Routledge
Taylor & Francis Group

Communication Monographs

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rcmm20

# Different pathways to identify moral framing from media content: A response to Hopp and Weber

Rong Wang & Wenlin Liu

**To cite this article:** Rong Wang & Wenlin Liu (2021) Different pathways to identify moral framing from media content: A response to Hopp and Weber, Communication Monographs, 88:3, 380-388, DOI: 10.1080/03637751.2021.1963516

**To link to this article:** https://doi.org/10.1080/03637751.2021.1963516

Published online: 24 Aug 2021.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rcmm20

COMMUNICATION MONOGRAPHS
2021, VOL. 88, NO. 3, 380–388
https://doi.org/10.1080/03637751.2021.1963516

INVITED COMMENTARY AND RESPONSE

# Different pathways to identify moral framing from media content: A response to Hopp and Weber

Rong Wang [a] and Wenlin Liu[b]

[a]Department of Communication, College of Communication & Information, University of Kentucky, Lexington KY, USA; [b]Jack J. Valenti School of Communication, University of Houston, Houston TX, USA

**ABSTRACT**

Moral framing is a mobilizing strategy in digital activism to raise awareness of a social issue. This essay discussed the strengths and weaknesses of multiple methods of extracting moral framing from media content. Two issues were highlighted from crowdcoding and automated coding using computational approaches: context sensitivity and the handling of coding discrepancies. We compared results from manual content coding and automated analysis using the eMFD, urging scholars to incorporate multiple methods to uncover the meanings behind each moral value. We argued that crowdcoding, computational approaches and manual coding are not exclusive of one another, and encouraged scholars to diversify their methodological toolbox and choose what to use based on the research context.

**KEYWORDS**

Moral framing; content analysis; moral foundation theory; digital activism

Guided by moral foundations theory (MFT) and information virality theory, our recent work (Wang & Liu, 2021) on the 2019–2020 Hong Kong anti-extradition bill protests used manual content analysis and semantic network analysis to identify what words from tweets could be mapped onto moral framing and what types of moral framing were more likely to go viral on social media. Drawing from the literature on social movement and collective action, we pointed out that existing studies have *primarily* operationalized moral framing as a unidimensional concept, that is, the overall degree to which a social issue is moralized in media content (p. 5). We further proposed a codebook to conduct manual content coding and detected the use of moral framing along the five dimensions of MFT (i.e., care, fairness, loyalty, authority, and sanctity). Our approach offered a nuanced understanding of the mobilization process of social movement participants.

Hopp and Weber (2021) provided a commentary on our work. In their commentary, Hopp and Weber (2021) pointed out that "the coding of moral foundations has a rich history in the discipline of communication and cognate social science fields, and is not confined to idiosyncratic, unidimensional codings of moral foundation" (p. 372). We agree with this argument and want to clarify that our statement about the prevalence of using unidimensional coding of moral foundations is situated in the social movement literature, which often measures the overall moralization of a particular social issue in a

**CONTACT** Rong Wang rong.wang@uky.edu
© 2021 National Communication Association

text (Mooijman et al., 2018). Our study thus fills the gap in the literature by offering a more contextualized interpretation of moral foundations and implications on how to leverage moral framing to mobilize support for collective action.

Hopp and Weber (2021) further offered the following critiques related to the use of manual content coding to identify moral framing: first, manual content analysis is less able to objectively capture moral values; second, expert coding masks more nuanced differences that may otherwise emerge during crowdsourced coding; and finally, lower intercoder reliability in content coding signals different coders' identities. They further proposed that crowd-coding or automated coding with the extended Moral Foundation Dictionary (eMFD) could serve as a better alternative. In light of these comments, we revisit the argument from existing literature that human judgment is still needed in identifying moral framing, as words cannot be understood correctly if taken out of context and analyzed or counted individually (Hoover et al., 2020). Below, we offer a conversation in response to their main arguments.

## Manual content analysis versus alternative approaches in the changing media landscape

Hopp and Weber's (2021) critique has cogently summarized several weaknesses associated with traditional manual content coding, particularly regarding its applicability in analyzing latent content from social media data. As they argued, manual content analysis is more likely to produce biased moral coding due to endogenous (e.g., factors inherent to the specific coding units, such as contexts of the text and text length) and exogenous (e.g., factors related to individual coders) factors. Furthermore, manual content analysis has largely relied on a binary coding approach (i.e., to code either presence or absence of a moral framing), which overlooks "*what* content in a text actually captures the moral information" (Hopp & Weber, 2021, p. 374, emphasis in original).

We fully agree with Hopp and Weber's (2021) critique of traditional content analysis. Such criticism is in line with the field's growing concern of exclusively relying on manual content coding to analyze digital trace data (e.g., Lewis et al., 2013; van Atteveldt et al., 2021). Other than the validity issues noted above, researchers have also pointed out the high costs and relatively low efficiency of manual content coding, which often relies on various sampling techniques to reduce the amount of data analyzed. These weaknesses all make the alternatives, including various computational content analysis approaches, seemingly more desirable.

The past few years have seen a rapid development of a family of computational methods that specialize in identifying and classifying content characteristics from a large text corpus. These methods include different combinations of dictionaries and the use of machine learning algorithms (Kennedy et al., 2021). Crowd-coding is often used in developing these dictionaries. For example, the eMFD dictionary, the method advocated by Hopp and Weber (2021) in their commentary, was developed through a set of rigorous procedures (see Hopp et al., 2021 for details) and represents one of the state-of-the-art tools to extract moral values from media texts. Its merits also lie in the measurement of both intensity (i.e., probability) and valence (i.e., vice or virtue) of each moral value, thus offering a more in-depth analysis of which aspect of a moral value (e.g., care versus harm) is present and to what degree.

However, even with such a meticulously developed method, we would like to note with caution that it may not be a perfect solution to the reliability and validity issues commonly associated with manual content analyses of texts. The dictionary-based approach often relies on word count and cannot capture intended sentiment, whereas human coders can detect the moral signal, which is particularly needed if background knowledge is required for case studies (Atari & Dehghani, 2021). Below, we outline a few points that are central in weighing the pros and cons of manual versus the two methods recommended by Hopp and Weber (2021): crowd-coding and automated computational content analysis. Our intention is not to draw conclusions on which approach performs better, but to provide anchor points for researchers to evaluate the strengths and weaknesses of each method and to decide what to use based on the context of the texts.

### *Context sensitivity*

The advantage of using crowd-based annotation is well elaborated in Hopp and Weber's (2021) commentary. To summarize, this approach leverages the "wisdom of the crowd," so that the same piece of content can be validated by multiple, potentially diverse crowd workers via intuition rather than a small group of homogenous experts, as the latter may be over-sensitized and produce biased coding. Although crowd-coding may be better at establishing objectivity, which is certainly crucial for quantitative content analysis, we would like to bring up an equally important issue, *context sensitivity*.

As elaborated by Kennedy et al. (2021), measuring morality in texts is a challenging task due to its complex linguistic nature. Fundamentally, detecting moral framing in text requires the understanding of contextual flexibility (i.e., a given word can have multiple meanings based on context). Being able to assign context-specific coding is challenging particularly when conducting implicit or latent content analysis, and this issue is further exacerbated by short text formats, such as tweets. Although certain crowd-coding methods have demonstrated a decent level of context recognition (Weber et al., 2018), the level of sensitivity is still not on par with that of human coding. And this is especially true when the text is too short or contains features such as sarcasm and humor (Hoover et al., 2020). As Simon (2001) lamented, "the computer is simply unable to understand human language in all its richness, complexity, and subtlety as can a human coder" (p. 87). Empirical evidence appears to support this claim. Hoover et al. (2020) assessed four different dictionaries and algorithms to annotate the moral sentiments of over 35,000 tweets. Their comparison suggested that each methodology showed strengths and weaknesses depending on the context of different domains (e.g., disasters vs. social movements). Not surprisingly, they found that the relevance of a particular moral foundation to a given domain was highly contingent on the domain's content. Furthermore, they argued that human coders need to be trained to address two sources of ambiguity: the appearance of multiple moral values appearing in one text, and the inference of the moral relevance intended by an author and the moral relevance encoded in their language. Therefore, we still need to be cautious to assume that a dictionary, even a well-established one, is universal and can be directly applied to texts from different contexts.

Regardless of the type of media text, a valid coding of latent content is always based on the accurate interpretation of the context (Krippendorff, 2018). Currently, it is still

reasonable to expect that a carefully executed manual coding protocol designed by experts and researchers outperforms most computational approaches (van Atteveldt et al., 2021). This is supported by the fact that manual coding is widely used to create the baseline in assessing the validity and reliability of machine-assisted coding. In addition, the likelihood of producing biases by using a computational approach is no less than that from using expert manual coding. Computational algorithms reflect biases of their makers and the data fed into them, which may well result in the reproduction or amplification of such biases (Noble, 2018).

Depending on the extent to which context sensitivity is prioritized in a research project, manual content analysis may still be preferred, as in our study (i.e., Wang & Liu, 2021). We examined a contentious social issue (i.e., police brutality that occurred in Hong Kong) that is intertwined with political tensions between mainland China and Hong Kong. In this specific context, content coding with experts is not necessarily noisier; instead, it helps to identify the rich context of moral framing (Graham et al., 2009). A dictionary-based approach faces the challenge of pragmatic validity as it is unlikely to capture sociolects in non-Western contexts, which has been termed as the "home-field disadvantage" (Atari & Dehghani, 2021, p. 16). Existing dictionaries have only been validated in a few non-English languages (see Alper et al., 2020 for a Turkey translation and Matsuo et al., 2019 for a Japanese translation). Although the tweets analyzed in our study were in English, the foreign context and a large number of non-native English speakers authoring the tweets (note that 67% of the accounts self-identified as Hong Kong residents) may pose limitations to the application of dictionary-based approaches.

### Coding discrepancies

Coding disagreement is inevitable in almost all forms of content coding, and how a method handles coding discrepancies has important consequences for establishing an acceptable level of coding reliability. Manual content analysis takes a *partial resolution approach* in the sense that coders will deliberately discuss coding discrepancies, and produce a revised codebook and/or adjusted coding. Disagreement is not expected to be fully resolved in manual content analysis. Various measures, such as Cohen's kappa or Krippendorff's alpha, are used to gauge the level of reliability.

There are well-documented limitations of relying on existing intercoder reliability measures, among which the issue of observation skewness has been frequently noted. Specifically, traditional intercoder reliability statistics perform poorly when used in highly skewed variables (Xu & Lorber, 2014), which explains why the intercoder reliability statistics of the *loyalty-betrayal* dimension and the *sanctity/degradation* dimension were particularly low in our study (i.e., Wang & Liu, 2021). Whereas Hopp and Weber (2021) suspected that it was due to the "deeper individual differences in their *loyalty-betrayal* intuitions" (p. 375), we believe it was attributed to the skewness of data, where most of the codings fell under one nominal category (i.e., not present). The skewness of data can punish intercoder reliability statistics, even when the percentage of coder agreement is high.

Most of the crowd-based coding relies on "majority vote" to resolve coding differences. The automated computational approach using dictionaries assigns each word a vector of possibilities so that the coding discrepancies are preserved rather than

overridden by majority vote (Hopp et al., 2021). Compared to using statistics to gauge coding reliability, the principle of majority vote can be problematic, as it still leaves crowd-coding subject to the influence of endogenous factors (Hoover et al., 2020). First, it assumes that regardless of coder qualifications, the majority always outperforms the minority. This argument actually counters the findings from the crowdsourcing literature that marginalized members of a crowd are more likely to contribute in a meaningful way (Jeppesen & Lakhani, 2010; Wang, 2020). As Lind et al. (2017) reflected on this issue, they similarly acknowledged that majority rule "assumes all crowd workers to be equally reliable, while it is evident that the quality and performance level differs within the group of workers" (p. 194). Furthermore, the majority vote rule assumes that there is a single "ground truth," which disregards any underlying reasons creating the discrepancies or the possibility that the coding protocol may be flawed so that it fails to capture the contextualized interpretation of the text. It also runs the risk of silencing marginal voices from non-Western cultures, resulting in a lack of pragmatic validity (Atari & Dehghani, 2021).

Obviously, neither manual nor crowd-based computational approaches offer a perfect solution to resolving coding discrepancies. The selection of methods ultimately depends on factors such as whether the existing coding protocols are adequately validated, the complexity of coding, estimated time and cost, and so forth. Considering the comparison discussed above, there is still significant merit in how manual content coding manages coding discrepancies. In particular, the manual coding approach allows one to use intercoder reliability statistics as a diagnostic tool to improve coding protocols, making coding less restricted to a set of otherwise flawed coding schemes.

## Comparing expert content coding results with eMFD scores

Applying eMFD to identify the moral loading of media content, we re-analyzed our data. As the moral values were used as independent variables in the negative binomial regression models, we followed Hopp et al.'s (2021) suggestion to use the single probability approach. The results (see Table 1) showed that among all five dimensions, *care* had the highest average probability score ($M = .08$, $SD = .06$), followed by *authority* ($M = .04$, $SD = .04$) and *fairness* ($M = .03$, $SD = .03$). This descriptive finding is generally aligned with Wang and Liu's (2021) findings that *care*, *fairness,* and *authority* were more prevalent in the dataset than the other two dimensions. However, given the small standard deviation scores identified, it also suggests that this approach does not capture the nuanced differences among the messages. Furthermore, the comparison of the results supported the argument from Atari and Dehghani (2021) that the dictionary approach is effective in measuring moral values over entire documents, but fails to measure values related to specific topics.

**Table 1.** The eMFD scores of the sampled 2013 Tweets.

| | Care probability | Fairness probability | Authority probability | Loyalty probability | Sanctity probability |
|---|---|---|---|---|---|
| Max | 0.45 | 0.27 | 0.29 | 0.27 | 0.26 |
| *S.D.* | 0.06 | 0.03 | 0.04 | 0.03 | 0.03 |
| *M* | 0.08 | 0.03 | 0.04 | 0.02 | 0.01 |

We included several tweets below from the data to show the discrepancy of the results across the two approaches. Using the eMFD approach, the following tweet returned with the highest probability for *care* (.45): "*Bloodless suicide?? Are u kidding me sir? #hongkongpolicebrutality #HongKongProtesters #HongKongPoliceTerrorists #HongKongPoliceTerrorism #FreeHongKong #StandWithHongKong #antichinazi https://t.co/AeLcA7WraX*." The *care* dimension of MFT captures intuitions of sympathy, compassion, and nurturance in response to caring for vulnerable populations, which this tweet did not contain. The content coding conducted in Wang and Liu (2021) identified this tweet as using the moral framing of *fairness*, due to its mention of "free Hong Kong" and "antichinazi." It is more about criticizing the unfair process of policing that should be about rights and justice. Furthermore, this tweet referenced joking/humor, which is a challenge for automated approaches that cannot identify the context. The expert coding thus offers a more refined interpretation of the moral framing in the context of criticizing police brutality in Hong Kong.

Another example to show the discrepancy is the following tweet: "*dogs are not weapons #hongkongpolicebrutality #StandWithHongKong https://t.co/DhLeAtwA9I*." We coded it as not having any moral framing, whereas the eMFD showed that its likelihood of loading to the *care* dimension was 0.22, which was much higher than the average probability (.07). Referring to the definition of the *care/harm* category, this tweet indicated the protesters' anger that the Hong Kong police used dogs as weapons; however, the text itself did not contain any notion of vulnerability or signs of suffering.

The eMFD approach identified the following tweet as having the highest probability of *fairness* (.27): "*If YOU give up Who protect our city #SOSHK #HongKongPolice #StandwithHK #hongkongpolicebrutality*." Wang and Liu (2021) coded it as not using any moral framing. The original conceptualization of *fair* emphasizes the violation of rights by a dominant party and the ability of the advocated cause to rectify the injustice. We do not think these components exist in the tweet, at least not in the content as currently written. We would have to infer that the police are arbiters of fairness in this context. An inference could just as likely be made for the foundation of authority (deferring to police as protective leaders). But, that context is not there.

The following tweet was identified as having the highest probability for authority (.29): "*Hong Kong protesters are cleaning mosque #StandwithHonKong #hongkongpolicebrutality #hongkongpolicebrutality https://t.co/zXHlq81uvc*." Wang and Liu (2021) coded it as not having a moral value. The dimension of *authority* captures the notion of traditional social hierarchies and maintaining social order, which makes individuals sensitive to signs of rank or status. A close examination of the tweet suggests that it does not contain any of these components, or it is simply too vague to be categorized.

The eMFD scores indicate probability. However, this approach does not offer a benchmarking method regarding the interpretation of results. For example, is there a threshold we can use to interpret that a particular message is morally relevant? The lower probability scores found in our dataset are consistent with the ones in Priniski et al. (2021), which also focused on the analysis of tweets. As Priniski et al. (2021, p. 3) summarized about using the eMFD approach, "While taking the mean activations for each foundation can be informative about the general trends in the data, Twitter data is rich and there is likely to be more complex structure to the data topology." Therefore, we ask: given the low loading of the scores, is it a sign that the eMFD does not

perform as well on shorter texts compared to longer ones, as the former does not have enough context sensitivity?

## Conclusions and future directions

Moral framing is conceptualized as a mobilizing strategy in digital activism (Wang & Liu, 2021). Thus, identifying the relevance of each moral value offers practical and theoretical implications regarding social media message design to build consensus and advocate for a collective agenda. With recent methodological advancements in extracting moral framing, in this response, we discussed their respective merits and criticisms. As Hoover et al. (2020) claimed, "human annotation remains the most accurate method for measuring moral sentiment in text" (p. 7), but computational approaches are often needed to analyze larger size samples to offset the cost related to human labor. Identifying moral framing is a delicate task. Consequently, we call for scholars to incorporate multiple methods to further uncover the meanings behind each moral value, accounting for reliability and validity concerns. For example, the most recent approach among non-dictionary-based methods is the Theorize-Annotate-Train-Scale-up (see Atari & Dehghani, 2021, for more detail). Another alternative is to combine automated coding with MFT dictionaries and human verification, which may offer an interpretation that fits better with a given study context and at a lower cost (Graham et al., 2009).

Automated coding using different dictionaries has been primarily used in longer texts, such as news articles and classic literature (see Priniski et al., 2021, for an exception). When applied to the analysis of short texts such as tweets, more discretion should be exercised as shown by the sample tweets presented in this article. Therefore, we also call for future research to compare the level of prediction accuracy when applying the dictionary-based approach to analyze texts of different lengths, under different contexts (e.g., social movement, breaking news, and natural disasters), and on different media platforms (e.g., Twitter, Facebook, and news websites).

Moral language is highly dependent on the context of the text (Garten et al., 2019). We acknowledge that there is no single "right" approach to conduct moral coding, yet we believe human coding helps to interpret the meaning of what is being said, how it is being said, and in what context. To conclude, we join the recent call from Atari and Dehghani (2021) that scholars need to diversify their methodological toolbox to move beyond a single model of measurement. We believe that human judgments of moral framing are still needed as they offer a valid method that takes linguistic and cultural context into account.

## ORCID

*Rong Wang* http://orcid.org/0000-0002-2307-709X

## References

Alper, S., Bayrak, F., Us, E. Ö., & Yilmaz, O. (2020). Do changes in threat salience predict the moral content of sermons? The case of Friday Khutbas in Turkey. *European Journal of Social Psychology*, *50*(3), 662–672. https://doi.org/10.1002/ejsp.2632

Atari, M., & Dehghani, M. (2021). Language analysis in moral psychology. In M. Dehghani & R. Boyd (Eds.), *The Atlas of language analysis in psychology* (Chap. 11). Guilford Press.

Garten, J., Kennedy, B., Sagae, K., & Dehghani, M. (2019). Measuring the importance of context when modeling language comprehension. *Behavior Research Methods*, *51*(2), 480–492. https://doi.org/10.3758/s13428-019-01200-w

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

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., Jeung, 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

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, *53*(1), 232–246. https://doi.org/10.3758/s13428-020-01433-0

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

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, *21*(5), 1016–1033. https://doi.org/10.1287/orsc.1090.0491

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

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage.

Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, *57*(1), 34–52. https://doi.org/10.1080/08838151.2012.761702

Lind, F., Gruber, M., & Boomgaarden, H. G. (2017). Content analysis by the crowd: Assessing the usability of crowdsourcing for coding latent constructs. *Communication Methods and Measures*, *11*(3), 191–209. https://doi.org/10.1080/19312458.2017.1317338

Matsuo, A., Sasahara, K., Taguchi, Y., & Karasawa, M. (2019). Development and validation of the Japanese moral foundations dictionary. *PloS One*, *14*(3), e0213343. https://doi.org/10.1371/journal.pone.0213343

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396. https://doi.org/10.1038/s41562-018-0353-0

Noble, S. U. (2018). *Algorithms of oppression*. New York University Press.

Priniski, J. H., Mokhberian, N., Harandizadeh, B., Morstatter, F., Lerman, K., Lu, H., & Brantingham, P. J. (2021). Mapping moral valence of tweets following the killing of George Floyd. *arXiv preprint arXiv:2104.09578*.

Simon, A. F. (2001). A unified method for analyzing media framing. In R. P. Hart & D. R. Shaw (Eds.), *Communication in U.S. election: New agenda* (pp. 75–89). Rowman and Littlefield.

van Atteveldt, W., van der Velden, M. A., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121–140. https://doi.org/10.1080/19312458.2020.1869198

Wang, R. (2020). Marginality and team building in collaborative crowdsourcing. *Online Information Review*, *44*(4), 827–846. https://doi.org/10.1108/OIR-09-2018-0269

Wang, R., & Liu, W. (2021). Moral framing and information virality in social movements: A case study of #HongKongPoliceBrutality. *Communication Monographs*, *88*(3), 350–370. https://doi.org/10.1080/03637751.2021.1918735

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

Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology*, *82*(6), 1219–1227. https://doi.org/10.1037/a0037489