EUROPEAN REVIEW OF SOCIAL PSYCHOLOGY
2020, VOL. 31, NO. 1, 254–286
https://doi.org/10.1080/10463283.2020.1796080

ARTICLE

Harm inflation: Making sense of concept creep

Nick Haslam[a], Brodie C. Dakin[a], Fabian Fabiano[a], Melanie J. McGrath[a], Joshua Rhee[a], Ekaterina Vylomova[a], Morgan Weaving[a] and Melissa A. Wheeler[b]

[a]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia; [b]Department of Management and Marketing, Swinburne University of Technology, Melbourne, Australia

ABSTRACT
"Concept creep" is the gradual semantic expansion of harm-related concepts such as bullying, mental disorder, prejudice, and trauma. This review presents a synopsis of relevant theoretical advances and empirical research findings on the phenomenon. It addresses three fundamental questions. First, it clarifies the characterisation of concept creep by refining its theoretical and historical dimensions and presenting studies investigating the change in harm-related concepts using computational linguistics. Second, it examines factors that have caused concept creep, including cultural shifts in sensitivity to harm, societal changes in the prevalence of harm, and intentional meaning changes engineered for political ends. Third, the paper develops an account of the consequences of concept creep, including social conflict, political polarisation, speech restrictions, victim identities, and progressive social change. This extended analysis of concept creep helps to understand its mixed implications and sets a multi-pronged agenda for future research on the topic.

ARTICLE HISTORY Received 20 September 2019; Accepted 10 July 2020

KEYWORDS Concept creep; conflict; harm; identity; morality

The idea of "concept creep" was introduced by N. Haslam (2016a) to describe a pattern of semantic inflation in some of the psychology's key concepts. Haslam argued that a set of related concepts had broadened their meanings over the past half-century, so that they now referred to a much wider range of phenomena than they did in earlier times. He noted that the meanings of academic concepts would not be expected to sit still, evolving in response to new evidence and changing theoretical fashions, but that it was important for psychologists to recognise and understand these historical changes.

As a case in point, Haslam examined the concept of "bullying". As initially defined in the developmental psychology literature in the 1970s, the term was

CONTACT Nick Haslam ✉ nhaslam@unimelb.edu.au Melbourne School of Psychological Sciences, University of Melbourne, Parkville VIC 3010 Melbourne, Australia

explicitly distinguished from general peer aggression and referred to direct aggressive behaviour that was intentional, repeated, and carried out in the context of power imbalance, where the perpetrator was more powerful than the victim in age, stature, or number. Over time, every one of these criteria was relaxed in the field of bullying research and theory. Increasingly it was recognised among adults in workplaces in addition to children in schoolyards, the requirements of intentionality and repetition were abandoned, and behaviour targeting those of equal or greater power was also commonly defined as bullying. Indirect, digitally mediated forms of aggression were acknowledged as "cyber-bullying", and increased acts of omission such as shunning were counted as bullying alongside acts of commission.

N. Haslam (2016a) presented several additional case studies of conceptual expansion. "Trauma" progressively broadened to include adverse life events of decreasing severity and those experienced vicariously rather than directly. "Mental disorder" came to include a wider range of conditions, so that new forms of psychopathology were added in each revision of diagnostic manuals and the threshold for diagnosing some existing forms was lowered (Fabiano & Haslam, in press;N. Haslam, 2016b). "Abuse" extended from physical acts to verbal and emotional slights, and incorporated forms of passive neglect in addition to active aggression. In the mid-20th century, "prejudice" referred to blatant antagonism to particular racial or ethnic groups, but in subsequent decades grew to include aversive, modern, benign, implicit, and unconscious attitudes towards an expanding set of disadvantaged and marginal identities.

After presenting these case studies of creeping concepts, N. Haslam (2016a) advanced four propositions. First, he distinguished two forms of semantic broadening. "Horizontal creep" occurs when a concept extends outward to refer to qualitatively new phenomena (e.g., "mental disorder" encompassing entirely new forms of behaviour and experience; "abuse" coming to include neglect), whereas "vertical creep" occurs when a concept's meaning extends downwards to refer to quantitatively less severe phenomena (e.g., "trauma" referring to vicarious or non-life-threatening experiences (N. Haslam & McGrath, in press); "prejudice" referring to subtle and ambiguous micro-aggressions (Lilienfeld, 2017)). The two forms of creep are not mutually exclusive but represent distinct dimensions – a concept might expand to refer to qualitatively new phenomena that are also less severe – and it may be difficult to determine whether a particular conceptual change is best described as vertical, horizontal, or both. Second, he argued that the case studies were all examples of a general pattern of conceptual expansion. Although it might be tempting to explain each concept's semantic shifts separately – the spread of mental disorder as evidence of "medicalisation" and the spread of prejudice as "political correctness", for example, a generalised explanation would be more parsimonious and might open up new ways of understanding the nature and drivers of conceptual change across the disparate conceptual domains. For example, a unified explanation might identify factors beyond the rise of medical discourse that accounts for why the concept of mental disorder has inflated, or point to factors that identify new ways of thinking about that rise as part of a broader pattern. Third, Haslam proposed that the common thread among the identified creeping concepts was *harm* – they all represented forms of harmful behaviour and experience or ways of being harmed – and any generalised explanation of concept creep therefore had to address the enlargement of specifically harm-related concepts. Finally, he suggested that concept creep was driven by a rising sensitivity to harm within at least some Western cultures, such that previously innocuous or unremarked phenomena were increasingly identified as harmful, and that this rising sensitivity reflected a politically liberal moral agenda.

In the years following the publication of the initial paper in 2016, "concept creep" has been widely deployed in political discourse, especially in the North American context (e.g., Campbell & Manning, 2018; Lukianoff & Haidt, 2018). N. Haslam (2016a) was at pains to emphasise that "concept creep" was a descriptive rather than critical term, intended to characterise a pattern of conceptual change rather than condemn it, and that the expansion of harm-related concepts was sure to have benefits as well as costs. However, it has come to be understood by some as a conservative or even reactionary idea. For example, it has occasionally been enlisted by combatants in the so-called "culture wars", highly politicised debates on social justice, oppression, and the limits on expression primarily centred on American college campuses, into narratives involving "fragility", "snowflakes", and the supposed decline of Western civilisation (Glancy, 2018).

In this context, it is timely to review the substantial programme of research on concept creep that we have been undertaking. Our review addresses three primary questions. First, we examine the characterisation of concept creep, presenting some theoretical refinements that update the original understanding of the phenomenon and a body of research that rigorously assesses historical semantic expansion in putative creeping concepts using the tools of computational linguistics. Second, we investigate the factors that may be causally responsible for concept creep, proposing several cultural, societal, and intentional or motivated sources of conceptual change, and review evidence for the proposition that sensitivity to harm has risen in recent decades. Third, we address the potential consequences of concept creep, both negative and positive. We then present an integrative model of concept creep and lay out an agenda for future research.

Characterising concept creep

Clarifying the boundaries of creep

The original presentation of concept creep (N. Haslam, 2016a) exemplified the phenomenon using case studies of six concepts drawn from the fields of clinical (addiction, mental disorder, trauma), developmental (abuse, bullying), and social psychology (prejudice). It argued that the common thematic element in these concepts was harm and noted that all of the concepts referred in some fashion to the negative or undesirable side of human experience. It further proposed that the semantic inflation that the concepts underwent had taken place "in recent decades" and illustrated accompanying rises in their relative frequency of use within the Google Books corpus from 1960 to 2005 to illustrate. This initial characterisation of concept creep was ambiguous or inaccurate in three respects, and subsequent evidence has clarified it.

First, it has sometimes been inferred that "concept creep" refers only to the six concepts that were first used to exemplify it. Instead, it should be made explicit that concept creep can refer to the semantic broadening of *any* harm-related concept in psychology or beyond. For example, we have examined a semantic change in "harassment" (Vylomova et al., 2019) and "hate" (N. Haslam & Murphy, 2020), and concept creep could also be explored in relation to such varied concepts as aggression, bipolar disorder, disability, misogyny, pain, racism, sexual assault, and violence, among many others. Harm-related concepts are semantically central to psychology, given its traditional foci on suffering, social conflict, and their amelioration. The primary claim of the theory of concept creep is not that every harm-related concept in psychology has demonstrated a semantic expansion in recent decades, but that there is a general tendency for this to be the case to a greater degree than for other concepts. It does not challenge the concept creep hypothesis if certain non-harm-related concepts expand their meanings, or if certain harm-related concepts do not. Determining whether there is indeed a general tendency for harm-related concepts to broaden semantically more than non-harm-related concepts is a focus of ongoing research.

Second, Haslam's (2016a) analysis of concept creep at times blurred the boundary between harm-relatedness and negativity. Harm is of course normally undesirable, and the six exemplary creeping concepts were all negative in this sense, but harm-related concepts as a semantic domain are not invariably undesirable. Within moral foundations theory (Graham et al., 2011), Harm is just one of five registers in which moral goodness as well as badness can be apprehended, and encompasses values and virtues involving Care. A dictionary of Harm foundation-related words contains positive terms such as compassion, empathy, protection, and safety, which are desirable because they shield against or otherwise mitigate harm. As concept creep is theorised to involve the semantic broadening of harm-related concepts, positive or negative, we might expect some or all of these desirable concepts to show evidence of expanded meanings, or to treat evidence against such expansion as grounds for qualifying the concept creep hypothesis. There is certainly anecdotal evidence that "safety" has extended its meaning in recent years to encompass protection from verbal and ideological as well as physical dangers (Campbell & Manning, 2018). In sum, "harm", not "negativity", is the central conceptual element when making sense of concept creep, a claim that is also supported by the finding, discussed at greater length later in the review, that people who endorse a harm-based morality tend to hold broad definitions of creeping concepts (McGrath et al., 2019).

Third, N. Haslam (2016a) was non-specific about when concept creep has occurred. As a process understood to be gradual it may be challenging to delimit when creep begins or ends, and it is highly unlikely that all harm-related concepts would display identically timed semantic expansions. Nevertheless, specifying more precisely when concept creep has occurred is important both to characterise it and to identify the concurrent historical trends that might be driving it. For example, one reading of concept creep, drawn from the work of Pinker (2011) on the decline of violence in the West, is that it is linked to the rights revolutions of the 1960s and 1970s. By this account, concepts of harm broadened as the result of a "civilising offensive": that is, previously tolerated forms of aggressive, domineering, and discriminatory behaviour became less socially acceptable at this time, and expanding concepts of what is harmful helped to define them as intolerable. If concept creep was a direct outgrowth of the rights revolutions, we would expect to see it accelerating when they were at the height of their influence and perhaps slowing again following their retreat in the 1980s.

Although definitive evidence is lacking, recent findings challenge this timeline. An analysis of semantic change in five harm-related concepts within a massive corpus of psychology articles (Vylomova et al., 2019), also discussed later in this review, found that change was greatest in every concept between the 1980s and the 1990s. Inspection of changes in the cultural salience of creeping concepts from 1960 to 2005 in N. Haslam (2016a) also consistently shows the steepest increases in the 1980s and 1990s, and Wheeler et al. (2019) similarly found the steepest rises in harm-related concepts in these decades. When characterising concept creep, it therefore now seems warranted to refer not to "recent decades" but to a process that has been particularly active in the 1980s and since, albeit not in lockstep for every concept. For example, N. Haslam and McGrath (in press) documented changes in the relative frequency of trauma-related concepts in the massive Google Books corpus from 1960 to 2008 as an index of their shifting cultural prominence. They found that "trauma" as a general concept rose most steeply in frequency during the 1980s but that specifically "psychological trauma" increased most sharply in the 1990s, and an ensemble of shared forms of trauma (i.e., "collective", "cultural" and "intergenerational" trauma) accelerated upwards in the 2000s. Thus, although concept creep and its cultural drivers may have been most evident from around 1980, specific concepts have distinctive creep trajectories during this period.

Explaining why concept creep may have gathered force since the 1980s remains extremely challenging, as many correlated historical changes took place around this time. Politically, the rise of neoliberal regimes across much of the Anglosphere (e.g., Reagan in the USA, Thatcher in the UK) may have led to a backlash focus on marginalised groups among traditionally liberal university researchers. Intellectually, the growing influence of critical theories originating in continental Europe may have driven attention to subtle forms of oppression. Culturally, the rise of post-materialist values favouring quality of life over materialist values of physical and economic security, which may have underpinned a growing concern with harm, was especially steep in the 1980s (Inglehart, 2008). Disentangling this skein of potential influences, and demonstrating how political, intellectual, and cultural factors produce semantic changes in concepts, is enormously difficult. However complex these factors may be, by locating an apparent historical inflexion point in the 1980s our recent work should help to resolve them.

Methods for assessing historical semantic change

The original analysis of creeping concepts presented in N. Haslam (2016a) was qualitative, based on a close reading of changes in the meaning of specific concepts within the psychological literature. This analysis was supported by quantitative analyses of changes in the relative frequency of related words in the Google Books corpus, a demonstration of "culturomics" (Michel et al., 2011). However, such frequency-based analyses do not assess changes in semantic breadth and thus cannot index concept creep directly. Evaluating semantic change in harm-related concepts in a systematic and quantitative manner is, therefore, an important task. Several computational methods for doing so have been developed in recent years, all based on changes in how language is used (Bybee, 2010).

Early computational attempts to evaluate semantic change examined large historical text corpora for evidence of changes between predefined periods in a word's frequency and in the words with which it is collocated. These approaches rested on the assumption that such changes in frequency and collocations reflect changes in word meanings and have been employed in numerous investigations (e.g., Heyer et al., 2009; Hilpert & Gries, 2016; Juola, 2003). Some have investigated n-grams (strings of words of length n) rather than single words. For example, Gulordava and Baroni (2011) used an n-gram corpus to compare language usage in the 1960s and 1990s by applying a distributional semantics approach that estimates the similarity between word meanings as a similarity of contexts in which they are used. Each word was represented as a high-dimensional sparse vector of its collocations with other (contextual) words, and historical change was estimated as cosine similarity between the corresponding word vectors from the two periods. The authors showed this similarity-based method to be superior to frequency-based methods for automatic detection of semantic change.

Although such quantitative analysis was prominent in corpus linguistics for many years, its limitations – primarily problems of generalisation – led to the development of models for language change that rely on dense representations of words (embeddings) that are obtained either from word co-occurrence statistics (count-based) or by training a model to predict a word from its context or vice versa (prediction-based). Prediction-based models, which are now dominant and have been shown to outperform count-based approaches (Kulkarni et al., 2015; Schlechtweg et al., 2019), involve training language models (such as word2vec, SGNS; Mikolov et al., 2013) incrementally for each subsequent time period ("epoch") and assessing cosine similarity between word vector representations in each epoch to track semantic changes so that the timing of changes can be located. Epoch-specific models can also be aligned post hoc using linear matrix transformations to evaluate the degree of change (Hamilton et al., 2016a). Hamilton et al. (2016b) further demonstrated the value of systematically examining nearest neighbours of a target word to evaluate semantic changes that are due to cultural shifts. One outcome of these technical advances has been the formulation and testing of laws of semantic change (e.g., Dubossarsky et al., 2017; Winter et al., 2014; Xu & Kemp, 2015), such as the relationship between rate of change and word frequency.

Applying computation methods to study concept creep

Inspired by these developments in lexical semantic change detection, Vylomova et al. (2019) applied some of the successful models to evaluate changes of five putatively creeping concepts: "bullying", "prejudice", "trauma", "harassment", and "addiction". Because concept creep is argued to originate in the recent academic literature of psychology and cognate fields, the analysis was carried out using a new corpus of abstracts from more than 800,000 articles published in 875 psychology journals between 1970 and 2019. Abstracts were chosen as the source of text because they provide a compact summary of the main contributions and intellectual context of the research reported in the articles.

Vylomova et al. (2019) first evaluated the change in semantic breadth of the five terms using a count-based method developed by Sagi et al. (2011) that employs latent semantic analysis. Semantic breadth was estimated as average cosine similarity between vector representations of the contextual usage of each word in each decade from the 1980s to the 2010s, where higher similarity represents lower semantic breadth (i.e., less disparate semantic vectors). The five concepts showed differing temporal trajectories over the four decades, although there was ample evidence of semantic broadening. "Addiction", for example, showed a consistency if there is a gentle increase in semantic breadth, whereas "harassment" broadened substantially from the 1990s onwards after narrowing from the 1980 to the 1990s.

Vylomova et al. then conducted a more detailed analysis, using the model proposed by Hamilton et al. (2016a). For each decade they first trained a prediction-based language model and then aligned the trained epoch-specific models using Procrustes. For each word, they extracted its most common collocation in each period of time and evaluated the dynamics of cosine similarity (i.e., how similarity between the word and its collocations changes over time). This analysis affords a finely detailed quantitative description of the shifting associations between a putatively creeping concept and its most common semantic contexts.

Figures 1 and 2 illustrate these analyses for "addiction" and "bullying", respectively. As Figure 1 shows, in the 1980s "addiction" had the highest similarity to terms related to substances such as "drug", "alcohol" and "heroin". In later decades, those similarities decline as the concept's strongest similarities shift towards behavioural terms such as "internet", "gaming" and "sexual", consistent with N. Haslam's (2016a) argument that addiction had inflated its meanings to encompass non-consummatory activities, as in so-called "behavioural addictions". Figure 2 shows that the meaning of "bullying" underwent a pattern of differentiated broadening. Its traditional associations with children and schooling remain relatively stable, but there is

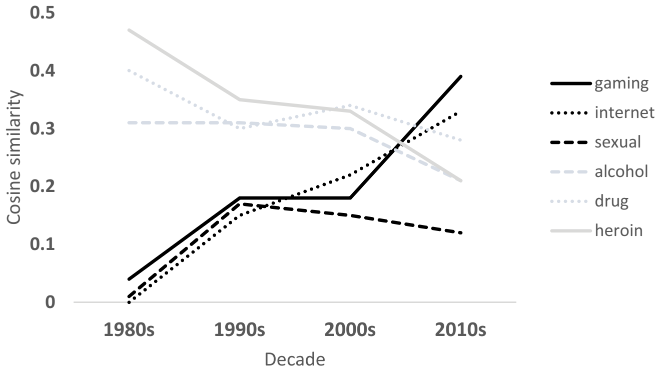

Figure 1. Changes in the semantic association of "addiction" with its most common nearest neighbours in psychology article abstracts by decade, 1980–2017, fromVylomova et al. (2019)

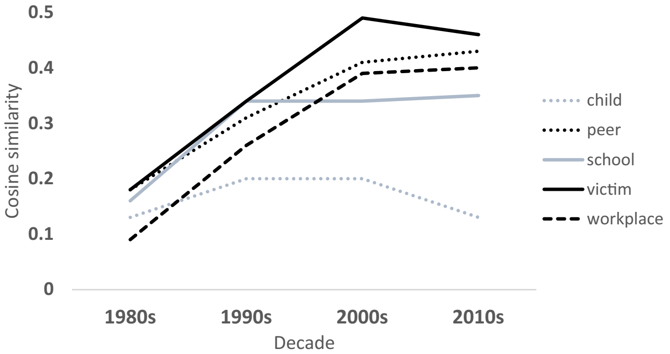

Figure 2. Changes in the semantic association of "bullying" with its most common nearest neighbours in psychology article abstracts by decade, 1980–2017, from Vylomova et al. (2019)

a growing emphasis on victimisation and on more adult- and organisation-focused meanings embodied by "workplace" and "peer". These trends are again in accordance with the argument that harm-related concepts are progressively acquiring new, broadened meanings in the academic discourse of psychology.

Further linguistic dimensions of concept creep

Findings such as those of Vylomova and colleagues support the general concept creep hypothesis that a variety of harm-related concepts have increased their semantic breadth over time. Further analyses are required to examine semantic changes in specific concepts in greater detail, and to explore new text corpora. The challenges involved in demonstrating historical semantic change computationally should not be under-estimated. However, computational analyses such as these do not exhaust the ways in which concept creep might be understood as a linguistic as well as a psychological phenomenon. Language, being a dynamic system, constantly changes in many ways: phonologically, morphologically, syntactically, and semantically, and linguists have a long history of studying language evolution and change. Some of these might shine a new light on concept creep.

Linguistic studies of diachronic (temporal) change processes have examined how words transform grammatically (grammaticalisation; Lehmann, 1985), how new words form, how other words become obsolete and fall out of use, and how existing words acquire new meanings or how they lose part of their meaning. The latter two changes are associated with semantic shifts – and are most relevant to the understanding of concept creep.

Bloomfield (1933) conducted the first comprehensive study of semantic shifts, defined as "innovations which change the lexical meaning rather than the grammatical function of a form" (p. 425), and developed a typology of nine forms or mechanisms of semantic change that is still used by researchers. Several of these may be particularly relevant to concept creep. First, Bloomfield distinguished narrowing (e.g., the Old English "mete", meaning food, becoming "meat", meaning edible flesh, a subset of food) and widening (e.g., the Middle English "bridde", meaning young birdling, coming to refer, as "bird", to birds of all ages). Second, Bloomfield presented litotes as a shift from a weaker to a stronger meaning (e.g., the pro-English "*kwalljan" [to torment] becoming the Old English "cwellan" [to kill]), whereas hyperbole is a shift in the opposite direction (e.g., the pre-French "*extonare" [to strike with thunder] becoming "astonish"). Finally, metaphor is a process whereby a word's meaning may extend to new, analogically linked meanings (e.g., the Germanic "biting" engendering "bitter", meaning harsh of taste).

Bloomfield's typology affords several ways to understand concept creep that might be explored in the future linguistic research. As a form of semantic expansion, some instances of creep might be interpreted as an expression of widening, hyperbole, or metaphoric extension. It could be argued that some instances of horizontal creep, in which a concept's meaning extends into qualitatively new phenomena (e.g., active "abuse" coming to refer also to passive neglect) might be viewed as examples of widening, and other instances of horizontal creep, where the new meaning is more clearly analogous to the original meaning (e.g., "cyber-bullying" vis-à-vis unmediated "bullying"), might be ascribed to metaphor. Hyperbole, in contrast, appears to be more germane to vertical creep, where a concept's meaning stretches to include less severe or intense phenomena (e.g., "trauma" being used to refer to relatively mild or vicariously experienced adversities).

The linguistic notion of "semantic bleaching" may also illuminate concept creep. This phenomenon occurs when concepts gradually lose semantic content (i.e., intensional features). For example, in earlier times "awesome" referred to events that induced awe, whereas it has come to refer to events that are merely positive, "awe" having been washed from the meaning. Although bleaching is generally understood as a form of semantic loss, the increased vagueness of bleached concepts creates a gain in the range of phenomena to which they may refer. Semantic bleaching can, therefore, be seen as a shift or change in the distribution of meaning rather than a simple loss (Hopper & Traugott, 1993). Arguably the relaxation of the original repetition criterion in recent definitions of bullying is an example of semantic bleaching that illustrates this point: if the concept of bullying loses this semantic feature, coming to refer to any antagonistic interpersonal behaviour directed towards someone of relatively low power, then bullying broadens to encompass single episodes of antagonism.

Linguistic concepts such as these should help to make sense of concept creep, as it is ultimately a claim about shifts in concept meanings. There is ample evidence from sociolinguistics that language change commonly reflects societal and cultural shifts of the sort implicated in concept creep (Blank & Koch, 2013; Kutuzov et al., 2018). Future research must examine which of the forms of semantic change best capture examples of concept creep, and also whether creep is best understood to involve gradual changes in a word's core meaning or more rapid changes in its nearest neighbours. This distinction, sometimes described as between linguistic drifts and cultural shifts, has been supported by the recent work of Hamilton et al. (2016b) and is amenable to computational linguistic study.

Causes of concept creep

The previous section introduced a sharpened analysis of the boundaries of concept creep by clarifying the domain of concepts proposed to have crept. In addition to this conceptual refinement, it presented evidence that harm-related concepts in psychology have indeed tended to undergo semantic inflation in the past half-century when semantic change is evaluated using methods drawn from computational linguistics. The section also suggests that the semantic changes involved may have been especially marked in the 1980s and since, and that in certain respects they resemble forms of semantic widening identified in the linguistic literature. These theoretical and empirical advances help to refine the characterisation of concept creep first sketched in N. Haslam (2016a), taking the evidence for concept creep well beyond its qualitative interpretation of conceptual change and quantitative analysis of word frequencies.

Improving how concept creep is characterised is an important basis for future research, but it is only a preliminary step towards explaining concept creep. If harm-related concepts have tended to broaden their meanings within psychology in the 1980s and since, especially if it can be shown that this broadening is at least somewhat distinctive to these concepts; then, we must ask what factors are responsible for that semantic inflation. No simple or definitive answer can be provided at present, and clarifying the complex, correlated, and multi-level factors involved in studying historical change processes is notoriously difficult. However, several candidates' causal influences can be sought in cultural shifts, societal changes, and intentional changes brought about by motivated political actors. In addition, we argue that cross-sectional evidence about the sorts of people who hold broader harm-related concepts may point to factors driving historical conceptual change.

Cultural causes

One of the most plausible explanations for concept creep is a growing sensitivity to harm in Western cultures, manifest in a rise in harm-based morality. If people have collectively become more sensitive to harm in their environments, we might expect them to identify a wider variety of phenomena as harmful. Expanded concepts of harm would serve this end. Evidence for this claim comes from a recent study by Wheeler et al. (2019), who examined trends in the use of moral language across the 20th century through the lens of the five domains of moral concerns proposed by Moral Foundations Theory (MFT; Graham et al., 2011). MFT, designed to categorise the automatic and intuitive emotional reactions that commonly occur in moral evaluation across cultures, identified five psychological systems (or foundations): Harm, Fairness, Ingroup, Authority, and Purity.

Each of the moral foundations is distinct, in that each embodies a separate set of associated concerns, virtues, and vices. The Harm foundation refers to issues of cruelty, the suffering of others, and the virtues of compassion, caring, and kindness. Fairness includes concerns of injustice, unfair treatment, reciprocity, equality, cheating, and individual rights. The Ingroup foundation covers loyalty and obligations for group membership, self-sacrifice, and betrayal. Authority is concerned with social order, an obligation to conform to hierarchical relationships, and obedience and respect for authority and tradition. The Purity foundation refers to contagion, both physical and spiritual, and encompasses concerns of sanctity, self-control, and the virtues of innocence and wholesomeness (Haidt & Graham, 2007; Haidt & Kesebir, 2010).

Using Google NGram Viewer, which allows users to gather relative word frequencies from the vast Google Books corpus of digitised books, we measured the proportion of moral foundation-related words within the corpus for the years 1900–2007. Foundation-specific terms were drawn from the Moral Foundations Dictionary (Graham et al., 2009), which was designed to include both positive "virtue" terms and negative "vice" terms, reflecting the valued and disvalued concepts of each of the five moral foundations. The dictionaries for each of the five moral foundations contained an average of 55 words and with one exception demonstrated excellent internal consistency. Trends in the relative frequency of each term in each year were standardised and then summed to generate average trajectories for each moral foundation, conceptualised as indices of its shifting cultural salience. We anticipated that the Harm foundation would show rising salience in recent decades, a requirement if rising cultural salience of harm underpins concept creep.

Figure 3 presents the trajectories for the five moral foundations, highlighting Harm, which shows a gentle decline until about 1960, punctuated by noticeable short-term rises around the two World Wars, and then rises steeply from about 1980. No other moral foundation demonstrates

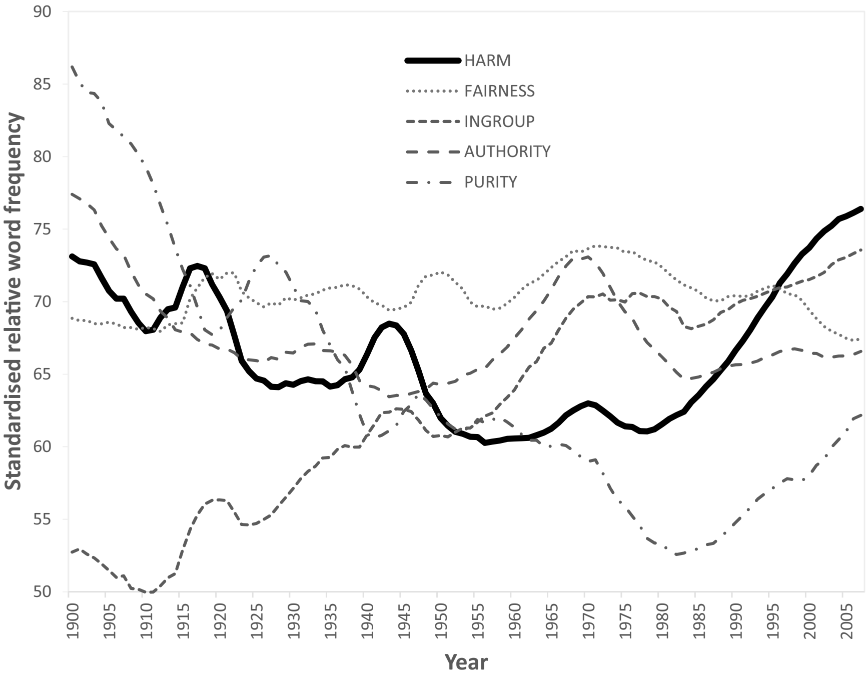

Figure 3. Changes in relative frequency of words associated with the five moral foundations in the Google Books corpus, 1900–2007, fromWheeler et al. (2019)

a similar rising trajectory in this period, suggesting that the upswing is specific to Harm. This sharp but steady increase in the late 20th century is consistent with N. Haslam's (2016a) proposition regarding concept creep and with the evidence of the computational analyses conducted by Vylomova et al. (2019) which identified the 1980s and 1990s as the period of greatest conceptual change. It is worth noting that positive and negative terms (i.e., virtues and vices) were highly correlated within the foundation of Harm, indicating that the late increase in harm sensitivity was not restricted to only negative terms (e.g., endanger, hurt, suffer) but also included terms relating to a morality of care (e.g., compassion, protect, peace). Although this analysis falls well short of demonstrating that large-scale cultural changes in the cultural salience of or sensitivity to harm are causally responsible for concept creep, it raises the plausibility of that proposition.

Societal causes

Broadening concepts of harm might reflect shifts in cultural values and preoccupations, but they could also spring from objective changes in social conditions. In particular, they might be due to decreases in the rates of harmful phenomena that people are exposed to in everyday life. It has been argued that rates of violence (e.g., homicide, war deaths) and other kinds of harm have been steadily declining for centuries (Pinker, 2011). While there are a variety of factors contributing to this decline, Pinker highlights that a central contributor is increased in empathy, self-control, and rationality in the modern era. The prevalence of severe harm is in decline because violence is becoming increasingly reprehensible to people worldwide.

However, despite the apparent decline in almost all kinds of violence, Pinker (2018) argues that people are if anything more focused on the prevalence of harm than in earlier times. Although he attributes some of this intensifying focus to the saturation and sensationalist coverage of contemporary media, psychological factors also contribute to it. We suggest that one reason why people continue to perceive harm as undiminished or even rising despite its reducing prevalence is because notions of what counts as "harm" inflate in response to that decreasing prevalence. In other words, as the rates of objective harm reduce, concepts of harm may expand to encompass new and previously innocuous phenomena, making harm appear as widespread as it ever was.

This possible mechanism linking reductions in the objective prevalence of harm to the semantic inflation of harm-related concepts has recently been investigated experimentally. In seven studies, Levari et al. (2018) found evidence of "prevalence-induced concept change", wherein people expanded their concepts in response to decreasing exposure to instances of those concepts. This was first demonstrated using simple perceptual phenomena. Participants were exposed to a long series of coloured dots on a screen that ranged from very purple to very blue and asked to categorise each dot as "blue" or "not blue". Participants in a condition which gradually decreased the proportion of blue dots began to classify dots as blue that they previously identified as purple, suggesting they adaptively expanded their concept of "blue" in response to seeing fewer blue dots. In subsequent studies, this same effect was replicated using more complex social phenomena. Participants who were exposed to decreasing prevalence of threatening faces on a screen grew more likely to categorise ambiguous faces as "threatening" and others asked to review the ethicality of research proposals were more likely to reject ethically neutral research proposals as unethical when the proportion of unethical proposals they read decreased.

Levari et al.'s (2018) findings imply that concept creep could be caused distally by the decreasing prevalence of the phenomena relevant to that concept. If crime and automobile accident rates decline, concepts of trauma may tend to broaden to include less serious harms, and if blatant expressions of bigotry recede the concepts of prejudice may extend to subtler manifestations. The implications of such conceptual changes may often be minor. However, both Pinker (2018) and Levari et al. (2018) point out that an unfortunate implication of this tendency to creep harm-related concepts in response to vanishing prevalence is that it can prevent us from acknowledging how much the prevalence of harms has declined. Determining whether concept creep is causally connected to downward shifts in the objective rate of harm in society, outside artificial experimental conditions, is therefore an important task for future research.

Concept creep as a motivated process

The preceding sections proposed that concept creep might be an unintended consequence of deeper cultural or societal changes. However, as several commentators have convincingly argued (Furedi, 2016; Sunstein, 2018), some examples of concept creep are surely the work of deliberate actors who might be called "expansion entrepreneurs". These actors actively seek the expansion of concepts to serve specific goals. In this section, we systematically explore the idea of concept creep as a motivated phenomenon by identifying some of the potential goals and incentives. We identify two broad domains of motivation that may drive expansion entrepreneurs: amplifying social problems; and importing moral, political or legal responses from already legitimised social issues.

The most well-established body of academic research on motivated concept creep appears in the context of collective action, where concept expansion can be used as a tactic to amplify the perceived seriousness of a movement's chosen social problem (Best, 1990; Charmaz et al., 2019). Movements for social change generally arise out of the group's identification of a social problem that the collective actors are driven to address (Van Zomeren, 2013; Wright et al., 1990). However, the definitions of these social problems rarely remain static as a social movement unfolds (Jenness, 1995), and groups may be motivated to expand or reframe the social problem in order to meet their strategic ends (Charmaz et al., 2019).

The process by which a social problem expands to cover a broader range of issues has been extensively documented within the context of real-world social movements by sociologists who describe the phenomenon as "domain expansion" (Best, 1990; Jenness, 1995). Such expansion can be effective means of enhancing the perceived seriousness of a social problem or threat by increasing the perceived prevalence of both "victims" and "perpetrators" (Jenness, 1995; see also Haidt, 2016). In one analysis, for example, Jenness (1995) analysed 32 campaigns against anti-gay/lesbian violence starting in the late 1980s, and found that the movement encouraged individuals affected by violence to share experiences that they may have otherwise considered too trivial or non-problematic; enabled the documentation of instances of anti-gay/lesbian violence that would not be classified as illegal; and garnered greater visibility to anti-gay/lesbian violence by proliferating reports on the high prevalence of the issue. Arguably, a similar strategic dynamic appears in recent activism that seeks to eradicate sexual harassment and promote gender equality. Where once sexual harassment referred to the use of threats or bribes to extort sex from employees (Best, 1999), in the age of #MeToo it commonly includes verbal and online behaviours. By this expanded definition, the prevalence of victims of sexual harassment was found to be 81% of American women in 2019 (Center on Gender Equity and Health, 2019), almost double the 43.6% rate reported by the Centres for Disease Controls and Prevention in 2015 (Smith et al., 2018) using a less expansive definition.

Another form of motivated concept creep, identified recently by Sunstein (2018), involves deliberately enlarging a concept so as to import the existing (negative) social or legal responses from its original narrower meaning into the new conceptual territory. This re-drawing of conceptual boundaries features in Sunstein's (2018) account of "opprobrium entrepreneurs", who seek to extend the opprobrium associated with an existing concept (e.g., bullying, prejudice) to the specific cases that they wish to condemn. Even if such extensions do not directly give rise to institutional remedies that apply to the narrower meaning, the stigma attached to the term may publicly tarnish the perpetrator. According to Sunstein (2018), the ultimate goal of opprobrium entrepreneurs is to trigger informational and reputational cascades against people holding views they oppose. If an opposed belief or expression can be labelled a form of "violence" or "hate", even if it does not rise to the level of legal definitions of those concepts, it may provoke the intensely moralised reaction normally recruited against other forms of violence or hate.

We have proposed that deliberate actors may expand concepts of harm either to amplify the perceived seriousness of social problems or to extend social, political or legal responses from already legitimated harms to new ones. It is uncertain how great such motivated processes play a role in concept creep and how they relate to the broader cultural and societal factors discussed previously (e.g., whether expansion entrepreneurs lead or merely capitalise on cultural trends favouring increased sensitivity to harm). However, we strongly suspect that concept creep is driven at least in part by motivated processes.

Individual differences in concept breadth

Concept creep is a historical phenomenon, and efforts to clarify its causes must, therefore, explore longitudinal dynamics. However, examining cross-sectional differences between people in concepts of harm may inform our understanding of concept creep in two ways. First, if harm-related concepts are expanding over time, people may adopt the broadened meanings to different extents as a function of their differing exposure to concept creep.

Second, the individual differences in political attitudes, personality traits, or demographic characteristics that are associated with holding broader concepts of harm may have implications for the historical drivers of concept broadening. With this rationale, we have carried out two studies of the individual difference correlates of harm-related concept breadth (HCB), defined as variations in the inclusiveness of concepts harm.

McGrath et al. (2019) developed a measure of HCB in which participants judged on a 1–6 scale whether abuse, bullying, prejudice, and trauma were present in vignettes that presented ambiguous or marginal examples of these concepts (five vignettes per concept). Against the possibility that concept breadth would be concept-specific, and for the proposition that they all share a common theme of harm (N. Haslam, 2016a), the subscales of the HCB measure were all positively correlated (mean *r*s =.33 and .35 in the two studies). Participants holding inclusive concepts of trauma also tended to hold inclusive concepts of prejudice, for example. The new scale had good reliability (α = .78 & .79 in Studies 1 & 2) as a measure of general HCB.

The two studies reported in McGrath et al. (2019) examined several potential correlates of the new HCB scale. In both studies, the strongest correlates of holding expansive concepts of harm were compassion-related trait values, left-liberal political attitudes, and forms of morality associated with both. However, several other distinctive associations were also found. In the first study of 276 American Amazon Mechanical Turk (MTurk) Workers, HCB was associated with affective and cognitive empathy, assessed by the Interpersonal Reactivity Index (Davis, 1983), a liberal political orientation, and most strongly of all with endorsement of harm- and fairness-based morality (*r*s = .44 & .40), as assessed by the Moral Foundations Questionnaire (Graham et al., 2011). Although endorsement of the harm-based morality covaried with liberalism and affective empathy, these three associations were the three independent predictors of HCB in a multiple regression analysis. Subsequent work (Jones & McNally, 2020) has replicated the association between liberalism and holding broader concepts of trauma, using a different measure of concept breadth, suggesting that this association is robust.

In the second study of 309 American MTurk Workers, constructs reflecting emotional concern for others and liberalism again emerged as strong correlates of holding broader concepts of harm (McGrath et al., 2019), which were also associated with sensitivity to perceiving injustice towards others, assessed by the Justice Sensitivity Inventory (Schmitt et al., 2010). However, in this study HCB was additionally associated with several variables that are less straightforwardly prosocial or political. HCB was found to correlate modestly with a tendency to see oneself as the victim of injustice ("victim sensitivity") and to greater feelings of personal vulnerability (a facet of Neuroticism), and a sense of personal entitlement (Brummel & Parker, 2015). The study also demonstrated that most of these individual difference variables were associated with HCB even after statistically controlling for a generalised tendency to hold more inclusive concepts. Interestingly, holding broader concepts of harm was not consistently associated with younger age in the two studies (Study 1 $r = -.02$, $p > .05$; Study 2 $r = -.21$, $p < .01$). Given that concept creep has been occurring gradually over recent decades, it might be expected that older individuals would tend to hold the relatively narrow concepts of harm that prevailed when they were growing up, but a weak negative correlation between HCB and age was only obtained in one study. Evidently, a simple generation-based interpretation of variations in the expansiveness of concepts of harm is untenable. More generally, although the samples in the two studies had substantial age ranges, their cross-sectional nature means they cannot directly inform us about processes of historical change, and their exclusively American composition further qualifies any general statements about how age is associated with HCB.

Our examinations of the correlates of harm concept breath to date have implications for popular narratives of concept creep. On one narrative, the expansion of concepts of harm represents a welcome and politically progressive increase in concern for the welfare of others, and especially the most vulnerable (e.g., Cikara, 2016). According to another narrative, concept creep is associated with "fragility" and "victimhood" and is most evident among younger people (e.g., Lukianoff & Haidt, 2018). Although both narratives find a degree of support in this work, our findings suggest that people who hold broader concepts of harm tend to demonstrate a strong concern for others and are not disproportionately young.

Consequences of concept creep

It is a truism that how we classify and interpret our experience is socially shaped, and that it influences how we behave and interact. It is also obvious that changes in the concepts we use to make sense of our experience are likely to induce changes in behaviour and social relations. This point has been made most forcefully by the philosopher Ian Hacking (1995), whose work on "looping effects" demonstrates how new concepts and classifications give rise to new social realities as people recognise themselves and others in new ways and form new identities and attitudes. Conceptual changes loop back to engender changed realities. It is therefore timely to ask what the effects of concept creep might be.

Social conflict

One probable effect of concept creep is widespread disagreement among people who hold harm-related concepts of differing breadth. If concepts of abuse, bullying, prejudice and the like are undergoing gradual expansion then people whose formative experiences are closer to and farther from recent shifts – whether as a function of age or social location (e.g., education) – may have discrepant thresholds for identifying harm, notwithstanding the evidence from McGrath et al. (2019) that age is only modestly associated with harm-related concept breadth. In effect, concept creep may stretch not only the definition of harm-related concepts but also the distribution of variants of those definitions within the population. The result of this stretching may be an increasing lack of consensus on moral issues and a widening penumbra of cases that are morally ambiguous or contestable.

The fact that people tend to understand harm so differently may therefore help to explain some yet to be determined proportion of public disagreement on morally charged social issues. Although moral disagreement is often attributed to differences in values, motives, or ideologies, some of it may be due to differences in the breadth of harm-related concepts. It is no wonder that people disagree fiercely if they are perceiving an ambiguous moral situation as either a heinous transgression (broad concepts of harm) or a minor peccadillo (narrow concepts of harm), or if childrearing practices considered normative by members of one generation are viewed as abusive by the next. These disagreements are likely to lead to conflicting opinions about whether problematic behaviour has occurred, how severe it is, what if anything should be done to punish "perpetrators", whether "victims" have legitimate standing as such, and whether institutional intervention is required to ensure justice for them. The fact that concept creep has taken place widens the range of social disagreements that arise. These conflicts may be especially challenging and intractable because, being based ultimately on understandings of "harm", they are intimately associated with moral and political commitments.

Moral typecasting and polarisation

As a result of its intimate connection to harm, concept creep and individual differences in harm-related concept breadth have the potential to contribute to social and political polarisation. One lens through which to understand this contribution is the theory of dyadic morality (Schein & Gray, 2016). This theory argues that the dyad of moral agent (the perpetrator of moral or immoral actions) and moral patient (the recipient of those actions) is central to morality and proposes that perceived harm is the basis for all moral judgement. When concepts creep, new and less severe forms of suffering or maltreatment come to be considered harmful. In this way, concept creep increases the range of actions and experiences that have moral relevance and thereby increases the number of people identified as moral agents and patients. We note in passing here that in identifying some experiences as less severe than others we are referring primarily to their subtlety and lack of intensity as single events rather than to the lack of severity of their effects. We do not deny that subtle events (e.g., ambiguous and deniable examples of racist behaviour) or repeated experiences (e.g., "death by thousand cuts" of street harassment) may have destructive outcomes.

Research in dyadic morality has identified the phenomenon of "moral typecasting" (Gray & Wegner, 2009). It demonstrates that moral patients are perceived as having a greater capacity for experience and sensitivity to pain, that moral agents are perceived as having a greater capacity for intention and responsibility, and that these agencies and patiency perceptions are inversely related. Being identified as a perpetrator of a harm results in diminished perception of patient-like qualities and amplification of agentic qualities. Inversely, for identified victims of harm, perceived agency is diminished, and perceived suffering is augmented. The upshot is that moral typecasting tends to generate polarised perceptions of cold and villainous perpetrators and deeply wounded and passive victims.

The dynamics of moral typecasting may thus contribute to aggravated punishment of moral transgressors because they are seen as being less sensitive to pain, and increased calls for third party protection of victims of moral harms because of their diminished capacity for personal responsibility. As concepts creep and individuals increasingly differ in the range of actions they believe are harmful, discrepant views regarding appropriate responses to these behaviours and experiences may further fuel polarisation. In essence, by broadening the range of situations in which harm is perceived, concept creep is likely to promote polarised views of perpetrators and victims, and by increasing differences between people in moral judgements of complex or ambiguous situations it is likely to deepen moral divisions.

In an illustrative study, Chan and Haslam (2019) examined the correlates of individual differences in the expansiveness of concepts of sexism. A sample of 201 MTurk workers rated whether 20 vignettes describing ambiguous potential instances of sexual harassment and gender discrimination were examples of these concepts. These ratings formed reliable scales for assessing the breadth of the two concepts. Participants then read another vignette describing a female employee encountering sexist behaviour perpetrated by male co-workers and rated her moral patiency (e.g., how upsetting the events were, how much she deserved compensation, whether a third party should intervene) and their moral agency (e.g., how responsible they were for their behaviour, how severe it was, how much they deserved punishment). They also completed the Moral Foundations Questionnaire (Graham et al., 2011) to assess endorsement of the five moral foundations and measures of political orientation and religiosity.

Chan and Haslam (2019) found that people holding broader concepts of sexual harassment and gender discrimination were more likely to endorse harm-based morality (*r*s = .27 & .39, *p*s <.01), consistent with our theoretical claim about the centrality of harm to concept creep and the findings of

McGrath et al. (2019) presented earlier. They were also somewhat more likely to have a liberal political orientation and to be female. In addition, participants holding broader concepts of sexism were more likely to see the woman and her male co-workers in the workplace sexism vignette in morally typecast ways (Gray et al., 2012). They saw the female victim as especially high in moral patiency (*r*s = .43 & .63, *p*s <.01) and the male perpetrators as especially high in moral agency (*r*s = .41 & .67, *p*s <.01). Importantly, the breadth of participants' concepts of sexism predicted these polarised judgements much more strongly (βs = .55 & .58, *p*s < .001) than their political orientation or gender, although more liberal participants also made significantly more sympathetic judgements of the woman (β = .16, p = .01) and nonsignificantly harsher judgements of the men (β = .09, p = .16). By implication, the breadth of people's concepts of harm may be a pre-eminent determinant of their moral judgements, potentially more influential than their political ideology or identity-based interests in some cases. Although this correlational study does not licence causal inferences, it is consistent with the view that concept creep leads ideas of harm to inflate, and that inflation may contribute to polarised moral judgements and resulting social conflict.

Speech codes and hate

The expansion of harm-related concepts has implications for acceptable self-expression and free speech. Creeping concepts enlarge the range of expressions judged to be unacceptably harmful, thereby increasing calls for speech restrictions. Expansion of the harm-related concepts of "hate" and "hate speech" exemplifies this possibility. N. Haslam and Murphy (2020) present evidence that the word "hate" has risen steeply in frequency of use both within the culture at large and in academic psychology. In the Google Books corpus, the relative frequency of the word dropped from 1920 to around 1980 then rapidly rose about 70% to the end of the corpus in 2008. In a corpus of over 500,000 psychology article abstracts from 1970 to 2018 the relative frequency of "hate" rose even more steeply, increasingly roughly threefold from the 1970s and 1980s to the 2010s. As was seen in the earlier findings of Vylomova et al. (2019) and Wheeler et al. (2019), both rises are steepest in the 1980s and since, as is true of other harm-related terms. This increase in the frequency of use commonly co-occurs with semantic broadening. In legal and political scholarship, "hate speech" has also changed its meaning in relation to harm in the recently proposed view that hate speech *constitutes* harm itself (Waldron, 2012) rather than it only *causing* harm to occur (Barendt, 2019). An obvious consequence for speech that is labelled "hateful" is that it usually becomes legally or socially disallowed from public discourse (Barendt, 2019). While we support the need for social and legal regulation on unmitigated free expression, we also note the inevitable trade-off between increasing speech prohibitions and people's autonomy for self-expression. A probable consequence of an expanding "hate speech" concept is increasing prohibitions on what is deemed acceptable belief expression and exchanging of ideas.

A second plausible consequence of creep in the concept of "hate" is the divisions it may sow in society. Concept creep does not occur uniformly or concurrently across the population but may be adopted more quickly by some individuals and groups than others, perhaps as a partial function of its social or political benefits to them or its alignment with their existing sympathies. A recent report on attitudes to free speech in the USA. (Ekins, 2017) revealed very large differences between people in concepts of "hate speech", with much of this variance being associated with differences in race, gender, age, and political ideology. Non-white, female, young, and politically liberal people were generally more likely to identify as hateful phenomena that others judged to be merely offensive or innocuous. There was also a tendency for more conservative people to identify "hate" more often in negative statement towards the USA, White people, and the police, implying that the link between concept creep and liberalism may not be invariant, but this tendency was weak relative to the more general pattern. This evidence of more conservative people holding a relatively low threshold for identifying "hate" when entities they value are targeted raises the possibility that there might be concepts where conservatism rather than liberalism is associated with greater concept breadth, contrary to the general pattern observed by McGrath et al. (2019).

Differing definitions of "hate", such as those documented by Ekins (2017), not only fuel disagreement about how specific episodes of controversial expression should be classified but also about their causes and remedies. Ekins found that groups who held more expansive concepts of "hate" were more likely to believe that problematic expressions sprang from the speaker's bad intentions, that they constituted acts of violence, and that there should be laws against them. Concept creep, and associated differences in the breadth of concepts of harm, may therefore have substantial implications for political disagreement over the limits of expression and may foster hostile and restrictive responses.

Concept creep and identity

A fourth domain in which concept creep is likely to have significant effects is its effects on identities. As Hacking's (1995) writings on looping effects and what he calls "dynamic nominalism" show, new concepts create new human kinds by altering the identities we ascribe to ourselves and others. As concept creep expands the range of phenomena that are perceived as harmful, it expands the range of people who are touched by harm and who may thereby come to identify as harmed or harmful.

The implications of concept creep for identity are particularly obvious in the field of mental disorder. Concepts of disorder have broadened considerably in recent decades, both through horizontal creep ("disease mongering") and vertical creep ("threshold lowering"). These enlargements of the field of psychopathology, sometimes referred to as diagnostic expansion or inflation (Frances, 2013; N. Haslam, 2016b), have taken place steadily over successive editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM), and increase the prevalence of diagnosable mental disorder in the population. As more people receive diagnoses, they are apt to adopt the disorder as part of their self-concept and to have their selfhood perceived through that lens by others who are aware of their diagnosis. Adopting a disorder identity may have mixed blessings for the person who receives a diagnosis. On the positive side of the ledger, it may offer clarity about previously confusing experiences, the hope of suitable treatment and recovery, and a community of fellow sufferers. On the negative side, however, this identity may foster a view that their problems are fixed rather than mutable, that they are damaged or broken, and that they cannot control their fate (N. Haslam & Kvaale, 2015).

The increasing prevalence of diagnoses may also lead to increases in the tendency for individuals to assume social identities that are defined in terms of their relevant disorder. While the adoption of social identities is generally thought to be beneficial for individual well-being (C. Haslam et al., 2012), a growing body of evidence suggests that increased identification may, in fact, be detrimental in the case of identification with mental illness (see Cruwys & Gunaseelan, 2016; Klik et al., 2019). For example, Cruwys and Gunaseelan (2016) demonstrate that diagnosed individuals who perceive depression to be a central part of their identity were more likely to internalise the symptom norms related to depression (e.g., "keep thinking negative and unhelpful thoughts"), in turn leading to lower psychological well-being. Thus, for some individuals, a disorder diagnosis may, in fact, become a self-fulfiling prophecy, where the resultant assumption of a disorder-based identity may lead to the internalisation and manifestation of group norms dictating the way that those with the disorder should think and behave. In addition to any such effects on identity, broadened concepts of mental disorder may have other consequences such as unnecessary treatment via over-diagnosis, and exposure to the stigma and discrimination that face many people with mental health conditions. Rapid apparent rises in the prevalence of particular conditions, which actually reflect broadened diagnostic criteria, may also prompt mistaken searches for new causal agents to account for the change. This dynamic may underpin the spurious attribution of rising rates of autism to vaccination, as well as other spurious explanations of questionable "epidemics" (e.g., anxiety, attention deficit hyperactivity disorder). Concept creep in the psychiatric arena therefore potentially has substantial implications.

Concept creep may also have implications for identity outside that arena. Mental disorder is just one form of creeping harm, and broadening concepts of abuse, bullying, prejudice, and trauma might also lead an increasing number of people not only to label their experiences in these terms but also to self-identify as victims of them (Branscombe et al. 1999; Ellemers & Barreto, 2006). As these concepts come to refer to new and less extreme phenomena than they did in earlier times, they become available to a wider range of people as ways to make sense of their experience. As with mental disorder concepts, the implications of identifying oneself as a victim of one of these experiences may be mixed, but it is credible that some implications may be negative. For example, as people tend to infer that "traumas" have severe and enduring effects that require professional intervention, identifying oneself as a victim of trauma – perhaps especially in the case of marginal examples of this creeping concept – may be detrimental. Seeing oneself as a permanently wounded victim of a trauma may be less conducive to resilience than interpreting the experience in less catastrophic terms. Recently, Jones and McNally (2020) provided experimental evidence consistent with this possibility. Participants exposed to a gruesome video clip were more likely to define it as a trauma and report negative emotions and subsequent post-traumatic symptoms if they held broader concepts of trauma.

It is debatable whether there has been a marked increase in the adoption of victim identities as some commentators have argued (Campbell & Manning, 2018). It can also be debated whether victim identities have adverse implications, such as promoting passivity, self-exculpation, moral grandstanding, and entitlement, or are primarily empowering (Cikara, 2016). However, concept creep would be a plausible partial explanation of these cultural shifts if they have even a germ of truth. It remains to be determined how changes in concepts of harm have altered how people make sense of their experiences with adversity, or how many people have incorporated those experiences in an identity as a harmed person.

Positive consequences of concept creep

Sometimes concept creep is presented in an exclusively negative frame, on the assumption that its effects are invariably damaging or the belief that concepts should remain static. However, a balanced evaluation of concept creep requires an exploration of its both negative and positive implications. To that end, we offer three positive consequences of the phenomenon. First, concept creep creates labels that can be useful in drawing attention to harms previously overlooked. Consider the vertical expansion of abuse to include "emotional abuse". Prior to this development, society lacked a conceptual framework to interpret the behaviour as harmful; it was "just part of life" – a problem without a name. The invention of the label resulted in a collective appreciation of maltreatment of intimates as abuse. It alerted the collective conscience to the harms associated with emotional abuse, delivering an essential tool for observers and victims to interpret, identify, and protest it. Similarly, although the broadening of the concept of mental disorder by the accretion of newly labelled conditions is often criticised as disease mongering, it enables people whose difficulties were previously ignored to seek support.

Second, concept creep can prevent harmful practices by modifying social norms. Social norms emerge, in part, due to normative expectations: the belief that behaviour is typically condemned or endorsed by one's community (Bicchieri, 2016). As psychological concepts of bullying, prejudice, and abuse hold normative weight, marking new phenomena with these labels indicates that they are norm violations. For example, the collective labelling of child neglect as "abuse" signals that the behaviour is communally rejected, establishing a social norm that child neglect is wrong. Because social norms are powerful determinants of behaviour, the conceptual extension of negative psychological concepts should prompt a decrease in newly problematised behaviours. Changing definitions of bullying that include social exclusion and antagonistic acts expressed horizontally rather than only downwards in organisational hierarchies may also entrench norms against the commission of destructive behaviour.

Finally, the expansion of psychology's negative concepts can motivate interventions aimed at preventing or reducing the harms associated with the newly categorised behaviours. For instance, the conceptual expansion of addiction to include "behavioural" addictions (e.g., gambling and internet addictions) has prompted a flurry of research into treatment options, which has found that a range of psychosocial treatments can be successfully used to treat gambling, internet, and sexual addictions (Yau & Potenza, 2015). If these interventions are successfully implemented, their introduction constitutes a cause for celebration, and concept creep can be credited with creating the label that motivated the response.

An integrative model

Figure 4 synthesises the links we have proposed between concept creep as a historical phenomenon and its causes, correlates, and consequences. Concept creep itself is understood to take two primary forms that may be conceptualised linguistically through mechanisms such as metaphor, semantic widening, hyperbole, and semantic bleaching. It is hypothesised to be driven by a combination of cultural changes reflected in rising sensitivity to harm, operating primarily from the 1980s and since; societal changes reflected in gradually declining rates of adversity and risk in Western nations; and motivated conceptual expansions in the service of social or political ends. Concept creep is hypothesised to be associated with systematic individual differences in tendencies to inflate harm, which research shows to involve a combination of prosocial tendencies, political liberalism, and personal vulnerability and sensitivity to threat. Creep is hypothesised to be in part responsible for an assortment of consequences, including the promotion of social disagreements and polarised moral judgements, drives to restrict expression, personal identities based on harm and victimisation, but also progressive social change.

Agenda for future research

Research on concept creep is in its early stages and substantial work is required to fill in many of the gaps and speculations identified in this review. Future studies must address a variety of lingering issues related to the characterisation, causes, and consequences of this phenomenon.

In regard to characterisation, considerable work is needed to refine computational linguistic analyses of the time course of semantic inflation of harm-related concepts, and to determine whether such inflation is indeed somewhat distinctive to harm-related concepts relative to others. Research is also needed to answer whether concept creep is primarily about harm as a thematic domain rather than negativity as a conceptual valence, and to clarify the relationship between rises in the use frequency (i.e., cultural salience) of concepts and rises in their semantic breadth. Studies must also examine semantic inflation in text corpora beyond psychology articles to clarify whether the same broadening has occurred in other academic disciplines and in the culture at large, and to explore how changes in academic discourse have disseminated into that culture. Research should examine broad trends across multiple creeping concepts as well as explore the historical trajectories of specific concepts (e.g., trauma, prejudice). In addition, researchers should examine possible semantic broadening of harm-related concepts in languages other than English. Finally, linguistic analyses should investigate how well examples of concept creep can be understood in terms of known forms of semantic change such as widening and hyperbole.

Future research on the causes of concept creep must specify more precisely the cultural changes that have contributed to the phenomenon, clarifying whether, where, and how any rising sensitivity to harm has manifested in our societies, and in the process attempt to locate when semantic broadening accelerated, noting lines of evidence pointing to the 1980s as a critical decade. Research should also ask whether the link between the reduced prevalence of a phenomenon and widening concepts of it demonstrated experimentally by Levari et al. (2018) is plausible as a mechanism for explaining widening concepts of harm. Finally, studies must resolve the extent to which concept creep reflects deliberate expansion entrepreneurship by politically or otherwise motivated actors.

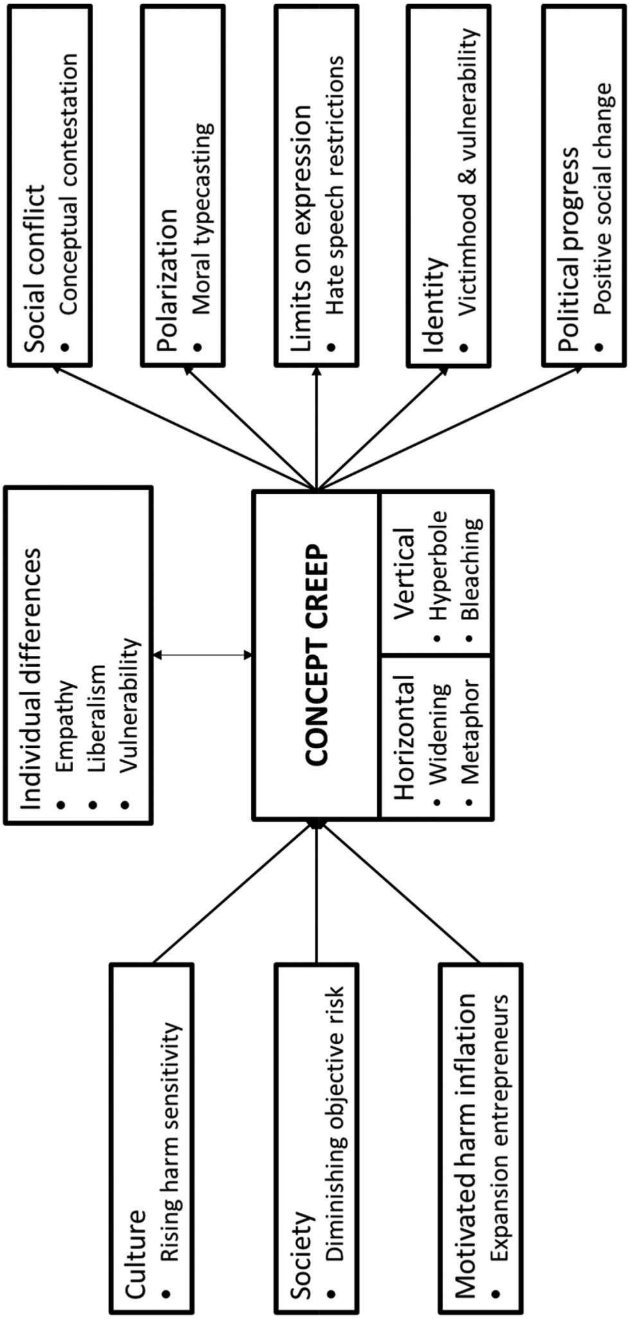

Figure 4. Integrative model of concept creep's proposed causes, correlates, and consequences.

In relation to the consequences of concept creep, further research should examine the extent to which the historical broadening of harm-related concepts does indeed contribute to social and political polarisation and disagreements about the interpretation of problematic events, such as whether or not they represent bullying, prejudice, or trauma. Studies are also urgently needed to clarify how inflating concepts of harm influence people's identities, whether victim-based identities are indeed on the rise as a result of such inflation, and whether such identities primarily engender vulnerability or empowerment in people who adopt them. More basically, it will be crucial to determine whether the legacy of concept creep is, on balance, predominantly positive or negative, and whether that question can even be answered in a neutral and unbiased way.

Many of these future research directions, especially in relation to concept creep's causes and consequences, are exploratory. The theory of concept creep does not make strong claims about which of several possible causal influences is most central or whether the consequences are primarily desirable or undesirable. However, some questions imply more stringent tests of the theory. If semantic broadening over the past half-century has not been greater among harm-related concepts than among other concepts on average, the theory is challenged. It is similarly challenged if close examination of concepts that have and have not broadened points to systematic exceptions to the prediction (i.e., groups of harm-related concepts that have not inflated or groups of harm-adjacent concepts [e.g., fairness-related] that have inflated), which might suggest that "harm" is too broad or narrow a concept to account for observed patterns of semantic expansion. The theory of concept creep would also be called into question if observed patterns of semantic broadening were found to be mere artefacts of increasing word frequency or if it was confined to academic discourse and not present in public discourse. A systematic investigation of semantic changes within psychology and beyond might enable more searching tests of concept creep theory, and serious consideration of exceptions to its claims might assist in refining it.

Conclusions

Concept creep is a phenomenon that has wide-ranging social and cultural ramifications. It is implicated in several of the most timely and contentious social issues in contemporary Western societies, and places psychology at the forefront of them as a source of many of our dominant concepts for making sense of harm. We believe that social psychology has a special role to play in accounting for the benefits, costs, and causes of creeping concepts. In taking on that role, however, social psychologists will have to engage deeply with theories, findings, methodologies, and scholars from other disciplines in overcoming the challenges of understanding historical, cultural, and linguistic change.

Funding

This work was supported by the Australian Research Council [DP170104948].

ORCID

Nick Haslam http://orcid.org/0000-0002-1913-2340
Melanie J. McGrath http://orcid.org/0000-0001-8632-218X
Joshua Rhee http://orcid.org/0000-0002-6245-7060
Ekaterina Vylomova http://orcid.org/0000-0002-4058-5459
Morgan Weaving http://orcid.org/0000-0002-8519-868X
Melissa A. Wheeler http://orcid.org/0000-0002-0319-1987

References

Barendt, E. (2019). What is the harm of hate speech? *Ethical Theory and Moral Practice*, *22*(3), 539–553. https://doi.org/10.1007/s10677-019-10002-0

Best, J. (1990). *Threatened children: Rhetoric and concern about child-victims*. University of Chicago Press.

Best, J. (1999). *Random violence: How we talk about new crimes and new victims*. University of California Press.

Bicchieri, C. (2016). *Norms in the wild*. Oxford University Press.

Blank, A., & Koch, P. (Eds.). (2013). *Historical semantics and cognition* (Vol. 13). Walter de Gruyter.

Bloomfield, L. (1933). *Language*. Rinehart and Winston.

Brummel, B. J., & Parker, K. N. (2015). Obligation and entitlement in society and the workplace. *Applied Psychology*, *64*(1), 127–160. https://doi.org/10.1111/apps.12023

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Campbell, B., & Manning, J. (2018). *The rise of victimhood: Microaggressions, safe spaces, and the new culture wars*. Palgrave Macmillan.

Center on Gender Equity and Health. (2019). *Measuring #MeToo: A national study on sexual harassment and assault*. University of California.

Chan, J., & Haslam, N. (2019). Broad concepts of sexism predict polarized moral judgments of victims and perpetrators. *Personality and Individual Differences*, *150*, 109488. https://doi.org/10.1016/j.paid.2019.06.031

Charmaz, K., Harris, S. R., & Irvine, L. (2019). *The social self and everyday life: Understanding the world through symbolic interactionism*. John Wiley & Sons.

Cikara, M. (2016). Concept expansion as a source of empowerment. *Psychological Inquiry*, *27*(1), 29–33. https://doi.org/10.1080/1047840X.2016.1111830

Cruwys, T., & Gunaseelan, S. (2016). "Depression is who I am": Mental illness identity, stigma and wellbeing. *Journal of Affective Disorders*, *189*, 36–42. https://doi.org/10.1016/j.jad.2015.09.012

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta control: Laws of semantic change and inherent biases in word representation models. In

Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 1136–1145), Copenhagen, Denmark.

Ekins, E. (2017). *The state of free speech and tolerance in America: Attitudes about free speech, campus speech, religious liberty, and tolerance of political expression*. Cato Institute.

Ellemers, N., & Barreto, M. (2006). Social identity and self-presentation at work: How attempts to hide a stigmatised identity affect emotional well-being, social inclusion and performance. *Netherlands Journal of Psychology*, *62*(1), 51–57. https://doi.org/10.1007/BF03061051

Fabiano, F., & Haslam, N. (in press). Diagnostic inflation in the DSM: A meta-analysis of changes in the stringency of psychiatric diagnosis from DSM-III to DSM-5. Clinical Psychology Review. http://doi.org/10.1016/j.cpr.2020.101889

Frances, A. (2013). *Saving normal: An insider's revolt against out-of- control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life*. William Morrow.

Furedi, F. (2016). The cultural underpinning of concept creep. *Psychological Inquiry*, *27*(1), 34–39. https://doi.org/10.1080/1047840X.2016.1111120

Glancy, J. (2018 September 2). Generation snowflake: Meet the professors who blame helicopter parents for coddling the minds of today's students. *The Sunday Times*. https://www.thetimes.co.uk/article/generation-snowflake-meet-the-professors-who-blame-helicopter-parents-for-coddling-the-minds-of-todays-students-mjwdxftx9

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520. https://doi.org/10.1037/a0013748

Gray, K., Young, L., & Young, L. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Gulordava, K., & Baroni, M. (2011). A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of the GEMS 2011 workshop on geometrical models of natural language semantics* (pp. 67–71). Edinburgh.

Hacking, I. (1995). The looping effect of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 351–383). Oxford University Press.

Haidt, J., & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 5, pp. 797–832). Wiley.

Haidt, J. (2016). Why concepts creep to the left. *Psychological Inquiry*, *27*(1), 40–45. https://doi.org/10.1080/1047840X.2016.1115713

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual* *meeting of the association for computational linguistics (Volume 1: Long Papers)* (pp. 1489–1501), Berlin, Germany.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing* (Vol. 2016, p. 2116), Austin, Texas, USA.

Haslam, C., Jetten, J., & Haslam, S. A. (2012). *The social cure: Identity, health and well-being*. Psychology press.

Haslam, N. (2016a). Concept creep: Psychology's expanding concepts of harm and pathology. *Psychological Inquiry*, *27*(1), 1–17. https://doi.org/10.1080/1047840X.2016.1082418

Haslam, N. (2016b). Looping effects and the expanding concept of mental disorder. *Journal of Psychopathology*, *22*, 4–9.

Haslam, N., & Murphy, S. C. (2020). Hate, dehumanization, and "hate". In R. J. Sternberg (Ed.), *Perspectives on hate: How it originates, develops, manifests, and spreads* (pp. 27–41). American Psychological Association.

Haslam, N., & Kvaale, E. (2015). Biogenetic explanations of mental disorder: The mixed blessings model. *Current Directions in Psychological Science*, *24*(5), 399–404. https://doi.org/10.1177/0963721415588082

Haslam, N., & McGrath, M. J. (in press). The concept creep of trauma. *Social Research*.

Heyer, G., Holz, F., & Teresniak, S. (2009). Change of topics over time-tracking topics by their change of meaning. *KDIR*, *9*, 223–228.

Hilpert, M., & Gries, S. T. (2016). Quantitative approaches to diachronic corpus linguistics. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 36–53). Cambridge University Press.

Hopper, P. J., & Traugott, E. (1993). *Grammaticalization*. Cambridge University Press.

Inglehart, R. F. (2008). Changing values among Western publics from 1970 to 2006. *West European Politics*, *31*(1–2), 130–146. https://doi.org/10.1080/01402380701834747

Jenness, V. (1995). Social movement growth, domain expansion, and framing processes: The gay/lesbian movement and violence against gays and lesbians as a social problem. *Social Problems*, *42*(1), 145–170. https://doi.org/10.2307/3097009

Jones, P. J., & McNally, R. J. (2020, May 11). *Does broadening one's concept of trauma undermine resilience?* https://doi.org/10.31234/osf.io/5ureb

Juola, P. (2003). The time course of language change. *Computers and the Humanities*, *37*(1), 77–96. https://doi.org/10.1023/A:1021839220474

Klik, K. A., Williams, S. L., & Reynolds, K. J. (2019). Toward understanding mental illness stigma and help-seeking: A social identity perspective. *Social Science & Medicine*, *222*, 35–43. https://doi.org/10.1016/j.socscimed.2018.12.001

Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. In *Proceedings of the 24th international conference on world wide web* (pp. 625–635). International World Wide Web Conferences Steering Committee.

Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic word embeddings and semantic shifts: A survey. In *Proceedings of the 27th international conference on computational linguistics* (pp. 1384–1397). Association for Computational Linguistics, Santa Fe, New Mexico, USA.

Lehmann, C. (1985). *Grammaticalization: Synchronic variation and diachronic change* (Vol. 20). Na.

Levari, D. E., Gilbert, D. T., Wilson, T. D., Sievers, B., Amodio, D. M., & Wheatley, T. (2018). Prevalence-induced concept change in human judgment. *Science*, *360*(6396), 1465–1467. https://doi.org/10.1126/science.aap8731

Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science*, *12*(1), 138–169. https://doi.org/10.1177/1745691616659391

Lukianoff, G., & Haidt, J. (2018). *The coddling of the American mind: How good intentions and bad ideas are setting up a generation for failure*. Penguin.

McGrath, M. J., Randall-Dzerdz, K., Wheeler, M. A., Murphy, S., & Haslam, N. (2019). Concept creepers: Individual differences in harm-related concepts and their correlates. *Personality and Individual Differences*, *147*, 79–84. https://doi.org/10.1016/j.paid.2019.04.015

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., & Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv Preprint arXiv:1301.3781.*

Pinker, S. (2011). *The better angels of our nature: Why violence has declined*. Viking.

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. Penguin.

Sagi, E., Kaufmann, S., & Clark, B. (2011). Tracing semantic change with latent semantic analysis. In K. Allen & J. A. Robinson (Eds.), *Current Methods in Historical Semantics* (pp. 161–183).De Gruyter Mouton.

Schein, C., & Gray, K. (2016). Moralization and harmification: The dyadic loop explains how the innocuous becomes harmful and wrong. *Psychological Inquiry*, *27*(1), 62–65. https://doi.org/10.1080/1047840X.2016.1111121

Schlechtweg, D., Hätty, A., Del Tredici, M., & Walde, S. S. I. (2019). A wind of change: Detecting and evaluating lexical semantic change across times and domains. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 732–746). Association for Computational Linguistics, Florence, Italy.

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The justice sensitivity inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research*, *23*(2–3), 211–238. https://doi.org/10.1007/s11211-010-0115-2

Smith, S. G., Zhang, X., Basile, K. C., Merrick, M. T., Wang, J., Kresnow, M., & Chen, J. (2018). *The National intimate partner and sexual violence survey (NISVS): 2015 data brief – updated release*. Centers for Disease Control and Prevention.

Sunstein, C. R. (2018). *The power of the normal.* Available at SSRN https://ssrn.com/abstract=3239204

Van Zomeren, M. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass*, *7*(6), 378–388. https://doi.org/10.1111/spc3.12031

Vylomova, E., Murphy, S., & Haslam, N. (2019). Evaluation of semantic change of harm-related concepts in psychology. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (pp. 29–34). Association for Computational Linguistics, Florence, Italy.

Waldron, J. (2012). *The harm in hate speech*. Harvard University Press.

Wheeler, M. A., McGrath, M. J., & Haslam, N. (2019). Twentieth century morality: The rise and fall of moral concepts from 1900 to 2007. *PLoS ONE*, *14*(2), e0212267. https://doi.org/10.1371/journal.pone.0212267

Winter, B., Thompson, G., & Urban, M. (2014). Cognitive factors motivating the evolution of word meanings: Evidence from corpora, behavioral data and encyclopedic network structure. In E. A. Cartmill, S. Roberts, H. Lyn & H. Cornish (Eds.), *The Evolution of Language* (pp. 353–360).World Scientific.

Wright, S. C., Taylor, D. M., & Moghaddam, F. M. (1990). Responding to membership in a disadvantaged group: From acceptance to collective protest. *Journal of Personality and Social Psychology*, *58*(6), 994–1003. https://doi.org/10.1037/0022–3514.58.6.994

Xu, Y., & Kemp, C. (2015). A computational evaluation of two laws of semantic change. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Pasadena, California, USA.

Yau, Y. H. C., & Potenza, M. N. (2015). Gambling disorder and other behavioral addictions. *Harvard Review of Psychiatry*, *23*(2), 134–146. https://doi.org/10.1097/HRP.0000000000000051