Methodological Appendix: The Public Legitimacy of EU Agencies

Dictionary Approach

We define four *a priori* dictionaries based on existing theory and central policy documents.

Legitimacy argument	Main sources
Evidence-based	Commission of the European Communities (2002); Majone (1997); Maggetti (2010); Bellamy (2010)
Legislator's command	Bellamy (2010); Maggetti (2010); Eriksen (2020)
Participation	European Commission White Paper on Governance (COM/2001/0428); Krick, Christensen, and Holst (2019)
Human rights	European Charter of Fundamental Rights

We aimed to define the dictionary before seeing the data we use for our analysis. Some tuning was nonetheless necessary. We had to remove words that made theoretical sense, but lacked the specificity necessary to pick up on our categories of theoretical interest. For instance, "control" is a core feature of the legislators' command argument, but impossible to use because it is too general (as in the term "border control"). The final dictionaries are given here:

Evidence-based terms

```
expert*
okvalificerad*
kvalificerad*
kvalifikation*
teknokrat*
kunnskap*
vetande
ekonom
ekonomen
ekonomer
analys*
analytiker*
motbevis*
bevis*
faktum
faktumet
professionell*
ovetenskap*
vetenskap*
"Falska nyheter"
"alternativa fakta"
metodologi*
teknik
```

teknisk
tekniska

Legislator's command terms

ansvarig*
folkvald*
riksdagen
grundlagen
delegera*
representant*
representation*
parlament*
rösta
röstade
röstat
folkomröst*

Public participation terms

deltaga*
medbestämmande*
civilsamhälle*
lekfolk
lekmän
lekman*
partnerskap
"civila samhället"
utfråging*
svårgenomskådligt
genomskådligt
transparens
öppenhet
inkluder*
intressegrupp*
konsultation*

Fundamental rights terms

människorätt*
människans värdighet
grundläggande rättighet*
mänsklig(a) rättighet*
ekonomisk(a) rättighet*
social(a) rättighet*
kulturell(a) rättighet*
jämlikhet*
jämställdhet*
rättssäkerhet*
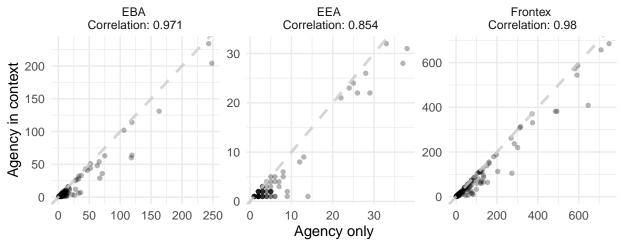fysisk integritet*
mental integritet*
humanitär*

Agency contexts

We defined a sparse set of context terms for each agency. The aim of the contexts are to capture the relative coverage of an agency in news articles about its policy field.

The context terms are selected to capture a common-sense idea of the agencies respective policy areas. The initial terms were tuned after looking at a sample of articles. For Frontex, we had to widen the initial search terms from a narrower focus on border control to a wider focus on migrants and refugees. For EEA, we initially searched only for "climate change" and "global warming." We found that most EEA articles in Sweden were about cars and air quality. So we included terms about air quality and emissions. EBA deals within a narrow domain which is well captured by a single word: "bank." We use the plural, indefinite and definite (bank**er**, bank**erna**) in order to avoid homonyms.

Plotting monthly articles mentioning an agency against montly articles mentioning an agency *and its context terms* shows that the sparse context terms capture well the articles in which the agencies are mentioned.

Agency vs agency–in–context

Hits for agency compared to hits for agency AND context terms.

Data Collection

We use the dictionaries to gather our final corpus. We download all articles matching {agency A_i} **and** {any term in legitimacy dimension D_j} AND {any term in context C_i}. This gives 12 non-mutually-exclusive sets: three agencies-in-context by four legitimacy dimensions. We combine these sets into one dataset. We used a script for parsing text files from the Retriever Mediearkivet database written in `Python` by [anonymized for peer review]. We ran the script from `R` through the `reticulate` package (Ushey, Allaire, and Tang 2020). This gave us a dataset with 8302 articles and 16 variables.

Table 2: Search terms for agencies and their respective contexts

Agency	Agency terms	Context terms (Swedish)	Context terms (Translation)
Frontex	Frontex	(migrant* OR flykting*)	(migrant* OR refugee*)
EEA	"Europeiska miljöbyrån"	(utsläpp OR luftförorening* OR klimatändring OR 'global uppvärmning')	(emissions OR 'air pollution' OR 'climate change' OR 'global warming')
EBA	"Europeiska bankmyndigheten"	(banker OR bankerna)	(banks or 'the banks')

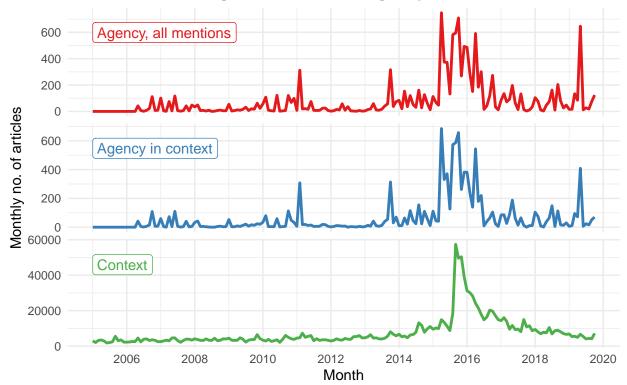Additionally, we download the monthly *frequency* of:

- Articles mentioning an agency
- Articles mentioning any of an agency's *context terms*.
- Articles mentioning an agency *and* any of its context terms.

These figures allow us to calculate an agency's relative coverage within a theoretically defined context. They also allow us to calculate the total number of legitimacy terms per article about an agency, without needing to download the full text of articles with zero hits on the legitimacy dictionaries.
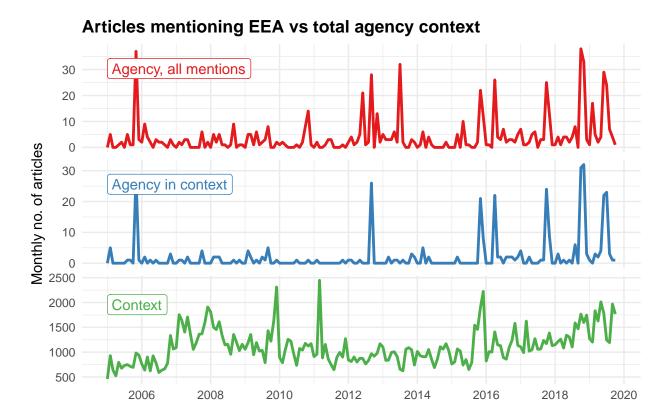
Descriptive statistics on agencies and their contexts

This section presents the monthly number of articles mentioning an agency, its context terms, and an agency *and* any of its context terms. Note that the scale of the plots' y-axes vary greatly.

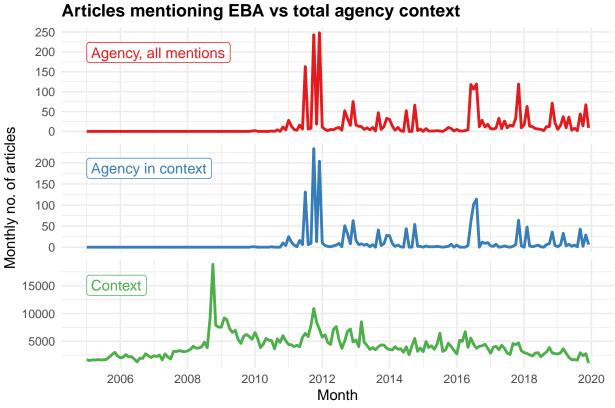Articles mentioning Frontex vs total agency context

Note: Y axes are different

Articles mentioning EEA vs total agency context

Note: Y axes are different

Articles mentioning EBA vs total agency context

Note: Y axes are different

Details on analysis

Term frequencies and TF-IDF

As a next step, we score each article on our legitimacy dimensions. We calculate two main measures: The term frequency ($tf_{d,t}$) counts the number of words from each legitimacy dimension in each article. The TF-IDF weighs the term frequency by how many documents the term appears in.

2. For each article d, count how many times it contains each term t in each dictionary. This gives every article an array of terms in four dictionaries, and their frequencies: $tf_{d,t}$
3. For each term, calculate its *inverse document frequency*, given by $idf_t = \log \frac{1+N}{1+df_t}$, where N is the total number of articles in the corpus and df_t is the number of articles containing term t.
4. For each article and term, calculate its TF-IDF by $tf_{d,t} \times idf_t$.
5. For each article, summarize the TF-IDF scores within each legitimacy dimension.
6. Also, summarize each article's total term frequency TF for each legitimacy dimension.

In order to obtain the words-per-article measure reported in the article's Figure 2, panel **A**, we calculated the total term frequency for each agency-legitimacy-dimension combination, divided by the total number of articles mentioning that agency. We also calculated the same measure by year in panel **B**.

Articles selected for qualitative coding

We select the 40 articles with the highest total TF-IDF on a legitimacy dimension for qualitative analysis. Some agencies have less than 40 articles in a legitimacy dimension. Moreover, since TF-IDFs are calculated from a limited set of integers, there are inevitably some articles with exactly equal TF-IDFs. When this is the case for

the 40th highest, we include all articles with that score. That gives more than 40 articles in some cells of the below table.

Table 3: Number of articles selected for qualitative analysis, by agency and legitimacy dimension.

Agency	Evidence-based	Fundamental rights	Legislator's Command	Participation
EBA	40	7	42	26
EEA	40	6	20	9
Frontex	40	41	40	45

References

Bellamy, Richard. 2010. "Democracy without democracy? Can the EU's democratic 'outputs' be separated from the democratic 'inputs' provided by competitive parties and majority rule?" *Journal of European Public Policy* 17 (1): 2–19. https://doi.org/10.1080/13501760903465256.

Commission of the European Communities. 2002. "Communication from the Commission on the operating framework for the European regulatory agencies." Brussels.

Eriksen, Erik Oddvar. 2020. "Depoliticisation and Its Legitimacy Problems." *ARENA Working Paper*, no. 3/2020. https://www.sv.uio.no/arena/english/research/publications/arena-working-papers/2020/wp-03-20.pdf.

Krick, Eva, Johan Christensen, and Cathrine Holst. 2019. "Between 'Scientization' and a 'Participatory Turn'. Tracing shifts in the governance of policy advice." *Science and Public Policy* 46 (6): 927–39. https://doi.org/10.1093/scipol/scz040.

Maggetti, Martino. 2010. "Legitimacy and Accountability of Independent Regulatory Agencies: A Critical Review." *Living Reviews in Democracy*.

Majone, Giandomenico. 1997. "The new European agencies: regulation by information." *Journal of European Public Policy* 4 (2): 262–75. https://doi.org/10.1080/13501769709696342.

Ushey, Kevin, JJ Allaire, and Yuan Tang. 2020. *Reticulate: Interface to 'Python'*. https://CRAN.R-project.org/package=reticulate.