

What is (and was) a person? Evidence on historical mind perceptions from natural language

Journal Article**Author(s):**

Ash, Elliott; Stambach, Dominik; Tobia, Kevin

Publication date:

2023-10

Permanent link:

<https://doi.org/https://doi.org/10.3929/ethz-b-000626557>

Rights / license:

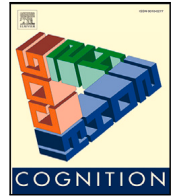
[Creative Commons Attribution 4.0 International](#)

Originally published in:

Cognition 239, <https://doi.org/10.1016/j.cognition.2023.105501>

Funding acknowledgement:

- The Ordinary Meaning of Law ()



Original articles

What is (and was) a person? Evidence on historical mind perceptions from natural language[☆]

Elliott Ash^{a,*}, Dominik Stambach^{a,1}, Kevin Tobia^{b,1}

^a *ETH Zurich, Switzerland*

^b *Georgetown University, United States of America*



ARTICLE INFO

Keywords:

Mind
Personhood
Agency
Experience
NLP
Historical psychology
Gender
Animals
Word embeddings

ABSTRACT

An important philosophical tradition identifies persons as those entities that have minds, such that mind perception is a window into person perception. Psychological research has found that human perceptions of mind consist of at least two distinct dimensions: agency (e.g. planning, deciding) and experience (e.g. feeling, hungering). Taking this insight into the semantic space of natural language, we develop a generalizable, scalable computational-linguistics method for measuring variation in perceived agency and experience in large archives of plain-text documents. The resulting text-based rankings of entities along these dimensions correspond to human judgments of perceived agency and experience assessed in blind surveys. We then map both dimensions of mind in historical English-language corpora over the last 200 years and identify two salient trends. First, we find that while women are now described as having similar levels of agency as men, they are still described as more experience-oriented. Second, we find that domesticated animals have gained higher attributions of experience (but not agency) relative to wild animals, especially since the rise of the global animal rights movement in the 1980s.

1. Introduction

What is a person? This longstanding philosophical question is at the root of modern moral, social, and legal debates. It fuels debates about animal rights (Sunstein & Nussbaum, 2005), corporate speech (Eliis, 2011), abortion (Little, 2008), brain death (Willke & Andrusko, 1988), and environmental protection (Gordon, 2019). Personhood attribution is also related to social phenomena including anthropomorphization (Epley et al., 2007), dehumanization (Harris & Fiske, 2006; Haslam, 2006; Haslam & Loughnan, 2014; Tarasenko-Struc, 2020), gender stereotyping (Hamilton, 1991), and objectification of men and women (Gray et al., 2011).

One influential answer points to the possession of a mind. Persons are those entities that have certain mental properties (e.g. Baker, 2000); or entities that are capable of acquiring those properties, or entities that are part of a kind (e.g. humans) whose members typically have those properties. On this view – since persons are entities with minds – gaining insight into human mind perception provides a window into human person perception.

In the study of mind perception, psychologists have distinguished between two different types of mental properties: agency (e.g. planning, deciding) and experience (e.g. feeling, hungering). This distinction has also been important in philosophy: Since Aristotle, philosophers have distinguished between agents – who act and cause events – and patients – who experience these actions and events (Irwin, 1986; Olson, 2019). Recent psychological research has begun to disentangle these agency and experience dimensions (Gray et al., 2007, 2011; Gray & Wegner, 2009; Strohminger & Jordan, 2021), finding that some entities are evaluated to be high in agency but low in experience (e.g. robots), while others are evaluated as high in experience but low in agency (e.g. kittens). Philosophers often connect agency with moral or legal obligations and experience with moral or legal protections, and further social-science research is needed on how these dimensions of mind are perceived in the real world.

This paper takes these ideas into the empirical space of natural language. Building on theories from linguistics (Dowty, 1991; Resinger et al., 2015) and psychology (Gray et al., 2007), we use the dimensions

[☆] We thank Joanna Bryson, Kurt Gray, Joshua Knobe, and Alexander Stremitzer for helpful feedback. Data and replication code are hosted at https://github.com/dominiksinsaarland/Evidence_Historical_Mind_Perceptions_Natural_Language (Ash et al., 2023). We acknowledge research support from ETH Zurich and the Swiss National Science Foundation including the Swiss National Science Foundation Spark Grant for “The Ordinary Meaning of Law,” CRSK-1-190713.

* Corresponding author.

E-mail addresses: ashe@ethz.ch (E. Ash), dominik.stambach@gess.ethz.ch (D. Stambach), kt744@georgetown.edu (K. Tobia).

¹ The authors contributed equally.

of mind concept as a lens into measuring perceptions expressed in text (see also Schweitzer & Waytz, 2021). The novelty of this work is in adapting a key technology from computational linguistics, namely word embedding models (Mikolov et al., 2013; Pennington et al., 2014), to produce scalable measures of these perceptions in large text archives. Through machine reading of large corpora, word embedding models can map words and concepts as vectors in a geometric space. We use these tools to learn robust dimensions in semantic space corresponding to agency and experience. Further, we score a large set of entities according to how they are described along the agency and experience 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 of relative agency and experience for this same set of entities. We show that the agency and experience rankings of entities by the language model match the respective rankings of entities by survey respondents. Further, the difference in the language scores for agency and experience match the difference in the respective survey scores. Thus, the language model scores properly capture a two-dimensional index of mind perceptions.

With the validated language-based measures in hand, we use them for a long-run historical analysis of mind 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/experience dimensions by decade from the 1820s through the 2010s. We then compute agency and experience scores over time for sets of relevant entities while accounting for sampling uncertainty in those measures, providing a view on mind perceptions even many decades into the past. Understanding these developments in language, and how they relate to changes in laws and culture, can help inform ongoing policy debates and indicate shifts in the boundaries of the moral universe.

The first historical analysis is to measure differences in the attributed agency and experience of men and women. In the early 1800s, men were depicted as having more agency in texts, while women were depicted as having more experience. 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 experience has been more persistent, however, and actually increased over the 1800s. Although the tendency to describe women as experiencers has decreased somewhat over the last 100 years, women are still more often described as experiencers than men.

Second, we consider the mind perceptions of domesticated animals over the same time period. Relative to wild animals, domestic animals have historically had similar levels of agency over the last 200 years. For experience, the trend is quite different. Over the time period, domesticated animals have become increasingly described as experiencing in texts, a trend that hastened starting in the 1980s with the establishment of PETA and the global animal rights movement. These cultural perceptions tend to move ahead 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.

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.

In philosophy, personhood is often connected with morality. 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 (Irwin, 1986; Olson, 2019). 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/experience dimensions also manifest 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.² 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 experience go together. Some entities have capacities for both choice and the ability to experience (e.g. human adults), while other entities have neither (e.g. inanimate objects). Experimental psychology, however, has questioned the one-dimensional view. In particular, Gray et al. (2007) provide evidence that humans understand agency and experience separately, with some entities having divergent perceptions. For example, a baby ranks high on the experience axis, but low on the agency axis. A robot ranks high on agency but low on experience.³

Thus, these studies conclude that the agency and experience “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 experience 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 & Wegner, 2009). More broadly, while personhood concepts are clearly relevant for a range of policy debates – e.g., animal rights (Sunstein & Nussbaum, 2005), corporate speech (Ellis, 2011), abortion (Little, 2008), and gender stereotypes (Hamilton, 1991) – those debates might be better informed by taking a two-dimensional approach which disentangles agency and experience.

To understand dimensions of mind in real-world contexts, the previous experimental survey approaches have some limitations. Some of these limitations can be circumvented by a complementary natural language approach. As such, one aim of our study is to build on prior experimental work, providing convergent evidence of the dimensions of mind. Concerning the limitations of experimental survey work, those studies are often conducted on a small sample of individuals, often college students in Western universities. Moreover, survey participants

² 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). Psychological research indicates that people's retributive sentiment sometimes leads them to extend punitive judgments to non-human animals (Goodwin & Benforado, 2015).

³ 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 experience and therefore low personhood.

may feel social pressure to report socially acceptable answers—such as gender-equitable attributions of agency and experience.⁴ Finally, a survey approach is generally limited to studying present cognition (e.g. agency/experience perceptions today, but not in the past).

Our approach addresses these limitations through analysis of mind perceptions 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, natural language can reveal implicit attitudes expressed in corpora, similar to those revealed by implicit association tests (Caliskan et al., 2017; Garg et al., 2018a). 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 experience perceptions. The quantitative historical results can provide invaluable context to ongoing policy debates on the expansion or restriction of moral rights and obligations.

These results are most relevant to the agency-experience dimensions in mind perceptions, yet our method could be extended to the broader scholarship that examines other dimensions. For example, Weisman et al. (2017) propose a three-dimensional “body–heart–mind” framework that fits survey data on the similarities and differences among mental capacities (as opposed to the agency-experience model, which works better in ranking entities). Another three-dimensional mind model, proposed by Tamir and Thornton (2018), divides up mental states along the dimensions of valence, social impact, and rationality (versus emotionality). Thornton et al. (2022) analyze these dimensions with word embedding models and show that the correlations of concept dimensions in the embedding space correspond to those assumed in the valence-impact-rationality theory. A promising task for future work would be, in line with our approach, to analyze perceptions of entities along the dimensions articulated in these other frameworks.

3. Measuring agency and experience perceptions in natural language

This section describes how we measure the agency and experience dimensions 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/moral associations on a large scale (Atari et al., 2022; Kozlowski et al., 2019; Wang & Inbar, 2021), to analyze emotionality and rationality in politics (Gennaro & Ash, 2021), and to discover bias in language (Brunet et al., 2019; Caliskan et al., 2017; Garg et al., 2018b).⁵

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).

⁴ Even incentive-compatible surveys, which reward participants for accurately estimating the answers of others (Prelac, 2004), might not fully capture implicit attitudes.

⁵ Two methods-oriented papers presenting similar approaches are Garten et al. (2018) and Watanabe (2021).

Table 1
Word lists for agency and experience 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
Experience	pain pleasure enraged afraid hunger thirsty sad proud embarrassed joy angry happy conscious aware experience imagine awake suffer enjoy desire

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 neighboring context.⁶ 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 experience, we construct semantic dimensions using twenty words, selected based on the previous psychology literature on dimensions of mind and mind perception (Gray et al., 2007; Strohminger & Jordan, 2021). Because we aimed to assess the dimensions of mind theory, we drew from the seminal paper in that literature; and because we wished to assess a large number of entities, we also drew from a paper that employed terms applicable to a wide range of entities. These pre-registered lists of words are shown in Table 1.⁷ To get an agency vector \vec{a} and an experience/patience 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 experience 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} .⁸ Fig. B.1 shows the most similar words to each pole. The resulting words are quite intuitive, with the agency words corresponding to actions, desires, and morality, and the experience words corresponding to feelings, knowledge, and ability to experience.

We use the same approach to measure how a given entity e scores along the agency and experience dimensions in language. The linguistic agency score $a_L(e)$ of an entity e in our framework (L is for “language

⁶ 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 (Table D.3). For full details and replication code, see Ash et al. (2023).

⁷ For a dictionary-based approach to these issues, see Schweitzer and Waytz (2021). That (contemporaneous) paper analyzes agency and experience as part of a broader analysis of mind perceptions. The paper provides a longer list of agency-related and experience-related words, which we will use in a robustness check (Table D.6). We prefer the embedding approach to the dictionary approach for two reasons. First, it places a lot of pressure on the design of the dictionary, because one would have to specify each word as indicating agency, experience, or neither. Our approach identifies a concept direction and is less strict in identifying the correct dictionary. Further, our approach can allow words to be partially related to agency or experience. For example, the word “agent” can mean chemical as well as an actor; the word embedding approach adjusts for these mixed meanings, while dictionaries cannot. See Appendix D.1 for more detail and results on a dictionary method alternative.

⁸ 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 dot product as our similarity metric.

model”) is the cosine similarity between the entity’s vector \vec{e} and the agency vector \vec{a} . Correspondingly, the experience score $p_L(e)$ is the cosine similarity with the experience 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. Fig. B.2 shows histograms for each variable. Table B.2 lists the top- and bottom-ranked entities along all four linguistic scores. For example, computer and office tend to have high agency and low experience, while puppy and ghost tend to have high experience and low agency (see also Table D.7 and Table D.8). Thus, similar to the survey measures in Gray et al. (2007), the two text-based mind dimensions can come apart and vary separately.⁹ We can see this statistically in Table B.3, showing pair-wise correlations between the language variables. Agency and experience are strongly correlated, but far from perfectly so. For example, for the GloVe embeddings, the correlation between agency and experience 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 experience matches up with human judgments. Building on recent work in experimental social psychology (Gray et al., 2007; Strohminger & Jordan, 2021), we developed a survey instrument asking respondents to implicitly score entities along an agency axis and an experience axis.¹⁰

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 experience 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 & Jordan, 2021) supplemented with an online list of the “most important English nouns” (Co., 2019).¹¹

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

⁹ The usefulness of the top- and bottom-ranked entities by language scores should not be overstated. We care most about the overall correlation between the language and survey rankings. Language measures, far more so than survey measures, are noisy, so pairwise rankings at the top and bottom of the distribution are less reliable.

¹⁰ We pre-registered our survey materials, including comprehension check questions and exclusion criteria, and our method of creating agency and experience measures from the survey data (<https://osf.io/e7x39>). Survey materials are included in the replication package (Ash et al., 2023).

¹¹ 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 (Table D.9). None of these words are used in the historical analysis.

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 experience 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 experience score $p_H(e)$ is the average of the ratings for that entity across all experience properties.

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 experience 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 Fig. C.1. Table C.1 shows the top and bottom-ranked entities along the survey ratings, and the pairwise correlations are shown in Table C.2. 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 experience 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 and 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 & 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 experience 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 entities. The full list of entities with group assignment is shown in Table B.1.

Fig. 1 reports our main validation results for the preferred measures of agency and experience from the survey and from the language model.¹² The figure shows scatter plots for the human and language measures of agency (Panel A) and experience (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)

¹² 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 experience words. In the human survey scores, our preferred specification uses the normalized scores that adjust for variation across properties.

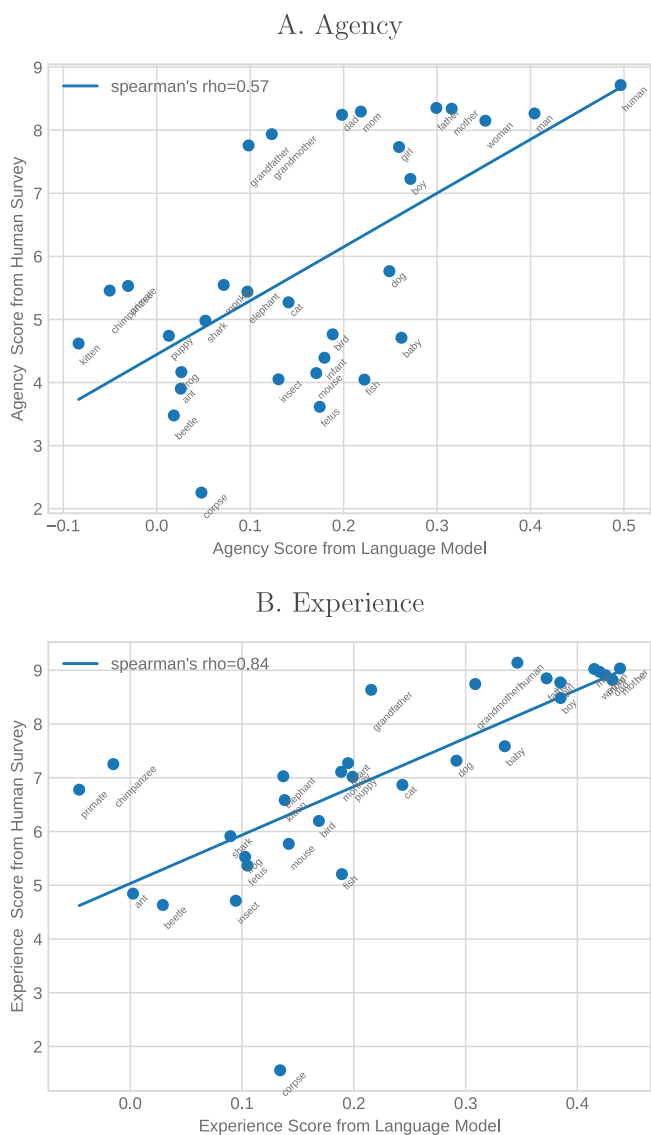


Fig. 1. Comparison of language and survey measures. Notes. Scatter plots of Agency and Experience scores, with human survey measures on the vertical axis and language measures on the horizontal axis. Plots include the pre-registered list of 31 entities, indicated by marker labels. Blue line is linear best fit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and 0.83 (experience). They are both highly statistically significant with $p < 0.005$.

Table D.1 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 experience 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 Table D.2, 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, Table D.3 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/experience rankings.

Besides the individual agency and experience scores, we would also like to check whether the language model can capture sums and differences of the survey’s agency and experience scores. We find that they can. Table D.4 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.¹³

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 experience (Gray et al., 2007; Strohinger & Jordan, 2021). From these previous surveys we take lists of agency and experience words, lists of entities, and associated entity rating contemporaneous for agency and experience. In both of these datasets, our word-embedding associations substantially recover the human-survey associations (Table D.5). Further, our validation results are even better when constructing the mind dimensions with longer lists of agency and experience words from Schweitzer and Waytz (2021) (Table D.6).

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

5. Historical analysis of mind dimensions

Linguistic measures of agency and experience 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 experience 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.

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.¹⁴ 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. As discussed in the previous section, we show in

¹³ Appendix D.1 shows that a dictionary-based metric for agency and experience, based on Schweitzer and Waytz (2021), performs poorly in ranking entities by the difference in the measures.

¹⁴ 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.

Tables D.2 and D.3 that our method using embeddings trained on recent decades of COHA produces results close to our baseline, while being robust to different word embedding algorithms and hyperparameters. Extrapolating from these results, we expect that our method assigns credible scores of agency and experience in the past and that COHA as a data source is well suited for our analysis.

For building diachronic word embeddings, we follow the method outlined 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.¹⁵ 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 experience 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 experience scores for each entity per decade.

We construct entity-level measures by calculating the mean μ and standard error σ across the 10 scores obtained by the different bootstrapped models. We use these statistics to test for statistical differences between entities. To compare groups of entities, we create group averages μ_i and μ_j for both groups i and j of the entity-level means for each of the 10 bootstraps. We then compute the differences in these entity-group means (denoted as μ_{diff}). For statistical inference, we construct 95% confidence intervals based on the mean and standard error of the 10 between-group differences ($\mu_{diff} \pm 1.96 * \sigma_{diff}$).

Historical gender differences in agency and experience. Our first historical analysis assesses variation in expressed attitudes about the agency and experience 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 & 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 experience. 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.

¹⁵ 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.

The gender differences in text-based agency and experience are reported in Fig. 2 Panel A. We find that in the earliest period (1820s–1850s), men were described as having higher agency and lower experience than women. Already in this early period, the relative male association with agency was shrinking, and there was not 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 (Rupp & Taylor, 1987). But then the difference flips 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 experience actually intensified in the early years, with the largest differences from 1870 through 1920. Since then, the relative female association with experience has gradually moderated toward zero. While it appears to still be shrinking, relative female experience has not yet disappeared. That result suggests there is still an operational stereotype in how women are represented as experiencers in texts.

To summarize, gender representations in terms of agency and experience have become more equal over the last two centuries. These results provide a long-run quantitative view on gender portrayals 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.

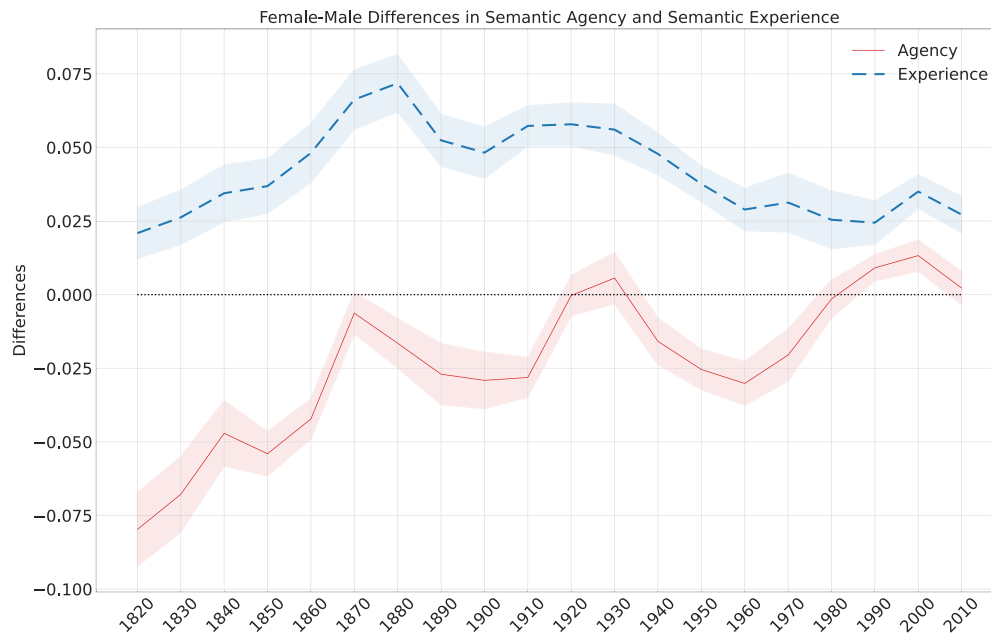
This pair of findings (compared to men, women are portrayed as equally agential, but still more experiential/patientic) resonates with some feminist debates, such as the clash between "power feminism" and "victim feminism" (Harris, 2011). The equal agency result is suggestive of natural language representations of power equality: today, women are described as active, deciding, powerful agents at rates similar to men. Yet the persistent experience 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 experiential 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. Moreover, it is important to keep in mind the complexities underlying the language-based measurements in our historical analysis. While our survey validation suggests that text *portrayals* reflect internal *perceptions*, there are other potential factors affecting the text. It is possible, for instance, that women obtained higher relative agency in certain time periods not solely due to changes in perceptions, but instead due to changes in their collective actions, or changes in the social and cultural norms governing their roles and behaviors. It is crucial to be mindful of these various factors when interpreting the results.

Evolution of agency and experience perceptions for domesticated animals. Our second historical analysis considers the perception of domesticated animals' agency and experience. How animals have been understood over time can enrich our understanding of attitudes about animals and what protections they deserve. Animals' agency (e.g. ability to do wrong) and experience (e.g. feeling pain) have been central to cultural, moral and legal debates about animal rights (Regan, 1983; Singer, 1973).

The history of animal rights in the United States is often traced to 1866, when Henry Bergh founded the American Society for the

A. GENDER DIFFERENCES IN PORTRAYALS OF AGENCY AND EXPERIENCE



B. AGENCY AND EXPERIENCE PORTRAYALS OF DOMESTICATED ANIMALS

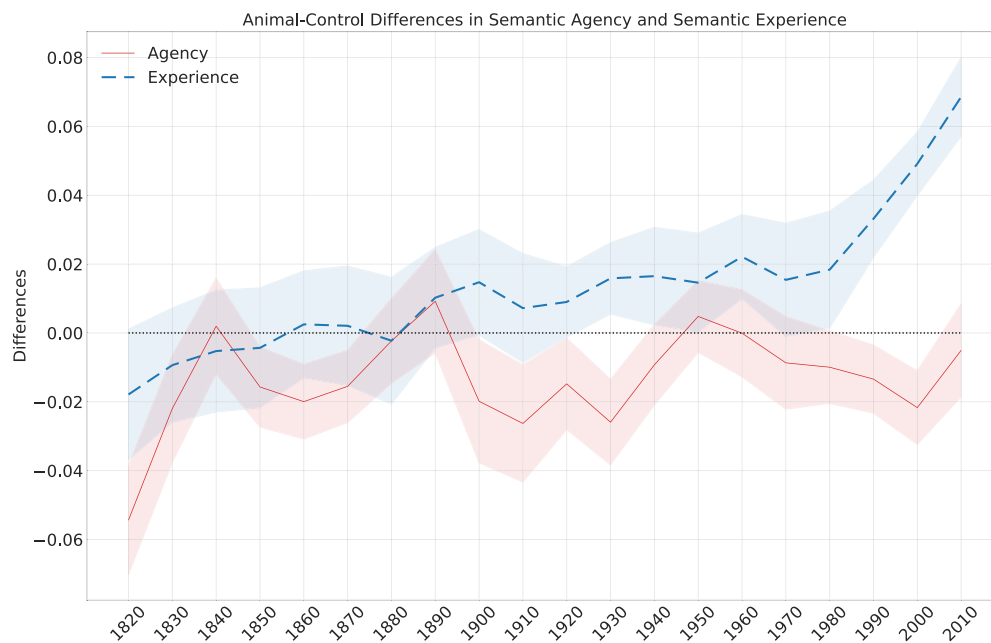


Fig. 2. Historical analysis: 200 years of agency and experience.

Notes. Time series of agency and experience differences, 1820s–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 wild animals as the control-group words. Agency is shown as red solid line, while experience is shown as blue dashed line. Confidence intervals (95%) produced by bootstrapped embeddings, as described in the text. Time series are smoothed by a moving average using a window size of 2. P-values for statistical tests are reported in [Table E.1](#). We show additional versions of this figure in [Appendix E](#). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Prevention of Cruelty to Animals (Zurlo & Goldberg, 1994). In 1892, the American Humane Association lobbied to prohibit unnecessary animal experiments; between 1896 and 1900, federal legislation to inspect animal laboratories was (twice) proposed and defeated (Zurlo & Goldberg, 1994). A second wave of animal rights activism began in the 1950s (Congress, 1958), accelerating in the 1970s, and continuing through today (Zurlo & Goldberg, 1994). This wave includes important philosophical works on animal rights, like Singer's *Animal*

Liberation (Singer, 1973) and Regan's *The Case for Animal Rights* (Regan, 1983), as well as the establishment of People for the Ethical Treatment of Animals (PETA) in 1980.

As in the gender analyses, we produce relative measures of agency and experience 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 the set of wild animals in our list: *rabbit, worm, mouse, pigeon, beetle, fox, chimpanzee, frog, chicken, turkey, bird, fish,*

elephant, monkey, ant, snake, primate, insect, shark, and bee.¹⁶ As before, we plot the domestic/wild differences in agency and experience over time with 95% confidence intervals.

Fig. 2 Panel B shows the results. As before, the red time series gives agency of domesticated relative to wild animals. We see no discernible trend in relative agency for these entities over the last 200 years. If anything, wild animals are portrayed as more agentic than domesticated animals, which perhaps makes sense given that “domesticated” literally means that an animal’s will has been tamed.

The relative portrayal of domestic animals as experiencers, in blue, looks quite different. Reflecting the historical rejection of animal rights in the 1800s and early 1900s, domesticated animals are not seen as experiencing much more than their wild counterparts. Yet there is a slow and steady increase over time. By the time the first animal welfare legislation is passed in 1958 (Congress, 1958), the domestic–wild difference is statistically significant. There is an especially striking upward bend in the series starting in 1980, the same year that PETA was founded. In this period, public interest in animal welfare and rights, and the adoption of plant-based diets, has increased significantly in Western countries. Accordingly, perceptions of animals as experiencing pain and other feelings has increased as well.

6. Conclusion

This paper has taken the distinction between agency and experience dimensions of mind – a subtle feature of human judgment – into the space of natural language. After validating that text associations reflect human perceptions on both dimensions, we show that agency and experience 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 social attitudes and perceptions in the realms of gender and domesticated animals. In particular, distinguishing these two dimensions clarifies important nuances in the historical evolution of both gender norms and animal rights. In both analyses, agency and experience moved independently and in historically meaningful ways.

Our quantitative analysis of natural language over time can be seen as a methodological proof of concept, answering recent calls for a historical psychology (Atari & Henrich, 2023; Muthukrishna et al., 2021). Other recent work has used word embeddings to study historical–psychological change. For example, Charlesworth et al. (2022) studies the positive versus negative valence applied to social groups over two hundred years (see also Garg et al., 2018b). Our study adds to this literature, demonstrating that word embeddings can detect even very subtle patterns (e.g. agency versus experience). As the calls for historical psychology explain, with linguistics tools in hand, researchers are no longer limited to gauging attitudes of today’s students at Western universities. As one of many promising future avenues, it could be interesting to compare perceptions of women’s agency across the political spectrum, using speeches by liberal and conservative politicians.

Still, there are some limitations of the method that could be addressed in future work. In particular, we assume two linear dimensions for agency and patience based on static word embeddings. Future work could deal with polysemy, for example, allowing for words to have multiple meanings based on the context (e.g. “feel” an object versus “feel” a sensation). Even better, more semantically rich representations of agency and patience, that take into account both explicit and implicit context, could be possible with sophisticated transformer-based models (e.g. Devlin et al., 2018; Stambach et al., 2022).

From a policy standpoint, a notable feature of the animal-rights analysis is that the change in mind perceptions preceded changes

in animal-rights legislation. There are, of course, multiple possible explanations of this pattern, and further research is needed to adjudicate among them; perhaps we are detecting individual advocacy of animals’ mental properties, rather than statements of social consensus about animals’ mental properties. Nevertheless, 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 experience scores, for these terms could help clarify subtle social attitudes and/or predict the direction of social and political change.

Data availability

Data will be made available on request.

Appendix 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 of 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 400 K words. We use the embeddings with size 300, which are available on the GloVe website¹⁷ and via Gensim (Rehurek & 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.¹⁸
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.¹⁹ This model provides embeddings for 3M words and phrases, however we only 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 FastText website.²⁰ These models were trained using the CBOW architecture with position weights, a window size of 5 and dimension of 300. While in principle close to word2vec,

¹⁶ In Fig. E.4, we show a similar analysis using a different control group: all entities in our survey which are not animals. We find that the trends are stable and these results are robust to the choice of the control group. Other perturbations to the word lists, such as counting chickens and turkeys as domesticated, also does not change the results.

¹⁷ <https://nlp.stanford.edu/projects/glove/>

¹⁸ <https://nlp.stanford.edu/projects/glove/>

¹⁹ <https://code.google.com/archive/p/word2vec/>

²⁰ <https://fasttext.cc/docs/en/crawl-vectors.html>



Fig. B.1. Semantic poles for agency and experience. Most similar words to our agency vector (Panel A), and our experience vector (Panel B). Embeddings computed using GloVe trained on Wikipedia and Gigaword.

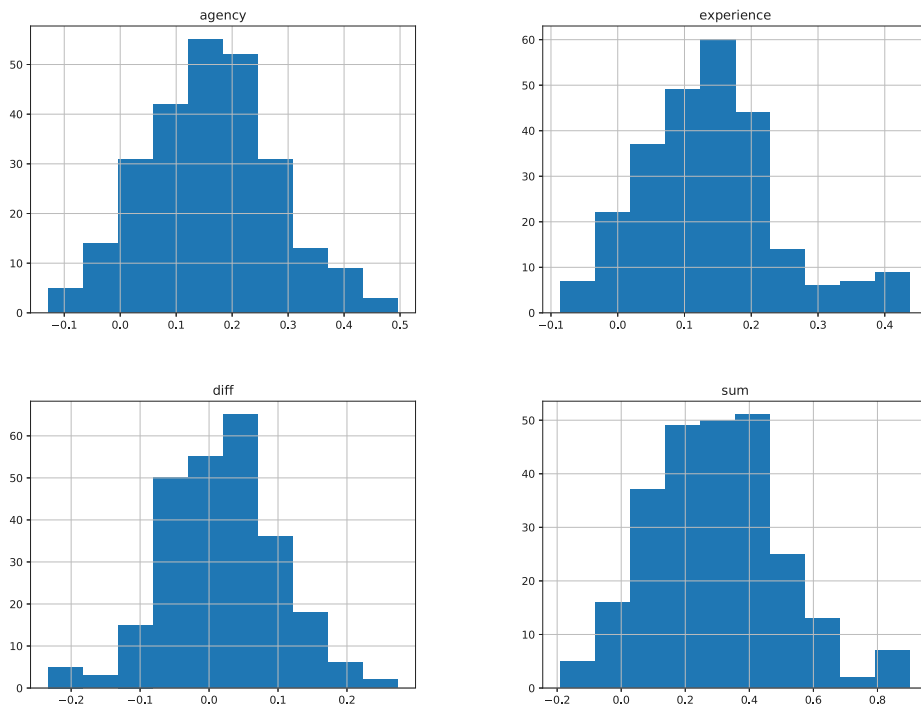


Fig. B.2. Distributions of language measures. Distributions of the language measure using GloVe embeddings trained on Wikipedia and the GigaWord corpus.

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. Table D.3 shows that the main hyperparameter choices (window size, vector

dimensionality) are not that important for capturing the agency and experience dimensions.

1. Word2Vec COHA—We use the Gensim implementation²¹ (Rehurek & 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.
2. GloVe COHA—We use the codebase in the GloVe github repository.²² We use all the default parameters: vectors with size 100,

²¹ <https://radimrehurek.com/gensim/models/word2vec.html>

²² <https://github.com/stanfordnlp/GloVe>

Table B.1
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 experience ratings were assessed in the survey, and for which we produce agency and experience 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).

Table B.2
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
experience	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.

Table B.3
Pairwise correlations: Language measures.

	Agency GloVe	Agency Word2Vec	Agency FastText	Experience GloVe	Experience Word2Vec	Experience FastText
Top 31 Entities	Agency GloVe	1.0	0.91	0.6	0.68	0.48
	Agency Word2Vec	-	1.0	0.54	0.78	0.31
	Agency FastText	-	-	1.0	0.72	0.55
	Experience GloVe	-	-	-	1.0	0.3
	Experience Word2Vec	-	-	-	-	1.0
	Experience FastText	-	-	-	-	-
Whole Survey	Agency GloVe	1.0	0.82	0.43	0.28	0.6
	Agency Word2Vec	-	1.0	0.31	0.59	0.42
	Agency FastText	-	-	1.0	0.46	0.49
	Experience GloVe	-	-	-	1.0	0.19
	Experience Word2Vec	-	-	-	-	1.0
	Experience FastText	-	-	-	-	-

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

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.²³ 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

recommended 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.

Appendix B. Language measures

See Figs. B.1 and B.2 and Tables B.1–B.4.

Appendix C. Survey measures

See Fig. C.1 and Tables C.1–C.3.

²³ <https://github.com/facebookresearch/fastText>

Table B.4

Entities with largest agency-experience difference in language score.

Entity	Rank agency - Rank experience	Score agency - Score experience
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 experience in language. Language scores were computed with GloVe embeddings pre-trained on Wikipedia and GigaWord.

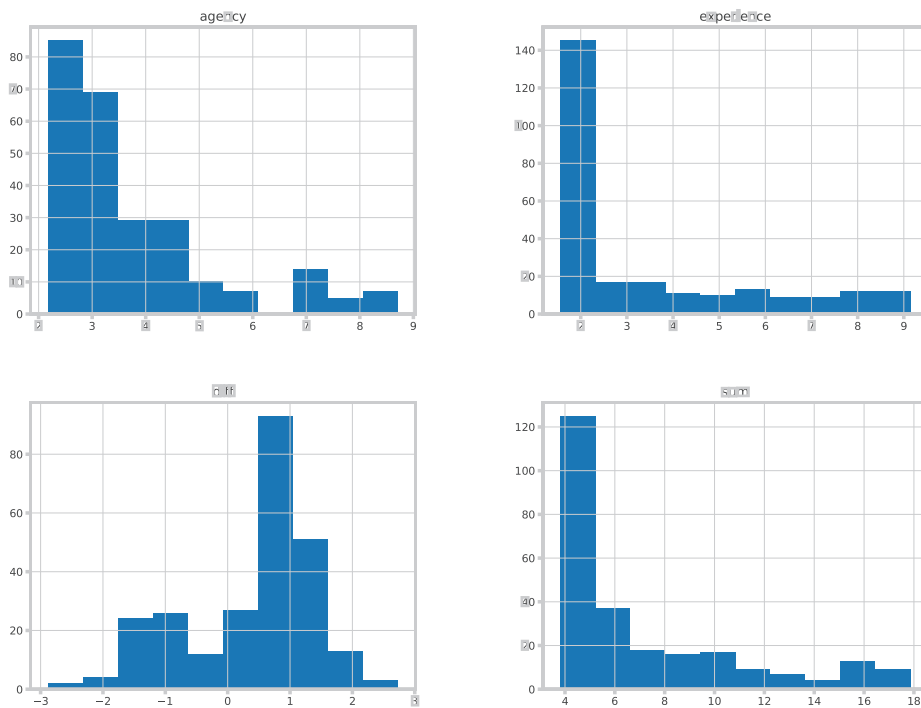


Fig. C.1. Distributions of survey measures. Distributions of the unadjusted survey ratings.

Table C.1
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
Experience	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, experience, agency-experience and agency+experience, based on the unadjusted survey ratings.

Table C.2
Pairwise correlations: Survey measures.

		agency	agency_adj	experience	experience_adj
Whole Survey	agency	1.0	1.0	0.83	0.83
	agency_adj	-	1.0	0.81	0.82
	experience	-	-	1.0	1.0
	experience_adj	-	-	-	1.0
		agency	agency_adj	experience	experience_adj
Top 31 Entities	agency	1.0	1.0	0.94	0.94
	agency_adj	-	1.0	0.93	0.94
	experience	-	-	1.0	1.0
	experience_adj	-	-	-	1.0

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

Table C.3
Entities with largest agency-experience rank difference in survey ratings.

Entity	Rank difference agency - Experience	Score difference agency - Experience
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 - experience rating obtained in the survey.

Table D.1Human validation of language measures: Main evaluation metrics (Spearman's Rho and p -value in brackets).

N Entities	Embeddings	Survey ratings	Agency	Experience
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.

Table D.2

Validation robustness to embedding pre-training corpus.

Algorithm	Training corpus	Agency	Experience
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)	.27 (.0)	.54 (.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 D.3

Validation robustness to context window size and dimensionality.

Window size	Vector size	Agency corr.	Experience 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 $p < 0.005$.

Appendix D. Comparing survey and language measures

D.1. Comparison to dictionary-based method

A contemporaneous paper by [Schweitzer and Waytz \(2021\)](#) introduces a dictionary-based method – the Mind Perception Dictionary (MPD) – for detecting words related to mind perceptions in text. In that dictionary, agency and experience are two of the several lists of words used. This section provides a brief comparison of our method to the dictionary method.

We produced a similar measure of agency and patiency by entity based on co-occurrence with MPD terms (within a ten-word window).

Fortunately, we find consistency in our word embedding scores of entities and those based on MPD co-occurrences. The entities in our survey and words in the mind-perception dictionary are highly correlated with attributed survey agency and experience scores, based on the texts. Hence, we have some validation of both measures since they agree.

We next compare our embedding scores to the dictionary scores in terms of their correspondence to human judgment. You can see those correlations for our baseline specification in the below table. The dictionary method is similar to the embedding method in terms of the measures on their own, and the sum. But for the difference between agency and experience, the dictionary method is much worse. Especially, this is concerning for the validity of the historical analysis.

Table D.4
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 experience, and agency plus experience), with the respective difference and sum of the survey measure. Uses unadjusted survey measures with 255 entities.

Table D.5
Correlation statistics for other surveys.

Entity rating survey	Embeddings	Agency/Experience words	Agency	Experience
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 experience words, as well as entities to assess. We consider the significant factors described in Gray et al. (2007) and Strohming and Jordan (2021) as alternative agency/experience 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 experience 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 experience wordlist contains ‘patience aware emotion feel experience’. All embedding models trained on the 2000 and 2010 decades of COHA

Table D.6
Human validation metrics using the mind perception dictionary.

N	Embeddings	Survey ratings	Agency	Experience
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 experience are derived from the Mind Perception Dictionary (Schweitzer & Waytz, 2021).

Correlation Dictionary	Agency	Experience	Sum	Difference
Word embeddings	0.4	0.36	0.38	0.41

Further, a limitation of the dictionary-based methods is sparsity. Such methods tend to have high precision if the dictionary is carefully compiled to prevent false positives. Yet they tend to have low recall because the concepts of interest (e.g. agency and patience) can be evoked at various levels of specificity.

We confirm this sparsity in our context. Half of the entities in our survey co-occur with less than 50 agency/experience words from the MPD dictionary in the two most recent decades in COHA. To put this in

context, the last two decades in COHA contain a total of 79Mio words. Thus, 50 co-occurrences is extremely rare and those special contexts are not likely to be representative of generic language or perceptions.

In our longitudinal analysis, we find this sparsity problem to be even more pronounced. We find that historically, in most decade, most entities in our list do not co-occur with more than 10 MPD terms. Given the size of COHA, these co-occurrences are extremely rare and not enough to generate confident insights. Word embeddings alleviate the sparsity problem.

For these reasons, we measure perceived agency and experience in language using embeddings methods instead of dictionary-based methods.

Table D.7
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.

Table D.8
Sorted entities by experience difference language and survey.

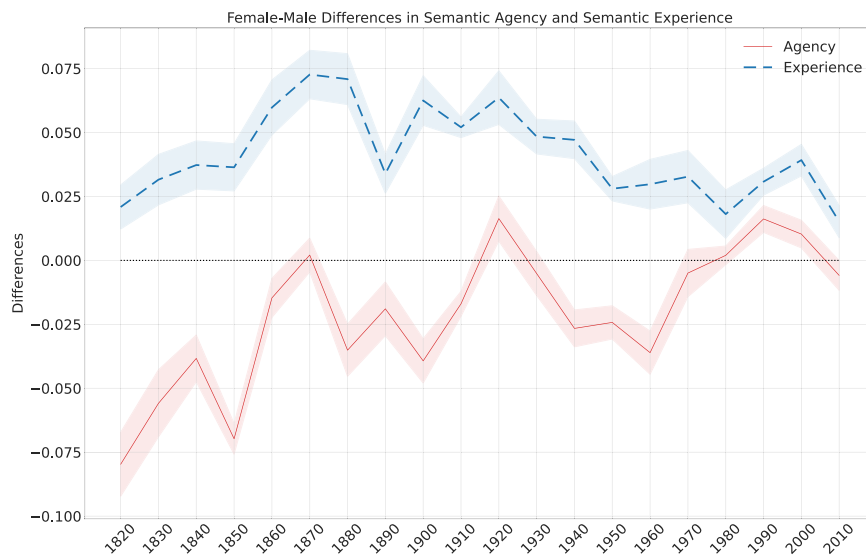
Entity	Experience rank language - Rank survey
chimpanzee	-15
primate	-11
human	-8
elephant	-5
grandfather	-4
shark	-4
man	-2
father	-2
infant	-2
grandmother	-2
monkey	-2
ant	-1
woman	0
kitten	0
frog	0
girl	1
mother	1
dog	1
baby	2
beetle	2
fetus	2
mom	2
puppy	3
bird	3
insect	3
mouse	4
dad	5
boy	5
cat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

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

Table D.9
Correlation of language and survey measures (excluding words which can also be verbs).

Survey dataset	N	Agency			Experience		
		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)

A. GENDER DIFFERENCES IN AGENCY AND EXPERIENCE PERCEPTIONS



B. AGENCY AND EXPERIENCE PERCEPTIONS ABOUT DOMESTICATED ANIMALS

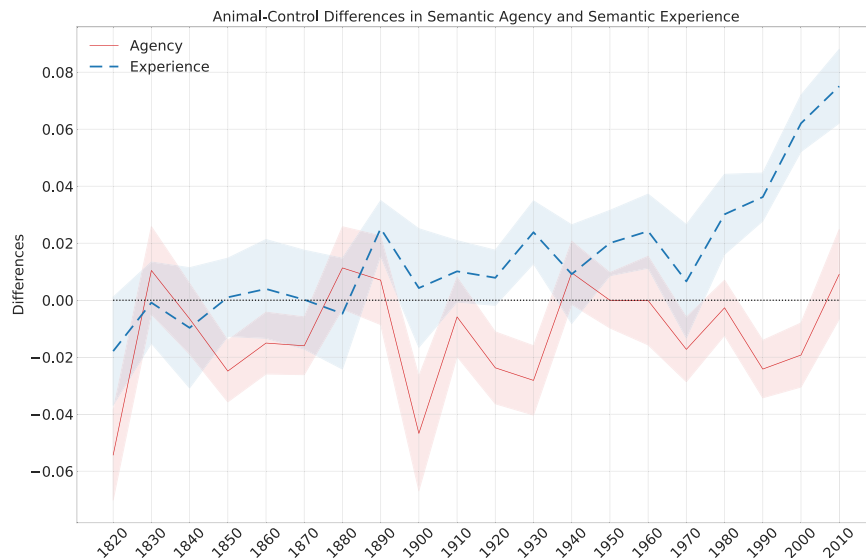
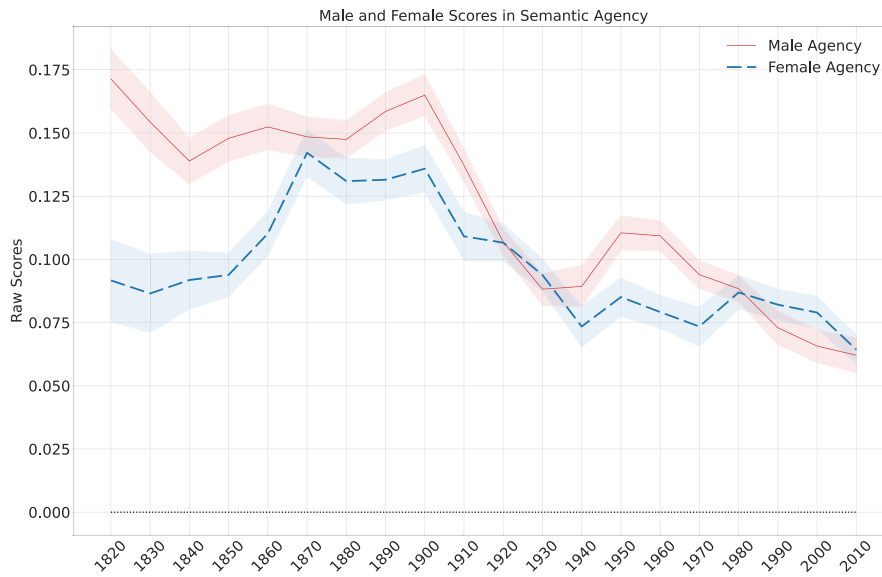


Fig. E.1. Historical analysis without smoothing
Notes. Main results without smoothed time series. Panel A shows the difference by gender (between a set of male words and set of female words). Panel B shows the difference for domesticated animals (a set of words referring to domesticated animals) and a list of wild animals as the control-group words. Agency is shown as red solid line, while experience is shown as blue dashed line. Confidence intervals (95%) produced by bootstrapped embeddings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A. PERCEIVED AGENCY SCORES BY GENDER



B. PERCEIVED EXPERIENCE SCORES BY GENDER

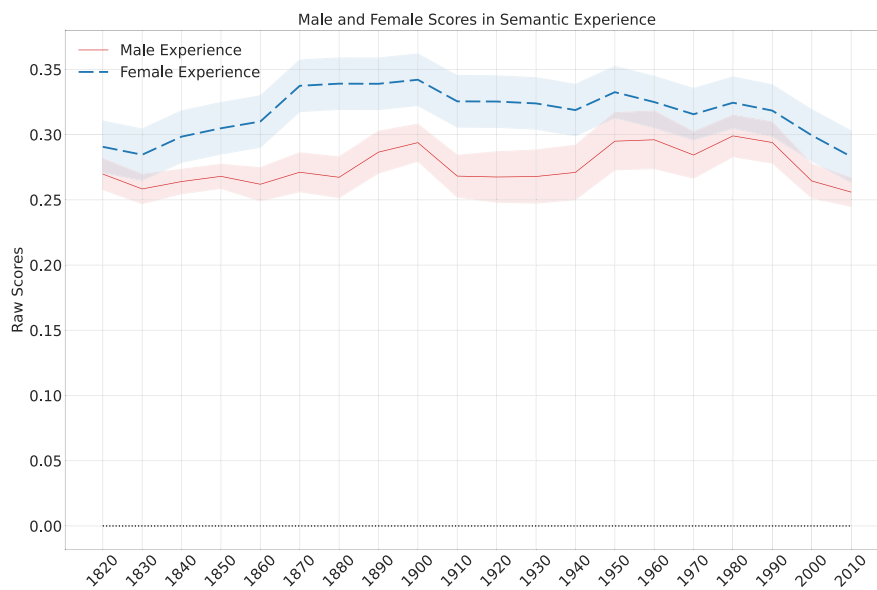


Fig. E.2. Historical analysis of gender, levels rather than differences.

Notes. Agency and experience scores by gender. Panel A shows perceived agency scores by gender; panel B shows perceived experience scores by gender. Female scores are shown in blue dashed line, while male scores are shown in red solid line. Confidence intervals (95%) produced by bootstrapped embeddings. Results smoothed via moving average using a window size of 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

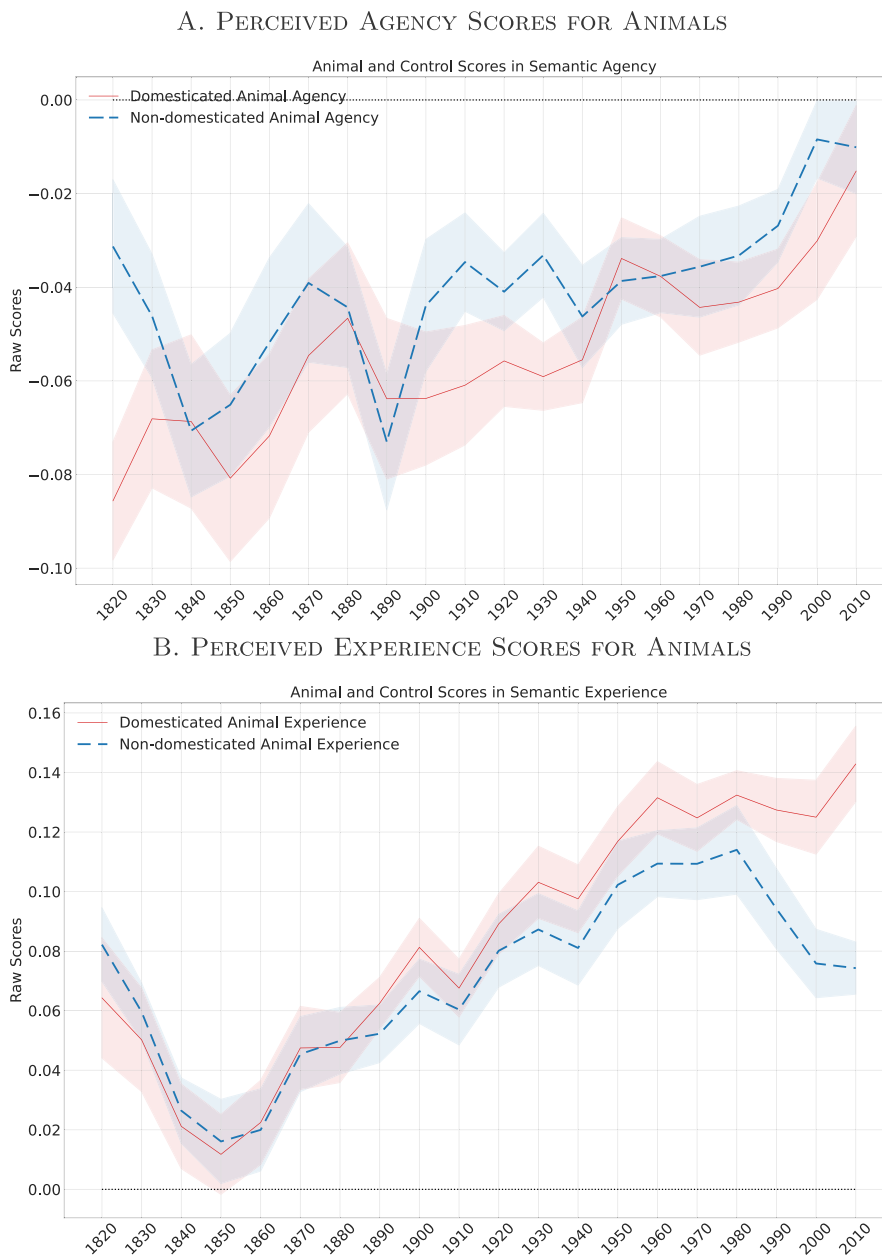
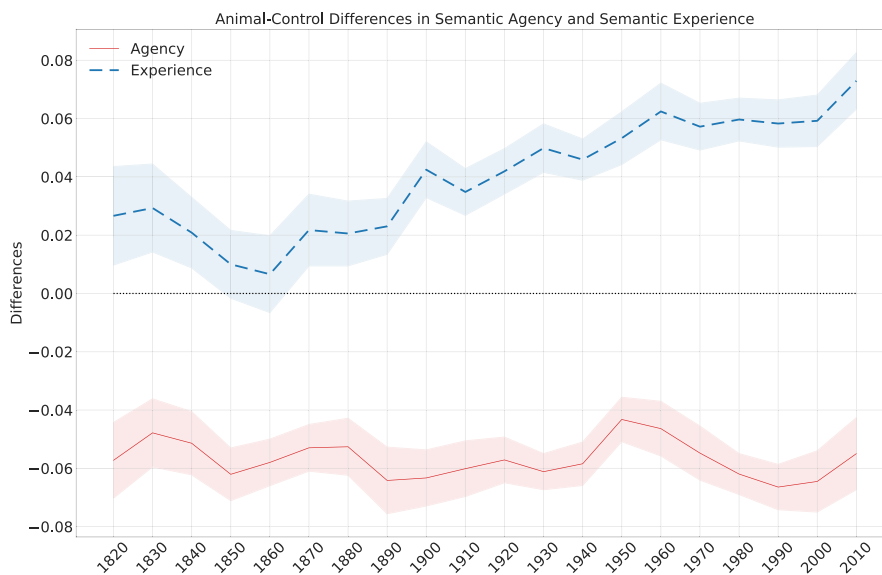


Fig. E.3. Historical analysis of animals, levels rather than differences.

Notes. Agency and experience scores for a set of domesticated and wild animals. Panel A shows perceived agency scores; panel B shows perceived experience scores. Scores for domesticated animals are shown in red solid line, while scores for wild animals are shown in blue dashed line. Confidence intervals (95%) produced by bootstrapped embeddings. Results smoothed via moving average using a window size of 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A. PERCEIVED AGENCY AND EXPERIENCE DIFFERENCES (SMOOTHED)



B. PERCEIVED AGENCY AND EXPERIENCE DIFFERENCES (NOT SMOOTHED)

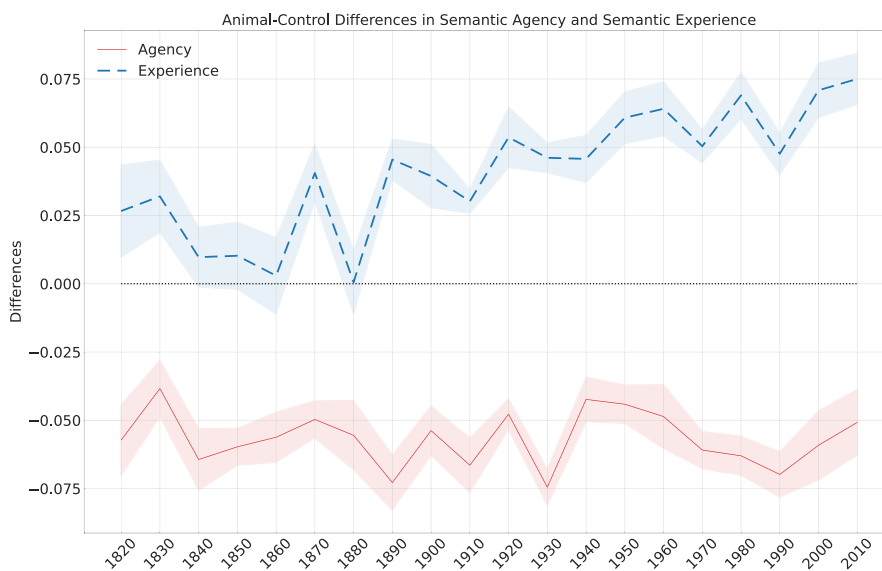


Fig. E.4. Historical analysis of domesticated animals, relative to all other entities.

Notes. Agency and experience scores for a set of domesticated animals and a control group (control consists of all non-animal entities in our survey. Scores for domesticated animals are shown in red solid line, while scores for wild animals are shown in blue dashed line. Confidence intervals (95%) produced by bootstrapped embeddings. Panel A shows results smoothed via moving average using a window size of 2, panel B shows a non-smoothed version. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table E.1
Statistical tests for longitudinal analysis.

(a) Female–male differences				
Year	Agency		Experience	
	Mean	p-value	Mean	p-value
1820	-0.08	<0.0001	0.021	0.0011
1830	-0.056	<0.0001	0.032	0.0001
1840	-0.038	<0.0001	0.037	<0.0001
1850	-0.07	<0.0001	0.036	<0.0001
1860	-0.015	0.0045	0.06	<0.0001
1870	0.002	0.5494	0.073	<0.0001
1880	-0.035	0.0001	0.071	<0.0001
1890	-0.019	0.0065	0.034	<0.0001
1900	-0.039	<0.0001	0.063	<0.0001
1910	-0.017	0.0001	0.052	<0.0001
1920	0.016	0.0056	0.064	<0.0001
1930	-0.005	0.2733	0.048	<0.0001
1940	-0.027	0.0001	0.047	<0.0001
1950	-0.024	<0.0001	0.028	<0.0001
1960	-0.036	<0.0001	0.03	0.0002
1970	-0.005	0.3224	0.033	0.0001
1980	0.002	0.3147	0.018	0.0046
1990	0.016	0.0002	0.031	<0.0001
2000	0.01	0.0048	0.039	<0.0001
2010	-0.006	0.0814	0.015	0.0011

(b) Domestic–wild animal differences				
Year	Agency		Experience	
	Mean	p-value	Mean	p-value
1820	-0.054	0.0001	-0.018	0.098
1830	0.01	0.2164	-0.001	0.91
1840	-0.007	0.3321	-0.01	0.392
1850	-0.025	0.0015	0.001	0.8878
1860	-0.015	0.0238	0.004	0.6652
1870	-0.016	0.013	0	0.9831
1880	0.011	0.1587	-0.005	0.6466
1890	0.007	0.3991	0.025	0.0007
1900	-0.047	0.0014	0.004	0.6964
1910	-0.006	0.4309	0.01	0.0988
1920	-0.024	0.0051	0.008	0.1459
1930	-0.028	0.0014	0.024	0.0022
1940	0.01	0.1238	0.009	0.3274
1950	0	0.9976	0.02	0.0076
1960	0	0.9876	0.024	0.0053
1970	-0.017	0.0154	0.007	0.5307
1980	-0.003	0.604	0.03	0.0023
1990	-0.024	0.0011	0.036	<0.0001
2000	-0.019	0.0085	0.062	<0.0001
2010	0.009	0.2862	0.075	<0.0001

Means and p-values from a t-test comparing the bootstrapped mean to zero, by year.

Appendix E. Additional results on longitudinal analysis

See Figs. E.1–E.4 and Table E.1.

References

Alatrash, R., Schlechtweg, D., Kuhn, J., & Schulte im Walde, S. (2020). CCOHA: Clean corpus of historical American english. In *Proceedings of the 12th language resources and evaluation conference* (pp. 6958–6966). Marseille, France: European Language Resources Association.

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

Ash, E., Stambach, D., & Tobia, K. (2023). Replication package: What is (and was) a person? Evidence on historical mind perceptions from natural language. https://github.com/dominiksinsaarland/Evidence_Historical_Mind_Perceptions_Natural_Language.

Atari, M., & Henrich, J. (2023). Historical psychology. *Current Directions in Psychological Science*, 32(2), 176–183.

Atari, M., Reimer, N. K., Graham, J., Hoover, J., Kennedy, B., Davani, A. M., Karimi-Malekabad, F., Birjandi, S., & Dehghani, M. (2022). Pathogens are linked to human moral systems across time and space. *Current Research in Ecological and Social Psychology*, 3, Article 100060.

Baker, L. (2000). *Cersons and bodies: A constitution view*. Cambridge University Press.

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

Brunet, M.-E., Alkalay-Houlihan, C., Anderson, A., & Zemel, R. (2019). Understanding the origins of bias in word embeddings. In K. Chaudhuri, & R. Salakhutdinov (Eds.), *Proceedings of machine learning research: Vol. 97, Proceedings of the 36th international conference on machine learning* (pp. 803–811). Long Beach, California, USA: PMLR, URL <http://proceedings.mlr.press/v97/brunet19a.html>.

Caliskan, A., Bryson, J. J., & Narayanan, A. (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.

Charlesworth, T. E. S., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from google books. *Proceedings of the National Academy of Sciences*, 119(28), Article e2121798119.

Chuan, C.-H., Tsai, W.-H. S., & Cho, S. Y. (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).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

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

Ellis, A. (2011). Citizens united and tiered personhood. *The John Marshall Law Review*, 44.

England, P., Levine, A., & Mishel, E. (2020). Progress toward gender equality in the United States has slowed or stalled. *Proceedings of the National Academy of Sciences of the United States of America*, 117.

Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114, 864–886.

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

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644.

Garten, J., Hoover, J., Johnson, K. M., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2018). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis: Distributed dictionary representation. *Behavior Research Methods*, 50, 344–361.

Gennaro, G., & Ash, E. (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.

Goodwin, G., & Benforado, A. (2015). Judging the goring ox: Retribution directed toward animals. *Cognitive Science*, 39(3), 619–646.

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

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

Gray, H., Gray, K., & Wegner, D. (2007). Dimensions of mind perception. *Science*, 315, 619.

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, 101, 1207–1220.

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

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., Leskovec, J., & Jurafsky, D. (2016). 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)* (pp. 1489–1501). Berlin, Germany: 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., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, 17(10), 847–853, PMID: 17100784.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 10, 252–264.

Haslam, N., & Loughnan, S. (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.

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

- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905–949.
- Levy, O., Goldberg, Y., & Dagan, I. (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., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Vol. 26, Advances in neural information processing systems* (pp. 3111–3119). Curran Associates, Inc..
- Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology*, 72, 717–749.
- 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., Danforth, C. M., & Dodds, P. S. (2015). Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLoS One*, 10(10), Article e0137041.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. 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., & Sojka, P. (2011). *Vol. 3, Gensim-python framework for vector space modelling*. (2), NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic.
- Resinger, D., Rudinger, R., Ferraro, F., Harman, C., Rawlins, K., & Van Durme, B. (2015). Semantic proto-roles. *Transactions of the Association for Computational Linguistics*, 3, 475–488.
- Rupp, L., & Taylor, V. (1987). *Survival in the Doldrums: The American Women's Rights Movement, 1945 to the 1960s*.
- Schweitzer, S., & Waytz, A. (2021). 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 (Washington, DC)*, 150(8), 1642–1672.
- Singer, P. (1973). *Animal liberation*.
- Stammbach, D., Antoniak, M., & Ash, E. (2022). Heroes, villains, and victims, and GPT-3-automated extraction of character roles without training data. arXiv preprint arXiv:2205.07557.
- Strohming, N., & Jordan, M. (2021). Corporate insecthood.
- Sunstein, C. R., & Nussbaum, M. C. (2005). *Animal rights: Current debates and new directions* (p. 352). Oxford University Press.
- Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, 22(3), 201–212.
- Tarasenko-Struc, A. (2020). On Inhumanity: Dehumanization and How to Resist It. <http://dx.doi.org/10.1093/pq/pqaa068>, pqaa068,
- Thornton, M. A., Wolf, S., Reilly, B. J., Slingerland, E. G., & Tamir, D. I. (2022). The 3d mind model characterizes how people understand mental states across modern and historical cultures. *Affective Science*, 3(1), 93–104.
- Wang, S.-Y. N., & Inbar, Y. (2021). Moral-language use by US political elites. *Psychological Science*, 32(1), 14–26.
- Watanabe, K. (2021). Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. *Communication Methods and Measures*, 15(2), 81–102.
- Weisman, K., Dweck, C. S., & Markman, E. M. (2017). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, 114(43), 11374–11379.
- Willke, J., & Andrusko, D. (1988). Personhood redux. *Hastings Center Report*, 18(5), 30–33.
- Zurlo, R., & Goldberg, A. (1994). *Animals and alternatives in testing: history, science, and ethics*.