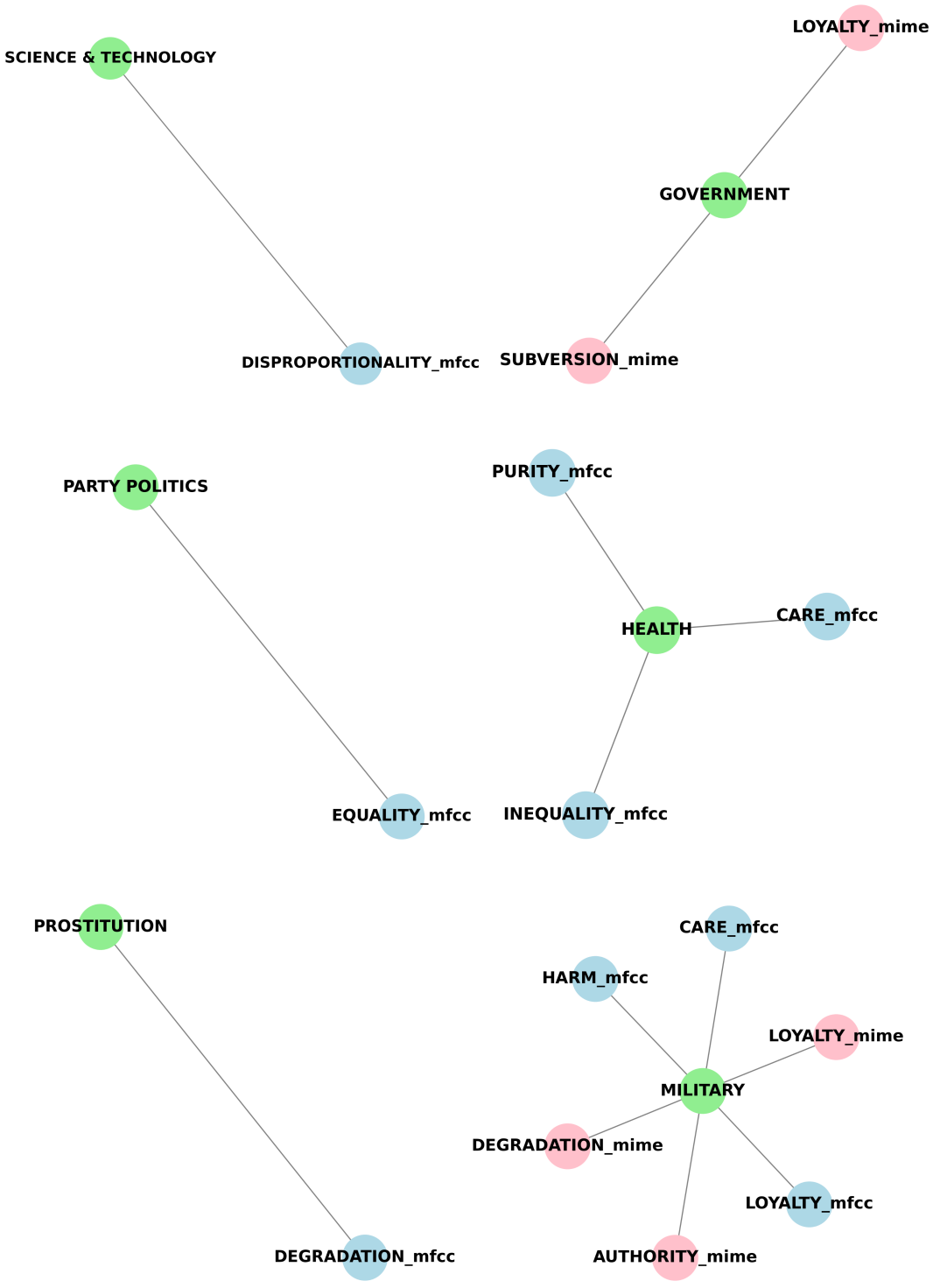
significant predictive dimension extracted from these analyses was observed at *p-value* < 0.005.

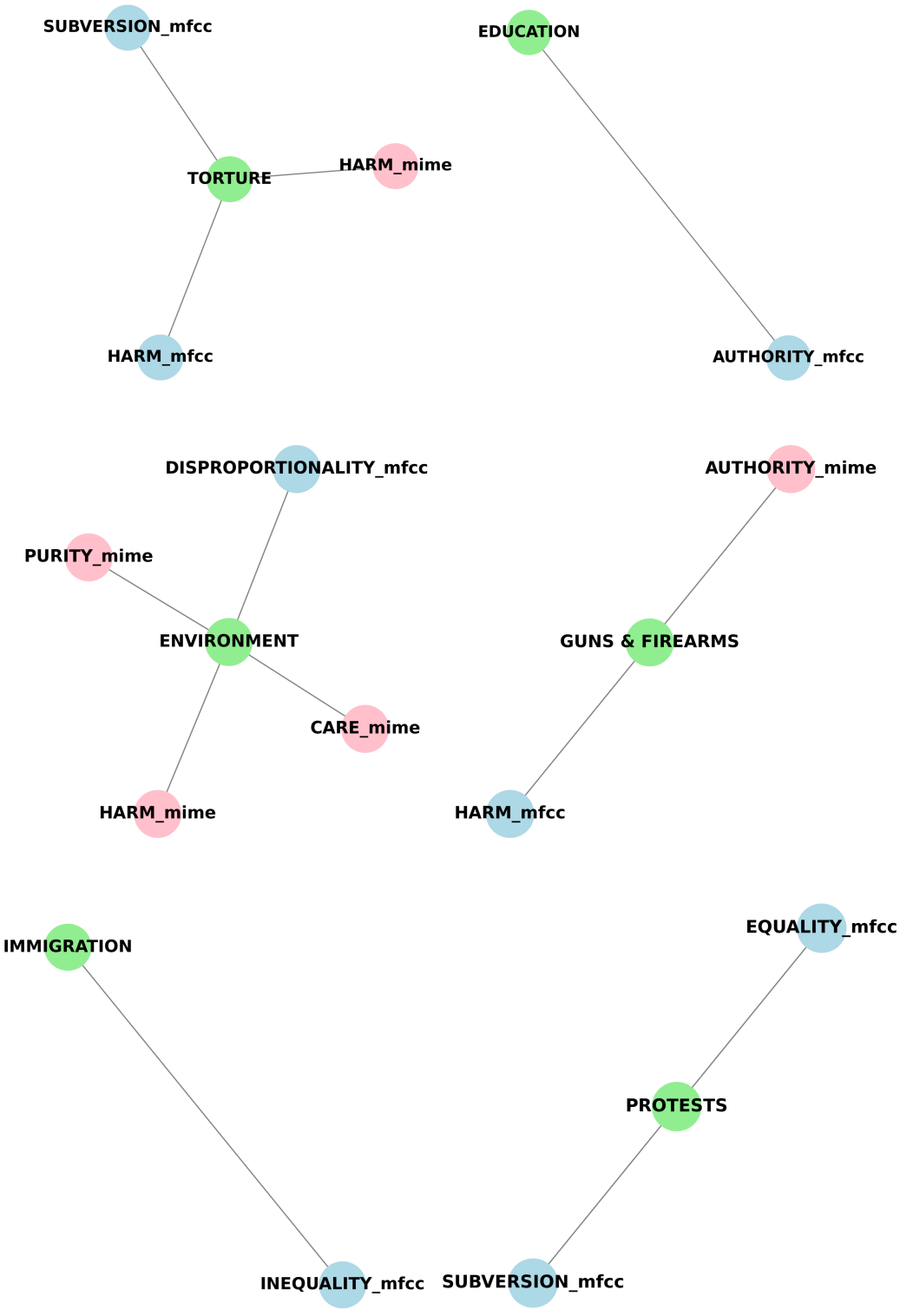
As demonstrated in Figure 9 (see Tables 10 and 11 in *SM* for exact coefficients), we observe that FDR-corrected significant relationships between specific MFCC-1 dimensional exemplification and 18 unique situational factors survived, suggesting that MFCC-1 ratings provide a comprehensive framework that is capable of robustly detecting moral signal across disparate contextual visual content. These include, for instance, (e.g., degradation → drugs & alcohol, prostitution; equality & subversion → protests; inequality → immigration). In contrast, we observe that FDR-corrected significant relationships between specific MIME dimensional exemplification and only 8 unique situational factors survived, suggesting that the validity of leveraging MIME ratings for MCA research remains limited given their empirical bias towards a small group of contextual visual content. In fact, we also observe that 6 out of the 8 situational factors amongst which MIME ratings outperform MFCC-1 ratings remain thematically similar (e.g., Government / Military / Torture / Guns & Firearms / Police System / Terrorism). This is not surprising given that within the theoretical framework of the *Model of Intuitive Morality and Exemplars*, the strength of "moral exemplification" within SMID images that thematically pertain to notions of violence and organizational hierarchy might represent moral constructs that MCA coders intuitively share a *semantic* understanding of. As shown in Figure 9, the "binding" dimensions remain most frequently predictive of such visual contexts (e.g., loyalty & subversion → government; loyalty, authority, & degradation → military; authority & subversion → police system). On the other hand, we observe substantial heterogeneity with respect to MFCC-1 dimensions' predictive capacities, highlighting that MFCC-1 ratings are not biased towards a particular

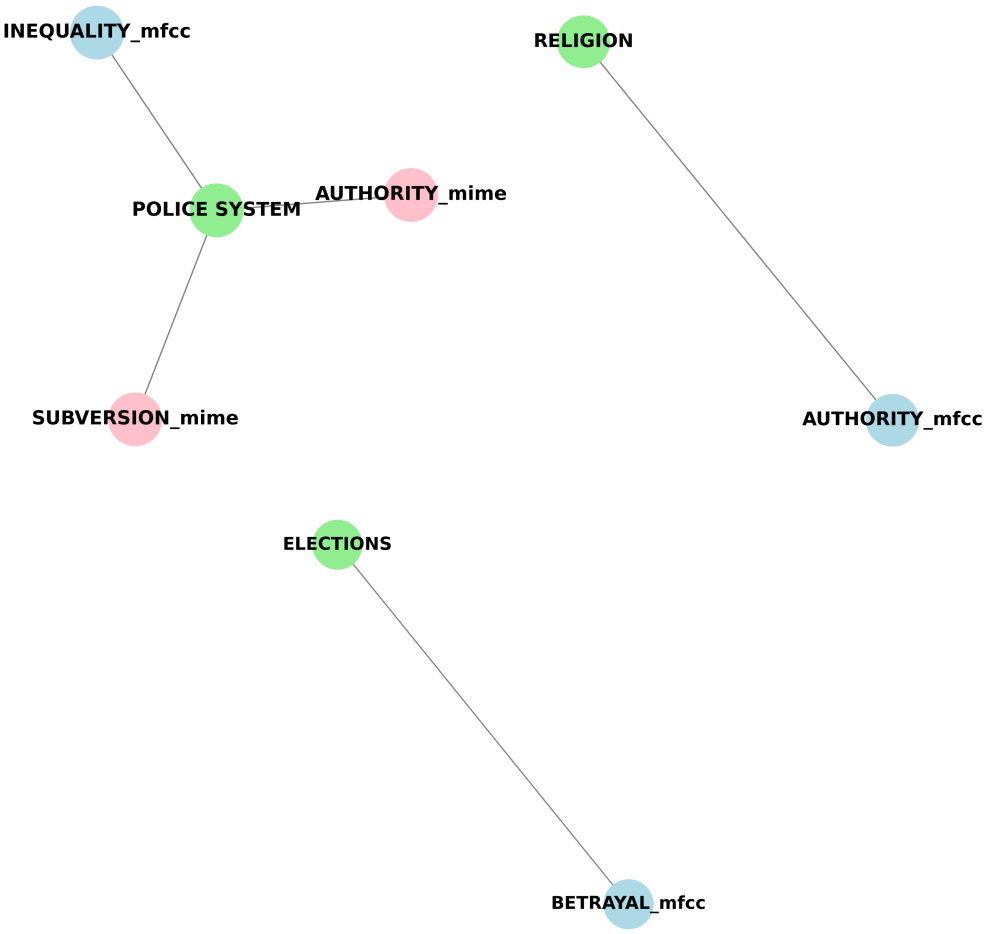
situational factor in visual content and that each dimension contributes a unique explanatory perspective within visual moral content.

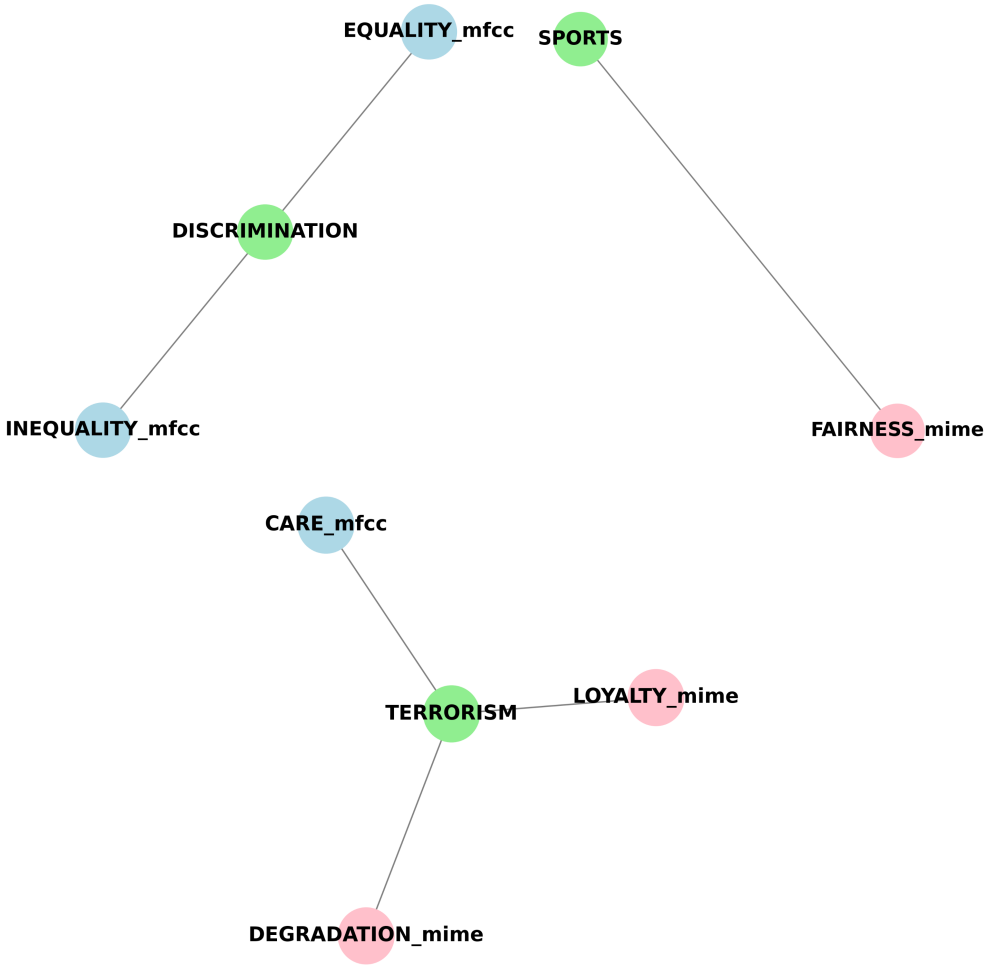
Figure 9. FDR-corrected significant relationships between moral dimension exemplification and situational factors. Green nodes indicate unique situational factors, blue nodes represent MFCC-1 derived moral dimensions, and pink nodes represent MIME derived moral dimensions.




SCIENCE & TECHNOLOGY
LOYALTY_mime
GOVERNMENT
DISPROPORTIONALITY_mfcc
SUBVERSION_mime
PARTY POLITICS
PURITY_mfcc
HEALTH
CARE_mfcc
EQUALITY_mfcc
INEQUALITY_mfcc
PROSTITUTION
CARE_mfcc
HARM_mfcc
LOYALTY_mime
MILITARY
DEGRADATION_mime
LOYALTY_mfcc
DEGRADATION_mfcc
AUTHORITY_mime


SUBVERSION_mfcc
TORTURE
HARM_mime
HARM_mfcc
EDUCATION
AUTHORITY_mfcc
DISPROPORTIONALITY_mfcc
PURITY_mime
ENVIRONMENT
CARE_mime
HARM_mime
AUTHORITY_mime
GUNS & FIREARMS
HARM_mfcc
IMMIGRATION
INEQUALITY_mfcc
EQUALITY_mfcc
PROTESTS
SUBVERSION_mfcc


INEQUALITY_mfcc
RELIGION
POLICE SYSTEM
AUTHORITY_mime
SUBVERSION_mime
AUTHORITY_mfcc
ELECTIONS
BETRAYAL_mfcc



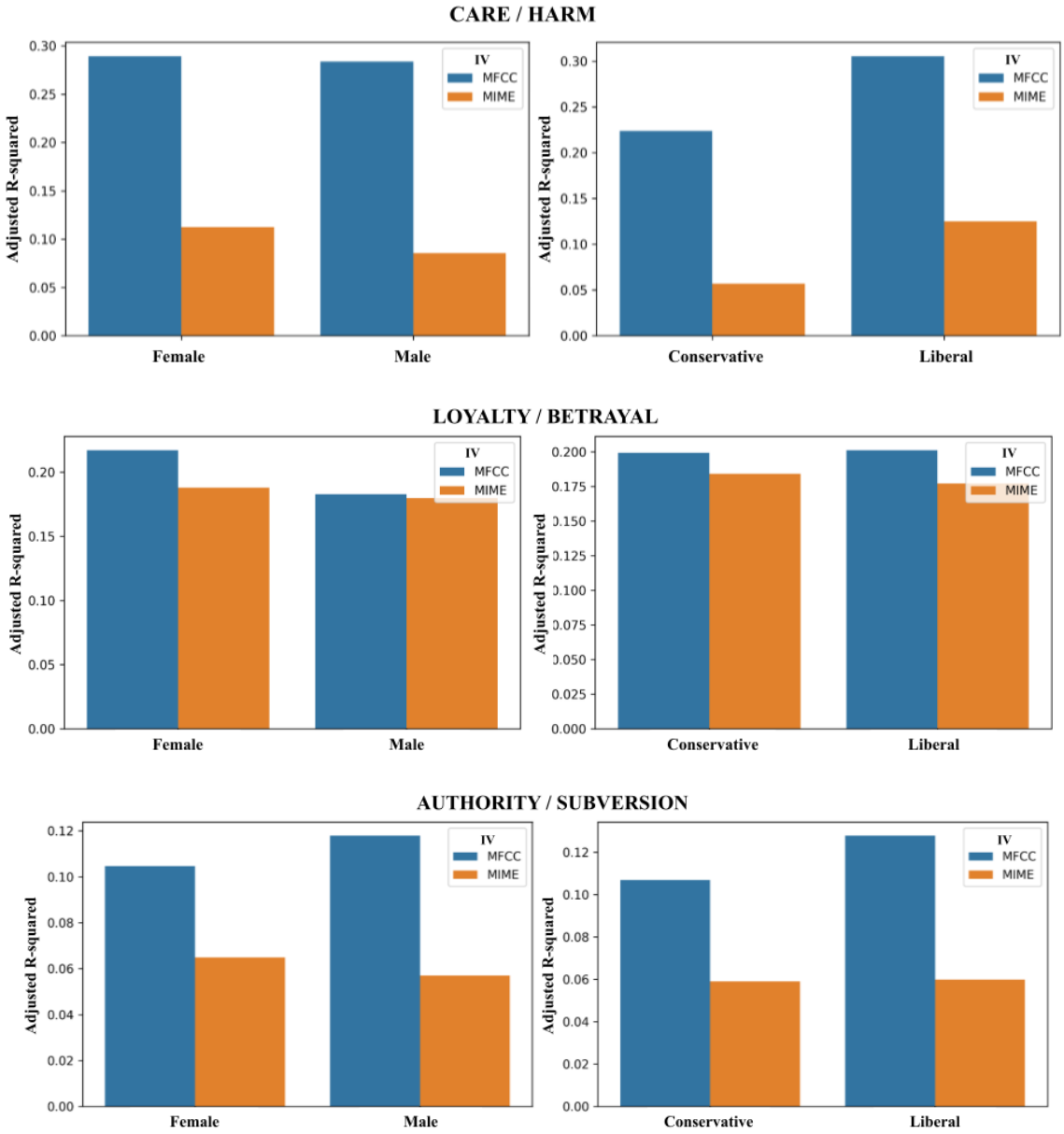
F. External Validity Analyses

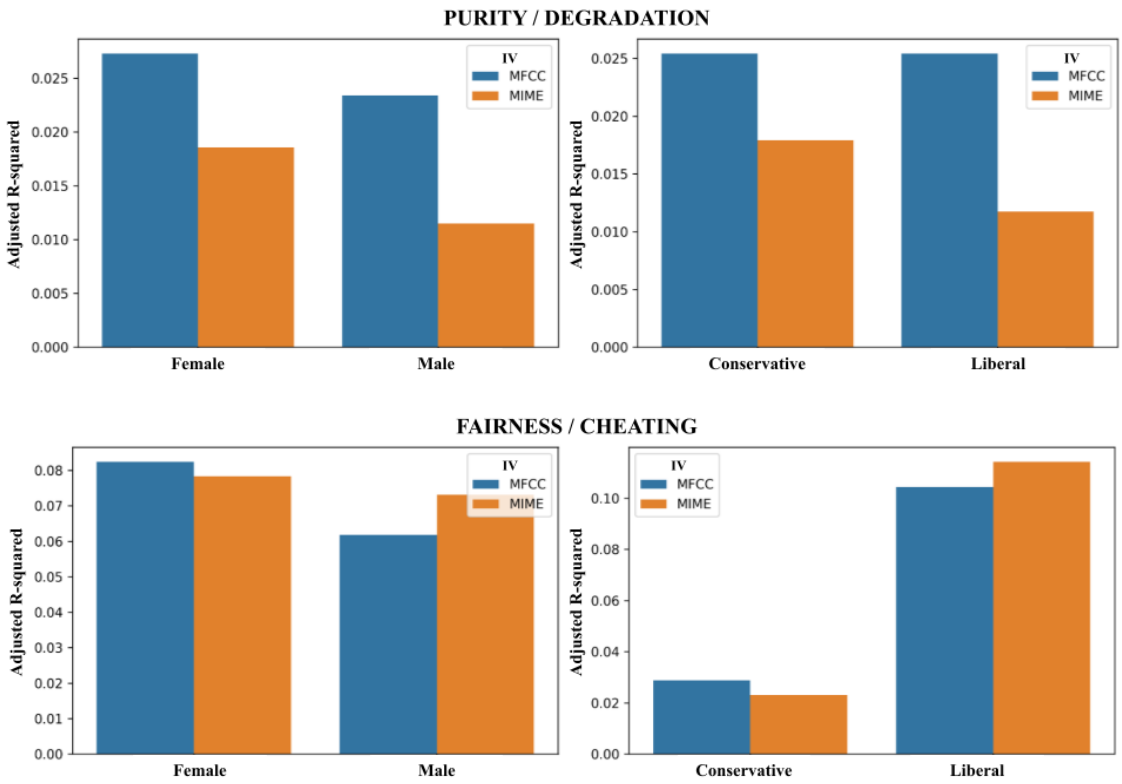
Our final set of *post-hoc* analyses investigated whether or not MFCC-1 exhibits greater external validity, as compared to MIME, when predicting within-foundation ratings generated via the out-of-sample crowd coders available within the original SMID dataset. The SMID database (n = 2941) provides rating information for each of the 5 MFT foundations on a 1 - 5 scale ("not at all" to "very much") across a crowd of coders (n = 2716). Each SMID image has been generally rated by 30 ~ 40 participants. SMID provides aggregated foundation-level ratings (e.g., Care / Harm) for each of the 5 MFT foundations, indicating the crowd - sourced respective moral signal in each image. We performed OLS

regression analyses for evaluating which dimensional ratings, i.e., MIME or MFCC-1, were significantly predictive of crowd-ratings as generated via SMID participants. To circumvent concerns around suppression effects, two separate models were constructed with the following template specification. The first set of independent variables were foundation-relevant MFCC-1 ratings (e.g., Care / Harm) while the second set of independent variables were their MIME counterparts. The dependent variables were the respective aggregated foundation-level ratings (e.g., Care / Harm) for each of the 5 MFT foundations, indicating the crowd - sourced respective moral signal in each image. Model performance was evaluated by observing the differentials between the respective adjusted *R-squared* measures acquired from each independent model. To ensure our results were consistent across heterogeneous demographics, we further measured these predictive effects of MFCC-1 / MIME on SMID ratings across the available subgroups of gender and political affiliation (i.e., Male / Female, Liberal / Conservative).

Figure 10 demonstrates the comparative external predictive capacities of individual measurement codebooks in predicting out of sample and crowd-sourced SMID codings. For instance, we observe that MFCC-1 ratings outperform MIME ratings for predicting crowd-sourced concerns related to Care / Harm (i.e., approximately 3 times greater explained variance), Authority / Subversion (i.e., approximately 2 times greater explained variance), Purity / Degradation (i.e., approximately 2 times greater explained variance), as well as Loyalty / Betrayal (i.e., marginal yet noticeable differences in explained variance). These effects persist robustly across gender and political affiliation subgroups demonstrating that these external predictive capabilities are not biased towards specific demographic subgroups (see Table 12 in *SM* for exact R_{adj} measures).

Figure 10. Comparative external predictive capabilities of MFCC-1 and MIME codebooks for predicting out of sample and crowd-sourced SMID ratings. Stratification performed along SMID participants' self-reported gender and political affiliations.





Interestingly, we observe that these effects were not observed for the Fairness / Cheating foundation and that for certain demographic subgroups (i.e., Males / Liberals) the comparative external predictive capabilities of MIME ratings outperformed MFCC-1 ratings. This is not a surprising finding, however, as SMID ratings for these specific moral constructs were extracted via an approach more aligned with our MIME expert coders (as opposed to our MFCC-1 expert coders who measured for the decomposed constructs of Equality / Inequality and Proportionality / Disproportionality). Regardless, we find it still quite impressive that MFCC-1 ratings, despite their usage of *semantically* different scales, yield explained variance measures comparable with that of MIME ratings, and for specific demographic groups even marginally superior (i.e., Females, Conservatives). Additionally, it is worth noting that SMID ratings were not expert-generated but instead crowd-sourced and the simple fact that our two female liberal coders from a large southwestern American

university were capable of predicting morality ratings extracted across a large and diverse sample of crowd-coders remains an impressive finding and indicative of the promise of refining measurement tools (as we initiate with MFCC-1) for future both *reliable* and *valid* MCA research.

IV. Discussion

The goal of this paper was to demonstrate the importance of leveraging multidimensional, as opposed to unidimensional, codebooks for MCA research. More specifically, this paper provides, to the best of our knowledge, the first piece of empirical evidence that supports the claim that multidimensional MCA codebooks yield comparatively greater ICR statistics. Our *post-hoc* analyses also demonstrate that these even exceed thresholds for acceptable reliability after implementing parameter relaxation procedures that take into account the *degree* of coder disagreement (i.e., slight versus severe). Additionally, this paper demonstrates the enhancement of discriminant, predictive, and external validities when using multidimensional MCA codebooks. In doing so, we believe that the current paper is positioned to accelerate MCA research that aims to measure visual moral signals in *comparatively* more reliable and valid ways than previously achieved. In the following discussion, we provide some explanations inspired from the cognitive bias literature that speculatively rationalize why comparatively weaker performance of unidimensional MCA codebooks was observed within our analyses. The discussion below is not meant to be a comprehensive review of the influence of cognitive biases on decision making (e.g., moral judgment or content coding tasks) but is intended to initiate conversations that evaluate whether or not the adoption of multidimensional MCA

codebooks can present a mechanism to mitigate psychological biases that can potentially emerge during routine MCA coder tasks.

While our reliability analyses predicated on exact agreement protocols were unable to approach "acceptable" thresholds, our *post-hoc* analyses demonstrated the interesting trade-off between *maximizing* coder agreement versus *minimizing* coder disagreement that MCA researchers should account for when using different codebooks. In particular, we observed that MIME ratings were consistently associated with severe disagreements while MFCC-1 ratings were more likely to exhibit slight disagreements. One explanation that we can speculatively provide for these findings is that unidimensional constructs, by virtue of their singularity, evoke binary coder perceptions and foster a cognitive environment where rating extremity thrives. While unidimensional constructs can indeed provide straightforward criterion for coder-task alignment, it's possible that they simultaneously heighten the risk of multiple cognitive biases that contaminate *the coding task itself.* In other words, we speculate that coders operating with unidimensional measurement tools remain susceptible to cognitive tendencies and mental shortcuts that inadvertently shape how they perceive and categorize content. For instance, unidimensionality in constructs can arguably elicit confirmation bias within coders (Nickerson, 1998) which results in content codings aligned with coders' preexisting *semantic,* and not necessarily *theoretical,* understandings of the construct. Accordingly, the simplicity of unidimensional constructs can encourage coders to cognitively prioritize *content features* that straightforwardly validate their previously-held schematic understandings. A theoretically-informed multidimensional approach, on the other hand, can dissuade such biases via the introduction of different

sub-dimensions that might not uniformly align with coders' *semantic* expectations, thereby, compelling them to evaluate their codings against specific benchmarks.

We also found that multidimensional MCA codebooks demonstrate greater discriminant validities. For instance, MFCC-1 ratings were capable of yielding robust discrimination between the virtuous and vicious representations of individual moral constructs, as theoretically predicted, as well as distinguish effectively between the three categories of autonomy, community, and divinity as posited within the taxonomic structure of ethics (Shweder et al., 1997). Likewise, we found that MFCC-1 ratings were capable of demonstrating unique associations between individual constructs and content categories, thereby, reflecting enhanced predictive validities which suggest the MFCC-1 ratings are not biased towards a particular content classification. One explanation that we can speculatively provide for these findings is that unidimensionality in constructs can foster anchoring biases (Furnham & Boo, 2011), where coders could potentially be categorizing content as either perfectly fitting the initial exemplars they were trained on, or not at all. This binary perspective can overlook nuances within the content that deviate from a rigid categorization inadvertently distorting the holistic representation of the construct. Similarly, coders could be influenced by the availability heuristic (Schwarz et al., 1991), leaning towards readily available examples that match the unidimensional construct. This bias might result in a skewed sample of content being coded, neglecting instances that do not immediately spring to mind, potentially leading to a bias towards specific categories of content being perceived as relevant (but not others). Multidimensional constructs counter these biases by prompting coders to consider every relevant dimension rather than relying solely upon intuitively

accessible exemplars thereby ensuring a more comprehensive and representative sample of coded content.

Finally, our analyses highlighted that MFCC-1 ratings demonstrate greater external validities, with only two female liberal coders from a large southwestern American university predicting morality ratings extracted across a large and diverse sample of crowd-coders. One explanation that we can speculatively provide for these findings is that the simplicity of MIME constructs can potentially contribute towards MIME coders inadvertently succumbing to the false consensus effect (Marks & Miller, 1987), assuming that their perspective on the unidimensional construct is widely shared. This can lead to an overestimation of the extent to which content aligns with the construct as well as overconfidence, leading them to believe they possess a definitive understanding of what constitutes a construct-code match. In contrast, the complex nature of multidimensional constructs necessitates that coders navigate complexity and acknowledge that diverse interpretations are possible, an objective which tempers *intuitive* inclinations towards specific referentials and prompts coders to strive for moderate codings that holistically encapsulate the obviously multifaceted nature of individual constructs.

The discussions outlined above are by no means exhaustive in nature (and neither are they meant to be). Regardless, it is indeed hoped that these arguments encourage debate surrounding the viewpoint that adoption of multidimensionality in MCA research can indeed present a mechanism to mitigate psychological biases that arguably emerge when coding traditional unidimensional MCA constructs. We argue that the nuanced, comprehensive, and multifaceted approach intrinsic to multidimensional MCA codebooks offer a counterbalance to the cognitive shortcuts and biases that unidimensional MCA codebooks inadvertently

encourage. Altogether, we believe that the deliberate consideration of codebook dimensionality will contribute to the methodological rigor and unbiased interpretation of content analysis, enabling a more accurate and holistic understanding of visual moral content.

V. Limitations and Future Directions

There are three major limitations inherent to this study that could benefit from corrections in future research. First, the assessment of MCA codebooks was executed by solely two coders for each codebook, potentially affecting the generalizability of the findings. Variations in coder intuitions, influenced by individual perspectives and interpretations, might have significant implications for the reliability and validity metrics derived from the analyses. It is important to acknowledge, however, that this study operates upon the premise that coder intuitions exert minimal influence on reliability and validity measures. The primary focus is placed on variations inherent to the measurement tool itself, i.e., the codebook, as the driving force behind these metrics. It is noteworthy that this assumption aligns with established practices in conventional moral content analysis research, where coder teams often operate in small groups. The absence of larger coder groups in the broader field suggests a degree of confidence in this assumption. While it is recommended for the sake of enhancing reproducibility that this study be extended to involve a larger set of coder pairs, the current study's assumption and design are consistent with prevailing literature norms.

Second, in order to ensure that coders maintained expertise within a specific format of MCA and to prevent potential confusion during the coding task of SMID 966 images, this study assigned each coder pair to a single codebook (either MIME or MFCC-1). However,

future investigations could benefit from adopting an experimental design in which coders are instructed to apply both the MIME and MFCC-1 codebooks alternatively. This approach would offer insights into the potential impacts of codebook ordering on coder behaviors and outcomes, as well as help control for biases specific to each coder pair. By systematically alternating between the two codebooks, researchers can evaluate potential discrepancies arising from codebook usage order, thus enhancing the robustness and comprehensiveness of the findings.

Finally, the multidimensional codebook proposed in this study (i.e., MFCC-1) is explicitly acknowledged to be preliminary in its formulation. Specific components within the codebook may necessitate revisions to refine their clarity and applicability. Notably, items characterized by more tangible or manifest attributes, such as "defiance against authorities," could potentially yield greater confidence in ratings, compared to items with more abstract or latent characteristics, like "displays of compassion towards others," which could invite subjective interpretations. While the current iteration of the MFCC-1 codebook was directly adapted from the MFQ-2 questionnaire, it remains imperative to recognize that room for improvement exists. Consequently, it is strongly recommended that forthcoming research endeavors take on the task of refining the codebook, incorporating lessons from the present study and building upon its insights to craft more refined and nuanced versions. This iterative approach to codebook enhancement will undoubtedly contribute to the advancement of MCA tools, demonstrate its potential of being leveraged in novel content-analytic research domains (e.g., Malik et al., 2022; Youk et al., 2023), as well as underscore our collective commitment to rigorous research practices for performing content analytic research at large.

References

1. Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment*, *38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

2. Aley, M., & Hahn, L. (2023). Altruistic and egoistic motivations of male and female characters in common sense media television programming for children. *Sex Roles*, 1-13. https://doi.org/10.1007/s11199-023-01372-y

3. Aley, M., Hahn, L., Tamborini, R., Goble, H., Zhang, L., Grady, S. M., & Baldwin, J. (2021). What does television teach children? Examining the altruistic and egoistic lessons in children's educational television. *Communication Reports*, *34*(2), 106–119. https://doi.org/10.1080/08934215.2021.1918738

4. Akinro, N., & Mbunyuza-Memani, L. (2019). Black is not beautiful: Persistent messages and the globalization of "white" beauty in African women's magazines. *Journal of International and Intercultural Communication, 12*(4), 308–324. https://doi.org/10.1080/17513057.2019.1580380

5. Askanius, T., & Keller, N. (2021). Murder fantasies in memes: fascist aesthetics of death threats and the banalization of white supremacist violence. *Information, Communication and Society, 24*(16), 2522–2539. https://doi.org/10.1080/1369118x.2021.1974517

6. Askanius, T. (2021). On Frogs, Monkeys, and Execution Memes: Exploring the Humor-Hate Nexus at the Intersection of Neo-Nazi and Alt-Right Movements in Sweden. *Television & New Media, 22*(2), 147–165. https://doi.org/10.1177/1527476420982234

7. Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology.* Advance online publication. https://doi.org/10.1037/pspp0000470

8. Barmaki, R. (2020). Explanations of blacks' criminality in America: 1630s-1950s. *Deviant Behavior, 41*(10), 1305–1329. https://doi.org/10.1080/01639625.2019.1614138

9. Barry, A. M. S. (1997). *Visual intelligence: Perception, image, and manipulation in visual communication*. SUNY Press.

10. Bast, J. (2021). Managing the image. The visual communication strategy of European right-wing populist politicians on Instagram. *Journal of Political Marketing,* 1–30. https://doi.org/10.1080/15377857.2021.1892901

11. Batova, T. (2021). “Picturing” Xenophobia: Visual Framing of Masks During COVID-19 and Its Implications for Advocacy in Technical Communication. *Journal of Business and Technical Communication, 35*(1), 50–56. https://doi.org/10.1177/1050651920958501

12. Bayrak, F., & Alper, S. (2021). A tale of two hashtags: An examination of moral content of pro‐ and anti‐government tweets in Turkey. *European Journal of Social Psychology*, *51*(3), 585–596. https://doi.org/10.1002/ejsp.2763

13. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

14. Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, *44*, 175–184. https://doi.org/10.1509/jmkr.44.2.175

15. Bennett, E. M., Alpert, R., & Goldstein, A. C. (1954). Communications through limited-response questioning. *Public Opinion Quarterly, 18*(3), 303-308. https://doi.org/10.1086/266520

16. Bertin, P., Marinthe, G., Biddlestone, M., & Delouvée, S. (2022). Investigating the identification-prejudice link through the lens of national narcissism: The role of defensive group beliefs. *Journal of Experimental Social Psychology, 98*, 104252. https://doi.org/10.1016/j.jesp.2021.104252

17. Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior.* Cambridge, MA: Harvard University Press.

18. Bouko, C., De Wilde, J., Decock, S., De Clercq, O., Manchia, V., & Garcia, D. (2021). Reactions to Brexit in images: a multimodal content analysis of shared visual content on Flickr. *Visual Communication, 20*(1), 4–33. https://doi.org/10.1177/1470357218780530

19. Bowen, E. A., & Capozziello, N. (2022). Faceless, nameless, invisible: a visual content analysis of photographs in U.S. media coverage about homelessness. *Housing Studies,* 1–20. https://doi.org/10.1080/02673037.2022.2084044

20. Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment.* New York: Basic Books.

21. Bowman, N., Lewis, R. J., & Tamborini, R. (2014). The morality of May 2, 2011: A content analysis of U.s. headlines regarding the death of Osama bin laden. *Mass Communication & Society*, *17*(5), 639–664. https://doi.org/10.1080/15205436.2013.822518

22. Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, 36, 317-346.

23. Brantner, C., Lobinger, K., & Wetzstein, I. (2011). Effects of Visual Framing on Emotional Responses and Evaluations of News Stories about the Gaza Conflict 2009. *Journalism & Mass Communication Quarterly, 88*(3), 523–540. https://doi.org/10.1177/107769901108800304

24. Braumüller, B., Emberger, D., & Hartmann-Tews, I. (2020). Gendered coverage of the Olympic Games in German print media: a longitudinal content analysis in the context of participation, success and disciplines. *European Journal for Sport and Society, 17*(4), 319–338. https://doi.org/10.1080/16138171.2020.1792086

25. Brown, E. K., & Silver, J. R. (2022). The moral foundations of crime control in American presidential platforms, 1968–2016. *Punishment & Society, 24*(2), 196–220. https://doi.org/10.1177/1462474520966979

26. Carolan, M. (2022). Presences and absences in food systems depictions: a systematic visual content analysis. *Food Security, 14*(5), 1347–1358. https://doi.org/10.1007/s12571-022-01295-8

27. Carnevale, J. J., Fujita, K., Han, H. A., & Amit, E. (2015). Immersion Versus Transcendence: How Pictures and Words Impact Evaluative Associations Assessed by the Implicit Association Test. *Social Psychological and Personality Science, 6*(1), 92–100. https://doi.org/10.1177/1948550614546050

28. Cingel, D. P., & Krcmar, M. (2020). Considering moral foundations theory and the model of intuitive morality and exemplars in the context of child and adolescent development. *Annals of the International Communication Association, 44*(2), 120-138. https://doi.org/10.1080/23808985.2020.1755337

29. Clark, S. (2020). Understanding contemporary images using iconography: Migration to the European Union and the representation of refugees and asylum seekers. *Journal of Visual Literacy, 39*(2), 111–124. https://doi.org/10.1080/1051144x.2020.1737908

30. CohenMiller, A. S., Koo, S., Collins, N., & Lewis, J. L. (2020). EXPOsing gender in science: a visual analysis with lessons for gender awareness and science diplomacy. *Gender, Technology and Development, 24*(2), 215–235. https://doi.org/10.1080/09718524.2019.1695519

31. Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

32. Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PloS One*, *13*(1), e0190954. https://doi.org/10.1371/journal.pone.0190954

33. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

34. de Waals, F. B. M. (1982). *Chimpanzee Politics: Power and Sex among Apes*. New York, NY: Harper & Row.

35. Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*(3), 434–449. https://doi.org/10.1007/s11747-011-0300-3

36. Esfandiari, M., Fridrich, B., & Yao, J. (2021). Visual content of Twitter during the 2018 protests in Iran: Analysis of its role and function. *Global Media and Communication, 17*(2), 213–230. https://doi.org/10.1177/17427665211002228

37. Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General, 145*(6), 772–787. https://doi.org/10.1037/xge0000165

38. Fahmy, S. (2004). Picturing Afghan Women: A Content Analysis of AP Wire Photographs During the Taliban Regime and after the Fall of the Taliban Regime. *Gazette (Leiden, Netherlands), 66*(2), 91–112. https://doi.org/10.1177/0016549204041472

39. Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*(1), 35-42. https://doi.org/10.1016/j.socec.2010.10.008

40. *Gallup, Inc. (2022). Gallup - workplace consulting & global research. Gallup.com; Gallup. https://www.gallup.com/*

41. Gazzaniga, M. S. (1998). The mind's past. In *The Mind's Past*. University of California Press.

42. Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, *25*(1), 46-69. https://doi.org/10.1111/comt.12048

43. Gervais, W. M., Shariff, A. F., & Norenzayan, A. (2011). Do you believe in atheists? Distrust is central to anti-atheist prejudice. *Journal of Personality and Social Psychology, 101*(6), 1189–1206. https://doi.org/10.1037/a0025882

44. Ghose, S., & Dou, W. (1998). Interactive functions and their impacts on the appeal of Internet presence sites. Journal of Advertising research, 38(2), 29-43. Retrieved from https://scholars.cityu.edu.hk/en/publications/publication(c024ac0c-49af-425a-9838-e46c218a9da3).html

45. Gibson, R., & Zillmann, D. (2000). Reading between the Photographs: The Influence of Incidental Pictorial Information on Issue Perception. *Journalism & Mass Communication Quarterly*, *77*(2), 355–366. https://doi.org/10.1177/107769900007700209

46. Goldfischer, E. (2018). "Peek-A-Boo, We See You Too": Homelessness and visuality in New York City. Environment and Planning D: *Society and Space, 36*(5), 831–848. https://doi.org/10.1177/0263775818768546

47. Grabe, M. E., & Bucy, E. P. (2009). *Image bite politics: News and the visual framing of elections*. Oxford University Press.

48. Graf, L. K. M., Mayer, S., & Landwehr, J. R. (2018). Measuring processing fluency: One versus five items. *Journal of Consumer Psychology: The Official Journal of the Society for Consumer Psychology*, *28*(3), 393–411. https://doi.org/10.1002/jcpy.1021

49. Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*(5), 1029–1046. https://doi.org/10.1037/a0015141

50. Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S. P., Ditto P. H. (2013). Moral foundations theory. In Devine P., Plant A. (Eds.), *Advances in experimental social psychology* (pp. 55–130). Elsevier. https://doi.org/10.1016/b978-0-12-407236-7.00002-4

51. Gray, K., & Wegner, D. M. (2009). Moral typecasting: divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*(3), 505–520. https://doi.org/10.1037/a0013748

52. Grizzard, M., Shaw, A. Z., Dolan, E. A., Anderson, K. B., Hahn, L., & Prabhu, S. (2017). Does repeated exposure to popular media strengthen moral intuitions?: Exploratory evidence regarding consistent and conflicted moral content. *Media Psychology*, *20*(4), 557–583. https://doi.org/10.1080/15213269.2016.1227266

53. Guo, L., & Vargo, C. (2020). "Fake News" and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 U.S. Presidential Election. *Communication Research*, *47*(2), 178–200. https://doi.org/10.1177/0093650218777177

54. Gwet, K. L. (2008). Computing inter‐rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61*(1), 29-48. https://doi.org/10.1348/000711006X126600

55. Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters.* Advanced Analytics, LLC.

56. Haßler, J., Kümpel, A. S., & Keller, J. (2021). Instagram and political campaigning in the 2017 German federal election. A quantitative content analysis of German top politicians' and parliamentary parties' posts. *Information, Communication and Society,* 1–21. https://doi.org/10.1080/1369118x.2021.1954974

57. Hahn, L., Tamborini, R., Prabhu, S., Klebig, B., Grall, C., & Pei, D. (2017). The importance of altruistic versus egoistic motivations: A content analysis of conflicted motivations in children's television programming. *Communication Reports, 30*(2), 67-79. https://doi.org/10.1080/08934215.2016.1251602

58. Hahn, L., Aley, M., Frank, A., Lawrence, C., & Lattimer, T. A. (2022). Examining the motivations of Walt Disney heroes and villains and their association with audience appeal and future film production. *Journal of Broadcasting & Electronic Media*, 1–21. https://doi.org/10.1080/08838151.2022.2120879

59. Hahn, L. (2021). Moral clarity decreases as viewer age increases: a content analysis of the moral values and reinforcement cues depicted in popular U.S. children's television. *Journal of Children and Media*, 1–20. https://doi.org/10.1080/17482798.2021.1943475

60. Hahn, L., Tamborini, R., Novotny, E., Grall, C., & Klebig, B. (2019). Applying moral foundations theory to identify terrorist group motivations. *Political Psychology*, *40*(3), 507–522. https://doi.org/10.1111/pops.12525

61. Haidt, J., & Joseph, C. (2008). The Moral Mind. In P. Carruthers et al. (Ed.). *The Innate Mind, 3*(367–392). Oxford University Press. New York. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

62. Hehman, E., Leitner, J. B., & Freeman, J. B. (2014). The Face–Time Continuum: Lifespan Changes in Facial Width-to-Height Ratio Impact Aging-Associated Perceptions. *Personality and Social Psychology Bulletin, 40*(12), 1624–1636. https://doi.org/10.1177/0146167214552791

63. Hester, N., & Gray, K. (2020). The Moral Psychology of Raceless, Genderless Strangers. *Perspectives on Psychological Science, 15*(2), 216–230. https://doi.org/10.1177/1745691619885840

64. Heuer, C. A., McClure, K. J., & Puhl, R. M. (2011). Obesity stigma in online news: a visual content analysis. *Journal of Health Communication, 16*(9), 976–987. https://doi.org/10.1080/10810730.2011.561915

65. Hoewe, J., Peacock, C., Kim, B., & Barnidge, M. (2020). The Relationship Between Fox News Use and Americans' Policy Preferences Regarding Refugees and Immigrants. *International Journal Of Communication, 14*, 21. Retrieved from https://ijoc.org/index.php/ijoc/article/view/12402/3040

66. Hoffmann, A., & Musch, J. (2019). Prejudice against women leaders: Insights from an indirect questioning approach. *Sex Roles, 80*(11–12), 681–692. https://doi.org/10.1007/s11199-018-0969-6

67. Hopp, F. R., & Weber, R. (2021). Reflections on extracting moral foundations from media content. *Communication Monographs*, *88*(3), 371–379. https://doi.org/10.1080/03637751.2021.1963513

68. Hopp, F. R., Fisher, J. T., & Weber, R. (2020). A graph-learning approach for detecting moral conflict in movie scripts. *Media & Communication, 8*(3). https://doi.org/10.17645/mac.v8i3.3155

69. Hoover, J., Portillo-Wightman, G., Yeh, L., Havaldar, S., Davani, A. M., Lin, Y., Kennedy, B., Atari, M., Kamel, Z., Mendlen, M., Moreno, G., Park, C., Chang, T. E., Chin, J., Leong, C., Leung, J. Y., Mirinjian, A., & Dehghani, M. (2020). Moral Foundations Twitter Corpus: A Collection of 35k Tweets Annotated for Moral Sentiment. *Social Psychological and Personality Science*, *11*(8), 1057–1071. https://doi.org/10.1177/1948550619876629

70. Krippendorff, K. (1970). Estimating the reliability, systematic error and random error of interval data. *Educational and psychological measurement, 30*(1), 61-70. https://doi.org/10.1177/001316447003000105

71. Inbar, Y., & Pizarro, D. A. (2022). How disgust affects social judgments. In *Advances in Experimental Social Psychology, 65*, 109-166. Academic Press. https://doi.org/10.1016/bs.aesp.2021.11.002

72. Ireri, K. (2022). Images from the battlefields: How Al-jazeera and BBC news sites visually framed the Libyan revolution. *Visual Communication Quarterly, 29*(1), 48–61. https://doi.org/10.1080/15551393.2021.1949997

73. Issa, S. (2022). Picturing atrocity: visual representations of ISIS in Arabic political cartoons. *Visual Studies, 37*(5), 475–490. https://doi.org/10.1080/1472586x.2020.1866441

74. Jardina, A., & Piston, S. (2021). Hiding in plain sight: Dehumanization as a foundation of white racial prejudice. *Sociology Compass, 15*(9). https://doi.org/10.1111/soc4.12913

75. Janoff-Bulman, R., & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review, 17*(3), 219–236. https://doi.org/10.1177/1088868313480274

76. Jones, J., & Himelboim, I. (2010). Just a guy in pajamas? Framing the blogs in mainstream US newspaper coverage (1999—2005). *New Media & Society, 12*(2), 271–288. https://doi.org/10.1177/1461444809342524

77. Joo, J., & Steinert-Threlkeld, Z. C. (2022). Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research, 4*(1). https://doi.org/10.5117/ccr2022.1.001.joo

78. Joo, J., Li, W., Steen, F. F., & Zhu, S. C. (2014). Visual persuasion: Inferring communicative intents of images. *Proceedings of the IEEE conference on computer vision and pattern recognition* (216–223). Retrieved from https://openaccess.thecvf.com/content_cvpr_2014/html/Joo_Visual_Persuasion_Inferring_2014_CVPR_paper.html

79. Kapidzic, S., & Herring, S. C. (2015). Race, gender, and self-presentation in teen profile photographs. *New Media & Society, 17*(6), 958–976. https://doi.org/10.1177/1461444813520301

80. Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Journal of Social Psychology, 38*(7), 1175–1183. https://doi.org/10.1002/ejsp.526

81. Kennedy, B., Atari, M., Davani, A. M., Hoover, J., Omrani, A., Graham, J., & Dehghani, M. (2021). Moral concerns are differentially observable in language. *Cognition, 212*, 104696. https://doi.org/10.1016/j.cognition.2021.104696

82. Kinney, N. T. (2006). Engaging in 'loose talk': Analyzing salience in discourse from the formulation of welfare policy. Policy Sciences, 38(4), 251–268. https://doi.org/10.1007/s11077-006-9009-4

83. Klebig, B., Hahn, L., Tamborini, R., Aley, M., Goble, H., & Baldwin, J. (2021). The influence of prevailing regional political narratives on intuitions featured in religious messages. *Communication Reports*, *34*(1), 37–49. https://doi.org/10.1080/08934215.2020.1869281

84. Korkmazer, B., De Ridder, S., & Van Bauwel, S. (2021). The visual digital self: A discourse theoretical analysis of young people's negotiations on gender, reputation

and sexual morality online. *DiGeSt Journal of Diversity and Gender Studies, 8*(1), 22–40. https://doi.org/10.21825/digest.v8i1.17608

85. Krause, A., & Bucy, E. P. (2018). Interpreting images of fracking: How visual frames and standing attitudes shape perceptions of environmental risk and economic benefit. *Environmental Communication, 12*(3), 322–343. https://doi.org/10.1080/17524032.2017.1412996

86. Krippendorff, K. (1989). Content analysis. In E. Barnouw, G. Gerbner, W. Schramm, T. L. Worth, & L. Gross (Eds.), *International encyclopedia of communication*, *1*(403-407). Oxford University Press. Retrieved from http://repository.upenn.edu/asc_papers/226

87. Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences, 98*(26), 15387-15392. https://doi.org/10.1073/pnas.251541498

88. Lang, A., Bailey, R. L., & Connolly, S. R. (2015). Encoding systems and evolved message processing: Pictures enable action, words enable thinking. *Media and Communication, 3*(1), 34-43. https://doi.org/10.17645/mac.v3i1.248

89. Lewis, R. J., & Mitchell, N. (2014). Egoism versus altruism in television content for young audiences. *Mass Communication & Society*, *17*(4), 597–613. https://doi.org/10.1080/15205436.2013.816747

90. Long, J. A., & Eveland, W. P., Jr. (2021). Entertainment use and political ideology: Linking worldviews to media content. *Communication Research*, *48*(4), 479–500. https://doi.org/10.1177/0093650218791011

91. Lovejoy, J., Watson, B. R., Lacy, S., & Riffe, D. (2016). Three decades of reliability in communication content analyses: reporting of reliability statistics and coefficient levels in three top journals. *Journalism & Mass Communication Quarterly, 93*(4), 1135–1159. https://doi.org/10.1177/1077699016644558

92. Lundberg, K. (2021). Visual criminology and lives lived in public space. *City, 25*(1–2), 108–128. https://doi.org/10.1080/13604813.2021.1885915

93. Malik, M., Hopp, F. R., Chen, Y., & Weber, R. (2021). Does regional variation in pathogen prevalence predict the moralization of language in COVID-19 news? *Journal of Language and Social Psychology*, *40*(5–6), 653–676. https://doi.org/10.1177/0261927x211044194

94. Malik, M., Hopp, F. R., & Weber, R. (2022). Representations of Racial Minorities in Popular Movies: A Content-Analytic Synergy of Computer Vision and Network Science. *Computational Communication Research, 4*(1). https://doi.org/10.5117/CCR2022.1.006.MALI

95. Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin, 102*(1), 72–90. https://doi.org/10.1037/0033-2909.102.1.72
96. Majavu, M. (2020). The 'African gangs' narrative: associating Blackness with criminality and other anti-Black racist tropes in Australia. *African and Black Diaspora: An International Journal, 13*(1), 27–39. https://doi.org/10.1080/17528631.2018.1541958

97. Marques, S., Mariano, J., Mendonça, J., De Tavernier, W., Hess, M., Naegele, L., Peixeiro, F., & Martins, D. (2020). Determinants of Ageism against Older Adults: A Systematic Review. *International Journal of Environmental Research and Public Health, 17*(7), 2560. https://doi.org/10.3390/ijerph17072560

98. McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods, 1*(1), 30. https://doi.org/10.1037/1082-989X.1.1.30

99. Milanesi, M., Kyrdoda, Y., & Runfola, A. (2022). How do you depict sustainability? An analysis of images posted on Instagram by sustainable fashion companies. *Journal of Global Fashion Marketing, 13*(2), 101–115. https://doi.org/10.1080/20932685.2021.1998789

100. Mortensen, M., & Trenz, H.-J. (2016). Media morality and visual icons in the age of social media: Alan kurdi and the emergence of an impromptu public of moral spectatorship. *Javnost - The Public, 23*(4), 343–362. https://doi.org/10.1080/13183222.2016.1247331

101. Mortensen, T. (2015). Visually assessing the first lady in a digital age: A study of Michelle Obama as portrayed by journalists and the White House. *Journal of Women, Politics & Policy, 36*(1), 43–67. https://doi.org/10.1080/1554477x.2015.985152

102. Morris, J. S. (2009). The daily show with Jon Stewart and audience attitude change during the 2004 party conventions. *Political Behavior*, *31*(1), 79–102. https://doi.org/10.1007/s11109-008-9064-y

103. Muylle, S., Moenaert, R., & Despontin, M. (2004). The conceptualization and empirical validation of web site user satisfaction. *Information & Management, 41*(5), 543–560. https://doi.org/10.1016/s0378-7206(03)00089-2

104. Napier, J. L., & Luguri, J. B. (2013). Moral mind-sets: Abstract thinking increases a preference for "individualizing" over "binding" moral foundations. *Social Psychological and Personality Science, 4*(6), 754–759. https://doi.org/10.1177/1948550612473783

105. Ng, R., & Lim-Soh, J. W. (2021). Ageism linked to culture, not demographics: Evidence from an 8-billion-word corpus across 20 countries. *The Journals of Gerontology. Series B, 76*(9), 1791–1798. https://doi.org/10.1093/geronb/gbaa181

106. Neuendorf, K. A. (2011). Content analysis—A methodological primer for gender research. *Sex Roles*, *64*(3–4), 276–289. https://doi.org/10.1007/s11199-010-9893-0

107. Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

108. Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research, 27*(3), 258–284. https://doi.org/10.1080/00909889909365539

109. Pettegrew, L. S. (1988). The importance of context in applied communication research. *Southern Speech Communication Journal, 53*(4), 331-338. https://doi.org/10.1080/10417948809372734

110. Pradantyo, R., Birk, M. V., & Bateman, S. (2021). How the visual design of video game antagonists affects perception of morality. *Frontiers in Computer Science, 3*. https://doi.org/10.3389/fcomp.2021.531713

111. Průchová Hrůzová, A., & Zápotocký, J. (2022). Seemingly innocent pictures. Visual representations of migrants and refugees circulating through the Czech Facebook nationalistic universe. *Visual Studies, 37*(4), 348–362. https://doi.org/10.1080/1472586x.2021.1953397

112. Radwin, L. E., & Cabral, H. J. (2010). Trust in nurses scale: construct validity and internal reliability evaluation. *Journal of Advanced Nursing, 66*(3), 683–689. https://doi.org/10.1111/j.1365-2648.2009.05168.x

113. Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*(1), 57–75. https://doi.org/10.1037/a0021867

114. Rebich-Hespanha S., & Rice R. E.. (2016). Dominant visual frames in climate change news stories: Implications for formative evaluation in climate change campaigns. *International Journal of Communication, 10*(1), 4830–62. Retrieved from https://escholarship.org/uc/item/2gj9n0c1

115. Rebich-Hespanha, S., Rice, R. E., Montello, D. R., Retzloff, S., Tien, S., & Hespanha, J. P. (2015). Image themes and frames in US print news stories about climate change. *Environmental Communication, 9*(4), 491–519. https://doi.org/10.1080/17524032.2014.983534

116. Rogers, B. (2020). #biguys and #biboys: The discursive production of bisexual men through Instagram's homonormative visual culture. *Journal of Bisexuality, 20*(4), 361–382. https://doi.org/10.1080/15299716.2020.1823298

117. Roper, R. L. (2019). Does gender bias still affect women in science? *Microbiology and Molecular Biology Reviews: MMBR, 83*(3). https://doi.org/10.1128/MMBR.00018-19

118. Rozin, P., Haidt, J., & McCauley, C. R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (2nd ed., 637–653). New York, NY: Guilford Press.

119. Sarstedt, M., Diamantopoulos, A., Salzberger, T., & Baumgartner, P. (2016). Selecting single items to measure doubly concrete constructs: A cautionary tale. *Journal of Business Research*, *69*(8), 3159–3167. https://doi.org/10.1016/j.jbusres.2015.12.004

120. Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Die Betriebswirtschaft, 69*(2), 211-227. Retrieved from https://www.proquest.com/scholarly-journals/more-less-comparison-single-item-multi-measures/docview/750491216/se-2

121. Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review, 22*(1), 32–70. https://doi.org/10.1177/1088868317698288

122. Schill, D. (2012). The visual image and the political image: A review of visual communication research in the field of political communication. *The Review of Communication, 12*(2), 118–142. https://doi.org/10.1080/15358593.2011.653504

123. Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195–202. https://doi.org/10.1037/0022-3514.61.2.195

124. Shweder, R. A., Much, N. C., Mahapatra, M. and Park, L. 1997. The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In *Morality and health*, 119–169. New York, NY: Routledge.

125. Scott, W. A. (1955). Reliability of content analysis: The case of nominal scaling. *Public Opinion Quarterly, 19*(3), 321–325. https://doi.org/10.1086/266577

126. Simon, T. F., Fico, F., & Lacy, S. (1989). Covering conflict and controversy: Measuring balance, fairness, defamation. *Journalism Quarterly, 66*(2), 427. Retrieved from

https://www.proquest.com/scholarly-journals/covering-conflict-controversy-measuring-balance/docview/1290761732/se-2

127. Slovic, P., Västfjäll, D., Erlandsson, A., & Gregory, R. (2017). Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. *Proceedings of the National Academy of Sciences, 114*(4), 640-644. https://doi.org/10.1073/pnas.1613977114

128. Soler, J. (1979). The semiotics of food in the Bible. In R. Forster & O. Ranum (Eds.), *Food and drink in history* (E. Forster & P. M. Ranum, Trans., 126 –138). Baltimore, MD: Johns Hopkins University Press. (Original work published 1973)

129. Soroka, S., Loewen, P., Fournier, P., & Rubenson, D. (2016). The impact of news photos on support for military action. *Political Communication, 33*(4), 563–582. https://doi.org/10.1080/10584609.2015.1133745

130. Stoet, G., & Geary, D. C. (2018). The Gender-Equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581–593. https://doi.org/10.1177/0956797617741719

131. Strabac, Z., & Listhaug, O. (2008). Anti-Muslim prejudice in Europe: A multilevel analysis of survey data from 30 countries. *Social Science Research, 37*(1), 268–286. https://doi.org/10.1016/j.ssresearch.2007.02.004

132. Tamborini R. (2013). Model of intuitive morality and exemplars. In Tamborini R. (Ed.), *Media and the moral mind* (43–74). Routledge. https://doi.org/10.4324/9780203127070

133. Tamborini, R., Hahn, L., Prabhu, S., Klebig, B., & Grall, C. (2017). The representation of altruistic and egoistic motivations in children's TV programming. *Communication Research Reports: CRR*, *34*(1), 58–67. https://doi.org/10.1080/08824096.2016.1227312

134. Tamborini R., Weber R. (2020). Advancing the model of intuitive morality and exemplars. In Floyd K., Weber R. (Eds.), *The handbook of communication science and biology* (456–469). Routledge. https://doi.org/10.4324/9781351235587

135. Tamborini, R., Hahn, L., Klebig, B., Walling, B. M., Kryston, K., & Aley, M. R. (2021). The representation of altruism and egoism in children's books. *Communication Studies*, *72*(2), 163–177. https://doi.org/10.1080/10510974.2020.1833356

136. Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology, 46*(1), 35-57. https://doi.org/10.1086/406755

137. Tukachinsky, R., Mastro, D., & King, A. (2011). Is a picture worth a thousand words? The Effect of Race-Related Visual and Verbal Exemplars on Attitudes and Support for Social Policies. *Mass Communication and Society*, 14(6), 720–742. http://doi:10.1080/15205436.2010.530385

138. Van Gorp, B. (2007). The constructionist approach to framing: Bringing culture back in. *The Journal of Communication, 57*(1), 60–78. https://doi.org/10.1111/j.0021-9916.2007.00329.x

139. Varava, K. (2016). Visual framing, stigmatization, race, and obesity: Examining television news presentation of stories about obesity and stories about health. *Communication Studies, 67*(5), 509–529. https://doi.org/10.1080/10510974.2016.1236348

140. Vargo, C. J., & Guo, L. (2017). Networks, Big Data, and Intermedia Agenda Setting: An Analysis of Traditional, Partisan, and Emerging Online U.S. News. *Journalism & Mass Communication Quarterly*, *94*(4), 1031–1055. https://doi.org/10.1177/1077699016679976

141. Wang, R., & Liu, W. (2021). Moral framing and information virality in social movements: A case study of #HongKongPoliceBrutality. *Communication Monographs*, 1–21. https://doi.org/10.1080/03637751.2021.1918735

142. Weber, R., Mangus, J. M., Huskey, R., Hopp, F. R., Amir, O., Swanson, R., Gordon, A., Khooshabeh, P., Hahn, L., & Tamborini, R. (2018). Extracting latent moral information from text narratives: Relevance, challenges, and solutions. *Communication Methods and Measures, 12*(2–3), 119–139. https://doi.org/10.1080/19312458.2018.1447656

143. Wendell, D. G., & Tatalovich, R. (2021). Classifying public policies with Moral Foundations Theory. *Policy Sciences*, *54*(1), 155–182. https://doi.org/10.1007/s11077-020-09399-8

144. Wollast, R., Lüders, A., Nugier, A., Khamzina, K., de la Sablonnière, R., & Guimond, S. (2022). Social dominance and anti-immigrant prejudice: A cross-national and prospective test of the mediating role of assimilation, multiculturalism, colour blindness, and interculturalism. *European Journal of Social Psychology*, 00, 1– 16. https://doi.org/10.1002/ejsp.2895

145. Yan, Y., & Bissell, K. (2014). The globalization of beauty: How is ideal beauty influenced by globally published fashion and beauty magazines? *Journal of Intercultural Communication Research, 43*(3), 194–214. https://doi.org/10.1080/17475759.2014.917432

146. Yang, S., Maloney, E. K., Tan, A. S., & Cappella, J. N. (2018). When visual cues activate moral foundations: Unintended effects of visual portrayals of vaping within

electronic cigarette video advertisements. *Human Communication Research, 44*(3), 223-246. https://doi.org/10.1093/hcr/hqy004

147. Youk, S., Malik, M., Chen, Y., Hopp, F. R., & Weber, R. (2023). Measures of Argument Strength: A Computational, Large-Scale Analysis of Effective Persuasion in Real-World Debates. *Communication Methods and Measures,* 1-23. https://doi.org/10.1080/19312458.2023.2230866

148. Zillmann, D. (1999). Exemplification theory: Judging the whole by some of its parts. *Media Psychology, 1*(1), 69-94. https://doi.org/10.1207/s1532785xmep0101_5

149. Zhang, Y. (2009). Individualism or Collectivism? Cultural orientations in Chinese TV commercials and analysis of some moderating factors. *Journalism & Mass Communication Quarterly, 86*(3), 630–653. https://doi.org/10.1177/107769900908600311

150. Zhou, Z., Sun, J., Pei, J., Peng, N., & Xiong, J. (2022). *A moral- and event-centric inspection of gender bias in fairy tales at A large scale*. https://doi.org/10.48550/ARXIV.2211.14358