

CORPUS MANUAL

Dialogic Density in the Books of William James and John Dewey

Alexander Klein, McMaster University

Study Overview

I employed a machine learning algorithm (the Stanford Named Entity Recognizer, or NER; see Finkel, Grenager, and Manning 2005) to examine the relative rates at which William James and John Dewey mention other persons in representative samples of their respective writings. The NER attempts to tag words and phrases in a corpus with either PERSON, ORGANIZATION, or LOCATION. I created two corpora, one for each author's key monographs and mono-authored collections. I then assembled databases for each book in each corpus. The databases collect every PERSON tag found using NER in each book. This allows me to model the relationship between the rate at which names of persons appear in each corpus per every 1000 words. I call this rate a corpus's "dialogic density."

Why model this relationship? In philosophy, pragmatists like James and Dewey both have a reputation as radicals who are more interested in changing the subject on old philosophical problems than in engaging. For experienced readers of James, however, this description might seem less apt, since his work is saturated in references to other persons, including to their ideas, theories, and data. This study aims to assess whether there are significant differences in the rates at which these two philosophers mention other figures in their writing. My hypothesis is that James is significantly more likely to reference other persons in his writing than is Dewey.

I further discuss the significance of this comparison between the relative rates at which James and Dewey mention others in (Klein Forthcoming).

Corpora Contents

There are two corpora in this study. The corpora consist of monographs and mono-authored essay collections by William James and John Dewey, respectively.

Because the study aims assess the reputation *in philosophy* that pragmatists have of being radicals who change the subject rather than engage old problems, I used the bibliographies from the *Stanford Encyclopedia of Philosophy* (SEP) entries on James and Dewey to construct my corpora. For each figure, I included in the relevant corpus every monograph and mono-authored collection referred to in that figure's main SEP page.

The SEP is the field's preeminent encyclopedia, and it routinely gets highly respected experts to write entries. That is true of the James and Dewey entries, which were written by Russell Goodman and David Hildebrand, respectively. I take the SEP to give a good indication of what

works informed philosophers have generally paid most attention to in James and Dewey's respective oeuvres.

William James Corpus

Following Goodman's SEP entry, my James corpus contains the full text of all books and essay collections published during James's lifetime, plus two posthumous publications—*Essays in Radical Empiricism*, which collects some of James's metaphysical essays mostly from 1904 – 1905, and *Some Problems of Philosophy*, a textbook on philosophy left unpublished when James died in 1910. Both works were published in 1911.

Note that Goodman's SEP entry on James discusses and refers to all these works. Two deserve a brief comment. Goodman discusses and refers to *Essays in Radical Empiricism*, but neglects to include that work in the bibliography, which looks to be an oversight. The same is true of *Psychology: The Briefer Course*. I have included both works in my study because they are discussed in the James SEP entry.

James's essay "La Notion de Conscience" appeared in *Essays in Radical Empiricism* (and indeed, was originally published) in French. It is excluded from my study, since the NER is trained on an English-language corpus.

Year	Work Title	Total Words
1890	<i>The Principles of Psychology</i>	564,071
1892	<i>Psychology: Briefer Course</i>	157,539
1897	<i>The Will to Believe and Other Essays in Popular Philosophy</i>	100,213
1899	<i>Talks to Teachers on Psychology and to Students on Some of Life's Ideals</i>	62,902
1902	<i>The Varieties of Religious Experience</i>	187,468
1907	<i>Pragmatism</i>	51,469
1909	<i>The Meaning of Truth</i>	58,067
1909	<i>A Pluralistic Universe</i>	61,746
1911	<i>Some Problems of Philosophy</i>	44,359
1911	<i>Essays in Radical Empiricism</i>	49,433
	SUM	1,337,267

John Dewey Corpus

My Dewey corpus includes all monographs or mono-authored collections published during Dewey's lifetime that are cited in David Hildebrand's SEP article, totaling 21 volumes. I did not include multi-authored volumes. Dewey lived much longer than James, and his publications are far more voluminous. Hildebrand's SEP entry is understandably selective, then, in the Dewey works cited. To take two examples, despite the wide range of Dewey's works Hildebrand does

cite, neither Dewey's 1888 *Leibniz's New Essays* nor Dewey's 1915 *German Politics and Philosophy* appear in his entry.

One would obviously get slightly different dialogic density scores if one used different criteria for sampling Dewey's writing (just as one would get slightly different scores if one sampled James's writing differently). However, I take Hildebrand's SEP bibliography to give us a good reflection of the Dewey volumes that have been most prominent in philosophy. Again, my study aims to elucidate James and Dewey's reputation specifically in philosophy, and Hildebrand's bibliography should give us a reliable tool for sampling Dewey's most philosophically prominent writing.

Note that Hildebrand discusses and refers to Dewey's 1910 essay collection *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought* but neglects to include that work in the bibliography, which looks to be an oversight. I have included this work in my study.

Since neither *Influence of Darwin* nor the 1916 collection *Essays in Experimental Logic* appear as a self-standing work in Dewey's *Collected Papers*, I reconstructed each work from the individual chapters as they appear in various places in *Collected Papers*.

Year	Work	Total Words
1887	<i>Psychology</i>	123,522
1891	<i>Outlines of a Critical Theory of Ethics</i>	57,220
1894	<i>Study of Ethics</i>	50,483
1899	<i>The School and Society</i>	38,301
1910	<i>How We Think</i>	62,861
1910	<i>The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought</i>	66,835
1916	<i>Democracy and Education</i>	139,803
1916	<i>Essays in Experimental Logic</i>	104,980
1920	<i>Reconstruction in Philosophy</i>	49,644
1922	<i>Human Nature and Conduct</i>	85,692
1925	<i>Experience and Nature</i>	125,517
1927	<i>The Public and its Problems</i>	53,400
1929	<i>The Quest for Certainty</i>	97,169
1930	<i>Individualism, Old and New</i>	30,110
1934	<i>A Common Faith</i>	21,996
1934	<i>Art as Experience</i>	135,947
1935	<i>Liberalism and Social Action</i>	23,040
1938	<i>Experience and Education</i>	21,802
1938	<i>Logic</i>	193,698

1939	<i>Freedom and Culture</i>	49,028
1939	<i>Theory of Valuation</i>	24,342
	SUM	1,555,390

How the Corpora Were Constructed

I used the Past Masters database to access the full text of the *Works of William James* and the *Collected Works of John Dewey* (Dewey 2003, James 2008). The text from each book was copied into a separate text file. I excluded any words written by the editors of either James's *Works* or Dewey's *Collected Works*. I included prefaces or introductions written by either James or Dewey. I excluded indexes and tables of contents, whether produced by James, Dewey, or their editors. My corpus thus only includes words written by James or Dewey. I did not trim, lowercase, remove stop-words, remove page numbers, or do any other pre-processing.

Dewey, John. 2003. *The Collected Works of John Dewey, 1882-1953 (2nd Release). Electronic Edition*. Edited by Boydston, Jo Ann and Larry Hickman. Charlottesville, Virginia, U.S.A.: InteLex Corp.

Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. 2005. "Incorporating Non-Local Information into Information Extraction Systems by Gibbs Sampling." *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*: 363-70.

James, William. 2008. *The Works of William James. Electronic Edition*. Edited by Burkhardt, Frederick H., Fredson Bowers, and Ignas K. Skrupskelis. Charlottesville, Virginia: InteLex Corp.

Klein, Alexander. Forthcoming. "Introduction." In *Oxford Handbook of William James*, edited by Alexander Klein. New York: Oxford University Press.