University of Reading

Justification: insights from corpora

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Grindrod, J. ORCID: https://orcid.org/0000-0001-8684-974X
(2024) Justification: insights from corpora. Episteme, 21 (3). pp. 794-818. ISSN 1750-0117 doi: 10.1017/epi.2022.44
Available at https://centaur.reading.ac.uk/108510/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1017/epi.2022.44

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

CrossMark

Episteme (2022), 1–25
doi:10.1017/epi.2022.44

episteme

ARTICLE

Justification: Insights from Corpora

Jumbly Grindrod

University of Reading, Reading, UK
Email: j.grindrod@reading.ac.uk

(Received 21 January 2022; revised 20 May 2022; accepted 13 October 2022)

Abstract

In this paper I use insights from exploratory analyses on large English language corpora to consider the extent to which there is a widely used ordinary notion of justification that attaches to beliefs. I will show that this has ramifications for one broad approach to theorising about justification – the folk justification approach. I will argue that the corpus-based findings presented pose a challenge to the folk justification approach insofar as they suggest that "justify" is not widely used talk about the justification of our beliefs. I will conclude by presenting the possible solutions to this challenge, and remarking on their feasibility.

Keywords: Corpus analysis; justification; ordinary language philosophy; experimental philosophy; metaphilosophy

1. The Explanandum for a Theory of Justification

It is sometimes said that justification lies at the heart of epistemological study. For instance, Fumerton states that "the concept of justification may be the most fundamental in epistemology" (2005: 204). Even today, where epistemology is a more varied domain than just the project of theorising about knowledge and justification, it is still fair to say that the latter is taken to be one of the central concepts. A key reason for thinking this is that justification is one of the key notions that distinguishes epistemology from the psychological topic of capturing how we do in fact form beliefs. With justification comes an element of normativity, a domain within which the question is not only how we do in fact form our beliefs, but how we ought to.

Corpus analysis is an empirical methodology that has been attracting increasing levels of interest across a range of disciplines, particularly with the rise of the so-called *digital humanities*. In this paper, I will show how this methodology can contribute to the more traditional epistemological investigation of justification. In particular, a corpus analysis of the lemma "justify" provides us with a new way of investigating our ordinary epistemic practices. Before I can make the case for this, however, some groundwork is required in order to distinguish between different kinds of theories of justification. Rather than distinguishing along familiar lines between those theories that posit some foundational level of justification and those that don't, or between theories that

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

have some internalist requirement and those that don't, I instead want to delineate in terms of what the various theories take the explanandum to be. One feature of much of the debate on justification is that there has been a lack of clarity on this matter, and so it is unsurprising that some have felt that the justification debate in epistemology is in some sense defective (Alston 1993; Cohen 2016). We can in fact distinguish between three broad approaches to justification that differ in terms of their explanandum, and the corpus analytical findings to be presented are relevant primarily to only one of them.

The first approach takes there to be a notion of justification that is already circulating within our community i.e. that notion that serves as the standard of epistemic evaluation by which we judge our own beliefs and the beliefs of others. A theory of justification, then, seeks to capture the ordinary notion of epistemic evaluation. I will call this the *folk justification approach*. One common way of embarking upon an investigation of folk justification is to attempt a conceptual analysis of justification, by providing a set of necessary and sufficient conditions for when a given belief is justified. The formulation of these conditions would be guided by consideration of hypothetical cases and whether or not beliefs in these cases would count as justified or not. This familiar approach would be to treat justification in the same way that knowledge has commonly been treated. Of course, conceptual analysis has faced a number of criticisms. For instance, certain experimental findings suggest that the kinds of intuitions that philosophers typically rely upon in conceptual analysis are in fact dependent upon extraneous factors such as ordering effects, socio-economic background, ethnic background etc. (Weinberg *et al.* 2001; Swain *et al.* 2008; Beebe and Buckwalter 2010).[1] This has led to a wider debate about the role of intuitions in philosophy and whether conceptual analysis is a suitable philosophical method (Williamson 2007; Cappelen 2012; Deutsch 2015). But whether conceptual analysis or the elicitation of intuitions is a suitable method in philosophy is not essential to the folk justification approach as it is understood here – what is essential is that there is some notion of epistemic evaluation that is part of our ordinary epistemic practices and that this notion is the explanandum for a theory of justification.

We can certainly find support for the folk justification approach in the literature. For instance, Goldman (1979) was explicit that he was seeking an account of the notion of justification as ordinarily used and that a theory of justification "will be a set of principles that specify truth conditions for the schema 'S's belief in p at time t is justified' i.e. conditions for the satisfaction of the schema in all possible cases" (Goldman 1979: 3). Many authors have since explicitly endorsed the view that justification is reflected in our ordinary epistemic judgments. To select just one further example, Ichikawa (2014: 188) states "our ordinary epistemic judgments respect a distinction between knowledge and non-knowledge justified belief". This would also be a natural way of understanding some of the key thought experiments regarding justification. For instance Lehrer's (1990: 163) Mr Truetemp case, BonJour's (1985: 38) clairvoyant case, and Lehrer and Cohen's (1983) new evil demon case are all thought experiments whereby we consider whether a given subject has a justified belief.[2] In relying on our ordinary intuitions in

[1]The results from Weinberg *et al.* and Swain *et al.* have failed to replicate, however (Kim and Yuan 2015; Seyedsayamdost 2015; Ziółkowski 2021).

[2]Lehrer phrases his discussion in terms of whether Mr Truetemp knows or not rather than whether he has a justified belief or not. But I take it as uncontroversial that some take the thought experiment to indicate something intuitive about justification i.e. that Mr Truetemp does not have justification for his beliefs.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

order to reach a verdict regarding these cases, the thought presumably is that we are able to employ justification as the ordinary notion of epistemic evaluation.[3]

Some have adopted a particularly linguistic form of the folk justification approach, by investigating the linguistic properties of "justify" and related word forms. For instance, Kvanvig and Menzel (1990) urge epistemologists to pay greater attention to the various different locutions in which "justify" appears, and so they provide a theory of a basic notion of justification that can then be used to explain the various locutions. More recently, Hawthorne and Logins (2020) have investigated the semantic gradability of "justified". They argue that "justified" is an absolute gradable adjective that is only derivatively associated with a scale, and so speakers and hearers must construct an appropriate scale on-the-fly depending on the concerns most salient to them. Certainly it is true that not all advocates of the folk justification approach choose to place such importance on the linguistic properties of sentences containing "justify", but it is worth keeping this possibility in mind as it will be pertinent to the discussion later in the paper.

In this paper, I aim to use insights from large corpora to pose a challenge to the folk justification approach. However, while the folk justification approach is common, it is not the only way in which justification is conceived of. Here I will consider two prominent alternatives. First, epistemic justification could be conceived of as a theoretical notion i.e. as whatever it is that one must add to true belief in order to obtain knowledge. This can often feel like the way we introduce new students to the notion of justification, by pointing out that true beliefs can fall short of knowledge if, for instance, they are just lucky guesses. As this approach develops justification in terms of its relation to knowledge, I will label it the *Theory of Knowledge* (TOK) approach. One problem with the TOK approach is that it is hard to make sense of the Gettier (1963) problem if Gettier cases are supposed to be cases of justified true belief that fall short of knowledge. Of course, one post-Gettier tactic is to claim that knowledge must be justified true belief plus some other property, and nothing I will say rules out that possibility. But notice that now in even claiming this, we seem to be envisioning justification in some way other than as what we add to true belief to get knowledge, and to that extent, we have moved away from the TOK approach. For instance, many view it as a requirement on a theory of justification that it is possible to have false justified beliefs, and this might constrain the theoretical moves available in taking knowledge to be justified true belief plus something else. But notice that in even taking on this requirement, we seem to have moved away from the TOK approach to some extent – we are placing restrictions on our theory of justification that don't arise from it being a constituent of knowledge.[4] A second problem with this approach is that some explicitly deny that

[3]This focus on justification has arguably become more prominent as perceived progress in the Gettier debate has stalled. As scepticism has become more widespread that a Gettier-proof definition of knowledge is possible, the thought is that it would be more worthwhile to shed light on the nature of justification as the property we typically try to ensure our beliefs have rather than focusing on the property that rules out the possibility of Gettier cases, particular and artificial as they tend to be. The rise of the knowledge-first programme has also encouraged attempts to theorise about justification in this manner, as the provision of an account of justification in terms of knowledge is sometimes taken to be a beneficial result of the programme (Bird 2007; Sutton 2007; Ichikawa 2014).

[4]Perhaps I am being unfair to the TOK approach here. Perhaps in considering Gettier cases we may come to realise that there are indeed two or more properties that knowledge has and that mere true belief lacks, and that we can label one of these as justification. Nothing I will say will rule out this possibility. But I do think that it is difficult to identify clear instances in the literature where this approach has been taken

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

justification is even necessary for knowledge.[5] Again, in order for this theoretical move to even make sense, it must be that we are envisioning justification as something other than as the difference between knowledge and true belief.

A second alternative to the folk justification approach is to develop a notion of justification from a theory of epistemic value. *Instrumentalism* about justification is roughly the view that those beliefs that are epistemically justified are those beliefs that are conducive to the epistemic goal or goals. If we take this as our starting point for a theory of justification, then it is conceivable that we could construct a theory of justification based on a theory of epistemic value. If we are able to identify what is epistemically valuable, and then identify the minimally acceptable ways of promoting that value in terms of belief-formation, then we would thereby have some account of epistemic permissibility that could serve well enough as a theory of epistemic justification. I won't say too much about this approach here, other than that while there are advocates of this approach in the literature (Alston 1985; BonJour 1985; Foley 1987), it is unclear the extent to which a pure version of this approach has been adopted i.e. one not combined with the previous two approaches mentioned. To take one example, Foley (1987, 2001, 2008) is an instrumentalist of this stripe who urges that we must theorise about epistemic rationality, with no initial regard for whether it has some conceptual connection to knowledge. Instead, we should try to capture epistemic rationality in its own right and its connections to other forms of rationality. However, in doing so, Foley takes it to be a success criterion for the theory that it captures "the everyday assessment of rationality of opinions, which tend to focus on whether individuals have been responsible in forming their opinions rather than on whether they have satisfied the prerequisites of knowledge" (Foley 2001: 214). To the extent that the theory is not merely sensitive to considerations about epistemic value, but also to the manner in which we evaluate beliefs in the everyday, the approach to justification described here is combined with the conceptual analysis approach.

In sum, we can distinguish between three positive approaches to justification: the folk justification approach, the TOK approach, and the instrumentalist approach. It is the folk justification approach that is of primary concern in this paper, and I don't take what I say here to be particularly damaging to the alternative approaches. But it is important to emphasise that *pure versions* of the alternative approaches are relatively rare in the literature. In the case of the TOK approach, Gettier famously argued that we shouldn't think that justification is the difference between knowledge and mere true belief, and for many this is one of the few points of agreement in epistemology. In the case of instrumentalism, there usually is appeal to some ordinary notion of epistemic evaluation, as we saw in the case of Foley. I haven't shown that pure versions of the alternative approaches are not viable; I merely want to emphasise that the folk justification approach is a very common one and that there is good reason to adopt it.

In this paper, I will show how corpus analysis can serve a theory of justification by shedding light on the extent to which there is an ordinary notion of justification. In particular, I will analyse the use of the lemma "justify" and its related word forms in large

rather than justification being conceived in some more ordinary sense. Thanks to an anonymous reviewer for pressing this point.

[5]Lewis (1996) is one example of someone who denies that knowledge requires justification, although he notes that some may define justification in the way I have described. It is often denied that justification is necessary for knowledge by externalists, who view justification as having some kind of internalist condition.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

English corpora.[6] After presenting a range of insights from a few different corpora, I will present a challenge that any advocate of the folk justification approach must meet. As a preview: the challenge for the folk justification approach is to give some account of how the ordinary notion of justification is spoken about. In this paper, I will explore the most straightforward possibility that speakers use "justify" to speak about epistemic justification. I will suggest that the evidence available from various corpora suggest that "justify" is not being widely used to talk about the epistemic justification of beliefs.

2. Corpus Analysis: Principles and Tools

Corpus linguistics attempts to generate linguistic insight on the basis of evidence drawn from linguistic corpora. The history of corpus linguistics is certainly an interesting one,[7] but it is fair to say that the approach has become much more powerful in the last few decades as large corpora – as well as the computational methods used to analyse those corpora – have become more available and more easily accessible. In terms of the study of meaning, it is often said that a starting principle within corpus linguistics is the distributional hypothesis: that there is a correlation between the meaning of a term and its distribution across a corpus, or as Harris (1954: 156) puts it:

> if we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.

This hypothesis could be interpreted in stronger or weaker senses. One strong way of interpreting the hypothesis is as claiming that, in the fullest sense, the meaning of a term can be captured purely by an account of its distribution across a suitably large corpus, and that a theory of meaning for a language should be distributional in nature.[8] It is something like this stronger sense that drives *distributional semantics*: a form of corpus linguistics where the meanings of terms are typically represented as vectors across a high dimensional space. For the purposes of this paper, a much weaker version of the distributional hypothesis is required. The thought is that we can gain insight into how a term is used by paying careful attention to (i) the kinds of texts the term is used in and (ii) the other terms that typically collocate with the term. In doing so, we can thereby draw tentative conclusions about the meaning of the term. This kind of inference will be used repeatedly throughout this paper, but I will be explicit at each stage of the kind of reasoning involved.

It should be noted that although this paper can be viewed as a form of experimental philosophy insofar as it will introduce empirical considerations from large corpora, the methodology used here is largely exploratory. That is to say, the conclusions drawn in this paper are not reached via statistical tests of significance concerning the relationship between a dependent and independent variable. This is not because such tests are not possible in corpus analysis, even if corpus analysis is particularly suited to exploratory

[6]A lemma is the base word form that one would typically find in the dictionary and that incorporates various different word forms. For instance, the lemma "know" groups together the word forms "know", "knew", "knows", "knowing" etc.

[7]See McEnery and Wilson (1996: Ch. 1) for an account of how the progress of corpus linguistics as a discipline was affected by the rapid development of the generative tradition led by Chomsky and others.

[8]See Grindrod (Forthcoming) for further discussion.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

investigations.[9] This does affect the kinds of conclusions that one can draw from the investigation, but I nevertheless hope to illustrate that such conclusions can prove useful within epistemology.

Something I hope to show in this paper is that corpus analysis can be a powerful tool for philosophical investigation, particularly considering the previously mentioned criticisms that more traditional methodologies have faced. Unlike the traditional armchair methods, corpus analysis depends primarily on patterns to be found in empirical data rather than on intuitions of the theorist. And unlike survey-based methods, there is no worry about ecological validity as corpora are recordings of authentic language use rather than elicited language use in an experimental setting. This is not to say that corpus analysis is always superior to alternative methodologies, but it is to say that if there are ways in which it can help shed light on philosophical questions, we absolutely should take advantage of that. In recent times, others have attempted to do precisely this (Hansen *et al.* 2019; Sytsma *et al.* 2019; Liao and Hansen 2022), and this paper follows their lead to that extent. However, one distinctive aspect of this paper is in the claim that not only do corpus findings raise a challenge for the folk justification approach, but that corpus analysis can also help the folk justification approach to answer that challenge.

Turning to the kinds of corpora that will be used, we require corpora that are representative of ordinary language use so that we can consider whether a notion of epistemic justification occurs within our ordinary linguistic practices. My focus will be only on English language use, and there are a few English language corpora that are readily available for analysis. I will use three:

(i) Corpus of Contemporary American-English (COCA)

COCA is a large corpus of over 1 billion words of contemporary American-English, collected between 1990–2019. It is evenly balanced across eight different genres of text (Spoken, Fiction, Magazines, Newspapers, Academic Writing, Web, Blogs, TV and movies). This not only ensures that the corpus is fairly representative, but also allows one to investigate how a term is used differently across different genres. Another attractive feature of COCA is that it is easy to explore using the host website (https://www.english-corpora.org/coca/).

(ii) British National Corpus (BNC)

The British National Corpus is a 100 million word text of British English collected from the late 20th century. It is smaller and older than COCA, and while it is separated into a few different sub-corpora, it has not been curated with the same balance across genres. The key benefit of using the BNC is that there is a great deal more information and meta-data about the documents and sub-corpora that make up the BNC, and so we have the opportunity to draw more inferences about the contexts in which a given term (such as "justify") is used, and the way in which the term is distributed across the corpus. This is particularly true given the *BNCWeb* interface made available by the University of Lancaster (http://corpora.lancs.ac.uk/BNCweb/), which I will make use of later in the paper.

[9]Kilgarriff (2005) argues that statistical significance testing should not be used in corpus analysis because the null hypothesis – that the relation between the independent and dependent variables is random – is never true. See Gries (2005) for a response.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

(iii) EnTenTen20

EnTenTen20 is a very large corpus of around 40 billion words of web-based English. It is compiled by scraping English text from the web up until 2020, and so there isn't the same deliberate balance across genres that one finds in COCA nor the detailed metadata of sub-corpora and documents that one finds in the BNC. Even if a large corpus is not necessarily a representative corpus, there is clearly some reason to think that such a large amount of text taken from the web, that is clearly not restricted to a particular genre or topic, would provide a good account of English language use. As we will see, another attractive feature of investigating EnTenTen20 is that this can be done via the program *SketchEngine*. SketchEngine is a text analysis program that comes with over 500 corpora across 90 languages pre-loaded onto it (including EnTenTen20). It is particularly useful for performing collocation analyses that take into account the grammatical relations of a term. For instance, if I want to find out the noun that most frequently serves as the object of the verb "catch", this is easily done (it is "eye", as in "caught my eye").

3. Initial Indications: Comparing the Frequency of "Justify" and "Know"

The folk justification approach is commonly employed with regard to knowledge or "know". That is, it is very common in analytic epistemology to give some account of the concept of knowledge that enjoys wider circulation within our epistemic community. And it seems plausible enough that there is a widely used notion often appealed to across a range of contexts and that could serve as the explanandum of an epistemological theory. Of course, there have been controversies over whether a conceptual analysis of the notion is possible, and over whether a reliance on the intuitions of philosophers in constructing such an analysis is appropriate. Here I will ignore those controversies and just focus on the initial step of selecting the ordinary notion of knowledge as the explanandum of a theory. We may wonder then whether that step could also be taken with justification. A worry one might have is that whereas "know" is one of the most common verbs used in English,[10] "justify" is used far less frequently. In COCA, the lemma "know" is the 39th most frequent lemma with a frequency of 2781.03 per million words, whereas the lemma "justify_v"[11] is the 2,643rd most frequent lemma (frequency 30.41 per mil). The noun "knowledge" is the 841st most frequent lemma (116.17 per mil) while the noun "justification" lies outside of the top 5,000 most frequent lemmas.[12] There clearly is a marked difference between the frequency of knowledge talk and the frequency of justification talk.

This is an initial indication that perhaps justification is not spoken of in ordinary discourse, and certainly not to the same extent that knowledge is spoken of. Others

[10]See Hansen *et al.* (2019) for a careful analysis of the frequency of "know" once discourse marker uses (e.g. "you know, I never watched that show") are taken into account. While such uses do account for a significant proportion of overall uses, the term is still one of the most frequently used verbs even after these uses are discounted.

[11]The lemma "justify" will group together the word forms "justify", "justified", "justifying" etc. However, "justified" is ambiguous between a verb (e.g. "John justified his son's belief in him") and an adjective (e.g. "Mungo had a justified belief"). A lemma only groups together word forms of the same part-of-speech category. So I will use the suffix "_v" to indicate I am talking about the verb and the suffix "_j" to indicate I am talking about the adjective.

[12]Access to the top 5,000 most frequent lemmas is given at https://www.wordfrequency.info/.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

have previously reached (or at least considered) the same conclusion via consideration of the philosophical literature. Plantinga (1990) notes that prior to Gettier, there were very few analyses of knowledge that appeal to justification and so suggests that "it is almost as if a distinguished critic created a tradition in the very act of destroying it" (Plantinga 1990: 45). In his recent data-driven history of philosophy, Brian Weatherson (2020) reaches a similar conclusion.[13] He uses topic-modelling on a dataset of philosophy journal articles taken from 12 prominent journals between 1876–2013. The result is 90 different topics that can be used to group the various articles. More specifically, topic-modelling will take the distribution of the words across the articles and, via the use of an unsupervised learning algorithm, create a number of topics such that each article can be given a probabilistic assignment according to how well it fits under each topic. The topics can be characterised according to a set of keywords that are more likely to turn up in articles under that topic than in the average article. Weatherson's study dives into each of the 90 topics, identifying the subject matter for that topic, and then charting its progress and popularity over time. He reasonably identifies one of the topics as about justification, with the following keywords: [believing, beliefs, epistemically, belief, justification, reliable, justified, reliability, epistemic, goldman, forming, believe, believed, accepting, warrant]. Charting the progress of the topic over time, Weatherson notes that the topic has very few papers from earlier years:

> The first paper that gets a topic probability above one-third is Harry Frankfurt! This tells us something interesting about the background to Gettier's 1963 paper. Just a few years before that paper, there was virtually no discussion of beliefs being justified. It wasn't that Gettier showed a familiar concept couldn't play a role in the analysis of knowledge. He effectively introduced the concept of justification. (Weatherson 2020: 2.76 Justification)

The comments from Plantinga and Weatherson are arguably unfair to Gettier himself, who cites examples of the kind of view he is targeting, but the important issue is whether the notion of justification really only started to be discussed in the philosophical literature in the middle of the twentieth century. As Weatherson suggests, if that is right, this may be initial reason to think that justification is a technical notion introduced by epistemologists.

On the other hand, it would be difficult to draw such conclusions from the state of the philosophical literature alone. After all, that a notion only starts to be discussed within philosophical circles from a given point in time doesn't tell us whether or not the notion was already in wider circulation prior to that time, just never discussed in philosophical contexts (as would obviously be the case with, for instance, *bullshit*). We also shouldn't place too much focus on historical considerations in considering whether the notion of justification is in wider circulation today. It is consistent with the notion only being introduced within the philosophical literature in the mid-twentieth century that it is in wider circulation today or even that it plays some crucial role in our ordinary epistemic lives. This possibility may seem far-fetched, but still it should lead us to turn our focus away from the philosophical history and towards the way the notion is currently employed in ordinary life.

[13]See also Dutant (2015) for a detailed argument that the justified true belief analysis of knowledge is not the traditional conception of knowledge, despite what was thought in the post-Gettier years.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Turning back to the corpus-based evidence, the frequency statistics cited above are really just indications as to the difference between "know" and "justify". It could be argued that "know" is really the outlier here in being a widely used term and that we shouldn't expect all terms amenable to conceptual analysis to be as frequent within a corpus as "know" is. Furthermore, once we look at the kind of terms with similar levels of frequency as the lemma "justify" (in COCA), we find perfectly ordinary terms such as "juice", "joy", "cousin", and "aunt". So while there is a marked difference in frequency between "justify" and "know", the frequency of the former in English corpora does not by itself suggest that it is not a widely circulated notion.

4. Distribution

It can often be useful to look not only at the frequency of a term but also the distribution of a term across a corpus. Some terms will be evenly spread, used at a consistent rate across all documents within the corpus. Others will have an uneven distribution, with higher frequencies in some documents, and lower frequencies or a zero frequency in others. If the term turns up in fewer documents, this is an indication that it is in some sense specialist – it may be that it is part of some dialect, that it concerns a particular subject matter, that it is a technical term for a discipline, or that it belongs to a particular register.

In general, the way to measure distribution across a corpus is to divide the corpus up into parts and measure the frequency of the term in each part. The result can be displayed visually, or it can be used to calculate a dispersion score, of which there are many different kinds available. For example, *Juilland's D* is a dispersion score between 0 and 1 with a score closer to 1 indicating that the term is more widely dispersed. Using the available frequency data on COCA from www.wordfrequency.info, we can easily access the Juilland's D for the lemmas "know" and "justify". "Know" has a Juilland's D of 0.96 while "justify" has a Juilland's D of 0.94. This is an initial indication that both terms are well-dispersed across the corpus, with little to pick between the two of them. However, there are reasons to take this with a pinch of salt. First, a high Juilland's D is very common among the most frequent 5,000 lemmas in COCA, with the average Juilland's D shown in Table 1.

Second, there is reason to think that Juilland's D does not always accurately reflect an uneven distribution when large corpora have been partitioned into many parts (Biber *et al.*, 2016). The scores given above are the result of partitioning COCA into 100 parts, and so it may be that a term like "justify" or "know" is in fact quite unevenly distributed but that this is not reflected by the score. Biber *et al.* (2016) recommend instead using Gries' DP (Gries 2008, 2019) and Gries himself has made this score (and a range of other dispersion scores) available for all word forms in the BNC.[14] Gries' DP (with a *lower score* indicating that the term is more widely distributed) for the various word forms for "justify" and "know" are shown in Table 2.

Here we can more clearly see a difference between the two lemmas, with a marked difference between the Gries' DP scores achieved across the various word forms. This suggests that "justify" is not widely distributed in the way that "know" is.

Moving away from quantitative measures, one natural way to investigate the dispersion of a term across a corpus is to rely upon a corpus that has been compiled with a balance across different kinds of text, as COCA has. For any given term in COCA, we

[14] http://www.stgries.info/research/dispersion/links.html.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 1. Average Juilland's D for 5,000 most frequent lemmas in COCA.

Mean	0.94
Median	0.95
Mode	0.96

Table 2. Gries' DP given for word forms of "know" and "justify" in the BNC.

Word form	Frequency	Gries' DP
known	25140	0.31
knows	8052	0.45
knowing	4710	0.50
know	118661	0.51
knowledge	14343	0.51
knew	23939	0.55
justified	2331	0.61
justify	2024	0.63
justification	1737	0.72
justifying	252	0.91
justifies	233	0.91

can look at the frequency of the term across the eight genres (Spoken, Fiction, Magazines, Newspapers, Academic, Web, Blogs, TV and Movies), and even drill down into sub-genres (e.g. the Blog genre consists in the following sub-genres: Academic, Argument, Fiction, Informational, Instructional, Legal, News, Personal, Promotional, Review, Miscellaneous).

If the distribution of a given term is heavily weighted towards a particular genre or even sub-genre, this may be good reason beyond the bare frequency to think the term is not widely circulated. In particular, if a term's distribution is heavily weighted towards academic writings, this may be good reason to think the term is largely used as a technical term confined to a specific discipline or disciplines. Focusing initially on the distribution of "justify" vs "know", there is a striking difference. The three most frequent genres for "know" are TV/Movies (6,021.49 frequency per mil), Spoken (4,501.43), and Fiction (1,851.46). The three most frequent genres for "justify" are Blog (28.85), Academic (28.46), and Web (24.88). The differences between the two are perhaps made clearer if we compare the top three genres for five terms that one would think are clearly commonly used and five terms that one would think are clearly technical terms within philosophy.

Among the terms that are plausibly widely used, certain genres recur frequently, such as TV/Movies and Spoken, while among the technical terms, the categories of Academic, Blog, and Web dominate (Table 3). Turning back to "know" and "justify", it seems plausible on the basis of this table to group "know" with the common terms while "justify" more plausibly groups with the technical terms. This kind of analysis

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 3. *Most frequent genres in COCA for common terms and philosophical terms.*

Term	1st Genre	2nd Genre	3rd Genre
Common terms			
Good	TV/Movies	Spoken	Blog
True	Web	Blog	Spoken
Mind	Fiction	TV/Movies	Web
Happy	TV/Movies	Spoken	Blog
Nice	TV/Movies	Spoken	Fiction
Technical terms			
Deontology	Academic	Blog	Web
Epiphenomenalism	Web	Academic	Blog
Phenomenology	Academic	Web	Blog
Satisficing	Academic	Blog	Web
Expressivism	Academic	Mag	Blog
Know	TV/Movies	Spoken	Fiction
Justify	Blog	Academic	Web

is admittedly coarse-grained, but I take it that the distribution of "justify" across the COCA corpus does provide reason to think that the term is not widely distributed. Instead, the term appears to be restricted to particular contexts. If we list the sub-genres for which "justify" has the highest frequency, we see clearly that the term is most common in academic contexts and other contexts with a higher register (Table 4).

We can gain further evidence if we turn to the BNC. As stated earlier, it is useful to turn to this corpus because we are able to analyse the distribution of an expression across many different categories, including written vs spoken, genre, domain, age of author, sex of author, age of audience, perceived level of difficulty, and more. The distribution of "justify" reveals several facts that all suggest that "justify", rather than being a term that is used commonly, in fact belongs to a formal register. First, as we have just seen with the COCA data, the ten genres in which "justify" has its highest frequency can all plausibly be thought of as formal contexts (Table 5).

Second, if we turn away from specific genres (for which in *BNCWeb* there are 46) and to the more general category of derived text type for which there are only 8 categories, the most frequent derived text types are as shown in Table 6.

It is notable here that academic prose, non-academic prose and biography, and newspapers form the top three, while fiction and verse and spoken conversation are the bottom two text types. This is further evidence that "justify" belongs to a formal register. *BNCWeb* also categorises the majority of the corpus documents in terms of perceived level of difficulty, and here again we find evidence that "justify" belongs to a formal register, as it becomes more frequent as the difficulty increases (Table 7).

These three considerations taken together, and particularly combined with the considerations from COCA, suggest that "justify" is a term largely confined to a formal

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 4. Most frequent sub-genres for "justify" in COCA.

Rank	Sub-genre	Words	Frequency	Freq per mil
1	Academic: Law/Political science	12,285,693	858	69.84
2	Academic: Philosophy/Religion	7,841,453	430	54.84
3	News: Editorial	4,797,501	224	46.69
4	Blog: Argumentative	68,736,491	2744	39.92
5	Academic: History	13,395,427	491	36.65
6	Web: Argumentative	54,970,418	1965	35.75
7	Academic: Miscellaneous	4,817,686	170	35.29
8	Blog: Academic	504,423	17	33.70
9	Magazine: Religion	5,553,662	165	29.71
10	Spoken: MSNBC	824,952	22	26.67
11	Academic Business	1,180,194	29	24.57
12	Web: Legal	3,683,079	88	23.89
13	Magazine: News/Opinion	27,676,740	650	23.49
14	Academic: Geography/Social Science	20,028,471	455	22.72
15	Academic: Humanities	16,203,015	365	22.53
16	Blog: Legal	423,009	9	21.28
17	Blog: Fiction	2,200,365	44	20.00
18	News: International News	6,323,096	125	19.77
19	Blog: Miscellaneous	32,935,088	635	19.28
20	News: Money	8,725,626	168	19.25

register and possibly a technical term, rather than a term that is used widely in ordinary discourse.[15]

Focusing as we have on the frequency and distribution of "justify" across corpora, we may wonder how exactly this spells trouble for the folk justification approach outlined earlier. I will return to this issue in greater detail later in the paper. But my opponent, anticipating the nature of the case, may want to object in two ways. First, if we have shown that "justify" belongs to a formal register, this is no clear barrier to an investigation of the term. After all, there is no barrier to taking some term widely used only within particular contexts and providing some account of its meaning. Second,

[15]A further interesting insight is that among the documents in the BNC for which the sex of the author is indicated, "justify" is much more commonly used among males (27.53 per million), than mixed authors (16.36) and female authors (12.48). The opposite is actually true for "know", with female proving to be the most common sex of author (1417.85), rather than male (737.49) and mixed (451.91). Perhaps there is an interesting insight to be gained here regarding the use of epistemic terms across different sexes or genders, or it may be that terms that skew towards more formal registers also skew towards male speakers because of the institutionalised sexism that still exists (and certainly existed when the BNC was compiled). However, while interesting, I will not explore this topic further.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 5. Most frequent genres for "justify" in BNC.

Rank	Genre	No. of words	No. of hits	Freq per mil
1	Non-academic: politics law education	4,703,303	384	81.64
2	Public debates, discussions, meetings	287,062	23	80.12
3	Lectures on humanities and arts subjects	51,510	4	77.65
4	Hansard/parliamentary proceedings	1,168,362	67	57.35
5	Academic prose: humanities	3,358,166	175	52.11
6	Broadsheet national newspapers: personal & institutional editorials & letters-to-the-editor	102,718	5	48.68
7	School essays	147,736	6	40.61
8	Academic prose: technology computing engineering	689,593	27	39.15
9	Commerce & finance, economics	3,807,342	142	37.3
10	Regional and local newspapers: commerce & finance	419,996	14	33.33

Table 6. Most frequent derived text types for "justify" in BNC

Rank	Derived text type	No. of words	No. of hits	Freq per mil
1	Academic prose	15,778,028	782	49.56
2	Non-academic prose and biography	24,178,674	517	21.38
3	Newspapers	9,412,174	162	17.21
4	Other spoken material	6,175,896	99	16.03
5	Other published written material	17,924,109	281	15.68
6	Unpublished written material	4,466,673	52	11.64
7	Fiction and verse	16,143,913	114	7.06
8	Spoken conversation	4,233,962	15	3.54

while the above evidence does suggest that "justify" is a formal term, there is also countervailing evidence that could be pointed to. For instance, if we take the technical terms used earlier, it is important to note that while "justify" does have a similar distribution to those terms, it is markedly more frequent (Table 8).

So while it might be that "justify" belongs to a formal register, it is certainly not as rare as other technical terms used in philosophical discourse. This also accords with intuition. It seems that we can use "justify" well enough in ordinary discourse with a non-philosophical audience without having to explain what it means as we would these other terms.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 7. Frequency of "justify" across levels of difficulty in BNC.

Perceived level of difficulty	No. of words	No. of hits	Frequency per million words
High	24,745,652	899	36.33
Medium	46,211,631	820	17.74
Low	16,946,288	189	11.15

Table 8. Frequency per million in COCA.

Word form	Freq per mil
Justify	16.35
Phenomenology	0.56
Deontology	0.05
Satisficing	0.04
Epiphenomenalism	0.01
Expressivism	0.01

At this stage it is important to note, however, that in considering whether the kind of justification that is of interest in epistemology is in wider circulation, we ultimately need to look beyond just the frequency and distribution of "justify". This is because justification in the sense that epistemologists are interested in attaches in the first place to beliefs and propositions (or other proposition-like objects), rather than, say, acts or emotions. Yet if we just look at the frequency and distribution of "justify", the numbers involved will naturally be inflated by uses where "justify" takes an action or an emotion as its object (e.g. "His anger was justified"). So we should consider whether there is evidence in the corpora of "justify" being used to talk about proposition-like items. This will be the topic of the next section.

5. Justifying What?

In order to investigate whether there is evidence of discourse surrounding justification of beliefs, we will need to go beyond considering mere frequency and distribution, and instead consider the *collocates* of "justify" i.e. the terms that frequently appear alongside the term. In doing so, we can make some grammatical distinctions. For instance, for the verb "justify", there are terms that will often appear as its subject, other terms that appear as its object, and adverbs that modify it etc. As mentioned earlier, the corpus analysis software *SketchEngine* is particularly useful for this task. As that is the case, I will largely perform analyses on the very large corpus *EnTenTen20*. Doing so is particularly useful when we are considering expressions and phrases that are used less frequently, as we have seen with "justify".

Our investigation thus far has largely ignored a fact that is arguably widely accepted within philosophy: that there are many different kinds of justification. One way to

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

distinguish between kinds of justification is to focus on the object of justification. We can distinguish between the justification of acts, beliefs, emotions, etc. A second way to distinguish between kinds of justification is to focus on the domain of justification. We can distinguish between moral justification, epistemic justification, prudential justification, legal justification etc. I am going to focus primarily on the object of justification, on what is being justified, as this is something that is amenable to a corpus-based analysis.

It is fair to say that epistemologists, in thinking about epistemic justification, have taken the justification of *belief* to be the central kind of case. This makes sense of course, given that the epistemic realm is concerned with belief-formation and knowledge. It may be that actions and possibly other objects can be epistemically justified, but it still seems fair to say that beliefs at least serve as the paradigmatic object of epistemic justification. So one way of considering whether epistemic justification is spoken of in ordinary discourse is to consider the extent to which justification of belief is spoken of and how this compares to justification of action or other objects. Alston has suggested that "the term 'justified' has been imported into epistemology from talk about voluntary action" (1993: 532). In doing so, he raises the possibility that it is only within epistemology that justification of belief is spoken of, and this is something we can investigate via a collocation analysis.

Before we do, there are two important preliminary points. First in performing a collocation analysis, it may initially seem sensible to focus on the terms that most frequently appear alongside our term of interest. So in investigating the object collocates of "justify_v", it may seem sensible to focus on the terms that *most frequently* appear as its object. One setback with this approach is that it does not take into account how common the collocate is across the corpus. A version of this issue arises with "justify_v". Taking into account all pronouns and nouns that serve as the object of "justify_v", the most common object is the pronoun "it" (e.g. "He couldn't justify it"). But this is likely not a result of the fact that the two terms stand in some particular relationship. Instead, this is just because "it" is an extremely common and flexible term.[16] For that reason, it is preferable to use an *association score* that better reflects the association between two terms. Collocates can then be found by finding those terms with the highest association score. SketchEngine provides the LogDice association score for collocates, for which more information can be found at Rychlý (2008). This is a score between 0 and 14. 14 would indicate that all occurrences of the first term occur with the second term. 0 would indicate that there is less than 1 occurrence of the two terms together per 16,000 occurrences of either term. LogDice can be used to rank collocates and also compare them. A difference in score of 1 indicates that the higher-scored term collocates twice as often. For our purposes, we find that LogDice solves our initial issue, with "it" receiving a comparatively low LogDice score (3.3) compared with terms we might expect to collocate specifically with "justify_v", such as "action" (7.4), "existence" (7.9) and "belief" (6.7).

The second preliminary point is that in considering the distinction between justification of a belief versus justification of an action, I will not consider whether certain terms are indicative of propositional or doxastic justification. It is an interesting further question whether the distinction can be teased out within a corpus, but it is not one I will explore here. I will just be interested in whether there is evidence of "justify" being

[16]"It" is the 13th most frequent term in Ententen20.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

applied to a belief, proposition, or other proposition-like object (even a claim or an assertion), and I will largely refer to this kind of justification as belief justification.

Focusing first on the top collocates that appear as the object of "justify_v" gives the results shown in Table 9.[17]

There are a few points of interest from Table 9. First, keeping the distinction between justification of belief and justification of action in mind, it is worth noting that both "belief" and "action" make it into the top 20 collocates of "justify_v", whether we calculate in terms of frequency or LogDice. Second, the list is dominated by action terms, with 15 of the 20 terms concerned with action ("mean", "action", "expense", "invasion", "war", "intervention", "killing", "investment", "violence", "cost", means", "decision", "expenditure", "murder", "purchase", "refusal"). This would lend support to the idea that "justify_v" is primarily used to talk about action. Third, as well as "belief", there are other terms indicative of the justification of belief, namely "claim" and "conclusion".

Regarding "claim", it is important here to note that the term is ambiguous between one sense akin to a proposition or assertion and another sense (more common in legal discourse) akin to a right to ownership (e.g. "seeking to justify his claim to the Liberian presidency"). We could attempt to filter out such uses in our query by appealing to syntactic properties. For instance, the ownership sense commonly uses prepositions "to" and "for" (e.g. "with the more detailed records being used to justify the claim *for* refund"). However, this kind of syntactic filtering will be inaccurate, particularly when claims are discussed without being explicit on what it is a claim for.[18] To get an idea of how many of the total instances of "justify_v" + "claim" are instances of "claim" in this ownership sense, a random sample of 300 instances was extracted from EnTenTen20 via SketchEngine, and were hand-coded according to whether they were clear instances of the ownership sense. Of that, 57 were clear instances of the ownership sense, amounting to 19% in total. This is a significant amount, but if we reduced the frequency of the collocate claim by that percentage, we would have an adjusted frequency of 4,542, which would mean it is still the 12th most frequent object collocate. All this is to say that while the ownership sense does make a significant contribution to the overall frequency of "justify_v" + "claim", it is not the reason why "claim" is one of the top object collocates of "justify".[19] So it is reasonable to think that "justify_v" + "claim" represents instances where belief justification is spoken of.

Turning to "justify_v" + "belief", we may wonder whether the fact that these are relatively closely associated is due to philosophical uses. One way of exploring this is by hand-coding a sample the way we just have with "justify_v" + "claim". Again, a random sample of 300 instances were extracted via SketchEngine, and each instance was hand-coded according to whether it was clearly an example of philosophical discourse. This was done primarily by looking out for other philosophical technical terms and phrases e.g. "justified true belief", "Gettier", "epistemology", "foundationalism" etc., but also on whether they have originated from websites known to be dedicated to philosophy e.g.

[17]Despite the previously-mentioned issue regarding "it", the list of collocates here have excluded pronouns anyway, as they tend not to give much insight into the meaning of a term. This only affects the frequency rankings, not the LogDice rankings (for which "it" and other common terms would be much lower down).

[18]The query would have poor *precision* (i.e. it would give back instances we are not interested in) even if it had good *recall* (i.e. it would give us all the instances we are interested in).

[19]Note that I am not committed to the idea that "claim" is only two-ways ambiguous. The above process of hand-coding the ownership sense of "claim" was only concerned with just that. It may be that the instances that remain are due to multiple other senses of the term.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 9. Collocate objects of "justify_v" in EnTenTen20.

Rank (LogDice)	Rank (Freq)	Collocate	Frequency	LogDice
1	7	existence	7423	7.86
2	3	mean	10247	7.75
3	1	action	19741	7.43
4	12	expense	4874	7.09
5	23	invasion	3073	7.07
6	6	war	7442	6.95
7	20	intervention	3410	6.78
8	30	killing	2344	6.74
9	10	investment	5312	6.72
10	14	belief	3973	6.7
11	15	violence	3900	6.69
12	2	cost	11870	6.6
13	19	conclusion	3411	6.58
14	26	means	2755	6.56
15	9	claim	5608	6.36
16	4	decision	9916	6.34
17	37	expenditure	1940	6.34
18	36	murder	2092	6.27
19	27	purchase	2501	6.26
20	57	refusal	1474	6.24

philpapers, philarchive, askphilosophers.org etc. Of the 300, 123 were clear instances of philosophical discussion, which amounts to 41% of the sample. It is also worth noting that religious discourse was also particularly prevalent within the sample. Of the remaining 177 instances, 61 were coded as clear cases of religious discourse (where clearly religious terms, such as "theism", "God", bible", etc. were present) (Table 10). This is likely to be partly due the fact that "belief" itself has some association with religious discourse – the most frequent modifier of "belief" being "religious" (LogDice 10.1).

Another way of getting an idea of how "justify_v" + "belief" is used is by looking at the collocates of the phrase. The most common modifiers of "justify_v" and the most common modifiers of "belief" among instances of "justify_v" + "belief" are given in Table 11.

A brief look at the terms that are being used to modify the phrase would strongly suggest that it is predominantly used in philosophical discourse.[20] It is important to keep in mind that we are dealing with low frequencies here, even in a corpus as

[20]It should be noted that although "foolish" appears as one of the key modifiers of "belief", a look at the 51 instances strongly indicates that this is almost entirely due to the book *Mistakes Were Made (But Not By Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts)* (Tavris and Aronson 2015), with only one of the 51 instances not citing that book.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 10. Sample of "justify" + "belief" from EnTenTen20 hand-coded to detect philosophical and religious discourse.

	Total	Percentage
Philosophical	123	41
Other	116	38.67
Religious	61	20.33

large as EnTenTen20. But it is still reasonable to take this as evidence that "justify_v" + "belief" is largely used only in philosophical discourse.

It is also notable that "true" is by far-and-away the most frequent modifier of "belief" on the list. Of course, this indicates discussion of whether or not a true belief is justified and an inspection of these confirms that they are instances of philosophical discourse where the tripartite theory of knowledge is being discussed.

6. Justify_j

Turning to the adjectival use of "justified", the nouns modified by the term with the highest LogDice scores are shown in Table 12.

There are a few important points to note. First, half the list in Table 12 is taken up with emotional terms rather than terms pertaining to action or belief ("anger", "outrage", "indignation", "suspicion", "paranoia", "grievance", "resentment", "distrust", "vengeance", "fear"). Second, the list is heavily populated with terms with negative connotations, with 15 of the 20 terms having a clearly negative connotation ("Sinner", "anger", "sinner", "outrage", "homicide", "indignation", "suspicion", "criticism", "paranoia", "slavery", "grievance", "resentment", "distrust", "vengeance", "fear"). Third, "belief" has the 8th highest LogDice score but is also the most frequent. However, there is good reason to think that the fact that "belief" is the most frequent is largely due to philosophical usages. From a random sample of 300, 273 were from clearly philosophical discourses, which amounts to 91%. An adjusted frequency that adjusted for philosophical instances would amount to 159, which would still keep "belief" as one of the top 20 collocates in terms of frequency. That "justified_j" + "belief" is largely used in philosophical discourse is also confirmed by looking at the terms that modify each of "justified_j" and "belief" across instances of "justified_j" + "belief" (Table 13).

Here again, we find that the modifiers are nearly all indicative that the phrase is being used in a philosophical context. We again find that "true" is the most frequent modifier of "belief", and again this appears to be entirely due to discussion of the tripartite theory of knowledge and the Gettier problem.

7. Discussion

It is time to consolidate these various data points and consider their epistemological repercussions. I will focus particularly on three key claims:

I. "Justified_j" is nearly never used to talk about beliefs outside of philosophical circles.

The evidence in support of this claim can be seen in Tables 12 and 13. In Table 12, we saw that the only belief or proposition term among the top 20 nouns modified by

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 11. Most frequent modifiers of "justify_v" and of "belief" among instances of "justify_v" + "belief".

	Modifiers of "justify"			Modifiers of "belief"	
	Frequency	LogDice		Frequency	LogDice
non-inferentially	6	8.370	nonbasic	4	6.490
epistemically	21	8.080	theistic	10	4.190
inferentially	4	7.110	foolish	51	4.170
rationally	26	4.140	justified	11	3.730
amply	5	1.650	perceptual	21	3.380
objectively	6	1.180	testimonial	4	3.240
logically	5	0.420	atheistic	4	3.080
			irrational	10	1.930
			phenomenal	6	0.710
			true	156	0.700
			racist	9	0.380
			empirical	11	0.080

"justified_j" was "belief", and further inspection of "justified_j" + "belief" revealed that the vast majority of instances (91%) were – from concordance analysis alone – clear instances of philosophical discourse.[21] Table 13 also supports this, as all of the modifiers for both "justified_j" and "belief" were indicative of philosophical discourse.

This is an interesting insight in its own right into the way "justified_j" is used, but it should also prove relevant to some forms of the folk justification approach. We saw earlier that some adopt a particularly linguistic form of the folk justification approach whereby linguistic properties of "justify" and related word forms are used to inform a theory of justification. Hawthorne and Logins (2020) are perhaps the clearest example of this where they have focused specifically on the gradability of "justified_j". However, given claim I, even if "justified_j" is a gradable term, it is highly questionable that we should take this into account when considering the epistemic justification of beliefs, as ordinary speakers do not seem to use "justified_j" to talk about beliefs or other proposition-like objects.

II. Justification talk primarily concerns something other than belief justification.

We saw with both "justify_v" and "justify_j" that the object terms they most frequently combine with do not refer to beliefs or other proposition-like objects. In the case of "justify_v", action terms dominate (Table 9), while in the case of "justify_j", emotional terms dominate (Table 12).[22] This perhaps lends some support to Alston's suggestion

[21]As a concordance analysis only gives the immediate textual context on either side of the token, it may well be that instances of philosophical discourse were not coded as such and that the percentage could be even higher.

[22]This difference between "justify_v" and "justify_j" is in itself an interesting finding that I do not have an explanation for.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Table 12. Terms modified by “justified_j” in EnTenTen20.

Rank (LogDice)	Rank (Frequency)	Noun	Frequency	LogDice
1	8	sinner	301	8.390
2	10	ancient	250	8.060
3	5	anger	533	6.550
4	12	sinner	186	6.440
5	11	outrage	207	6.080
6	19	homicide	150	5.940
7	32	self-defense	66	5.900
8	1	belief	1753	5.890
9	28	indignation	76	5.620
10	44	favouritism	41	5.590
11	15	suspicion	171	5.560
12	3	criticism	615	5.400
13	31	paranoia	67	5.330
14	16	slavery	163	5.240
15	24	grievance	93	5.230
16	29	resentment	71	5.110
17	25	dismissal	93	5.080
18	37	distrust	51	5.070
19	42	vengeance	44	4.880
20	7	fear	485	4.870

Table 13. Most frequent modifiers of “justified_j” and “belief” among instances of “justified_j” + “belief”.

	modifier of “justified_j”			modifier of “belief”	
	frequency	LogDice		frequency	LogDice
epistemically	20	8.290	non-inferentially	5	5.960
inferentially	4	8.060	non-basic	3	4.930
rationally	6	2.040	priori	9	3.470
			true	675	2.810
			facie	4	2.090
			undefeated	5	2.010
			inconsistent	3	0.350

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

that justification-talk was transferred over from talk of voluntary action. It is perhaps tempting to take Alston's suggestion one step further and claim that it is only within philosophical discourse that the transfer has been made. This would be continuous with the possibility mentioned earlier in the discussion of Plantinga and Weatherson that it is only after Gettier that philosophers explicitly discussed the notion of justification. But the findings here do not suggest that talk of belief-justification is restricted to philosophical circles, even if it is far more frequent there. Notably, as Table 9 indicates, "justified_v" does collocate with "conclusion", "belief", and "claim", and even once we adjust for the possession usage of "claim", and for the fact that many instances of "justified_v" + "belief" are philosophical usages, a considerable number of instances still remain. So the evidence here does not suggest that justification of beliefs or other proposition-like objects only occurs within philosophical discourse. But it does suggest that justification-talk is primarily not about belief justification.

III. "Justify" is a high register term.

Finally, we saw a range of evidence from both COCA and BNC to suggest that "justify" belongs to a higher register, and so is largely used in more formal contexts. This includes, but is not limited to, academic contexts, as Tables 3–7 indicate.

III is perhaps the most important finding of this investigation, as it forms the basis of a challenge for the folk justification approach. Echoing a response considered earlier, it might be thought that the fact that "justify" is a high register term poses no particular challenge to the folk justification approach, provided that we focus on the *high register folk* i.e. the practices of the people that take part in the high register contexts in which the term is used. At this stage however, it is important to revisit the basic motivation behind the folk justification approach. The thought is that there is some standard for belief evaluation that is widely circulated within our epistemic community, and that this plays a central role in our epistemic lives – we make efforts to have beliefs that are justified, and when others have formed unjustified beliefs, we take them to have behaved improperly in some sense. As Booth states:

> We feel that we are obliged to believe in accordance with the available evidence … and only when we do so are we epistemically justified in having a particular belief; such that we feel, for that reason, that the beliefs of Holocaust deniers, creationists, and members of the flat earth society (for example) are epistemically unsalutary. In short, we feel we ought to have justified beliefs and we want to know how. (Booth 2011: 40)

Given the importance that epistemic justification is thought to have in our epistemic lives, the idea that this property is only picked out within certain high register contexts seems implausible. We now have a tension between the following three claims:

(1) Epistemic justification is a central part of our epistemic lives, such that it is important across a wide range of contexts.
(2) Use of the word "justify" is largely restricted to high register contexts.
(3) "Justify" is the primary term used to pick out epistemic justification.

It has been a working hypothesis of this investigation that 3 is true, but at this stage, the most natural way for the folk justification theorist to respond is to reject 3. That is, they

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

could concede that using "justify" to talk about justification is something that largely occurs in certain higher register contexts and particularly philosophical contexts, but they should claim that there are other terms or phrases that are used to talk about epistemic justification.

This move will have a couple of repercussions. First, any attempt to give an account of epistemic justification via consideration of the meaning properties or syntactic properties of the word "justify" requires a great deal further motivation before it can be considered plausible, contrary to the projects we saw at the start of the paper from the likes of Goldman, Kvanvig and Menzel, and Hawthorne and Logins. Second, there is now a particular challenge faced by the folk justification theorist: to identify how exactly the ordinary notion of epistemic justification is spoken of, if not via the use of "justify". Note that I describe this as a *challenge* rather than as an *objection*, as I do think that there are routes to overcome it. Nevertheless, it is a challenge that needs to be answered. In what follows, I will outline the clearest available responses.

One option will be to claim that ordinary speakers in fact primarily use *knowledge discourse* to talk about the justification of their beliefs. After all, as mentioned at the start of this investigation, knowledge discourse is much more prevalent, with "know" consistently figuring as one of the most common verbs across English language corpora. The idea would be that rather than saying something as cumbersome as "her belief that p was justified", an ordinary speaker may instead say "she knew that p". Of course, if we are to allow *false* justified beliefs, then this approach would require some way of talking about them via knowledge discourse despite the fact that knowledge attributions are typically thought to be factive. But even assuming that there is some way round this issue, the more important point here is that we seem to have reverted back to the TOK approach to justification. That is, if we claim that justification is really spoken of in terms of knowledge, it seems we have no grounding in our ordinary epistemic practices with which to understand justification that is independent of our ordinary knowledge discourse. Instead, the role of justification would just be to understand some aspect of our knowledge. Again, as stated earlier, I am not claiming here that the TOK approach is implausible, the point is that this does not look like a way of saving the folk justification approach.

An alternative approach is to claim that ordinary epistemic agents typically use some other term or phrase to talk about the justification of their beliefs. Drawing inspiration from the philosophical literature, we might consider candidates such as "warrant", "rational", "reasonable" etc. Frankly, I am sceptical that either "warrant" or "rational" could serve the role of being the ordinary term used to refer to epistemic justification, although I will not seek to show that here other than to note that both terms are less frequent than "justify".[23] It may be that justification is spoken of in terms of evidence or reason, and I think this is certainly a promising avenue to explore. One issue that may arise in considering this possibility is whether in ordinary discourse we rely upon a notion that bears a threshold, such that we can distinguish between the set of justified beliefs and the set of unjustified beliefs. That is, while we no doubt will find ordinary talk of people's *reasons* or *evidence* for believing, having *some* reason/evidence seems consistent with having reason/evidence insufficient for justified belief in the philosopher's sense. There are ways in which we might indicate in conversation that some threshold has been crossed such that a belief is indicated as justified, perhaps with

[23]In EnTenTen20, "justify" has a frequency of 24.35 per mil, "warrant" has a frequency of 18.00 per mil, and "rational" has a frequency of 12.21 per mil.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

phrases like "reasonable", "good reason/evidence", or "sufficient reason/evidence". But one possibility to consider here is that as far as our ordinary practices are concerned, and quite apart from the role that justification plays with regard to knowledge, it may be that justification is really a purely scalar term. This would then have ramifications on the kinds of epistemic evaluation we can hold each other to – without a minimum threshold of justification, it becomes less straightforward to reach a verdict that someone has conducted themselves acceptably or unacceptably.

So there are options available, and it may be tempting to think that as a result, this challenge is not particularly pressing for the folk justification theorist provided that one or more of these terms is used in ordinary discourse to talk about justification. In particular, we might think that there is not a single term or phrase that is used to talk about justification, that instead speakers select from a long menu of possible phrases in order to engage in this form of epistemic evaluation. That may be right, and if so this is something that future corpus analysis can help reveal. However, the folk justification approach would face particular pressure from a pluralism about epistemic evaluation, akin to Alston (1993). That is, if our ordinary practices suggest that we appeal to a number of different notions in evaluating our beliefs, why not take that at face value and take there to be (in Alston's terms) a range of epistemic desiderata rather than a single notion of epistemic justification? And embracing this pluralism would itself prove somewhat revisionary in the way that folk justification is usually approached. For instance, it would mean that we shouldn't ask whether the new evil demon case, the Truetemp case, or the clairvoyant case are instances of justified belief, for that would conflate a number of different issues. Instead, we should ask whether the subject knows, whether they have good reason, good evidence, reliability etc.

These are further avenues that need to be explored in order for the folk justification approach to meet this challenge. Hopefully what has become clear in outlining them is that corpus analysis can play a role in this: by considering the frequencies, collocates, and distributions of the phrases discussed above across various corpora and text types, we can start to build a picture of our ordinary practices surrounding epistemic evaluation. And, as stated earlier, the particular benefit of corpus analysis is that it provides a way of inspecting our folk practices that is far less reliant on intuitive judgments and is also based on genuine language use rather than elicited language use in experimental settings.

8. Conclusion

In this paper, I have attempted to use corpus analytical methods on large English corpora to investigate the lemmas "justify_v" and "justified_j", and in doing so I have raised a challenge to the folk justification approach: to identify the way in which epistemic justification is spoken of ordinarily. This is necessary to make good of the idea that this form of epistemic evaluation is something that plays a central role in our epistemic lives. I hope to have shown that corpus analysis can be used to raise interesting philosophical questions, thus continuing a theme one finds in recent work in experimental philosophy. In the final section, I also hope to have shown that corpus analysis will not just be a tool to raise such challenges, but should also be a tool to resolve them.[24]

[24]Thanks to Sylvia Jaworska for her tremendous help with the corpus analytical methods used in this paper. Thanks also to Nat Hansen for reading an earlier version of this paper.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

References

Alston W.P. (1985). 'Concepts of Epistemic Justification.' *The Monist* **68**(1), 57–89.

Alston W.P. (1993). 'Epistemic Desiderata.' *Philosophy and Phenomenological Research* **53**(3), 527–51. doi: 10.2307/2108079.

Beebe J.R. and Buckwalter W. (2010). 'The Epistemic Side-Effect Effect.' *Mind and Language* **25**(4), 474–98. doi: 10.1111/j.1468-0017.2010.01398.x.

Biber D., Reppen R., Schnur E. and Ghanem R. (2016). 'On the (Non)utility of Juilland's D to Measure Lexical Dispersion in Large Corpora.' *International Journal of Corpus Linguistics* **21**(4), 439–64. doi: 10.1075/ijcl.21.4.01bib.

Bird A. (2007). 'Justified Judging.' *Philosophy and Phenomenological Research* **74**(1), 81–110. doi: 10.1111/j.1933-1592.2007.00004.x.

BonJour L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Booth A.R. (2011). 'The Theory of Epistemic Justification and the Theory of Knowledge: A Divorce.' *Erkenntnis* **75**(1), 37–43. doi: 10.1007/s10670-010-9264-9.

Cappelen H. (2012). *Philosophy Without Intuitions*. Oxford: Oxford University Press.

Cohen S. (2016). 'Theorizing about the Epistemic.' *Inquiry* **59**(7–8), 839–57. doi: 10.1080/0020174X.2016.1208903.

Deutsch M.E. (2015). *The Myth of the Intuitive: Experimental Philosophy and Philosophical Method*. Cambridge, MA: MIT Press. http://ebookcentral.proquest.com/lib/reading/detail.action?docID=3339982.

Dutant J. (2015). 'The Legend of the Justified True Belief Analysis.' *Philosophical Perspectives* **29**, 95–145. doi: 10.1111/phpe.12061.

Foley R. (1987). *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.

Foley R. (2001). 'The Foundational Role of Epistemology in a General Theory of Rationality.' In A. Fairweather and L. Zagzebski (eds), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, pp. 214–30. Oxford: Oxford University Press.

Foley R. (2008). 'An Epistemology that Matters.' In P.J. Weithman (ed), *Liberal Faith: Essays in Honor of Philip Quinn*, pp. 43–55. Notre Dame, IN: University of Notre Dame Press.

Fumerton R. (2005). 'Theories of Justification.' In P.K. Moser (ed.), *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press.

Gettier E.L. (1963). 'Is Justified True Belief Knowledge?' *Analysis* **23**(6), 121–3. doi: 10.2307/3326922.

Goldman A.I. (1979). 'What is Justified Belief?' In G.S. Pappas (ed.), *Justification and Knowledge: New Studies in Epistemology*, pp. 1–23. Dordrecht: Springer. doi: 10.1007/978-94-009-9493-5_1.

Gries S.Th. (2005). 'Null-hypothesis Significance Testing of Word Frequencies: A Follow-up on Kilgarriff.' *Corpus Linguistics and Linguistic Theory* **1**(2), 277–94. doi: 10.1515/cllt.2005.1.2.277.

Gries S.Th. (2008). 'Dispersions and Adjusted Frequencies in Corpora.' *International Journal of Corpus Linguistics* **13**(4), 403–37. doi: 10.1075/ijcl.13.4.02gri.

Gries S.Th. (2019). 'Analyzing Dispersion.' In M. Paquot and S.Th. Gries (eds), *A Practical Handbook of Corpus Linguistics*, pp. 99–118. Cham: Springer. doi: 10.1007/978-3-030-46216-1_5.

Grindrod J. (Forthcoming). 'Distributional Theories of Meaning.' In D. Bordonaba Plou (ed.), *Experimental Philosophy of Language*. Cham: Springer.

Hansen N., Porter J.D. and Francis K. (2019). 'A Corpus Study of "Know": On the Verification of Philosophers' Frequency Claims About Language.' *Episteme* 1–27. doi: 10.1017/epi.2019.15.

Harris Z.S. (1954). 'Distributional Structure.' *Word* **10**(2), 146–62. doi: 10.1080/00437956.1954.11659520.

Hawthorne J. and Logins A. (2020). 'Graded Epistemic Justification.' *Philosophical Studies* **178**(6), 1845–58. doi: 10.1007/s11098-020-01512-0.

Ichikawa J.J. (2014). 'Justification is Potential Knowledge.' *Canadian Journal of Philosophy* **44**, 184–206. doi: 10.1080/00455091.2014.923240.

Kilgarriff A. (2005). 'Language is Never, Ever, Ever, Random.' *Corpus Linguistics and Linguistic Theory* **1**(2), 263–75. doi: 10.1515/cllt.2005.1.2.263.

Kim M. and Yuan Y. (2015). 'No Cross-Cultural Differences in the Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001.' *Episteme* **12**(3), 355–61. doi: 10.1017/epi.2015.17.

Kvanvig J.L. and Menzel C. (1990). 'The Basic Notion of Justification.' *Philosophical Studies* **59**(3), 235–61. doi: 10.1007/BF00355744.

Lehrer K. (1990). *Theory of Knowledge,* 2nd edition. London: Routledge.

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press

Lehrer K. and Cohen S. (1983). 'Justification, Truth, and Coherence.' *Synthese* **55**(2), 191–207. doi: 10.1007/BF00485068.

Lewis D. (1996). 'Elusive Knowledge.' *Australasian Journal of Philosophy* **74**(4), 549–67. doi: 10.1080/00048409612347521.

Liao S. and Hansen N. (2022). '"Extremely Racist" and "Incredibly Sexist": An Empirical Response to the Charge of Conceptual Inflation.' *Journal of the American Philosophical Association*. doi: 10.1017/apa.2021.46.

McEnery T. and Wilson A. (1996). *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Plantinga A. (1990). 'Justification in the 20th Century.' *Philosophy and Phenomenological Research* **50**, 45–71. doi: 10.2307/2108032.

Rychlý P. (2008). 'A Lexicographer-Friendly Association Score.' In *RASLAN. Recent Advances in Slavonic Natural Language Processing*, 6–10. Corpus ID: 27713791.

Seyedsayamdost H. (2015). 'On Normativity and Epistemic Intuitions: Failure of Replication.' *Episteme* **12**(1), 95–116. doi: 10.1017/epi.2014.27.

Sutton J. (2007). *Without Justification*. Cambridge, MA: MIT Press.

Swain S., Alexander J. and Weinberg J.M. (2008). 'The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp.' *Philosophy and Phenomenological Research* **76**(1), 138–55. doi: 10.1111/j.1933-1592.2007.00118.x.

Sytsma J., Bluhm R., Willemsen P. and Reuter K. (2019). 'Causal Attributions and Corpus Analysis.' In E. Fischer and M. Curtis (eds), *Methodological Advances in Experimental Philosophy*. London: Bloomsbury.

Tavris C. and Aronson E. (2015). *Mistakes Were Made (But Not By Me): Why we Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*. Hawthorne, CA: Mariner Books.

Weatherson B. (2020). *A History of Philosophy Journals*. Available at: http://www-personal.umich.edu/~weath/lda/index.html.

Weinberg J.M., Nichols S. and Stich S. (2001). 'Normativity and Epistemic Intuitions.' *Philosophical Topics* **29**(1 & 2): 32.

Williamson T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell Publishing.

Ziółkowski A. (2021). 'The Stability of Philosophical Intuitions: Failed Replications of Swain et al.' *Episteme* **18**(2), 328–46. doi: 10.1017/epi.2019.20.

Jumbly Grindrod is a Lecturer in Philosophy at the University of Reading. His research focuses on issues in philosophy of language and epistemology, and particularly where the two intersect on the matter of context-sensitivity. More recently, his research has particularly focused on how corpus linguistics can shed new light on philosophical questions. He has published in a number of philosophical journals, including *Mind and Language*, *Ergo*, *Inquiry*, and *Metaphilosophy*.

Cite this article: Grindrod J (2022). Justification: Insights from Corpora. *Episteme* 1–25. https://doi.org/10.1017/epi.2022.44

https://doi.org/10.1017/epi.2022.44 Published online by Cambridge University Press