ETH zürich

ETH Library

# Dimensions of Mind in Semantic Space

**Working Paper**

**Author(s):**
Ash, Elliott; Stammbach, Dominik; Tobia, Kevin

**Publication date:**
2021-11

**Permanent link:**
https://doi.org/https://doi.org/10.3929/ethz-b-000530211

**Rights / license:**
In Copyright - Non-Commercial Use Permitted

**Originally published in:**
Center for Law & Economics Working Paper Series 2021(14)

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

ETH zürich

# Center for Law & Economics Working Paper Series

Number 14/2021

## Dimensions of Mind in Semantic Space

**Elliott Ash**
**Dominik Stammbach**
**Kevin Tobia**

November 2021

All Center for Law & Economics Working Papers are available at
lawecon.ethz.ch/research/workingpapers.html

# Dimensions of Mind in Semantic Space

Elliott Ash, Dominik Stammbach, Kevin Tobia*

November 19, 2021

**Abstract**

Recent psychology research has established that people do not employ a simple unidimensional scale for attributions of personhood, increasing from non-sentient rocks to mentally complex humans. Rather, there are two personhood dimensions: agency (e.g. planning, deciding, acting) and patiency (e.g. feeling, desiring, experiencing). Here we show that this subtle distinction also occurs in the semantic space of natural language. We develop computational-linguistics tools for measuring variation in agency and patiency in language and validate the measures against human judgments. To demonstrate the usefulness of the method, we map both dimensions of personhood in historical English-language corpora over the last 200 years. First, we show that while women are now described as having similar levels of agency as men, they are still described as more patient-like. Second, we show that domesticated animals have gradually gained higher attributions of patiency, but not agency, over time.

## 1 Introduction

What is a person? This longstanding philosophical question is at the root of modern moral and legal issues. It fuels debates about animal rights (Sunstein and Nussbaum, 2005), corporate speech (Ellis, 2011), abortion (Little, 2008), brain death (Willke and Andrusko, 1988), and environmental protection (Gordon, 2019). Personhood attribution is also related to social phenomena including anthropomorphization (Epley et al., 2007), dehumanization

*Author info: Elliott Ash, ETH Zurich, ashe@ethz.ch; Dominik Stammbach, ETH Zurich, dominik.stammbach@gess.ethz.ch; Kevin Tobia, Georgetown University, kt744@georgetown.edu. We thank Joanna Bryson, Joshua Knobe, Alex Stremitzer, and Daniel M. Wegner for helpful feedback. We acknowledge research support from ETH Zurich and the Swiss National Science Foundation.



Electronic copy available at: https://ssrn.com/abstract=3959847

(Harris and Fiske, 2006; Haslam, 2006; Haslam and Loughnan, 2014; Tarasenko-Struc, 2020), gender stereotyping (Hamilton, 1991), and objectification of men and women (Gray et al., 2011).

Since Aristotle, philosophers have distinguished between agents – who act and cause events – and patients – who experience these actions and events (Olson, 2019; Irwin, 1986). Recent psychological research has begun to disentangle these agent and patient dimensions (Gray et al., 2007; Gray and Wegner, 2009; Gray et al., 2011; Strohminger and Jordan, 2021), finding that some entities are evaluated to be high in agency but low in patiency (e.g. robots), while others are evaluated as high in patiency but low in agency (e.g. kittens). Given the deep links of agency with legal obligations, and patiency with legal protections, further social-science research is needed on how these dimensions of mind are perceived in the real world.

Notwithstanding their established relevance, these dimensions of mind have not been analyzed outside the lab. Without a survey, perceptions about agency and patiency cannot be measured directly. The task of this paper is to measure them indirectly – from large corpora of natural language. Thus, we supplement the psychological approach with complementary tools from natural language processing (Brunet et al., 2019). Building on linguistic (Dowty, 1991; Resinger et al., 2015) and psychological theory (Gray et al., 2007), we use the dimensions of mind theory as a lens into perceptions expressed in natural language (Schweitzer and Waytz, 2020).

More specifically, we adapt a key technology from computational linguistics, word embedding (Mikolov et al., 2013; Pennington et al., 2014). Through machine reading of large corpora, word embedding models can map words and concepts as vectors in a geometric space. In this paper, we use these tools to learn dimensions in semantic space corresponding to agency and patiency. Further, we score a large set of entities according to how they are described along the agency and patiency dimensions in everyday language usage.

To guide the design of the language measures and to validate their usefulness, we run an online survey to assess perceptions on relative agency and patiency for this same set of entities. We show that the agency and patiency rankings of entities by the language model match the respective rankings of entities by survey respondents. Further, the difference in the



Electronic copy available at: https://ssrn.com/abstract=3959847

language scores for agency and patiency match the difference in the respective survey scores. Thus, the language model scores properly capture a two-dimensional index of personhood perceptions.

With the validated language-based measures in hand, we use them for a long-run historical analysis of personhood perceptions over the last 200 years. Using the Corpus of Historical American English (COHA) (Davies, 2010), a genre-balanced historical corpus, we train embeddings and construct agency/patiency dimensions for each decade since the 1820s. We then compute agency and patiency scores over time for sets of relevant entities while accounting for sampling uncertainty in those measures.

The first historical analysis is to measure differences in the attributed agency and patiency of men and women. In the early 1800s, men were depicted as having more agency in texts, while women were depicted as having more patiency. The difference in agency shrank over the 1800s, with women having relatively high agency during the Suffragette's Movement (1910s-1920s) and the Women's Liberation Movement (1970s-1990s). The association of women with patiency has been more persistent, however, and actually increased over the 1800s. That has moderated over the last 100 years but women are still described as more patient-like than men in texts.

Second, we consider the mind perceptions of domesticated animals over the same time period. Historically, animals have scored quite low on the agency spectrum, and that has not changed at all over 200 years. For patiency, the trend is quite different. Starting in the late 1800s, domesticated animals have become increasingly patient-like in texts, a trend that continues today. These cultural perceptions about domesticated animals seem to have shifted long before the assignment of legal protections, suggesting that personhood concepts, as measured in language, may serve as harbingers of future legal and policy change.

The rest of the paper proceeds as follows. Section 2 provides a conceptual background on the dimensions of mind perception. Section 3 lays out our method for measuring those dimensions from natural language corpora. In Section 4, we validate the measures against human surveys. Section 5 reports the historical analysis, and Section 6 concludes.



Electronic copy available at: https://ssrn.com/abstract=3959847

## 2 Dimensions of Mind

An influential set of philosophical theories holds that personhood is fundamentally about having a *mind* (Chappell, 2011; Irwin, 1986). Persons are entities that have certain mental properties: the ability to feel pain, make plans, or act morally. Intuitively, this view seems to correctly identify persons. Are adult humans persons? Yes. Are rocks? No. Are dogs? Sort-of.

Personhood is closely connected with moral responsibility. To identify "persons" is to identify who we might hold responsible for their actions and who might be objects of moral concern. Since Aristotle, philosophers have distinguished between moral agents and moral patients (Olson, 2019; Irwin, 1986). Agents act and cause events, while patients experience, often as the objects of others' actions.

As a simple example, consider the story of King Kong. When Kong captures Ann and carries her up the Empire State Building, he is depicted as an agent. The story invites us to judge this choice as a bad one and to judge Kong as the morally responsible person, deserving condemnation and/or punishment. Conversely, Ann is presented as a moral patient; she is *being* carried. We are inclined to judge that she deserves better treatment or recompense for the attack she experiences.

The agency/patiency parallel also arises in law. When evaluating who should compensate whom in a multi-vehicle accident, law identifies both the set of agents (e.g., all drivers potentially contributing to the accident) and the set of patients (e.g., all injured persons who could potentially receive compensation). If an animal, or a weather event, had contributed to the accident, that could not imply legal responsibility because they are not understood as agents – persons who make moral decisions.[1] Similarly, if a wild animal had been injured in an accident, that would not imply compensation because it is not a patient – that is, it does not have the dimension of mind for experiencing a moral harm or injury.

On a one-dimensional view of personhood, agency and patiency go together. Some entities have capacities for both choice and experience (e.g. human adults), while other entities have neither (e.g. inanimate objects). Recent work from experimental psychology, however, has

[1] Although that has not always been true. There is a record of animals being tried and sued for alleged wrongdoings, as late as the nineteenth century (Girgen, 2003).



Electronic copy available at: https://ssrn.com/abstract=3959847

questioned the one-dimensional view. In particular, Gray et al. (2007) provide evidence that humans understand agency and patiency separately, with some entities having divergent perceptions. For example, a baby ranks high on the patiency axis, but low on the agency axis. A robot ranks high on agency but low on patiency.[2]

Thus, these studies conclude that the agency and patiency "dimensions of mind" can come apart. An important open question is what this means for law and for society. Sharpening this point, some parallel experimental evidence has shown that intuitions about agency and patiency inform moral judgement; one study found that entities seen as moral patients are less likely to be seen as capable of acting as agents (e.g. performing good or bad actions) (Gray and Wegner, 2009). More broadly, while personhood concepts are clearly relevant for a range of policy debates – e.g., animal rights (Sunstein and Nussbaum, 2005), corporate speech (Ellis, 2011), abortion (Little, 2008), and gender stereotypes (Hamilton, 1991) – those debates might be better informed by taking two-dimensional approach which disentangles agency and patiency.

To understand dimensions of mind in real-world contexts, the previous approaches from experimental psychology have some limitations. Those studies are often conducted on just a small sample of individuals, generally college students in Western universities. Even broadening to a more representative sample of individuals would be limited, for example because survey participants may feel social pressure to report socially acceptable answers – such as gender-equitable attributions of agency and patiency.[3] Finally, even if these problems were minor, the survey approach would still be ineffective for examining agency/patiency perceptions in the past.

Our approach addresses these limitations through analysis of personhood concepts in natural language. Rather than build representative samples of survey respondents, we build representative corpora of documents. These documents contain expressed representations of the attitudes of a given social milieu (Kozlowski et al., 2019) – here, English-language discourse in the United States. As argued in recent related work using similar methods,

[2]Similarly, Strohminger and Jordan (2021) focuses on corporations, which are considered legal "persons" in some jurisdictions (Ellis, 2011). Yet in the surveys, corporations are evaluated to have low patiency and therefore low personhood.

[3]Even incentive-compatible surveys, which reward participants for accurately estimating the answers of others (Prelac, 2004), might not fully capture implicit attitudes.



Electronic copy available at: https://ssrn.com/abstract=3959847

natural language can reveal implicit attitudes expressed in corpora, similar to those revealed by implicit association tests (Caliskan et al., 2017; Garg et al., 2018). Hence, our language measures can address the issue of distortions in survey responses arising from perceived social pressure. Finally, unlike surveys, which cannot be given to people in the past, our natural language tools can be transported back as far as documents are available. Thus, we can quantify 200 years of agency and patiency perceptions. The quantitative historical results can provide invaluable context to ongoing policy debates on the expansion or restriction of moral rights and obligations.

# 3 Measuring Agency and Patiency Perceptions in Natural Language

This section describes how we measure the agency and patiency dimensions of personhood from text. The starting point is a technology from computational linguistics called word embedding. Word embedding refers to a family of algorithms which assigns to each word in a vocabulary a low-dimensional dense vector, where the direction of the vector in geometric space encodes semantic and syntactic information. Words with similar meaning tend to co-locate in this space, and linear directions encode analogical concepts. The canonical example is that the operation *king* - *man* + *woman* produces a vector close to the word *queen*. Recent work has used word embeddings to examine cultural associations and categories on a large scale (Kozlowski et al., 2019), to analyze the emotionality and rationality of speeches (Gennaro and Ash, 2021), and to discover bias in language (Caliskan et al., 2017).

There are a number of algorithms used to train word embeddings – e.g. Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014), and FastText (Bojanowski et al., 2016). The core intuition behind these algorithms is the same, summarized by the "distributional hypothesis" that "a word is characterized by the company it keeps" (Firth, 1957). Computationally, this means that word embedding training objectives work by sampling words from large collections of text, and then learning vectors that predict a word's neigh-



Electronic copy available at: https://ssrn.com/abstract=3959847

Table 1: Word Lists for Agency and Patiency Dimensions

| Dimension | Word List |
|---|---|
| Agency | change cause aim intentional harm help decide control moral plan communicate think choose deliberate create guilty responsible act do express |
| Patiency | pain pleasure enraged afraid hunger thirsty sad proud embarrassed joy angry happy conscious aware experience imagine awake suffer enjoy desire |

boring context.[4] As a result, words appearing in similar contexts tend to have vectors with a high dot product and high cosine similarity. Appendix A provides a run-down of all of the embedding approaches used. The trained embedding model provides a vector for each word observed in the pre-training corpus.

We now use the embeddings to map the two dimensions of mind. For both agency and patiency, we construct semantic dimensions using twenty representative words, selected based on the previous psychology literature on personhood and mind concepts (Gray et al., 2007; Strohminger and Jordan, 2021). These pre-registered lists of words are shown in Table 1.[5] To get an agency vector $\vec{a}$ and a patiency vector $\vec{p}$, we average across the vectors of all words in the respective category.

To gain a first qualitative sense of what is captured by the constructed agency and patiency dimensions, we ranked each word $w$ in the vocabulary by its proximity to both poles. Proximity is measured using the standard approach – cosine similarity of the associated vector $\vec{w}$ to each vector pole $\vec{a}$ and $\vec{p}$.[6] Appendix Figure B1 shows the most similar words to

[4]Our baseline specifications uses the context window and vector dimensionality hyperparameters from the pre-trained models (see Appendix A). We show that changing these hyperparameters does not affect the quality of the embeddings according to our human survey validation (Appendix Table D10).

[5]For a dictionary-based approach to measuring agency and patiency, see Schweitzer and Waytz (2020). That paper provides a longer list of agency-related and patiency-related words, which we will use in a robustness check (Appendix Table D13). As with the previous papers, we choose patiency words pertaining to the ability to experience, rather than having rights or deserving care (moral patiency). Thus we implicitly assume that experience and moral patiency go together.

[6]Cosine similarity is defined as $\text{sim}(\vec{u}, \vec{v}) = (\vec{u} \cdot \vec{v}/(||\vec{u}||||\vec{v}||)$, the dot product between two vectors scaled by the magnitude of the vectors (see e.g. Levy et al., 2015). Our results are similar using the non-normalized



Electronic copy available at: https://ssrn.com/abstract=3959847

each pole. The resulting words are quite intuitive, with the agency words corresponding to actions, desires, and morality, and the patiency words corresponding to feelings, knowledge, and experiences.

We use the same approach to measure how a given entity $e$ scores along the agency and patiency dimensions in language. The linguistic agency score $a_L(e)$ of an entity $e$ in our framework ($L$ is for "language model") is the cosine similarity between the entity's vector $\vec{e}$ and the agency vector $\vec{a}$. Correspondingly, the patiency score $p_L(e)$ is the cosine similarity with the patiency vector $\vec{p}$. Besides the individual scores, we are also interested in the sum of these measures $a_L(e) + p_L(e)$, providing a single combined personhood score for $e$, as well as the difference $a_L(e) - p_L(e)$, giving a relative agency score.

Additional information and summary statistics on the language measurements are shown in Appendix B. Appendix Figure B2 shows histograms for each variable. Appendix Table B2 lists the top- and bottom-ranked entities along all four linguistic scores. For example, `computer` and `office` tend to have high agency and low patiency, while `puppy` and `ghost` tend to have high patiency and low agency (see also Appendix Table D14). Thus, similar to the survey measures in Gray et al. (2007), the two text-based mind dimensions can come apart and vary separately. We can see this statistically in Appendix Table B3, showing pair-wise correlations between the language variables. Agency and patiency are strongly correlated, but far from perfectly so. For example, for the GloVe embeddings, the correlation between agency and patiency is just 0.28 in the set of 255 entities.

# 4 Validation Against Human Judgment

The next step is to validate whether our language-based measure of agency and patiency matches up with human judgments. Building on recent work in experimental social psychology (Gray et al., 2007; Strohminger and Jordan, 2021), we developed a survey instrument asking respondents to implicitly score entities along an agency axis and a patiency axis.[7]

dot product as our similarity metric.

[7]We pre-registered our survey materials, including comprehension check questions and exclusion criteria, and our method of creating agency and patiency measures from the survey data (https://osf.io/e7x39).



Electronic copy available at: https://ssrn.com/abstract=3959847

**Survey Design.** Each survey participant was randomly assigned to one "property" condition. There were forty primary property conditions, corresponding to the twenty agency words (e.g. "act") and twenty patiency words (e.g. "hunger"). A participant was asked to evaluate a randomly assigned collection of entities along the assigned property, on a scale from 1 (low) to 10 (high). For example, a participant assigned to the hunger condition would assess whether various entities – e.g., frogs, men, and robots – are low or high in the capacity to hunger. For each participant, 100 entities were randomly sampled from a larger list of 255 entities, which were drawn from existing surveys (Gray et al., 2007; Strohminger and Jordan, 2021) supplemented with an online list of the "most important English nouns" (Co., 2019).[8]

We initially aimed to recruit 3510 participants. Following our pre-registered analysis plan, we excluded from all analyses any participant who: (a) failed any of the comprehension check questions, (b) failed the CAPTCHA, or (c) sped through the entire study in less than two minutes. We also pre-registered that if those exclusions resulted in a loss of over 700 participants (20 percent), we would recruit an additional set of participants, equal to the number excluded. After the initial data collection, 1023 participants failed the comprehension criteria, so we aimed to recruit an additional 1023 participants to take the same survey and pooled the successful responses with the original survey. In this second round, 1005 participants were successfully recruited and 311 participants were excluded based on the same criteria, so the final sample had 3181 responses.

**Survey Measurements of Agency and Patiency Perceptions.** We produced the following pre-registered scores for entities based on the survey responses. For a given entity, the baseline agency score $a_H(e)$ is the average of the survey ratings for that entity across all agency properties ($H$ is for "human survey"). Similarly, a baseline patiency score $p_H(e)$ is the average of the ratings for that entity across all patiency properties.

[8] A number of words in this noun list are polysemous and contain relatively common verb senses: sail, fish, drop, face, stem, line, stick, pin, brush, head, ring, dress, lock, spring, picture, drain, board, house, watch, hand, match, book, ship, store, train. In the survey, it is clear we are referring to the noun. But our language model mixes up the noun and verb form. What this means in practice is that the polysemous words score artificially high in agency. Following the pre-analysis plan, our main results include these nouns. When we exclude them, the validation results improve (Appendix Table D16). None of these words are used in the historical analysis.



Electronic copy available at: https://ssrn.com/abstract=3959847

Second, we computed a set of adjusted scores, which were designed to make the different properties more comparable. We took each property (e.g. "feeling hunger") and normalized the ratings for that property across all responses by subtracting mean and dividing by the standard deviation. Then the agency score for an entity is the average across all responses for that entity of the normalized agency scores. Likewise, the adjusted patiency metric is the average of the property-normalized responses for an entity. We found that the baseline scores and adjusted scores are highly correlated and provide similar validation results.

As done with the linguistic scores, we are interested in the sum of the human scores $a_H(e) + p_H(e)$, as well as their difference $a_H(e) - p_H(e)$. Additional detail and summary statistics for the human-survey measures are recorded in Appendix C. Distributions of the ratings are shown in Appendix Figure C3. Appendix Table C5 shows the top and bottom-ranked entities along the survey ratings, and the pairwise correlations are shown in Appendix Table C6. The adjusted and unadjusted survey ratings produce almost identical rankings of the entities.

**Comparing Language-Model and Human-Survey Measurements.** Now we use the human-survey ratings for agency and patiency to validate the corresponding similarity scores generated using the language model. For a given set of entities $e \in E$, we have the language model measurements $L_e$ and the human survey measurements $H_e$. The scales of the language and human metrics are not comparable are not very meaningful on their own, so we compare them by the rankings of entities produced by each scale. Formally, we compute the Spearman's rank correlation coefficient $\rho(L_e, H_e)$, which is bounded between $-1$ (perfect anti-correlation of ranks) and 1 (perfect correlation of ranks). For statistical significance, we report a p-value from a t-test of the null hypothesis that the variables are uncorrelated (Kokoska and Zwillinger, 2000).

Besides the full list of 255 entities assessed in the survey, we compute correlations for pre-registered entity subsets designed to be representative of the different agency and patiency concepts from the literature (Gray et al., 2007). In the main text, we report results for the smallest list of 31 entities for greater interpretability. In the appendix we report results for the full set of 255 entities, as well as for an intermediate list (also pre-registered) of 58



Electronic copy available at: https://ssrn.com/abstract=3959847

entities. The full list of entities with group assignment is shown in Appendix Table B1.

Figure 1 reports our main validation results for the preferred measures of agency and patiency from the survey and from the language model.[9] The figure shows scatter plots for the human and language measures of agency (Panel A) and patiency (Panel B), including a line of best fit. In both graphs, we see a clear positive relationship. According to the rank correlation, the language measures are highly predictive of the human measures, with coefficients of 0.58 (agency) and 0.83 (patiency). They are both highly statistically significant with $p < 0.005$.

Appendix Table D8 reports correlation statistics for a number of alternative specifications, holding the training corpus constant. For example, when including the full set of 255 entities, the GloVe language model reproduces human agency judgments with a correlation of 0.4 and patiency judgments with a correlation of 0.36 (both highly statistically significant). We obtain similar rank correlations when using the unadjusted human-survey scores that do not normalize across properties. Next, the similarity rankings produced by other embedding algorithms (Word2Vec and FastText) are quite good in reproducing human judgments, but not as good as GloVe.

In Appendix Table D9, we see that the quality of embeddings depends on the pre-training material. Embeddings trained on a broader, less-well-curated corpus (Common Crawl) get worse performance in the human validation. Performance with our newly trained embeddings using recent decades of the Corpus of Historical American English (Davies, 2010) is similar to the baseline using pre-trained models. Holding the corpus constant, Appendix Table D10 shows that varying the main hyperparameters for the embedding models – window size and vector dimensionality – has a relatively mild influence on the validation performance. There is some indication that shorter context-window sizes produce slightly better agency/patiency rankings.

Besides the individual agency and patiency scores, we would also like to check whether the language model can capture sums and differences of the survey's agency and patiency

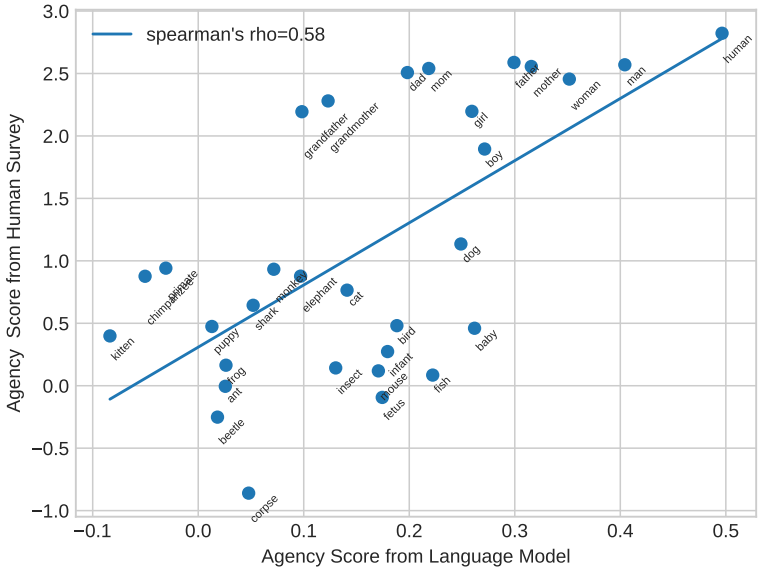
[9] In the language model scores, our preferred specification uses GloVe embeddings pre-trained on Wikipedia and Gigaword, with dimensions constructed using our pre-registered list of 20 agency words and 20 patiency words. In the human survey scores, our preferred specification uses the normalized scores that adjust for variation across properties.



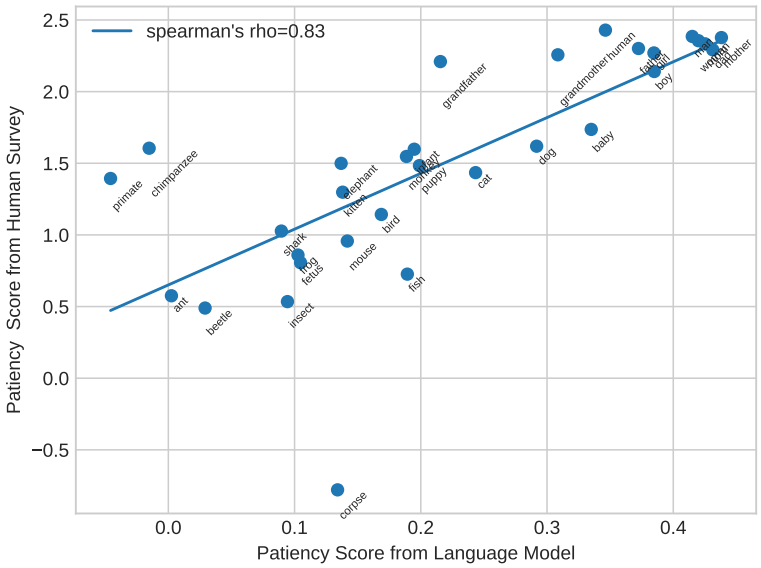
Electronic copy available at: https://ssrn.com/abstract=3959847

Figure 1: Comparison of Language and Survey Measures

A. Agency




B. Patiency




Notes. Scatter plots of Agency and Patiency scores, with human survey measures on the vertical axis and language measures on the horizontal axis. Plots include the pre-registered list of 31 entities, indicated by marker labels. Blue line is linear best fit.



Electronic copy available at: https://ssrn.com/abstract=3959847

scores. We find that they can. Appendix Table D11 shows that the unadjusted sum of the survey scores has a correlation of 0.38 with the unadjusted sum of the language scores. The corresponding survey score difference and language score difference have a correlation of 0.41, using our preferred setting with GloVe embeddings.

For robustness, we show that these results are not driven by our survey design choices, as the language approach can reproduce the results of two other surveys on agency and patiency (Gray et al., 2007; Strohminger and Jordan, 2021). From these previous surveys we take lists of agency and patiency words, lists of entities, and associated entity ratings for agency and patiency. In both of these datasets, our word-embedding associations substantially recover the human-survey associations (Appendix Table D12). Further, our validation results are even better when constructing the personhood dimensions with longer lists of agency and patiency words from Schweitzer and Waytz (2020) (Appendix Table D13).

To summarize, statistical language associations about personhood concepts expressed in corpora reflect human associations expressed in surveys. The measurements produced by a language model capture psychological semantic concepts of agency and patiency, as well as their sum and difference. Thus we gain confidence that the resulting language metrics can be used in empirical analysis. These results illustrate the importance of human validation in constructing linguistic measures of psychological concepts such as agency and patiency.

# 5 Historical Analysis of Mind Dimensions

Linguistic measures of agency and patiency dimensions provide a way for social scientists to address important but unanswered empirical questions about cultural attitudes towards different entities. For example, quantitative analysis of historical corpora can illuminate changes in people's agency and patiency perceptions over time: Are men described as more "agent-like" than women? If so, has that difference narrowed over time? Psychological survey methods can provide insight into contemporary attitudes, but they typically cannot address such historical questions. In this section, we analyze the historical evolution of mind dimensions over the last two centuries.



Electronic copy available at: https://ssrn.com/abstract=3959847

**Training Historical Embeddings.** The starting point is the Corpus of Historical American English (COHA) (Davies, 2010). COHA contains 141,000 documents (475 million words of text) from the 1820s through the 2010s. It is designed to be balanced by genre (fiction, magazine, news, academic, and other non-fiction) over time.[10] It is over 50 times larger than other comparable historical corpora of American English.

This corpus is ideal for building diachronic word embeddings that can evolve over time. Following the method in Hamilton et al. (2016), we trained a separate set of word embeddings for each decade. To reduce noise and make similarity measures more comparable over time, we then sequentially align each embedding matrix with the one from the previous decade.[11] We maintain a consistent vocabulary, where words are required to appear at least ten times in each decade.

Word embedding vectors, and the resulting similarity scores, are statistical estimates and therefore potentially sensitive to sampling. To allow for uncertainty in the resulting vectors and measurements, we use a bootstrapped approach to training based on Antoniak and Mimno (2018). Specifically, we trained embeddings 10 times for each decade. For each decade's corpus of $n$ sentences, $n$ sentences were sampled with replacement at each bootstrap iteration. Sequential across-decade alignment was performed in parallel within each bootstrap iteration.

**Measuring Agency and Patiency Over Time.** We compute similarity scores for each decade and each bootstrapped model, following the formulas discussed in Section 3. With 10 different embedding matrices for each decade, we assign 10 agency scores and 10 patiency scores for each entity per decade.

We construct entity-level measures as the mean and standard error across the 10 scores. We use these statistics to test for statistical differences between entities. To compare groups

[10] Unlike Google Books, for example, which would be biased by changing compositions of genres over time, such as increasing prevalence of scientific documents (Pechenick et al., 2015). For robustness, we made our time series graphs separately by genre. The results are much noisier but show that the long-run trends hold across genres. They are not driven by a single genre, nor by changes in composition of genres.

[11] More specifically, the 1830s model is aligned to the 1820s model, and then the 1840s model is aligned to the already-aligned 1830s model, and so on. Alignment is performed using Procrustes matrix alignment (Hamilton et al., 2016). We also tried the suggestion by Hamilton et al. (2016) of training embeddings incrementally, with the first decade trained from scratch and each subsequent decade being fine-tuned from the previous decade's. That did not change the results.



Electronic copy available at: https://ssrn.com/abstract=3959847

of entities, we create group averages of the entity-level means for each of the 10 bootstraps, then compute the differences in these entity-group means. For statistical inference, we construct 95% confidence intervals using the mean and standard error of the 10 between-group differences.

**Historical Gender Differences in Agency and Patiency Perceptions.** Our first historical analysis assesses variation in expressed attitudes about the agency and patiency of men and women over the last 200 years. This is an interesting period to explore the relationship between language-usage and conventional narratives of women's rights. For example, scholars describe several "waves" of feminist thought and activism. According to that traditional narrative, the first wave emerged in the mid-1800s, with the Seneca Falls Convention (1848), and continued through the early 1900s. The traditional story presents a large gap before the second wave, which begins in the 1960s to 1990s, although other scholars contest that gap (Rupp and Taylor, 1987). A "third wave" is often identified in the early 1990s to present, and some have suggested an emerging "fourth wave" today (Rampton, 2008).

For this analysis, we are interested in the gender difference in agency, as well as in patiency. We produce the measures using gendered words from our survey lexicon – *man*, *boy*, *father*, *dad*, and *grandfather* for male; *woman*, *girl*, *mother*, *mom*, and *grandmother* for female. As discussed above, we produced the mean group differences by decade for each bootstrap sample. We then plot the gender differences over time with 95% confidence intervals constructed using the standard error of the mean.

The gender differences in agency and patiency perceptions are reported in Figure 2 Panel A. We find that in the earliest period (1820s-1850s), men were described as having higher agency and lower patiency than women. Already in this early period, the relative male association with agency was shrinking, and there wasn't much of a gender difference in agency by the late 1800s. Interestingly, around the time of the Suffragette's Movement and first-wave feminism, where women claimed the right to vote, the gender difference in agency briefly flips sign (although not statistically significantly). Subsequently, there is something of a gender-norms retrenchment where men again have higher agency in the 1940s-1960s, consistent with the traditional narrative in which that period was devoid of feminist activism



Electronic copy available at: https://ssrn.com/abstract=3959847

(Rupp and Taylor, 1987). But then these perceptions flip again in the heyday of second-wave feminism and the Women's Liberation Movement – women actually had statistically higher agency in the 1990s, before reverting to no difference at all in the 2010s.

Meanwhile, the relative female association with patiency actually intensified in the early years, with the largest differences from 1870 through 1920. Since then, the relative female association with patiency has gradually moderated toward zero. While it appears to still be shrinking, relative female patiency has not yet disappeared. That result suggests there is still an operational stereotype in how women are represented as patients in texts.

To summarize, gender representations in terms of agency and patiency have become more equal over the last two centuries. These results provide a long-run quantitative view on gender perceptions in the United States, adding to the literature on long-run progress toward social and legal gender equality (England et al., 2020). Reflecting changes in women's roles and in their relative power in politics and the economy, women have become less disproportionately represented in texts as experiencers and are today portrayed equally with men as actors.

This pair of findings (compared to men, women are portrayed as equally agential, but still more patientic) resonates with modern feminist debates. For example, Anita Harris describes the clash between "power feminism" and "victim feminism" (Harris, 2011). The equal agency result is suggestive of natural language representations of power equality: today, women are described as active, deciding, powerful agents at rates similar to men. Yet the persistent patiency gap suggests the influence of some old gender stereotypes in written texts: women are more often described as feeling, understanding, or scared experiencers.

Of course, there could be many legal, cultural, and other factors contributing to this long-run narrowing of both the agential and patientic gender gaps. One interesting possibility is that there are likely more female writers in the modern sample. Unfortunately, COHA does not have authorship information. Analyzing the mechanisms of this change is an important area for future work.

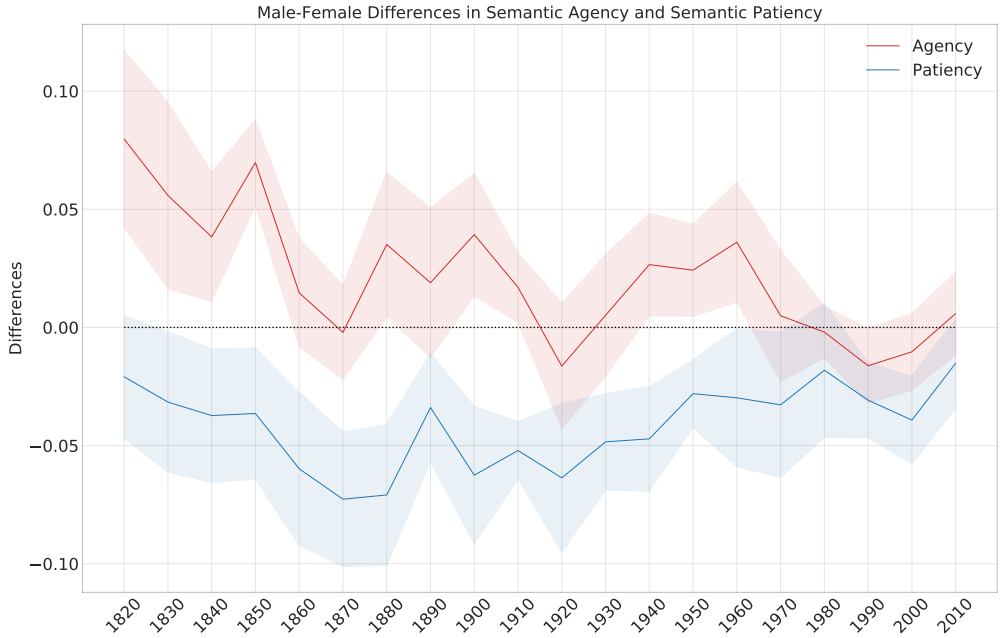
**Evolution of Agency and Patiency Perceptions for Domesticated Animals.** Our second historical analysis considers the perception of domesticated animals' agency and patiency. How animals have been understood over time can enrich our understanding of at-



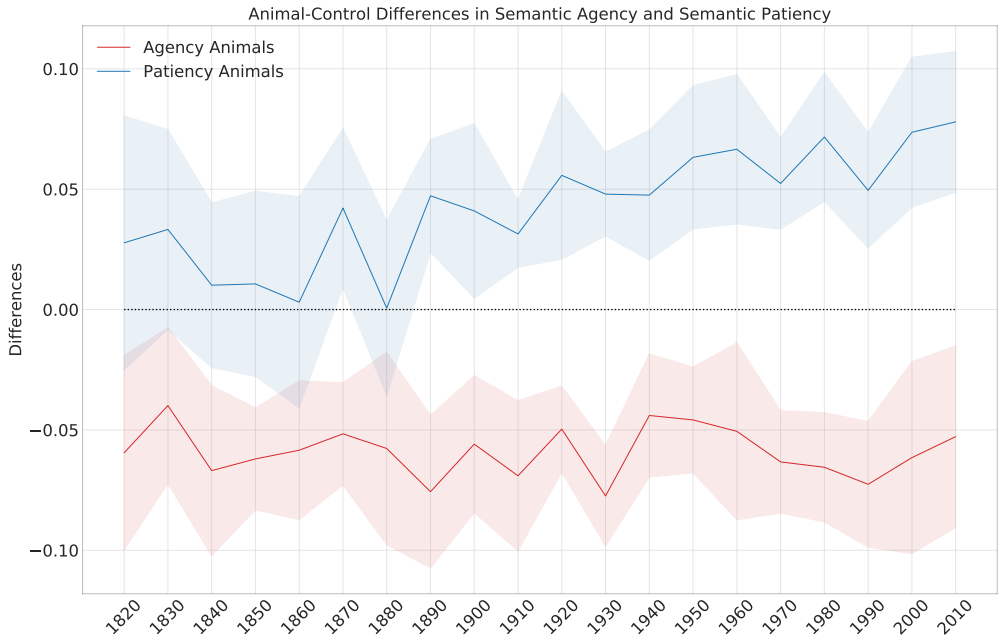
Electronic copy available at: https://ssrn.com/abstract=3959847

Figure 2: Historical Analysis: 200 Years of Agency and Patiency Perceptions

A. Gender Differences in Agency and Patiency Perceptions




B. Agency and Patiency Perceptions about Domesticated Animals




Notes. Time series of agency and patiency differences, 1920s-2010s. Panel A shows the difference by gender (between a set of male words and set of female words). Panel B shows the difference for domesticated animals (a set of words referring to domesticated animals) and a list of control-group words. Agency is shown in red, while patiency is shown in blue. Confidence intervals (95%) produced by bootstrapped embeddings, as described in the text.



Electronic copy available at: https://ssrn.com/abstract=3959847

titudes about animals and what protections they deserve. Animals' agency (e.g. ability to do wrong) and patiency (e.g. feeling pain) have been central to cultural, moral and legal debates about animal rights (Singer, 1973; Regan, 1983).

The history of animal rights in the United States is often traced to 1866, when Henry Bergh founded the American Society for the Prevention of Cruelty to Animals (Zurlo and Goldberg, 1994). An important wave of social and legal activism took place over the next forty years. In 1892, the American Humane Association lobbied to prohibit unnecessary animal experiments. And between 1896 and 1900, federal legislation to inspect animal laboratories was (twice) proposed and defeated (Zurlo and Goldberg, 1994). A long second wave of animal rights activism began in the 1950s (Congress, 1958) and continues through today (Zurlo and Goldberg, 1994; Regan, 1983; Singer, 1973).

As in the gender analyses, we produce relative measures of agency and patiency for the set of domesticated animals in our entity list: *cat*, *dog*, *cow*, *pig*, *kitten*, *lamb*, *sheep*, *horse*, and *goat*. As the comparison entity group, we use all other entities in our survey lexicon, excluding any other animals. As before, we plot the animal-control differences in agency and patiency over time with 95% confidence intervals.

Figure 2 Panel B shows the results. In red, we see that agency for these entities has been more-or-less constant over the last 200 years. For patiency, however, there has been a steady increase for animals, relative to control entities, starting around the end of the nineteenth century. One large jump occurs between 1860 and 1890, mirroring social and legal efforts conducted by the American Society for the Prevention of Cruelty to Animals and American Humane Association. That upward trend has continued through today. As expressed in representative historical corpora, domesticated animals are now described to have greater capacity to feel and experience.

It is perhaps historically dissonant that empathy for the experience of animals increased in parallel with the development of industrialized mass animal farming in the early 1900s. In particular, the increasing patiency language for animals preceded major political action by decades. In the United States, it was not until 1955 that the first major organization lobbied for humane slaughter legislation and not until 1958 that the first animal welfare legislation was passed (Congress, 1958). Hence, the cultural attitudes revealed by our method do not



Electronic copy available at: https://ssrn.com/abstract=3959847

necessarily reflect policy priorities.

# 6 Conclusion

We find that the distinction between agency and patiency dimensions of mind – a subtle feature in human judgment – is also encoded in natural language. Just as surveys show that these dimensions are evaluated separately by human subjects, so does our language analysis show that they are encoded separately in written texts. This convergent evidence from naturally occurring language should increase our confidence that people do in fact perceive agency and patiency as separate dimensions of mind.

The natural language analysis demonstrates that these agency and patiency concepts are not just constant features of psychology. They are socially and culturally embedded quantities that can vary over time and across cultures. In particular, our historical analysis shows that these agency and patiency dimensions are not merely a product of today's culture; they evolve over two centuries of human writing. Substantively, the analysis has enriched our understanding of gender and personhood attributions.

In particular, using two dimensions of mind (rather than one) clarifies important nuances in the historical evolution of both gender norms and animal rights. In both analyses, agency and patiency moved independently and in historically meaningful ways. A two-dimensional model of personhood concepts was needed to reveal these interesting patterns, and a one-dimensional model would have missed them.

Our quantitative analysis of natural language can be seen as a methodological proof of concept. We can take the dimensions-of-mind theory into the real world. With these linguistics tools in hand, researchers are no longer limited to gauging attitudes of today's students at Western universities. We have already shown how the method could be applied across time. But it could also be applied across space, for example to compare perceptions in different countries or regions or languages. Finally, personhood perceptions might vary along other social or cultural dimensions, such as politics or religion. Relative to surveys, measures from naturally occurring language may be less susceptible to distortion by social pressures (Prelac, 2004). These are important areas for future work.



Electronic copy available at: https://ssrn.com/abstract=3959847

From a policy standpoint, a notable feature of the animal-rights analysis is that the change in mind perceptions significantly preceded changes in animal-rights legislation. This finding raises the possibility that today's naturally occurring language might help predict tomorrow's social and political changes. Insofar as social attitudes enable social movements or political actions, natural language can reveal the most promising areas for organization by providing a window into people's latent attitudes. As one example for future study, consider the case of machines, algorithms, and artificially intelligent entities (Chuan et al., 2019). Beyond surveying people, analyzing the naturally occurring language, including agency and patiency scores, for these terms could help clarify subtle social attitudes and/or predict the direction of social and political change.



Electronic copy available at: https://ssrn.com/abstract=3959847

# References


Alatrash, R., D. Schlechtweg, J. Kuhn, and S. Schulte im Walde (2020, May). CCOHA: Clean corpus of historical American English. In *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille, France, pp. 6958–6966. European Language Resources Association.

Antoniak, M. and D. Mimno (2018). Evaluating the stability of embedding-based word similarities. *TACL 6*, 107–119.

Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov (2016). Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Brunet, M.-E., C. Alkalay-Houlihan, A. Anderson, and R. Zemel (2019, 09–15 Jun). Understanding the origins of bias in word embeddings. In K. Chaudhuri and R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning*, Volume 97 of *Proceedings of Machine Learning Research*, Long Beach, California, USA, pp. 803–811. PMLR.

Caliskan, A., J. J. Bryson, and A. Narayanan (2017). Semantics derived automatically from language corpora contain human-like biases. *Science 356*(6334), 183–186.

Chappell, T. (2011). On the very idea of criteria for personhood. *Southern Journal of Philosophy 49*(1), 1–27.

Chuan, C.-H., W.-H. S. Tsai, and S. Y. Cho (2019). Framing artificial intelligence in american newspapers. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 339–344.

Co., T. (2019). Learn 600 of the most important english nouns.

Congress (1958). Humane slaughter act.

Davies, M. (2010). The corpus of historical american english (coha).

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language 67*, 547–619.

Ellis, A. (2011, 11). Citizens united and tiered personhood. *The John Marshall law review 44*.

England, P., A. Levine, and E. Mishel (2020). Progress toward gender equality in the united states has slowed or stalled. *PNAS 117*.

Epley, N., A. Waytz, and J. Cacioppo (2007, 11). On seeing human: A three-factor theory of anthropomorphism. *Psychological review 114*, 864–86.

Firth, J. R. (1957). *A synopsis of linguistic theory*.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences 115*(16),




Electronic copy available at: https://ssrn.com/abstract=3959847

E3635–E3644.

Gennaro, G. and E. Ash (2021). Emotion and reason in political language. *Center for Law & Economics Working Paper Series 2021*(02).

Girgen, J. (2003). The historical and contemporary prosecution and punishment of animals. *Animal L. 9*, 97.

Gordon, G. J. (2019, Nov.). Environmental personhood. *Columbia Journal of Environmental Law 43*(1).

Grave, E., P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov (2018). Learning word vectors for 157 languages.

Gray, H., K. Gray, and D. Wegner (2007, 03). Dimensions of mind perception. *Science (New York, N.Y.) 315*, 619.

Gray, K., J. Knobe, M. Sheskin, P. Bloom, and L. Barrett (2011, 11). More than a body: Mind perception and the nature of objectification. *Journal of personality and social psychology 101*, 1207–20.

Gray, K. and D. M. Wegner (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology 96*(3), 505–520.

Hamilton, M. C. (1991). Masculine bias in the attribution of personhood: People = male, male = people. *Psychology of Women Quarterly 15*(1), 393–402.

Hamilton, W. L., J. Leskovec, and D. Jurafsky (2016, August). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, pp. 1489–1501. Association for Computational Linguistics.

Harris, A. (2011). Not waving or drowning: Young women, feminism, and the limits of the next wave debate. *Outskirts*.

Harris, L. T. and S. T. Fiske (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science 17*(10), 847–853. PMID: 17100784.

Haslam, N. (2006, 02). Dehumanization: An integrative review. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc 10*, 252–64.

Haslam, N. and S. Loughnan (2014). Dehumanization and infrahumanization. *Annual Review of Psychology 65*(1), 399–423. PMID: 23808915.

Irwin, T. H. (1986). Aristotelian actions. *Phronesis 31*(1), 68–89.



Electronic copy available at: https://ssrn.com/abstract=3959847

Kokoska, S. and D. Zwillinger (2000). *CRC standard probability and statistics tables and formulae*. Crc Press.

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019, Sep). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review 84*(5), 905–949.

Levy, O., Y. Goldberg, and I. Dagan (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the association for computational linguistics 3*, 211–225.

Little, M. O. (2008). Abortion and the margins of personhood. *Rutgers Law Journal 39*, 331–348.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, Volume 26, pp. 3111–3119. Curran Associates, Inc.

Olson, E. T. (2019). Personal Identity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.). Metaphysics Research Lab, Stanford University.

Pechenick, E. A., C. M. Danforth, and P. S. Dodds (2015). Characterizing the google books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PloS one 10*(10), e0137041.

Pennington, J., R. Socher, and C. D. Manning (2014). Glove: Global vectors for word representation. In *In EMNLP*.

Prelac, D. (2004). A bayesian truth serum for subjective data. *Science 306*.

Rampton, M. (2008). Four waves of feminism. *Pacific Magazine*.

Regan, T. (1983). *The Case for Animal Rights*.

Rehurek, R. and P. Sojka (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 3*(2).

Resinger, D., R. Rudinger, F. Ferraro, C. Harman, K. Rawlins, and B. Van Durme (2015). Semantic proto-roles. *Transactions of the Association for Computational Linguistics 3*, 475–488.

Rupp, L. and V. Taylor (1987). *Survival in the Doldrums: The American Women's Rights Movement, 1945 to the 1960s*.

Schweitzer, S. and A. Waytz (2020). Language as a window into mind perception: How mental state language differentiates body and mind, human and nonhuman, and the self from others. *Journal of Experimental Psychology: General*.

Singer, P. (1973). *Animal Liberation*.

Strohminger, N. and M. Jordan (2021). Corporate insecthood.



Electronic copy available at: https://ssrn.com/abstract=3959847

Sunstein, C. R. and M. C. Nussbaum (2005). *Animal Rights: Current Debates and New Directions*. Oxford University Press.

Tarasenko-Struc, A. (2020, 10). On Inhumanity: Dehumanization and How to Resist It. *The Philosophical Quarterly*. pqaa068.

Willke, J. and D. Andrusko (1988). Personhood redux. *Hastings Center Report 18*(5), 30–33.

Zurlo, R. and A. Goldberg (1994). *Animals and Alternatives in Testing: History, Science, and Ethics*.



Electronic copy available at: https://ssrn.com/abstract=3959847

# Appendices

## A Training Word Embedding Models

For the analysis in this paper, we use word embeddings obtained by using different algorithms and experiment with embeddings constructed using various plain-English corpora. Our preferred setting is GloVe embeddings pre-trained on 6B tokens from Wikipedia and GigaWord. We assess robustness to using other popular embedding algorithms by Mikolov et al. (2013) and Bojanowski et al. (2016). In the following, we describe the various embedding models in more detail.

**Previously trained models.** We considered the following available pre-trained word embedding models in our experiments.

1. GloVe GigaWord (Pennington et al., 2014) – These embeddings have been trained on the English Wikipedia (a dump from 2014) and GigaWord 5, an archive if news wire text. The resulting corpus has 6B tokens. All tokens are lowercased and a window size of 10 is used. The model provides embeddings for 400K words. We use the embeddings with size 300, which are available on the GloVe website[12] and via Gensim (Rehurek and Sojka, 2011). We downloaded the word vectors from Gensim.

2. GloVe CommonCrawl (Pennington et al., 2014) – These vectors are also trained with GloVe, but on the bigger Common Crawl corpus consisting of 840B cased tokens. This process results in vectors for 2.2M words. We only considered the lowercased version of words in our experiments. The model is also trained with a window size of 10 and vectors have size 300. We downloaded the embeddings from the GloVe website.[13]

3. Word2Vec GoogleNews (Mikolov et al., 2013) – We consider the word2vec embeddings trained with the skip-gram architecture which we downloaded from the Word2Vec website.[14] This model provides embeddings for 3M words and phrases, however we only

[12] `https://nlp.stanford.edu/projects/glove/`
[13] `https://nlp.stanford.edu/projects/glove/`
[14] `https://code.google.com/archive/p/word2vec/`



Electronic copy available at: https://ssrn.com/abstract=3959847

consider the lowercased versions of words. The embeddings are trained on part of Google News dataset (ca. 100B tokens) using a window size of 10 and each vector has size 300.

4. FastText CommonCrawl (Grave et al., 2018) – We also experiment with FastText embeddings pre-trained on Common Crawl and Wikipedia (ca. 840B tokens). We downloaded the model from the the FastText website.[15] These models were trained using the CBOW artchitecture with position weights, a window size of 5 and dimension of 300. While in principle close to word2vec, there are 2 important differences. First, the embeddings are trained on 8x more data than the Google skip-gram embeddings. Secondly, input to the model are not words, but character n-grams of length 5, and the objective is to predict neighboring n-grams in a window size of 5. The embedding of a word is then the aggregation of the character 5-grams appearing in the word.

**Newly trained models.** Additionally, we trained word vectors with GloVe, word2vec and FastText. We train the respective embeddings on a consistent corpus of 80M words from the Corpus of Historical American English, using documents for the years 2000-2019. Following Alatrash et al. (2020), we drop the `@` token, which is a placeholder in COHA due to copyright reasons.

The hyperparameters are indicated below and use standard choices from the algorithm defaults and associated applications. Appendix Table D10 shows that the main hyperparameter choices (window size, vector dimensionality) are not that important for capturing the agency and patiency dimensions.

1. Word2Vec COHA – We use the Gensim implementation[16] (Rehurek and Sojka, 2011) for training the Word2Vec models. The default parameters are the CBOW architecture, vectors of size 100 and the model is trained for 5 epochs. We increase the window size from 5 to 8 and increase the minimum count of a word to appear in the corpus from 5 to 10 to get more meaningful embeddings.

[15] `https://fasttext.cc/docs/en/crawl-vectors.html`
[16] `https://radimrehurek.com/gensim/models/word2vec.html`



Electronic copy available at: https://ssrn.com/abstract=3959847

2. GloVe COHA – We use the codebase in the GloVe github repository.[17] We use all the default parameters: vectors with size 100, a minimum count of 10 for a word to appear in the vocabulary and a window size of 15, and training the models for 15 epochs.

3. FastText COHA – We train our fasttext models using the codebase from the FastText github repository.[18] Again, we use default hyper-parameters provided by the authors, that is we use the skip-gram architecture, train vectors of size 100, set the minimum count of words to be included to 5. We use the recommemnded minimum and maximum length of character n-grams of 3 and 6, and the size of the context window is set to 5. The model is trained for 5 epochs.

[17] `https://github.com/stanfordnlp/GloVe`
[18] `https://github.com/facebookresearch/fastText`

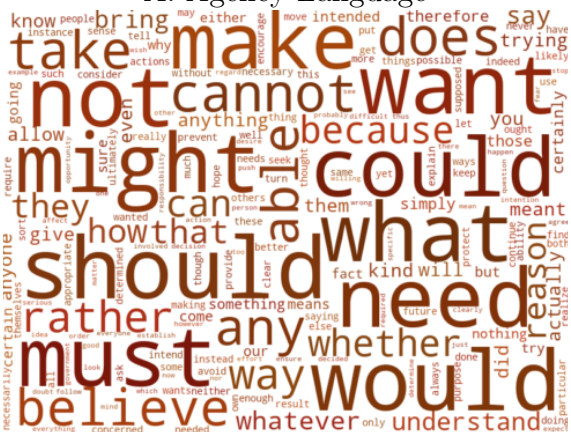


Electronic copy available at: https://ssrn.com/abstract=3959847

# B Language Measures

Figure B1: Semantic Poles for Agency and Patiency

A. Agency Language




B. Patiency Language




Most similar words to our agency vector (Panel A), and our patiency vector (Panel B). Embeddings computed using GloVe trained on Wikipedia and Gigaword.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table B1: List of Entities by Pre-Registered Group

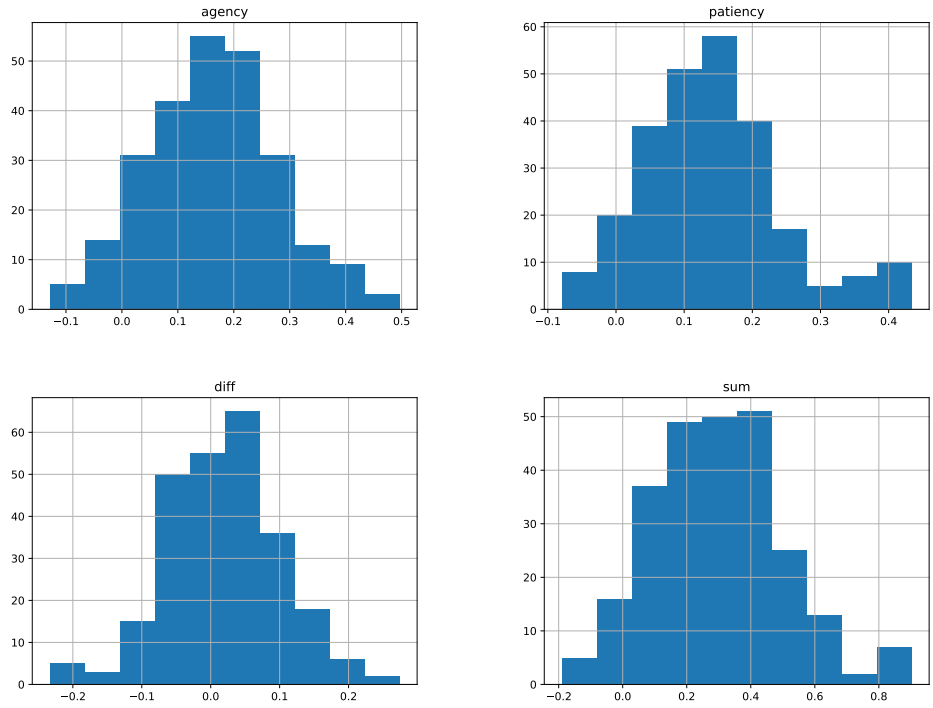
| | |
|---|---|
| Top 31 Entities | human man woman boy girl father mother dad mom grandfather grandmother baby infant fetus corpse dog puppy cat kitten frog ant fish mouse bird shark elephant beetle insect chimpanzee monkey primate |
| Top 58 Entities | human man woman boy girl father mother dad mom grandfather grandmother baby infant fetus corpse dog puppy cat kitten frog ant fish mouse bird shark elephant beetle insect chimpanzee monkey primate car rock hammer computer robot god angle ghost puppet pigeon chicken rabbit fox turkey pig cow horse sheep lamb cucumber lettuce potato cabbage chocolate coffee tea butter |
| All Entities | infant baby puppy fetus kitten rabbit worm pig mouse lamb sheep sail pigeon beetle fox cow chimpanzee fly rate frog chicken slaves turkey cat bird horse fish goat elephant monkey ant fowl snake dog stomach primate chin skin insect shark nose sponge lettuce drop toe face ghost tail leaf boy jews italians cabbage neck tooth root berry immigrants stem ear mexicans nut throat branch coat knee cheese cake apple orange potato circle catholics japanese trousers cushion rock button heart hat africans line frame corpse lip girl pocket hair drawer arabs stamp nail irishmen floor spade bee jewel cucumber feather basin tea indians box island thumb skirt butter puppet arch rod tongue muslims curtain wing collar grandfather stick egg ball plate bucket angle sock square muscle stocking shelf foot bottle cup bath receipt grandmother tray bell pot pin glove brush kettle british table boot head nerve basket ring spoon whip cart woman parcel ticket whistle card net thread tree dress shirt eye bag door screw coffee comb band lock station chess leg spring arm flag garden cord bone rail finger bulb picture drain shoe plough mother chocolate man chain fork carriage mom board mouth street dad hook umbrella brick horn house seed pipe cloud bed knot pump key window wall father watch star scissors pencil wire blade roof brain needle hand hammer moon human oven brake map boat bridge clock pen match book camera knife wheel town library prison office engine god army ship car farm store plane gun church train school robot sun computer hospital |

This table shows the full list of 255 entities for which agency and patiency ratings were assessed in the survey, and for which we produce agency and patiency language scores based on the word embeddings. The top 31 entities and top 58 entities are pre-registered subsets of entities based on the previous literature (Gray et al., 2007).



Electronic copy available at: https://ssrn.com/abstract=3959847

Figure B2: Distributions of Language Measures




Distributions of the language measure using GloVe embeddings trained on Wikipedia and the GigaWord corpus.

Table B2: Top- and Bottom-Ranked Entities by Language Scores

| Measure | Top 10 Entities | Bottom 10 Entities |
|---|---|---|
| agency | human face hand key man god heart board office house | tray chimpanzee sock lettuce spade cabbage plough kitten kettle cucumber |
| patiency | mother god dad watch face mom woman heart man boy | scissors comb spade parcel primate kettle cucumber tray plough basin |
| diff | key board map computer branch office stem rate umbrella line | mother girl bed ghost grandmother puppy stomach mom kitten dad |
| sum | face human god hand man heart watch woman mother father | drawer chimpanzee lettuce basin primate spade tray kettle plough cucumber |

Language scores obtained using GloVe embeddings trained on Wikipedia and the GigaWord corpus.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table B3: Pairwise Correlations: Language Measures

| | | Agency GloVe | Agency Word2Vec | Agency FastText | Patiency GloVe | Patiency Word2Vec | Patiency FastText |
|---|---|---|---|---|---|---|---|
| Top 31 Entities | Agency GloVe | 1.0 | 0.91 | 0.6 | 0.68 | 0.48 | 0.74 |
| | Agency Word2Vec | - | 1.0 | 0.54 | 0.78 | 0.31 | 0.76 |
| | Agency FastText | - | - | 1.0 | 0.72 | 0.55 | 0.53 |
| | Patiency GloVe | - | - | - | 1.0 | 0.3 | 0.72 |
| | Patiency Word2Vec | - | - | - | - | 1.0 | 0.61 |
| | Patiency FastText | - | - | - | - | - | 1.0 |
| | | Agency GloVe | Agency Word2Vec | Agency FastText | Patiency GloVe | Patiency Word2Vec | Patiency FastText |
| Whole Survey | Agency GloVe | 1.0 | 0.82 | 0.43 | 0.28 | 0.6 | 0.53 |
| | Agency Word2Vec | - | 1.0 | 0.31 | 0.59 | 0.42 | 0.73 |
| | Agency FastText | - | - | 1.0 | 0.46 | 0.49 | 0.34 |
| | Patiency GloVe | - | - | - | 1.0 | 0.19 | 0.64 |
| | Patiency Word2Vec | - | - | - | - | 1.0 | 0.64 |
| | Patiency FastText | - | - | - | - | - | 1.0 |

Spearman rank correlations between language scores. Language scores obtained using embeddings trained on the 2000 and 2010 decades of COHA.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table B4: Entities with Largest Agency-Patiency Difference in Language Score

| Entity | Rank Agency - Rank Patiency | Score Agency - Score Patiency |
|---|---|---|
| puppy | -13 | -0.19 |
| dad | -10 | -0.23 |
| kitten | -10 | -0.22 |
| mom | -8 | -0.21 |
| grandmother | -8 | -0.19 |
| grandfather | -6 | -0.12 |
| cat | -4 | -0.1 |
| monkey | -4 | -0.12 |
| mother | -3 | -0.12 |
| girl | -1 | -0.13 |
| corpse | -1 | -0.09 |
| boy | 0 | -0.11 |
| frog | 0 | -0.08 |
| chimpanzee | 0 | -0.04 |
| elephant | 1 | -0.04 |
| beetle | 1 | -0.01 |
| woman | 1 | -0.07 |
| primate | 2 | 0.02 |
| infant | 2 | -0.02 |
| man | 3 | -0.01 |
| dog | 3 | -0.04 |
| father | 3 | -0.07 |
| ant | 3 | 0.02 |
| baby | 3 | -0.07 |
| mouse | 4 | 0.03 |
| shark | 4 | -0.04 |
| bird | 6 | 0.02 |
| fish | 7 | 0.03 |
| insect | 8 | 0.04 |
| human | 8 | 0.15 |
| fetus | 9 | 0.07 |

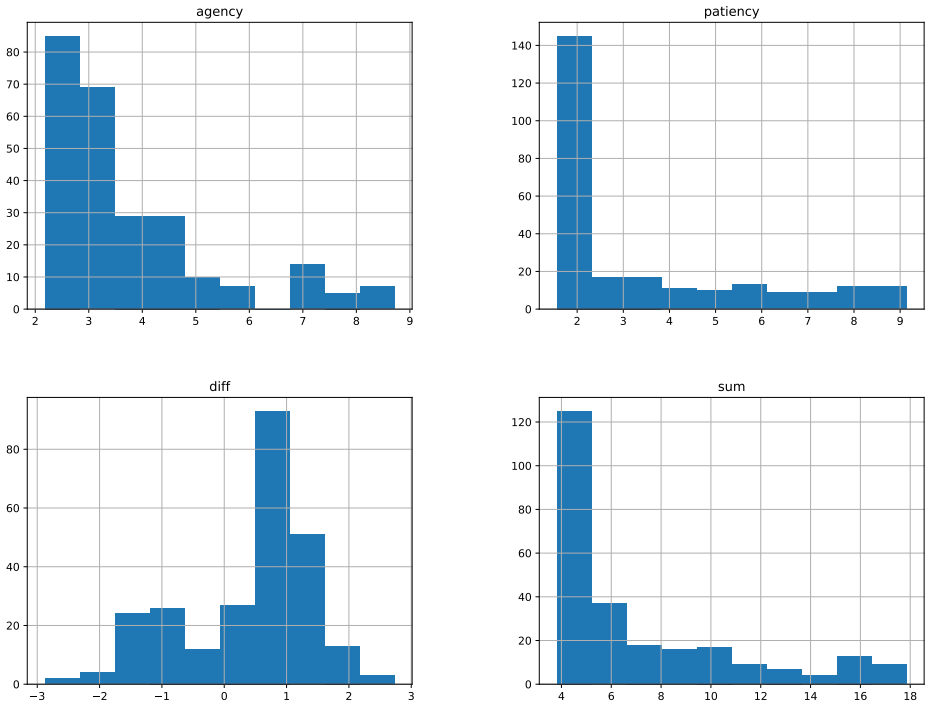
Top 31 Entities sorted by highest rank difference of our language measure in ascending order. In the third column, we show the difference in cosine similarity between agency and patiency in language. Language scores were computed with GloVe embeddings pre-trained on Wikipedia and GigaWord.



Electronic copy available at: https://ssrn.com/abstract=3959847

# C Survey Measures

Figure C3: Distributions of Survey Measures




Distributions of the unadjusted survey ratings.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table C5: Top- and Bottom-Ranked Entities by Survey Ratings

| Measure | Top 10 Entities | Bottom 10 Entities |
|---|---|---|
| Agency | human father mother mom man dad woman grandmother grandfather girl | stamp trousers cushion sponge rock cheese circle lettuce corpse drop |
| Patiency | human mother man woman mom father dad girl grandmother grandfather | sock circle bag ticket umbrella square bell hook chess corpse |
| Difference | computer hospital robot sun train gun ship school plane store | lamb rabbit fetus pig slaves chimpanzee kitten puppy baby infant |
| Sum | human mother man mom father woman dad grandmother girl grandfather | sock cheese bucket cushion square rock stamp drop circle corpse |

Top and bottom ranked entities, out of our set of 255 assessed entities, for agency, patiency, agency-patiency and agency+patiency, based on the unadjusted survey ratings.

Table C6: Pairwise Correlations: Survey Measures

| | | agency | agency_adj | patiency | patiency_adj |
|---|---|---|---|---|---|
| Whole Survey | agency | 1.0 | 1.0 | 0.83 | 0.83 |
| | agency_adj | - | 1.0 | 0.81 | 0.82 |
| | patiency | - | - | 1.0 | 1.0 |
| | patiency_adj | - | - | - | 1.0 |
| | | agency | agency_adj | patiency | patiency_adj |
| Top 31 Entities | agency | 1.0 | 1.0 | 0.94 | 0.94 |
| | agency_adj | - | 1.0 | 0.93 | 0.94 |
| | patiency | - | - | 1.0 | 1.0 |
| | patiency_adj | - | - | - | 1.0 |

Pairwise Spearman rank correlations for the survey measures. The adjusted measures produce almost identical rankings.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table C7: Entities with Largest Agency-Patiency Rank Difference in Survey Ratings

| Entity | Rank Difference Agency - Patiency | Score Difference Agency - Patiency |
|---|---|---|
| infant | -9 | -2.88 |
| baby | -9 | -2.88 |
| fetus | -3 | -1.75 |
| woman | -3 | -0.82 |
| puppy | -2 | -2.28 |
| man | -2 | -0.76 |
| girl | -2 | -1.04 |
| kitten | -1 | -1.96 |
| mouse | -1 | -1.62 |
| mother | -1 | -0.69 |
| corpse | 0 | 0.7 |
| boy | 0 | -1.26 |
| ant | 0 | -0.94 |
| human | 0 | -0.43 |
| beetle | 0 | -1.15 |
| chimpanzee | 0 | -1.8 |
| fish | 0 | -1.16 |
| elephant | 1 | -1.59 |
| dad | 1 | -0.58 |
| mom | 1 | -0.61 |
| frog | 1 | -1.36 |
| grandmother | 1 | -0.81 |
| grandfather | 1 | -0.88 |
| dog | 1 | -1.55 |
| cat | 2 | -1.6 |
| bird | 3 | -1.43 |
| insect | 3 | -0.66 |
| monkey | 3 | -1.56 |
| father | 4 | -0.5 |
| shark | 5 | -0.93 |
| primate | 6 | -1.25 |

Top 31 Entities sorted by highest rank difference of the survey measure in ascending order. In the third column, we show the raw agency rating - patiency rating obtained in the survey.



Electronic copy available at: https://ssrn.com/abstract=3959847

# D Comparing Survey and Language Measures

Table D8: Human Validation of Language Measures: Main Evaluation Metrics

| N Entities | Embeddings | Survey Ratings | Agency | Patiency |
|---|---|---|---|---|
| 31 | glove | adjusted | .67 (0.0) | .75 (0.0) |
| 31 | glove | unadjusted | .67 (0.0) | .76 (0.0) |
| 31 | word2vec | adjusted | .33 (0.07) | .67 (0.0) |
| 31 | word2vec | unadjusted | .32 (0.078) | .68 (0.0) |
| 31 | fasttext | adjusted | .26 (0.159) | .71 (0.0) |
| 31 | fasttext | unadjusted | .25 (0.179) | .72 (0.0) |
| 58 | glove | adjusted | .55 (0.0) | .54 (0.0) |
| 58 | glove | unadjusted | .53 (0.0) | .54 (0.0) |
| 58 | word2vec | adjusted | .35 (0.008) | .48 (0.0) |
| 58 | word2vec | unadjusted | .33 (0.012) | .47 (0.0) |
| 58 | fasttext | adjusted | .49 (0.0) | .56 (0.0) |
| 58 | fasttext | unadjusted | .49 (0.0) | .55 (0.0) |
| 255 | glove | adjusted | .4 (0.0) | .36 (0.0) |
| 255 | glove | unadjusted | .39 (0.0) | .35 (0.0) |
| 255 | word2vec | adjusted | .23 (0.0) | .51 (0.0) |
| 255 | word2vec | unadjusted | .23 (0.0) | .51 (0.0) |
| 255 | fasttext | adjusted | .35 (0.0) | .38 (0.0) |
| 255 | fasttext | unadjusted | .35 (0.0) | .38 (0.0) |

This table shows the main evaluation statistics comparing the human survey ratings checks comparing the human All word embedding algorithms trained on the 2000 and 2010 decades of COHA.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table D9: Validation Robustness to Embedding Pre-Training Corpus

| Algorithm | Training Corpus | Agency | Patiency |
|---|---|---|---|
| GloVe | Wiki + GigaWord (6B tokens) | .41 (.0) | .5 (.0) |
| GloVe | CommonCrawl (840B tokens) | .37 (.0) | .52 (.0) |
| GloVe | COHA (80M tokens) | .4 (.0) | .38 (.0) |
| Word2Vec | GoogleNews Corpus (100B tokens) | .05 (.42) | .3 (.0) |
| Word2Vec | COHA (80M tokens) | .23 (.0) | .51 (.0) |
| FastText | CommonCrawl (840B tokens) | .1 (.12) | .52 (.0) |
| fasttext | COHA (80M tokens) | .35 (.0) | .4 (.0) |

Rank correlations using embeddings pretrained on different corpora. Our preferred setting consists of GloVe embeddings trained on Wikipedia and GigaWord. See Appendix A for descriptions of each corpus. Uses adjusted survey measures with 255 entities.

Table D10: Validation Robustness to Context Window Size and Dimensionality

| Window Size | Vector Size | Agency Corr. | Patiency Corr. |
|---|---|---|---|
| 5 | 50 | .38 (0.0) | .47 (0.0) |
| 5 | 100 | .4 (0.0) | .49 (0.0) |
| 5 | 200 | .38 (0.0) | .41 (0.0) |
| 10 | 50 | .4 (0.0) | .45 (0.0) |
| 10 | 100 | .42 (0.0) | .39 (0.0) |
| 10 | 200 | .4 (0.0) | .38 (0.0) |
| 15 | 50 | .4 (0.0) | .42 (0.0) |
| 15 | 100 | .39 (0.0) | .37 (0.0) |
| 15 | 200 | .37 (0.0) | .36 (0.0) |

Robustness of correlation results to different Window sizes and Vector Sizes. All experiments are conducted using GloVe embeddings trained on the 2000 and 2010 decades of COHA. Spearman's R is displayed for rank correlation between our language measure and the whole survey. All results are highly significant with $p < 0.005$.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table D11: Language and Survey Rank Correlations for Summed and Differenced Measures

| Embeddings | Difference | Summed |
|---|---|---|
| GloVe | .41 (0.0) | .38 (0.0) |
| Word2Vec | .44 (0.0) | .4 (0.0) |
| FastText | .25 (0.001) | .4 (0.0) |

Spearman rank correlations for the difference and sum of the language measure (agency minus patiency, and agency plus patiency), with the respective difference and sum of the survey measure. Uses unadjusted survey measures with 255 entities.

Table D12: Correlation Statistics for Other Surveys

| Entity Rating Survey | Embeddings | Agency/Patiency Words | Agency | Patiency |
|---|---|---|---|---|
| GGW Survey | glove | GGW | .73 (0.005) | .26 (0.394) |
| GGW Survey | word2vec | GGW | .38 (0.201) | .21 (0.482) |
| GGW Survey | fasttext | GGW | .59 (0.035) | .15 (0.616) |
| Corp. Insecthood Survey | glove | Corp. Insecthood | .27 (0.15) | .26 (0.168) |
| Corp. Insecthood Survey | word2vec | Corp. Insecthood | .31 (0.102) | .28 (0.147) |
| Corp. Insecthood Survey | fasttext | Corp. Insecthood | .26 (0.168) | .16 (0.411) |

We show that our measure is not over-engineered towards our word list and robust to other choices of agency and patiency words, as well as entities to assess. We consider the significant factors described in Gray et al. (2007) and Strohminger and Jordan (2021) as alternative agency/patiency lists (column 3). For the Gray Gry Wegner Survey (GGW), we used a wordlist derived from the significant factors in the respective paper. The agency wordlist for the GGW survey consists of `"control morality memory emotion recognition plan communicate think"`, the patiency wordlist of `"hungry afraid pain pleasant angry desire personality conscious proud embarrassed joyful"`. For the Corp. Insecthood Survey, we use words derived from the questions posed in the survey as our wordlists. The agency wordlist contains the words `"agency think judge reason decide act"`, the patiency wordlist contains `"patiency aware emotion feel experience"`. All embedding models trained on the 2000 and 2010 decades of COHA



Electronic copy available at: https://ssrn.com/abstract=3959847

Table D13: Human Validation Metrics Using the Mind Perception Dictionary

| N | Embeddings | Survey Ratings | Agency | Patiency |
|---|---|---|---|---|
| 31 | glove | adjusted | .74 (0.0) | .75 (0.0) |
| 31 | glove | unadjusted | .73 (0.0) | .75 (0.0) |
| 31 | word2vec | adjusted | .47 (0.008) | .73 (0.0) |
| 31 | word2vec | unadjusted | .46 (0.009) | .74 (0.0) |
| 31 | fasttext | adjusted | .46 (0.01) | .66 (0.0) |
| 31 | fasttext | unadjusted | .45 (0.011) | .66 (0.0) |
| 58 | glove | adjusted | .61 (0.0) | .7 (0.0) |
| 58 | glove | unadjusted | .59 (0.0) | .69 (0.0) |
| 58 | word2vec | adjusted | .39 (0.003) | .59 (0.0) |
| 58 | word2vec | unadjusted | .38 (0.003) | .58 (0.0) |
| 58 | fasttext | adjusted | .51 (0.0) | .59 (0.0) |
| 58 | fasttext | unadjusted | .5 (0.0) | .59 (0.0) |
| 255 | glove | adjusted | .49 (0.0) | .56 (0.0) |
| 255 | glove | unadjusted | .49 (0.0) | .56 (0.0) |
| 255 | word2vec | adjusted | .3 (0.0) | .56 (0.0) |
| 255 | word2vec | unadjusted | .3 (0.0) | .56 (0.0) |
| 255 | fasttext | adjusted | .4 (0.0) | .42 (0.0) |
| 255 | fasttext | unadjusted | .41 (0.0) | .42 (0.0) |

This table shows the main evaluation statistics comparing the human survey ratings checks comparing the human All word embedding algorithms trained on the 2000 and 2010 decades of COHA. Wordlists for agency and patiency are derived from the Mind Perception Dictionary (Schweitzer and Waytz, 2020).



Electronic copy available at: https://ssrn.com/abstract=3959847

Table D14: Sorted Entities by Agency Difference Language and Survey

| Entity | Agency Rank Language - Rank Survey |
|---|---|
| primate | -15 |
| chimpanzee | -15 |
| grandmother | -11 |
| grandfather | -11 |
| kitten | -9 |
| monkey | -9 |
| puppy | -8 |
| mom | -7 |
| dad | -6 |
| shark | -5 |
| elephant | -5 |
| father | -3 |
| mother | -1 |
| frog | -1 |
| human | 0 |
| cat | 0 |
| girl | 2 |
| ant | 2 |
| man | 3 |
| dog | 3 |
| beetle | 3 |
| woman | 4 |
| boy | 5 |
| bird | 6 |
| corpse | 7 |
| insect | 8 |
| mouse | 9 |
| infant | 9 |
| fetus | 14 |
| baby | 14 |
| fish | 17 |

Top 31 entities sorted by ascending rank difference from our agency language measure compared to the agency survey measure. Language scores computed using GloVe embeddings pre-trained on Wikipedia and GigaWord.



Electronic copy available at: https://ssrn.com/abstract=3959847

Table D15: Sorted Entities by Patiency Difference Language and Survey

| Entity | Patiency Rank Language - Rank Survey |
|---|---|
| chimpanzee | -15 |
| primate | -11 |
| human | -8 |
| elephant | -5 |
| grandfather | -4 |
| shark | -4 |
| man | -2 |
| father | -2 |
| infant | -2 |
| grandmother | -2 |
| monkey | -2 |
| ant | -1 |
| woman | 0 |
| kitten | 0 |
| frog | 0 |
| girl | 1 |
| mother | 1 |
| dog | 1 |
| baby | 2 |
| beetle | 2 |
| fetus | 2 |
| mom | 2 |
| puppy | 3 |
| bird | 3 |
| insect | 3 |
| mouse | 4 |
| dad | 5 |
| boy | 5 |
| cat | 6 |
| corpse | 8 |
| fish | 10 |

Top 31 entities sorted by ascending rank difference from our patiency language measure compared to the patiency survey measure. Language scores computed using GloVe embeddings pre-trained on Wikipedia and GigaWord.

Table D16: Correlation of Language and Survey Measures (excluding words which can also be verbs)

| Survey Dataset | N | Agency | | | Patiency | | |
|---|---|---|---|---|---|---|---|
| | | glove | fasttext | word2vec | glove | fasttext | word2vec |
| 255 entities | 230 | 0.45 ( 0.0 ) | 0.2 ( 0.003 ) | 0.12 ( 0.061 ) | 0.56 ( 0.0 ) | 0.6 ( 0.0 ) | 0.35 ( 0.0 ) |
| Top 31 | 30 | 0.63 ( 0.0 ) | 0.45 ( 0.012 ) | 0.01 ( 0.938 ) | 0.85 ( 0.0 ) | 0.84 ( 0.0 ) | 0.68 ( 0.0 ) |
| Top 58 | 57 | 0.53 ( 0.0 ) | 0.44 ( 0.001 ) | 0.12 ( 0.392 ) | 0.64 ( 0.0 ) | 0.75 ( 0.0 ) | 0.49 ( 0.0 ) |



Electronic copy available at: https://ssrn.com/abstract=3959847