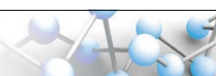ORIGINAL ARTICLE

bioethics    WILEY

# Research integrity codes of conduct in Europe: Understanding the divergences

Hugh Desmond[1,2] | Kris Dierickx[1]

[1]Center for Biomedical Ethics and Law, KU Leuven, Leuven, Flanders, Belgium

[2]Department of Philosophy, University of Antwerp, Antwerpen, Belgium

**Correspondence**
Prof. Dr. Hugh Desmond, Department of Philosophy, University of Antwerp, Rodestraat 2, Antwerpen 2000, Belgium.
Email: hugh.desmond@uantwerpen.be

## ABSTRACT

In the past decade, policy-makers in science have been concerned with harmonizing research integrity standards across Europe. These standards are encapsulated in the European Code of Conduct for Research Integrity. Yet, almost every European country today has its own national-level code of conduct for research integrity. In this study we document in detail how national-level codes diverge on almost all aspects concerning research integrity—except for what constitutes egregious misconduct. Besides allowing for potentially unfair responses to joint misconduct by international collaborations, we argue that the divergences raise questions about the envisaged self-regulatory function of the codes of conduct.

**KEYWORDS**
European code of conduct, professionalism, research integrity, research methodology, science policy, self-regulation

## 1 | BACKGROUND

In one of the first surveys of research integrity (RI) standards, the European Science Foundation concluded that there was a 'wide range of approaches' across European countries and that there was a need for 'harmonized standards across Europe'.[1] This was the original rationale for the European Code of Conduct for Research Integrity (ECoC), which was first published in 2011 and updated in 2017.

Harmonization does not mean uniformity. The harmonization the ECoC aimed to provide was that of a reference point for a 'common understanding of the demands of research integrity'.[2] Such a common understanding was intended to be compatible with national-level differences: 'However, unlike the fundamental values of

scientific integrity and the violation thereof, which have a universal character, [poor and inappropriate] practices[3] may be subject to different national traditions, legislative regulations or institutional provisions'.[4]

Similarly, the 2017 ECoC explicitly aims to be a reference point that 'allows for local or national differences in its implementation'.[5]

We call this the (European) *core versus national periphery* model of harmonization: the 'core' aspects of RI—the principles, the good practices, and the definition of misconduct—must be specified by Europe-wide standards set by the ECoC, but the 'peripheral' aspects—i.e. what counts as merely a questionable research practice may vary from country to country.

In this article we will seek to map patterns of divergence and convergence across national-level codes and guidelines in Europe,

[1]ESF (European Science Foundation). (2008). Stewards of integrity: Institutional approaches to promote and safeguard good research practice in Europe. European Science Foundation. Retrieved from http://www.enrio.eu/wp-content/uploads/2017/03/StewardOfIntegrity.pdf, p. 50.

[2]ESF-ALLEA (European Science Foundation and All European Academies) (2010). A European code of conduct for research integrity: Background paper. Retrieved from https://allea.org/wp-content/uploads/2015/09/A-European-Code-of-Conduct-for-Research-Integrity_final.10.10.pdf

[3]For instance: 'questionable procedures for obtaining informed consent, insufficient respect and care for participants in the research, improper research design and carelessness in observation and analysis, unsuitable authorship or publishing practices, and reviewing and editorial derelictions'. Ibid: 14.

[4]Ibid: 14.

[5]ALLEA (All European Academies) (2017). The European code of conduct for research integrity. Retrieved from http://www.allea.org/wp-content/uploads/2017/04/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017.pdf

and seek to inquire whether these patterns are compatible with the core-periphery model of harmonization. Since the publication of the background document to the first ECoC, over a decade has passed, and as we will document in this paper, there has been a proliferation of national-level research integrity codes of conduct since then. It is timely to revisit the original intention of harmonization (as formulated by the core-periphery model), and inquire to what extent this is currently realized.

Note that the general impact of the ECoC is not in question. It is the standard for all projects funded by the European Commission, including those by the European Research Council and those within the Horizon 2020 and successor Framework Programmes. While more difficult to quantify, it has also undoubtedly stimulated national policy-making efforts. Nonetheless, it is one thing to state that the values of RI should be held in common, and another for there to actually be widespread agreement, let alone unanimity, on the core aspects of RI.

Why does harmonization matter? We will normatively evaluate observed divergences in light of the following two questions:

1. Do the divergences allow for unfair responses to joint misconduct by international collaborations?
2. Do the divergences undermine the credibility of RI codes of conduct as self-regulatory tools?

The first question regarding fairness is the most straightforward one: if different national-level codes of conduct define misconduct differently, then—in principle—not all researchers in an international collaboration may be held accountable (in the same way, or at all) for joint misconduct. Even slight divergences in the formulation of what constitutes research misconduct may have outsized impact: when allegations of misconduct are tried in courts of law, judges may use codes of conduct as 'soft' legal documents to guide their judgment.[6] Even when university bodies investigate allegations of misconduct, accused researchers may hire the services of a lawyer or team of lawyers, and then the precise wording of codes of conduct become crucial (personal experience of author 2 [KD] as member of research integrity commissions). This strong rationale to align definitions of misconduct has been noted by others,[7] and one of the research aims of this paper is to help map divergences that may be problematic for reasons of fairness.

The second question is more subtle and concerns what we call the *credibility* of a guideline or code of conduct. We flesh this out in more detail in the Discussion section, but it basically refers to how some voices are sceptical of the efficacy of self-regulation, and believe codes of conduct to primarily be window-dressing, presenting a morally agreeable façade to the rest of society while hiding a harsher reality of competition and prestige-maximization. The belief is that the primary purpose is to create a *perception* that integrous behaviour is valued, and not to promote actual integrous behaviour. Thus, sceptics would not be surprised by, for instance, a large divergence in the fundamental values listed in RI codes: after all, for window-dressing it is less important which values are listed, and more important the fact that *some* values are present. In this way we will evaluate existing divergences according to whether they are consistent with the sceptical view that codes of conduct are not credible as effective self-regulatory tools.

In sum, in this paper we investigate the state of the harmonization of national-level codes and guidelines, where 'harmonization' is understood in terms of the core-periphery model. While harmonization does not mean complete uniformity across national-level regulatory documents, for reasons of methodological soundness, we will nonetheless investigate patterns of replication of the ECoC across national-level documents. The detailed patterns of divergence and convergence will allow for conclusions to be drawn about harmonization in the European context, as well as about potential problems arising from the lack of harmonization.

## 2 | METHODS

### 2.1 | Methodological desiderata

This is not the first study aimed at detailing differences in national-level codes and guidelines on RI.[8] There is an obvious, content-related reason for undertaking a new study, and that is that the regulatory situation for RI in Europe has continued to change at a fast pace. A second European code of conduct was published in 2017, while countries such as France, Estonia, The Netherlands, Italy, and the UK have published new codes or guidelines in the past few years (see Supporting Information).

However, there are also methodological reasons for undertaking a new study, and an examination of previous studies helps identify two methodological desiderata. The first is that previous studies depended on interpretative decisions about the meanings of values and virtues. For instance, Godecharle et al. (2013) distinguish between 'honesty' and 'openness or open communication', even though—at least, by dictionary definitions of these words—they are near-synonyms. An example in Aubert Bonn et al. (2017) is the way in which similar virtues are lumped together: thus 'Openness; Verifiability' form one category, and 'Objectivity; Scrupulousness; Transparency' another. However, with equal justification, 'transparency' could have been categorized together with 'openness' and 'verifiability'.

[6]Kovács, A., Tóth, T., & Forgács, A. (2016). The legal effects of European soft law and their recognition at national administrative courts. *ELTE Law Journal, 2*, 53–70.

[7]e.g. Fanelli, D. (2011). The black, the white, and the grey areas: Towards an international and interdisciplinary definition of scientific misconduct. In T. Mayer & N. H. Steneck (Eds.), *Promoting research integrity in a global environment* (pp. 79–90). World Scientific Publishing.

[8]Aubert Bonn, N., Godecharle, S., & Dierickx, K. (2017). European universities' guidance on research integrity and misconduct: Accessibility, approaches, and content. *Journal of Empirical Research on Human Research Ethics, 12*(1), 33–44. https://doi.org/10.1177/1556264616688980; Godecharle, S., Nemery, B., & Dierickx, K. (2013). Guidance on research integrity: No union in Europe. *The Lancet, 381*(9872), 1097–1098. https://doi.org/10.1016/S0140-6736(13)60759-X

Such interpretations are not necessarily problematic if reasons are given, but this was not the case. Without such explication, it raises the question if another researcher, looking at the same basic data, might come to different conclusions about the patterns of divergence. For instance, if a large divergence between national-level codes is noted, to what extent does this reflect a genuine lack of harmonization, or simply a divergence resulting from interpreting 'honesty' and 'openness' as different values? Given this threat to reproducibility, we can stipulate as a methodological desideratum that, when aiming to understand the differences between different national-level approaches, one should avoid, as much as possible, controversial or unsupported interpretative decisions on the basic data—i.e. the words used by guidelines and codes of conduct.

A second, related methodological desideratum is that previous studies either did not, or only insufficiently, discuss the importance of the observed differences. For instance, if code A lists 'transparency' as a virtue, and code B lists 'openness', then most would find this to be harmless variation on basically the same virtue. Why should differences actually matter? In this study we anticipate this issue by focusing the discussion on the core-periphery model of harmonization.

## 2.2 | Search methodology

The purpose of the search methodology was to represent national-level approaches to research integrity (RI) by a single document that can be considered as the leading document in the national context. In this way the comparative analysis of the ECoC and the national-level documents was considerably simplified; we will argue later that this simplification does not call the conclusions into question.

### 2.2.1 | Initial collection

For the initial collection we cast the net widely and included all national-level 'regulatory documents' directly pertaining to RI. A 'regulatory document', as we understand it, can in principle refer to a document regulating both the actions of individuals (e.g. an ethics code) or the actions of institutions. Thus, a regulatory document can refer to any of the following: codes of conduct, guidelines, policy documents, laws (statutes, charters), and even more descriptive documents such as survey reports, meeting reports, and position papers. We included a document if it contained substantial *normative* position statements on any one of the following: (a) the principles underlying research integrity (e.g. honesty), (b) behaviours constituting good research practice, (c) behaviours constituting research misconduct, (d) a plea for the importance of research integrity for science and society.

For the initial collection of regulatory documents, we used the following six independent search methods for each member of the 32 EFTA countries (EU28 + Norway, Switzerland, Liechtenstein,

Iceland). This means that if, for instance, five documents were found with method 1, these were double checked with methods 2 through 6. We used these six independent methods to minimize the probability of missing an important regulatory document.

1. Exhaustive search of following websites, if available:
   a. national research council;
   b. national agency on research integrity;
   c. national scientific fund;
   d. national academy of science.
2. Search of the websites of academy members of ALLEA.
3. Search for "integrity" and "<integrity translated into local language>" of websites of prominent universities in that country—whenever possible, websites in original language (through Google translate).
4. A search by means of an internet search engine (Google) with search terms (("research integrity" OR "scientific integrity" OR "science integrity") AND "<name of country>").
5. Search of the resources for that country listed on the website of the European Network of Research Integrity Officers (http://enrio.eu).
6. Resources listed on European Science Foundation 'Stewards of Integrity' document (URL: http://digital.csic.es/bitstream/10261/8663/1/StewardsOfIntegrity.pdf).

This initial collection was completed in March 2019 and updated through August 2020.

### 2.2.2 | Selection

The selection consisted of three steps. In the first step, we selected candidates for the national 'leading' regulatory documents for inclusion in the comparative study. We considered a document to be a 'candidate leading document' if it contained (a) authoritative formulations of principles of RI, AND (b) definitions of good practices, AND (c) definitions of misconduct. In many national-level contexts, this procedure was sufficient to identify the national leading document. However, some countries proved to have multiple potentially leading documents. For instance, in Denmark and Norway, both a law and a national code of conduct can be considered leading. In the UK and France, there are multiple national codes of conduct (authored by Research Councils UK or UK Research Integrity Office; Centre Nationale de la Recherche Scientifique or Agence Nationale de la Recherche), as well as national documents aimed primarily at institutions (Concordat in the UK, Charter in France: see Supporting Information).

When this was the case, in the second step we chose a single candidate leading document based on following additional criteria:

- We prioritized documents aimed at guiding individual researchers over those designed for guiding institutions (when the latter seek to author an institution-specific code or guideline). The reason for
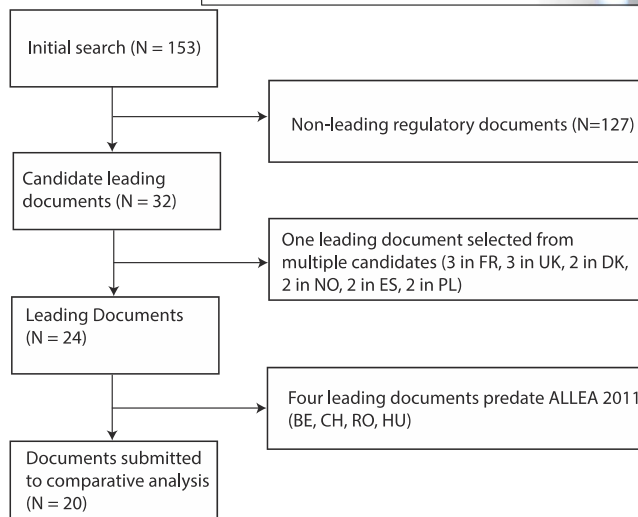
**FIGURE 1** Flowchart for the initial search and selection of national regulatory documents for inclusion in the comparative analysis

this was to facilitate a comparison between each country's leading document and the ALLEA code, which is primarily focused on individual researchers.

- If the previous criterion yielded a tie, we prioritized documents with detailed statements on the constitutive elements of RI over those with few detailed statements on the elements of RI.
- If the tie was still not broken, we then prioritized the document with the most institutional signatories.

This decision-making process allowed us to select a single candidate leading document for each country.

In a final step, we presented this choice, together with our justification, to local RI experts so they could verify our choice (or offer corrections if necessary). A person was deemed a local expert if he or she (a) had been appointed as a contact person by a national research council, a national agency for research integrity, a national scientific fund, or a national academy; OR (b) was a member of ENRIO (European Network of Research Integrity Officers); OR (c) was part of an ERC or Horizon2020-funded project on research integrity, OR (d) was a member of EARMA (European Association of Research Managers and Administrators). The initial search and selection of documents are summarized in Figure 1.

## 2.3 | Methodology of content analysis

In light of the methodological problems faced by previous studies (see Section 2.1), in this study we opted for a very minimalist method of content analysis.[9] This consisted of simply verifying whether national-level documents *replicated* the ECoC in the following three areas: (a)

principles of RI, (b) definition of good research practices, and (c) definition of research misconduct. 'Replication' was defined as the literal copying of the principles listed by the ECoC, the categories of good research practices, and the categories of misconduct. Thus, for instance, the 2017 version of the ECoC lists four values of RI: reliability, honesty, respect, accountability. By contrast, the leading regulatory document in Estonia lists six principles (or categories of principles): freedom, responsibility, honesty and objectivity, respect and caring, justice, and openness and cooperation. The leading regulatory document in The Netherlands contains five values: honesty, scrupulousness, transparency, independence, responsibility. We categorized both Dutch and Estonian documents as 'non-replications', thus abstracting away from the obvious overlap between all three lists.

In focusing on the actual words used rather than their meaning we consciously avoided attempting to quantify any conceptual overlap between documents, since this would have led to verbal disputes of dubious importance (see Section 2.1). For instance, we were able to avoid questions such as: Is 'scrupulousness' the same as 'reliability'? Is 'responsibility' the same as 'accountability'? 'Transparency' and honesty'?

The strength of this minimalist methodology is that the obtained results are very robust; the weakness is that it places a greater onus on normatively evaluating the results. (By contrast, when normative presuppositions are built into describing codes of conduct, then interpreting the resulting differences is more straightforward.) The measure of literal replication will categorize both semantically similar terms and dissimilar terms as 'non-replications', and this means that, especially with regard to RI principles, widespread non-replication could be compatible with widespread semantic agreement.[10] Nonetheless, as discussed in Section 4.2, non-replications raise important questions about the purposes of RI codes. Hence the importance of the Discussion section, where we will argue that the observed divergences are important, and potentially threaten the core-periphery model. Nonetheless, the results are such that future research may in principle, using a different normative framework, evaluate the results differently.

## 3 | RESULTS

### 3.1 | Evolution of national regulatory documents

Twenty-four countries (out of a total of 32) were found to have a leading regulatory document on RI. Figure 2 shows how many countries that lacked a leading national-level regulatory document in 2012 possessed one in 2020. Moreover, among those without such a document today, two (Bulgaria, Luxemburg) explicitly adopt the European Code of Conduct. A further two (Greece and Slovenia) have stated the intention to develop a national-level framework. For four countries—Malta, Liechtenstein, Cyprus, and Iceland—no statement concerning national-level framework could be found. However,

---

[9]Rössler, P. (2013). Comparative content analysis. In F. Esser & T. Hanitzsch (Eds.), *The handbook of comparative communication research* (pp. 459–468). New York: Routledge. https://doi.org/10.4324/9780203149102-39

[10]Issues of translation can also lead to non-replication despite relative semantic similarity.

FIGURE 2  Countries with leading regulatory documents concerning research integrity (RI) in 2012 (left) and 2020 (right). The left map is based on the data in[43]



**2012**

**2020**

| | = | Leading national-level code not found | | = | Leading national-level code present |
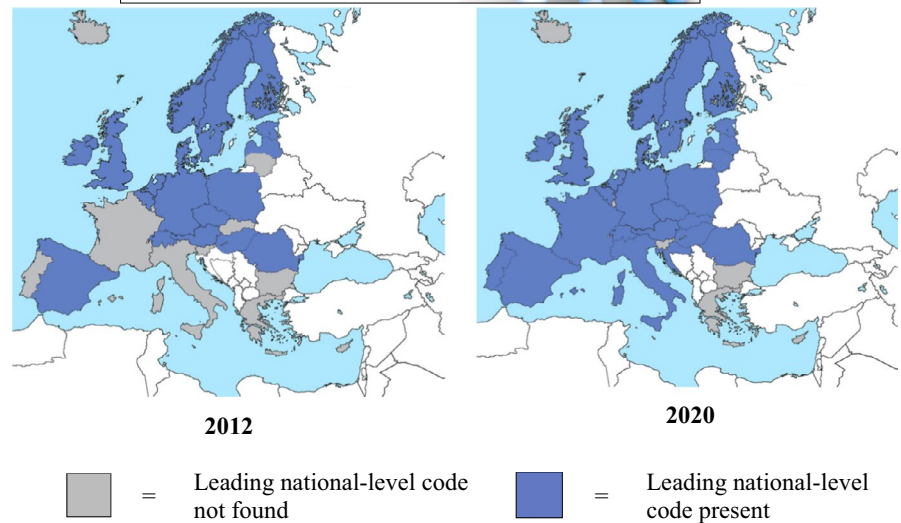
institutional-level RI-regulatory documents exist in Malta, Iceland, and Cyprus. No institutional code was found for the three Liechtenstein higher education institutions,[11] but given the fact that the University of Liechtenstein, the largest institution in Liechtenstein, is small by international standards (1,200 students), it is perhaps not surprising that an explicit code has not (yet) been deemed necessary.

In any case, we concluded from the search that, with the exception of Liechtenstein, all countries in Europe have a de facto leading regulatory document, whether that is the ECoC itself, a specific national code or guideline, or, for small countries, a code or guideline in a large university.

## 3.2 | Differentiation of functions

Our investigation of the regulatory documents corroborates ESF's prior finding that there is a broad variance in types of regulatory document.[12] In other words, there is a wide variety between different groups of envisaged users (all researchers, PhD students, supervisors, policy-makers) as well as different envisaged uses (pragmatic rules for quick consultation vs. in-depth discussion of scientific methodology and values in science). Since this will be one of the obvious explanations to be discussed (in Section 4), we also aimed to map some of this variance.

A difficulty here is that the codes of conduct themselves do not always explicitly state what their envisaged uses are, nor what their target users are. Yet it is clear that the envisaged uses are different. The longest document (that of Sweden, *c.* 44,000 words) is 50 times longer than the shortest (that of Norway, fewer than 900 words): whereas the latter contains only concise, actionable rules and advice, the former contains in-depth discussions of, for instance,

various possible approaches to defining research misconduct. Contrasting with both of these is the Danish Code of Conduct for Research Integrity (approximately 5,000 words), which is structurally very similar to the new Danish Research Misconduct etc. Act (of April 26, 2017).

With this example in mind, one can look at 'orders of magnitude' differences in word count as a rudimentary indication that different envisaged uses are at work (Figure 3). The preceding discussion suggests a rough tri-partite categorization of envisaged uses.

- Portable vademecum, containing actionable advice. Paradigmatic example: Norwegian leading document.
- A quasi-legal document, containing detailed descriptions of good practices and categories of misconduct. Paradigmatic example: Danish leading document.
- An academic exploration, containing substantial philosophical, sociological, and legal background to research integrity. Paradigmatic example: Swedish leading document.

Note that the boundaries between these categories are vague, and that a more detailed categorization of envisaged uses may be possible. However, even this rudimentary categorization will help structure the normative evaluation of the observed differences between leading regulatory documents.

## 3.3 | Patterns of replication

Four of the 24 leading national-level regulatory documents pre-date the first ECoC (2011), and thus could not reflect the impact the core-periphery model has had on national-level documents. (This is especially relevant with regards to the problem of credibility, in Section 4.2.) The comparative analysis focused on the 20 remaining documents, of which 10 were published in the period 2011–2016, and 10 in the period after 2017. We found that the formulations of the ECoC—whether ECoC 2011 or ECoC 2017, and whether concerning the values of RI,

---

[11]See https://www.liechtenstein.li/en/education/higher-education/

[12]ESF, op. cit. note 1.

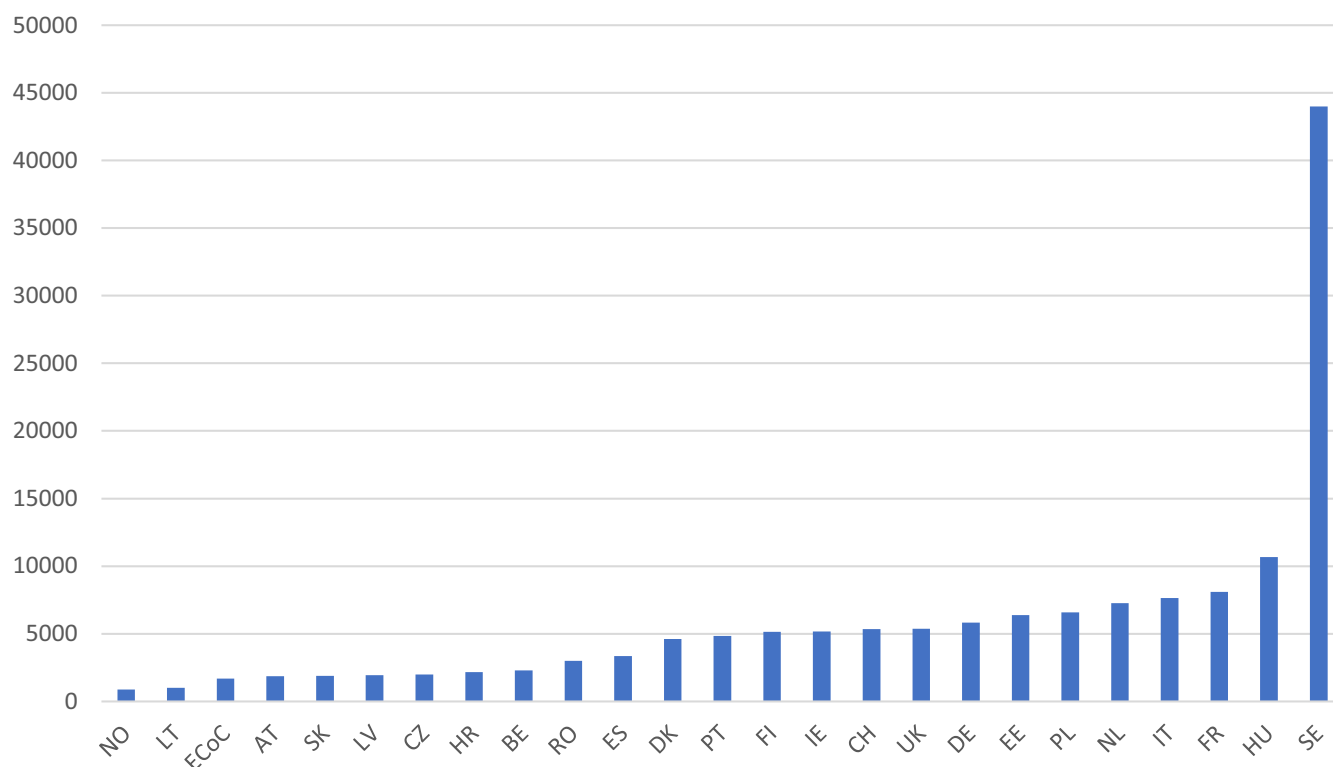[43]Godecharle et al., op. cit. note 8, pp. 1097–1098.

Word Count



**FIGURE 3** The longest regulatory document analysed in this paper was over 50 times longer than the shortest. Even if one cuts out the two largest and two smallest outliers, the variation in length is still one order of magnitude. This in itself is a strong indication that different documents are intended for different uses. (Preambles but not tables of contents, annexes or appendices, or references are included in the word count.)

definitions of misconduct, or definitions of good practices—are almost never replicated by national-level documents. Of the 60 points of comparison with the ECoC (three per country), the formulations of an ECoC were replicated only twice: the Irish code replicates the values listed in the 2011 ECoC, and the Portuguese code replicates the values listed in the 2017 ECoC. This bird's-eye view is summarized in Figure 4, which reveals only the Portuguese and Irish leading documents to replicate core aspects of RI in the relevant ECoC.

However, partial overlaps also matter. Note that 'non-replication' is a broad category that contains both partial overlap and a complete lack of overlap. Supporting Information Tables S1 and S2 give, as much as possible, the literal wording in the leading regulatory documents. For instance, three of the six principles listed by the Austrian leading regulatory document are adopted from the ECoC principles (reliability, impartiality, fairness). By contrast, none of the principles listed by the Estonian leading document overlaps with the ECoC principles (exercising methodological doubt; designing good experiments; managing data; proper use of funding).[13]

---

[13]Note that 'designing good experiments' could be taken as a principle that is very close to reliability. However, it is not the same, because reliability might be applicable to research in theoretical sciences or in humanities, where no experiments are involved. Moreover, the 'goodness' of an experiment may cover many dimensions, of which reliability is only one. As a reminder, for such reasons we have consciously avoided making judgments of this sort (i.e. whether 'reliability' and 'designing good experiments' refer to the same principle). See Section 2.1.

We separately mapped the partial overlaps, and how these have changed over time. Figures 5 and 6 show how probable a principle or category of misconduct listed in the ECoC would be duplicated by a national-level document. Figure 5 shows that 'honesty' is the only principle to be frequently listed by national-level documents, especially after 2017. Figure 6 shows that fabrication, falsification, and plagiarism (FFP) is the only category of misconduct that is universally duplicated across national-level documents (and indeed the consensus on FFP as a form of misconduct is truly global, with the Federal Policy on Research Misconduct and the Singapore Statement also referring to FFP). The first version of the ECoC was vague about other forms of misconduct, and its two other categories of 'minor misdemeanours' and 'failure to meet ethical and legal requirements' are never adopted by national-level documents. The second version of the ECoC was much more detailed in this regard, but the categories it introduced were not widely adopted. Of the 13 other categories of misconduct, only 'conflict of interest', 'manipulating authorship' and 'misrepresentation' were cited by more than one national-level document.

We also mapped how the 'grain' of the taxonomy of core RI elements varied across national-level documents. By 'grain' we mean the extent to which detailed distinctions were made. Thus the 2011 ECoC is relatively 'fine-grained' with regards to RI principles (eight principles listed), but 'coarse-grained' with regards to categories of misconduct (three categories). By contrast, the
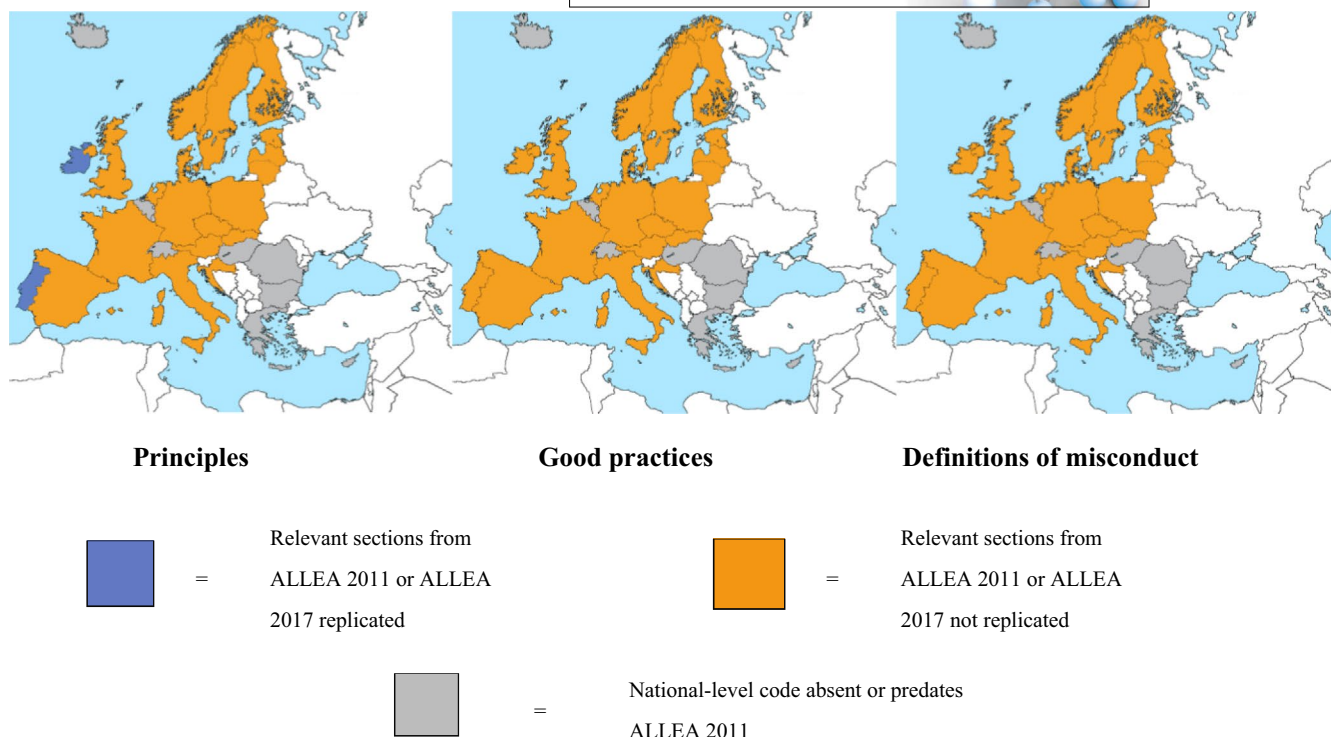
**Principles**  **Good practices**  **Definitions of misconduct**

= Relevant sections from ALLEA 2011 or ALLEA 2017 replicated

= Relevant sections from ALLEA 2011 or ALLEA 2017 not replicated

= National-level code absent or predates ALLEA 2011

**FIGURE 4** National-level regulatory documents almost never replicate the European Code of Conduct for Research Integrity (ECoC) with regards to either principles of research integrity (RI; left), definitions of good practices (centre), or definitions of misconduct (right). For details, see Supporting Information
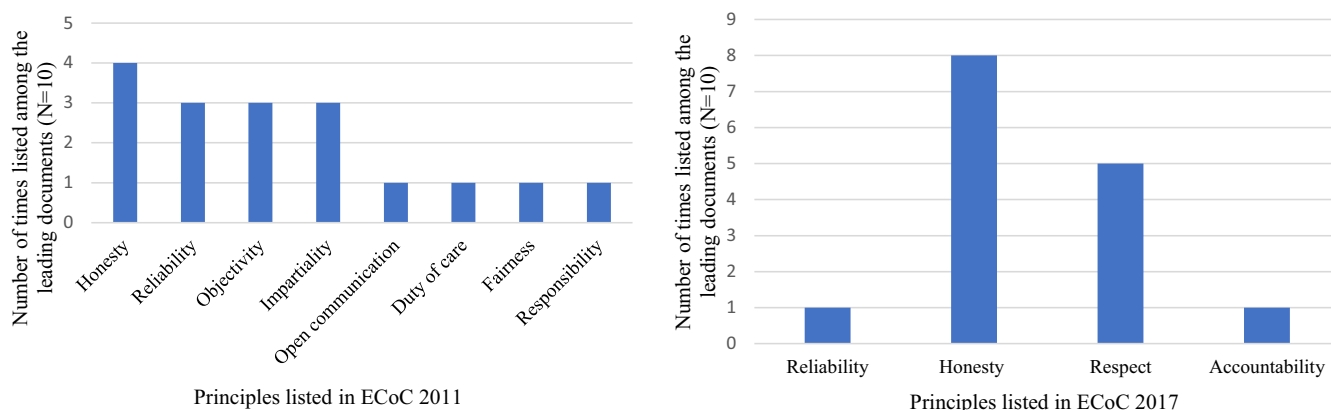


**FIGURE 5** Number of times a principle listed in the European Code of Conduct for Research Integrity (ECoC) is also listed in a leading national regulatory document

2017 ECoC is coarse-grained with regards to RI principles (four principles listed), but fine-grained with regards to categories of misconduct (14 categories).

Figure 7 gives an overview of the results of this exercise, and Figure 8 shows the same results as a more detailed distribution of the number of listed core RI elements. The first takeaway is that the number of elements listed per dimension of RI (principles/good practices/misconduct) varies greatly. For instance, the fine-grained approach to RI principles adopted by the 2011 ECoC is not followed by a majority of national-level documents: most list three or four principles vs. the ECoC's eight principles.

It is interesting to see the trends in these data, and global changes between the two periods of 2011–2016 and 2017–present. We caution against reading too much into these data: the averages and standard deviations are not statistically significant given the low sample sizes (N = 11 including the ECoC). In other words, if one or two national-level documents would have listed very different numbers of RI principles or types of misconduct, then a very different picture would emerge.

Nonetheless, with this caveat in mind, in Table 1 one can see how the granularity of the all leading documents (plus ECoC) actually evolved over time.
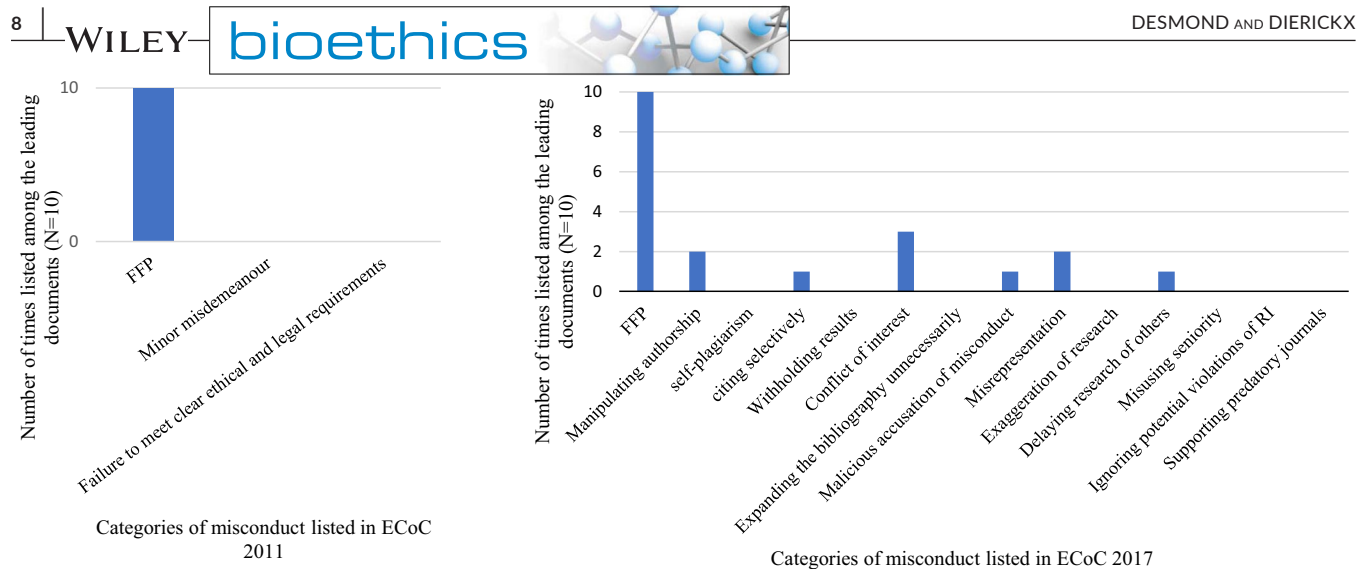
**FIGURE 6** Number of times a category of misconduct listed in the European Code of Conduct for Research Integrity (ECoC) is also listed in a leading national regulatory document
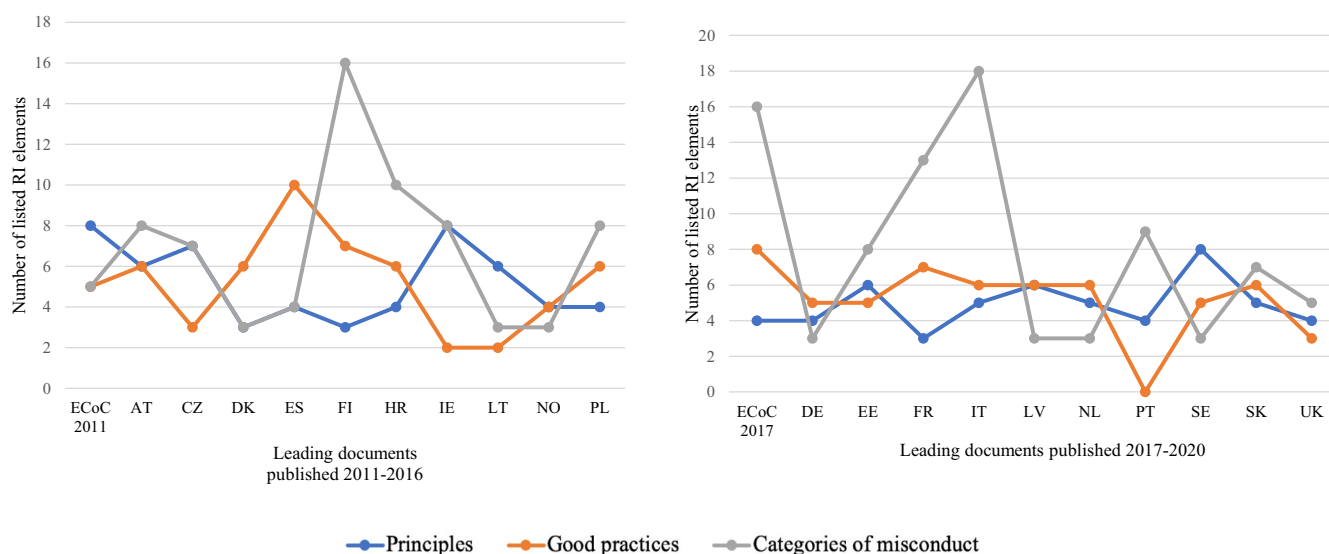


**FIGURE 7** The number of listed research integrity (RI) elements fluctuates across leading national regulatory documents

With respect to RI principles, there was a weak convergence on listing four to five principles. With respect to good practices, there was a stronger convergence on listing five to six categories of good practice. Finally, and most interestingly, leading regulatory documents are now, on average, more detailed as to the types of behaviour that are in violation of the code (under four to over five types). Moreover, there is an increased disparity in the approaches of different documents. Some leading documents opted, with the 2017 ECoC, to list a large number (10+) of categories of misconduct, while others opted for a less detailed approach, and almost half (four out of 10) of the leading documents for this period only list three types of misconduct. The result is a flattened distribution with a large right skew (bottom graph in Figure 8).

To sum up the results: not a single national-level leading document entirely adopts the ECoC's formulation of the core elements of RI. There is a large divergence in the principles and categories of

misconduct listed: the only unambiguous consensus is that FFP count as misconduct. However, this consensus is global,[14] and is not a particularly surprising finding. With regards to principles of RI, the only candidate for a quasi-consensus would be honesty. However, also here the consensus seems to be broader than the European context alone, given how 'honesty' is the very first principle listed by the Singapore Statement on Research Integrity.[15] We will now discuss consequences of these findings for the envisaged European model of harmonization.

---

[14]WCRI (World Conference on Research Integrity). (2010). Singapore statement on research integrity. Retrieved from https://wcrif.org/documents/327-singapore-state ment-a4size/file; OSTP (Office of Science and Technology Policy, Executive Office of the President). (2000). Federal policy on research misconduct. *Federal Register.* Retrieved from https://www.govinfo.gov/content/pkg/FR-2000-12-06/pdf/00-30852.pdf
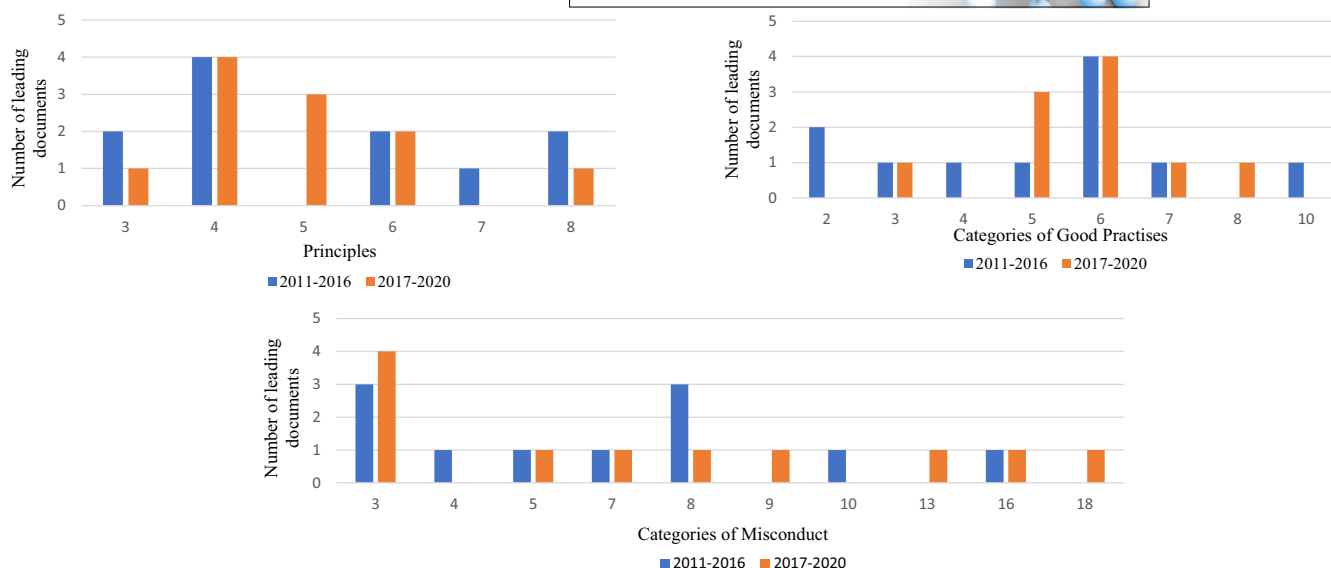
[15]WCRI, op. cit. note 14.

**FIGURE 8** The distributions of the number of core research integrity (RI) elements across leading national regulatory documents

**TABLE 1** Time evolution of the granularity of European Code of Conduct for Research Integrity (ECoC) and leading national-level regulatory documents

| | 2011–2016 | 2017–2020 |
|---|---|---|
| Number of RI principles | Average: 4.9 Standard deviation: 1.9 | Average 4.9 Standard deviation: 1.4 |
| Number of good practices | Average 5.2 Standard deviation 2.4 | Average: 5.7 Standard deviation 1.3 |
| Number of categories of misconduct | Average 6.8 Standard deviation: 3.9 | Average: 8 Standard deviation: 5.5 |

## 4 | DISCUSSION

While the large divergences documented in the previous section do not necessarily falsify the core-periphery model of harmonization, they do call it into question. If the divergences are compatible with a 'common understanding' of the core aspects of research integrity,[16] then it is not obvious at all what this common understanding amounts to precisely—beyond categorizing FFP as a violation.

Could one not sidestep this question about 'common understanding' by construing it simply as the *recognition* that research integrity matters rather than actually agreeing about what research integrity entails? Thus, the documented variances could be cast as harmless differences of opinion: what matters is that the importance of integrity is recognized, and all leading national-level documents do this. However, if only the recognition matters, why then the construction of a relatively detailed code of conduct would be unnecessary. Instead, research bodies could put out a joint statement affirming the importance of RI. Since the trouble has been taken to construct substantive codes of conduct, we cannot simply construe the 'common understanding' of RI as a recognition of the importance

of RI. There must be a shared understanding about substantive content as well.

How can we evaluate the divergences mapped in the previous section, and how can we distinguish between divergences that are harmless, and those that are genuinely problematic and put pressure on the core-periphery model? Here we will evaluate the divergences according to two dimensions: threats to fairness, and threats to credibility. If the harmonization between national-level leading regulatory documents is to be genuine, then such documents should give no grounds for unfair treatment of joint misconduct by international collaborations, and they should make clear that they are truly aimed at guiding researcher behaviour (as opposed to being merely window-dressing). We will argue that, with regards to fairness, harmonization is clearly lacking, and that with regards to credibility, harmonization is genuinely ambiguous.

### 4.1 | Fairness

It would be unfair for a researcher to be judged more harshly than his or her collaborator for joint misconduct, simply because they are affiliated with an institution in country X rather than country Y. This fairness issue was anticipated by the ECoC 2011, which states that potential misconduct should be investigated in the

---

[16]ESF-ALLEA, op. cit. note 2, p. 4.

country/institute of the project leader.[17] However, this passage has been deleted in the 2017 version of ECoC, most likely because of the serious legal obstacles faced by an institution or committee in country X investigating a researcher at a different institution in country Y.

Do the current divergences between national-level documents avoid giving grounds for this type of unfairness?[18] While all examined documents categorize FFP as violations, there is a large divergence in what additional categories of behaviour constitute misconduct or 'questionable practice'. The 2017 ECoC lists a total 15 categories of misconduct (including FFP), whereas, for instance, the German leading regulatory document only explicitly lists FFP as an example of research misconduct. Even among the documents adopting a coarse-grained approach to defining misconduct, large differences can be seen, even in the strategy of definition. For instance, the Dutch leading regulatory document also only explicitly prohibits FFP, but instead of listing types of prohibited behaviour, it seeks to define expected standards of behaviours that can be used to judge if a particular behaviour is problematic or not. If one would include failures of cited standards as different types of conduct, the Dutch leading regulatory document in fact takes the most fine-grained approach to defining misconduct, with 22 categories in total.

Thus, in situations where the precise wording of the definition of misconduct would matter—i.e. when allegations of misconduct end up in courts of law, or when accused researchers hire lawyers for representation in internal disciplinary investigations—it seems that different codes of conduct could lead to different judgments. The harmonization of national-level leading regulatory documents is *not* such that no grounds are given for this type of potential unfairness.

We surmise that the trend towards more detailed definitions of categories of misconduct (see Table 1 and Figures 6–8) is an implicit recognition of the importance of anticipating legal challenges to internal disciplinary investigations. Nonetheless, more work remains to be done not just to add increasing detail in the definition of misconduct, but also to reflect on what is the best way of defining misconduct, and what jurisprudential approach is most appropriate for the goals of RI (see Section 4.3).

Besides the definition of the categories of misconduct, a second, and perhaps more fundamental source of potential unfairness lies in divergences between how FFP-behaviour is evaluated. While we did not systematically map this for the current study, reading and analysing the documents was sufficient for finding examples that different codes do not give the same grounds for findings of FFP. We identified two issues: responsibility (whether a researcher is blameworthy is for misconduct) and culpability (the extent to which a researcher is blameworthy).

With regards to responsibility: some codes explicitly hold all co-authors of a fraudulent research paper responsible, while many codes are silent on this issue. For instance, the Austrian Agency for Scientific Integrity (OeAWI) state in its guidelines that all co-authors of a publication are jointly responsible for it.[19] By contrast, the main regulatory documents in France[20] do not contain a provision similar to the OeAWI document.[21] Divergences in such provisions would have concrete consequences for cases of FFP in co-authored papers.

With regards to culpability: in various areas of the law, it is common to distinguish between levels of intent in order to determine the proportionate sanction.[22] Codes of conduct have often adopted this legal framework, most explicitly in the Federal Policy on Research Misconduct (USA), which states that a finding of misconduct (defined as FFP) must be shown to have been 'committed intentionally, or knowingly, or in reckless disregard of accepted practices'.[23] This explicitly excludes negligent actions as potential violations of research integrity standards, i.e. the individual who violated RI standards *did not* know but *should have* known those standards.[24]

In the European context it has not been possible to take a similarly uniform line with regards to culpability, due to divergent legal contexts between European countries, and this is one reason why the 2017 ECoC is silent on levels of intent and culpability (Maura Hiney, co-author of the 2017 ECoC, personal communication). Yet this means that some codes of conduct include negligent violations of the code as potential misconduct (e.g. DE, AT, FI, NO), while others purposely restrict research misconduct to behaviours with a conscious intention to deceive (e.g. IE, UK).

In sum, there is potential for unfair treatment of alleged misconduct by international collaborations with regards to (a) the categories of what behaviours count as misconduct, (b) the extent to which researchers are responsible for the misconduct of their collaborators, (c) the conditions under which researchers are culpable for their own behaviour.

---

[17]ESF-ALLEA (European Science Foundation and All European Academies). (2011). The European code of conduct for research integrity. Retrieved from http://www.allea.org/wp-content/uploads/2015/07/Code_Conduct_ResearchIntegrity.pdf

[18]Note that such unfairness can always arise due to different legal or institutional contexts; the question at stake here is whether the divergent formulations in the codes and guidelines give any grounds to unfair judgments of joint international misconduct.

[19]OeAWI (Austrian Agency for Scientific Integrity). (2015). OeAWI guidelines for good scientific practice. Retrieved from https://www.cdg.ac.at/fileadmin/main/documents/Sonstige_Dokumente/160418_OeAWI_Richtlinien_Broschuere_DE_EN.pdf

[20]COMETS (CNRS Ethics Committee). (2017). Integrity and responsibility in research practices. Retrieved from http://www.cnrs.fr/comets/IMG/pdf/guide_2017-en.pdf; CNRS & INRA. (2015). French National Charter for Research Integrity. Retrieved from http://www.cnrs.fr/comets/IMG/pdf/french_national_charter__research_integrity_29jan2015_-2.pdf

[21]Compare these to the latest version of the ICMJE recommendations, where it is only specified that 'an author should be able to identify which co-authors are responsible for specific other parts of the work' and that 'authors should have confidence in the integrity of the contributions of their co-authors'. Hence it stops short of stipulating that authors should take responsibility for the integrity of the contributions of co-authors.

[23]OSTP, op. cit. note 15, p. 76262.

[22]Yaffe, G. (2010). Intention in Law. In T. O'Connor & C. Sandis (Eds.), *A companion to the philosophy of action*. Oxford: Wiley-Blackwell.

[24]For a fuller overview of this issue, see Desmond, H. (2020). Professionalism in Science: Competence, Autonomy, and service. *Science and Engineering Ethics*, *26*(3), 1287–1313. https://doi.org/10.1007/s11948-019-00143-x

## 4.2 | Credibility

The second issue concerns ethics more directly (rather than the law): the credibility of the ethical content in codes of conduct. As mentioned in the first section, with 'credibility' we have in mind how there is considerable scepticism about the effectiveness of self-regulation, not only by corporations[25] but also by professions.[26] After all, codes of conduct are one of the main tools for self-regulation, since they guide the scientific community in regulating their own behaviour, including dealing with their own cases of misconduct. How credible are codes of conduct as effective tools for self-regulation? This is increasingly controversial. Some have been calling for more robust external regulation, for instance in the form of an increased criminalization of scientific misconduct.[27] Such sceptics would argue that codes of conduct are primarily window-dressing, hiding a harsher, competitive reality. In this view, the only way to regulate credit-maximizers is not to appeal to their sense of ethics, but to change their calculus by means of retribution and punishment.

This consideration can be situated in the broader discussion about what precisely the purpose is of professional codes of ethics, and what the reasons are for complying with them.[28] It is also a particular instance of the yet broader (and perennial) philosophical problem of how ethics can become subordinated to dynamics of power and prestige.[29]

Mindful of these broader implications, for our purposes here we will understand 'window-dressing' in the following way: if the ethical content of codes of conduct is merely window-dressing, then its primary function is to create a *perception* to the outside world that a community cares about ethical behaviour. By implication, the regulatory function to guide the *actual* behaviour of the members of the community is either secondary or non-existent. In other words, if a code is mere window-dressing, then it is more important *that* ethical values are present rather than *which* precise ethical values are listed in the code.

This has two implications for the content of codes. The first is that the listed values in the resulting ethical frameworks can be expected to be rather generic and ambiguous with regards to what precise behaviours are being promoted. This is, for instance, how Giacomini et al.[30] understand the concept of window-dressing in their investigation of the ethical content of Canadian health care policies. By means of a qualitative content analysis, they conclude:

> Currently, many ethics frameworks seem detached from policies, floating alongside in special text boxes or appendices. They are also set apart by vague, platitudinous language; imperatives are typically expressed in terms that are inspirational, positive, nonspecific, and non-operational. (...) Some degree of flexibility is appropriate and necessary in ethical analysis. However, expansive ethical terms with little definition or operationalization become weak rhetoric, and apparent agreements (e.g., to include "accountability" among guiding principles) may mask fundamentally different aims (e.g., involving public participation vs. transparent reporting).[31]

We cannot rule out an analogous finding out with regards to RI codes of conduct. The values listed by the 2017 ECoC that are most commonly duplicated by national-level documents—'honesty' and 'respect', see Figure 5—are arguably precisely those that are applicable to any economic activity or interpersonal relationship. These values could be deemed generic insofar they are not specifically applicable to the activity of scientific research, and hence one could legitimately ask what precise guiding function they have for the activity of scientific research (rather than any economic activity or any interpersonal relationship).

However, a systematic testing of the principles of RI for their precise information content is outside the scope of the minimalist methodology used in this study. Moreover, while we admit that the principles in the surveyed documents are often generic, we would not wish to go so far as Giacomini et al. in suggesting they are 'weak rhetoric', since many of the ethical values are connected to good practices and categories of misconduct. Nonetheless, testing the genericity of ethical content in codes of conduct is a legitimate research question and could be subject to future study.

Here we discuss a second textual implication of 'window dressing', namely that, if the ethical content is not intended to play a genuine regulatory function, one can expect different codes to list very different ethical values without any discernible rationale for the divergence. Thus, if ethical content is mere window-dressing, then it would not matter if a code listed 'honesty, respect, integrity' or 'accountability, accuracy, and truthfulness'—as long as there was

[25]Short, J. L., & Toffel, M. W. (2010). Making self-regulation more than merely symbolic: The critical role of the legal environment. *Administrative Science Quarterly*, 55(3), 361–396. https://doi.org/10.2189/asqu.2010.55.3.361

[26]Carvalho, T., Correia, T., & Serra, H. (2018). Professions under suspicion: What role for professional ethics and commitment in contemporary societies? *Sociologie, Problemas e Práticas*, 88, 9–25. https://doi.org/10.7458/SPP20188814795

[27]Collier, R. (2015). Scientific misconduct or criminal offence? *CMAJ : Canadian Medical Association Journal*, 187(17), 1273–1274. https://doi.org/10.1503/cmaj.109-5171; Sovacool, B. K. (2005). Using criminalization and due process to reduce scientific misconduct. *The American Journal of Bioethics: AJOB*, 5(5), W1–W7. https://doi.org/10.1080/15265160500313242

[28]Spielthenner, G. (2015). Why comply with a code of ethics? *Medicine, Health Care and Philosophy*, 18(2), 195–202. https://doi.org/10.1007/s11019-014-9594-5; Dobson, J. (2005). Monkey business: A neo-Darwinist approach to ethics codes. *Financial Analysts Journal*, 61(3), 59–64. https://doi.org/10.2469/faj.v61.n3.2728

[29]Evans, J. H. (2012). *The history and future of bioethics: A sociological view*. Oxford University Press; Callahan, D. (2005). Bioethics and the culture wars. *Cambridge Quarterly of Healthcare Ethics*, 14(4), 424–431. https://doi.org/10.1017/s0963180105050577

[30]Giacomini, M., Kenny, N., & DeJean, D. (2009). Ethics frameworks in Canadian health policies: Foundation, scaffolding, or window dressing? *Health Policy*, 89(1), 58–71. https://doi.org/10.1016/j.healthpol.2008.04.010

[31]Ibid: 67.

*some* ethical content to create the perception of that the research community cares about integrity.

Note the direction of the inference in the previous paragraph: from window-dressing to divergence. Hence, conversely, given the observed divergences in ethical values, we cannot simply conclude from this that the leading regulatory documents are intended as window-dressing. There are other possible explanations for the divergence that we must consider. We limit the discussion to three alternative explanations: (a) different envisaged uses of the leading regulatory documents, (b) different legal and institutional contexts, (c) honest disagreement about a complex subject matter. Each of these explanations is a possible response to the sceptical stance.

### 4.2.1 | Different envisaged uses of the documents?

The first explanatory factor to consider is that the content of the different regulatory documents may have been strongly influenced by different envisaged uses (see Figure 3). For instance, the Norwegian General Guidelines for Research Ethics is only a single page and contains concise statements about the nature of science, and advice for action. By contrast, the Swedish Good Research Practice is over 50 times longer than the Norwegian document, and contains detailed discussions about different ways of defining misconduct, and philosophical reflections on the difference between law and ethics. It is clear that these two documents have very different envisaged uses by researchers. Another example is the way in which some guidelines contain directives on institutional responsibilities, such as The Netherlands Code of Conduct for Research Integrity, the Croatian Code of Ethics, or the UK's Concordat to Support Research Integrity. By contrast, other codes like the Code of Ethics for Scientific Research in Belgium do not emphasize institutional responsibilities to the same extent, and thus mainly target individual researchers.

National-level codes may also wish to be more precise than the ECoC. For instance, among violations the ECoC includes 'exaggerating the importance and practical applicability of findings'.[32] However, since the ECoC does not define what constitutes 'exaggeration', and since exaggeration is otherwise a context-dependent and slippery term, authors of national-level documents may have believed it necessary to add more concrete detail. This seems to have been the case for the Dutch leading regulatory document, where relatively precise standards are offered (see discussion in Section 4.1).

Nonetheless, while the envisaged use could explain some divergence—for instance the detail with which misconduct is defined—at stake here is the observed divergence in principles and values. Moreover, the universal character of RI values is a fundamental presupposition of the core-periphery model,[33] and is a presupposition we also make for the purposes of this study.[34] Thus what principles

are listed should not depend on whether the document is intended as a proto-legal document or an ethical statement. In sum, divergence in envisaged uses can be a sufficient explanation for the divergence in listed RI principles.

### 4.2.2 | Different legal and institutional contexts?

A second explanatory factor is that each national-level regulatory document is designed to be finely attuned to specific needs or legal constraints within the national context. As mentioned in the first section, the background paper to the 2011 ECoC acknowledges explicitly that national-level documents could and even should diverge.

However, while this factor surely can explain some divergence in content—for instance, divergence in understandings of responsibility and culpability—also this factor hardly can explain much of the observed divergence with regards to ethical values. That there are differences in 'national traditions' and 'legislative regulations' between European countries is undeniable, but once again, this potential explanation of the observed differences is likewise insufficient given the universal character of RI principles. Even across different national traditions and legislative regulations, the fundamental principles of research integrity remain unchanged.

### 4.2.3 | Honest disagreement?

A final, and in our view the most plausible, explanation for the divergence is honest disagreement between the authors of the various codes and guidelines. The nature of research integrity is, after all, a very complex issue. It depends crucially on what 'good' science is, and defining the latter presupposes an answer to all of the following, difficult questions: What is good methodology? What are good collegial and collaborative relationships between scientists? How much can scientists be influenced by societal and moral considerations while keeping their integrity?[35]

Thus, for instance, three examples of a different listing of RI principles are: 'generally valid ethical values and thoughtfulness, respect, courtesy, and honesty' (SK), 'dignity, responsibility, equity, correctness, diligence' (IT), or 'honesty, responsibility, scientific integrity' (FR). If this is due to honest disagreement, then authors simply have a different idea of what the principles of RI are. This is plausible given the complexity of the nature of RI.

A reason in support of this explanation is that the authors of national-level documents can be assumed to be experts on RI in their national context, and hence one can assume that any deviation, no matter how small, reflects a conscious deliberation. These experts are presumably familiar with the ECoC, hence even a small substitution of

---

[32]ALLEA, op. cit. note 5, p. 8.

[33]ESF-ALLEA, op. cit. note 2, p. 14.

[34]Questioning and justifying this presupposition would require a different, more philosophical paper.

[35]For a physicist or biologist, the answer to that question might seem easy (i.e. there should be zero influence), but for an economist or sociologist the issue is not so straightforward (see e.g. Sen, A. (1987). *On ethics and economics*. Oxford: Blackwell).

'honesty' for 'transparency' can be assumed to be reflect a different understanding of what precisely 'research integrity' amounts to. Even hypothetically, if they chose to disregard the ECoC and wished to construct a new code or guidelines from the bottom up, this would still be a conscious choice insofar it would reflect a dissatisfaction with the ECoC. (By contrast, if we drop the assumptions that the authors are experts who have deliberated over deviations from the ECoC, we end up with the sceptical window-dressing hypothesis, where it matters *that* a code is written rather than *how* the code is written.)

However, if the divergence is due to honest disagreement, then the sceptical hypothesis is not strongly undermined. The honest disagreement then would point to a lack of common understanding about what precisely the regulatory function of RI principles is. What difference does it make to researcher behaviour whether to list 'respect, courtesy, or honesty' vs. 'dignity, responsibility, equity'? The sophisticated sceptic could acknowledge that the intentions of each individual author may have been noble, and that the disagreement may have been honest, but that, since they did not coordinate with each other to clarify what precisely the regulatory function is of RI principles, authors unwittingly acted to provide window-dressing without ensuring that actual researcher behaviour would be optimally impacted. The fact that codes have sometimes been authored in response to high-profile scandals of research fraud[36] only strengthens the sceptical view that codes are about managing perceptions rather than about changing behaviour.

In this respect, the contrast between the divergence on RI principles and the convergence on FFP provisions is striking. FFP provisions are understood to play a clear regulatory function, and this is reflected by a total lack of divergence across various national-level documents on this issue (even if they do on the more detailed issues of responsibility and culpability). FFP provisions matter, but it is less clear whether and how RI principles matter. The sceptic would point to this issue, and would explain the state of honest disagreement between authors of codes/guidelines as due to the fact that RI principles do not play any guiding influence on researcher behaviour, and so authors are free to pursue their own intuitions, unconstrained by objective regulatory function.[37]

We do not hold the sceptical view, but do believe it should be taken seriously. The current drive for increasing attention to research integrity issues would be partially self-defeating if the upshot were that the sceptical view would be strengthened. Moreover, this sceptical view is far from being a fringe conspiracy theory: it has underpinned the increased external regulation of other professions such as medicine or law in recent decades, for instance through new public management.[38] We would also like to note the long tradition in sociology and philosophy of taking a sceptical-deflationary view towards professional codes of conduct—and even towards ethics as such.[39]

### 4.2.4 | Summary

Ultimately, while we are most sympathetic to the explanation that the divergences are due to honest disagreement, we do acknowledge that the current textual state of RI codes does not allow one to satisfactorily falsify the sceptical hypothesis (i.e. that codes of conduct are window-dressing). In most codes of conduct, no reason is given for diverging from the ECoC on core aspects of RI—the divergence is often not even acknowledged. If the divergences reflected reasoned and conscious disagreements about RI, then one would expect some preamble stating the presuppositions of the code of conduct. Yet, few had such preambles, as is reflected in Figure 3, where the distribution of word counts is clearly skewed towards the shorter word counts. The predominant goal is providing clear, actionable advice, but since the end result is that different clear and definitive advice is given, some sceptics may seize upon this to undermine the credibility of RI codes of conduct as such.

### 4.3 | The need for a normative framework of RI codes and guidelines

The previous discussion underlines the need for a more systematic normative framework that can be used to evaluate descriptive differences in RI codes and guidelines. The two issues discussed here—fairness in dealing with allegations of misconduct, and the threat that codes of conduct are viewed as mere window-dressing—suggest two dimensions that such a normative framework would incorporate.

The first dimension would be a jurisprudential one that would help evaluate legal consequences of RI codes. The need for such a framework can also be observed in previous studies. For instance, Godecharle et al. introduce a distinction between norms and values in order to analyse content of codes of conduct, understanding the latter as 'universal' but the latter as 'embedded in a specific context: situation, time, and place'.[40] They use this distinction to categorize codes on a value-focused vs. norm-focused continuum, with as an

[36]DFG (German Research Foundation). (2013). Sicherung guter wissenschaftlicher Praxis [Safeguarding good scientific practice]. Retrieved from http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf, p. 64.

[37]The sceptic would respond similarly to the idea that the authors of national-level leading documents deviate from the ECoC in order to inject some originality into the documents. When it concerns FFP provisions, originality is not only not necessary, but also undesirable. If originality is pursued with regards to RI principles, this is only a reflection of the fact that there is no objective regulatory function of this part of codes/guidelines, and the sceptic would charge that the originality is tolerated because the real function of codes/guidelines is to provide window-dressing. (We thank an anonymous referee for bringing our attention to the possibility of originality.)

[38]Carvalho, T., & Correia, T. (2018). Editorial: Professions and professionalism in market-driven societies. *Professions and Professionalism, 8*(3), e3052. https://doi.org/10.7577/pp.3052

[39]Foucault, M. (2012). *Discipline and punish: The birth of the prison.* Knopf Doubleday Publishing Group; Goldstein, J. (1984). Foucault among the sociologists: The 'disciplines' and the history of the professions. *History and Theory, 23*(2), 170–192. https://doi.org/10.2307/2505005; Larson, M. S. (1977). *The rise of professionalism: A sociological analysis.* University of California Press.

[40]Godecharle, S., Nemery, B., & Dierickx, K. (2014). Heterogeneity in European research integrity guidance: Relying on values or norms? *Journal of Empirical Research on Human Research Ethics, 9*(3), 79–90. https://doi.org/10.1177/1556264614540594

example of the former the Code of Ethics for Scientific Research in Belgium, and an example of the latter the Danish Code of Conduct for Research Integrity. Moreover, they understand the distinction between values and norms to correspond to the distinction between morality-based approaches and legalistic approaches.

Yet, while it may be useful in some contexts to distinguish between 'legalistic' and 'ethics-focused' approaches to research integrity and research misconduct,[41] codes of conduct must nonetheless cover both approaches. When findings of misconduct are sanctioned by integrity commissions, researchers often have the legal right to challenge such findings. Thus, the question is what legal implications a code of conduct has, not whether the code of conduct avoids legal terminology.

Such considerations lead to jurisprudential questions about what the purpose is of RI legislation and of the sanctioning of research misconduct: is it, for instance, retribution and deterrence, or is it rehabilitation? These questions matter for how fine-grained the categories of misconduct will be. Thus, if the aim of sanctioning is retribution or deterrence, then the sanction will harm the misbehaving researcher (in a way proportional to the transgression); hence, to avoid a chilling effect on scientific communities, the evidential bar for a misconduct finding would need to be relatively high. This implies that many of the less serious forms of misconduct would not be categorized as transgressions. By contrast, if the aim of sanctioning is rehabilitation or education, then the sanction does not necessarily harm the misbehaving researcher, and lesser forms of misconduct could in principle be sanctioned without a chilling effect on scientific communities.

While a further exploration of questions would lead us far beyond the scope of the current paper, this brief discussion illustrates the need for a normative framework with a jurisprudential dimension that identifies the 'proper aim' of the sanctioning of misconduct. Such a framework would not only help evaluate current codes of conduct for their legal implications, but would also help in shaping future codes of conduct.

The second dimension is sociological. Ethical choices are not made in isolation from the large system of incentives (grants, tenure, promotion, etc.) that may not always promote integrous behaviour. What system of incentives is desirable? How should responsibilities be organized? These questions pertain to what social structures are to be deemed as desirable, and today a large choice facing the scientific community is whether to promote professional social structures, where researchers and communities of researchers follow their own judgment, or bureaucratic social structures, where researchers follow procedures and processes that are defined by non-researchers (e.g. corporate leaders or administrators). In the first case, individuals must navigate a system of incentives that is not always oriented towards the 'service ideal';[42] in the second case, individuals are averted as much as possible from such dilemmas.

Previous studies sometimes inadvertently depend on this distinction in their normative evaluation of differences between codes of conduct. Thus, in their analysis of ethical frameworks used by health care policies, Giacomini et al. distinguish between values or themes that lie on a continuum between being 'more like ethical principles' and 'more like policy processes or goals'.[43] Thus, policy processes seek to strengthen bureaucratic structures; ethical principles (as service ideals) serve to strengthen professional structures.

The sociological dimension helps evaluate existing codes of conduct, as well as inform future codes of conduct. If the goal is to strengthen professionalism, then listing generic values such as 'respect' or 'honesty' are likely insufficient to actually impact researcher behaviour. Take honesty for instance: honest scientific communication does not mean turning scientific papers into chronicles of all details regarding the research process. There is a distinction between honesty and pedantry; yet, at what point does the selection of data and insights stop being honest, and start being biased or ideological? In true Aristotelian fashion, 'honesty' is a virtue situated in the middle between the extremes of pedantry and bias or deceit, but simply listing 'honesty' in codes of conduct does not actually guide researchers in *how* to be honest.

In a bureaucratic approach, values such as honesty are not intended to actually impact behaviour: what matters is how such values are fleshed out in terms of procedures and processes (e.g. to pre-register research protocols, keeping data for future reproduction efforts, etc.). By contrast, in a professionalism approach, values such as honesty are intended to directly impact behaviour, but they would need to be fleshed out in terms of research goals (e.g. truth, understanding), and would leave the details of how to achieve those goals in particular circumstances up to the individual judgment of researchers. Some groundwork in this regard has been done about the desired social structures of the scientific community,[44] but more remains to be said, and the current study underlines the need for a normative framework with not just a jurisprudential dimension, but also a sociological one, such that the sociological implications of a codes of conduct can be evaluated.

## 5 | CONCLUSION

The ECoC was intended to help harmonize research integrity standards across Europe, by identifying the 'core' elements of RI: especially the definition of misconduct and the universal principles of RI. In this study we investigated whether this 'core-periphery' model of harmonization is realized by examining patterns of replication of the ECoC. We found that the only unambiguous consensus concerns FFP provisions; we also found a relatively large consensus that the value of 'honesty' is central to research integrity. Otherwise, with regards to less egregious forms of misconduct, or with regards to other principles of RI, the consensus is not obviously present.

[41]See also Pennock, R. T. (2019). *An instinct for truth: Curiosity and the moral character of science*. The MIT Press.

[42]Freidson, E. (2001). *Professionalism, the third logic: On the practice of knowledge*. University of Chicago Press.

[43]Giacomini et al., op. cit. note 29, fig. 1 (p. 64).

[44]Desmond, H. (op. cit. n. 24).

This presents some scope for injustice: divergences between national-level documents give grounds for diverging evaluations of joint misconduct by international collaborations. However, most insidiously, we argued that it gives ammunition to sceptics of self-regulation, who believe that the primary purpose of a code of conduct is to provide window-dressing, i.e. a perception that integrity matters but while actual behaviour is driven by the cold competitive calculus of incentive maximization. It is not a coincidence that there is such an unambiguous consensus regarding FFP: findings of FFP, not coincidentally, have very large repercussions for reputations, careers, and even livelihoods, and hence the precise wording of what allows for a finding of FFP is important. Does this mean that the precise wording of RI principles is thought not to matter? While we cautioned against drawing this conclusion (another explanation would be honest disagreement about the principles of RI), we also argued that it cannot be ruled out altogether, and that this alone is problematic enough to be highlighted.

In this way, while this study targeted the European regulatory situation, it ultimately raised fundamental questions about how codes of conduct should be constructed and about what precise impact we want these codes to have. What are precisely the values that are specific to the activity of scientific research? What legal approach to sanctioning research misconduct is appropriate? These questions must be addressed if the 'common understanding' aimed at by the original ECoC is to be achieved in any substantial way.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Hugh Desmond: (a) conceived the idea for an overview study, (b) designed the methodology for documents and identified the core-periphery model, (c) collected the documents, (d) selected the leading documents, (e) analysed and visually represented the differences between documents, (f) interpreted the differences, and (g) wrote the successive drafts of the paper.

Kris Dierickx: (a) conceived the idea for an overview study, and (b) commented on important intellectual content in successive drafts of the paper.

## ORCID

*Hugh Desmond* https://orcid.org/0000-0002-4822-923X

## AUTHOR BIOGRAPHIES

HUGH DESMOND is professor of philosophy at the University of Antwerp and postdoctoral researcher at the Centre of Biomedical Ethics and Law, KU Leuven. He works on a spectrum of issues spanning philosophy of science, moral psychology, social epistemology, and ethics. Website: http://www.hughdesmond.net

KRIS DIERICKX is professor of biomedical ethics at the Centre of Biomedical Ethics and Law, at the school of medicine, KU Leuven Belgium. His research focuses on research integrity & ethics and the ethical challenges in new medical technologies.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.