Automated Content Analysis of Cultural Diversity Perspectives in Annual Reports (DivPAR): Development, Validation, and Future Research Agenda

Joep Hofhuis1, Pytrik Schaafraad2, Damian Trilling2, Nastasia Luca1, and Bastiaan van Manen1

1Erasmus Research Centre for Media, Communication, and Culture (ERMeCC), Erasmus University Rotterdam
2Amsterdam School of Communication Research (ASCoR), University of Amsterdam

Objective: In this article, we present a digital tool (Diversity Perspectives in Annual Reports [DivPAR]) for automated content analysis of annual reports, designed to identify the presence of three cultural diversity perspectives—the Moral, Market, and Innovation perspectives—based on earlier work by Ely and Thomas (2001). Method: In Study 1, we describe the development and validation of the instrument, through an iterative procedure in which manual annotation of independent subsamples (n = 24, 25) by human coders was compared to the computer coding in subsequent rounds, until sufficient agreement was reached. In Study 2, we illustrate the type of data that the script generates, by analyzing the prevalence of the three perspectives in annual reports of 55 Dutch organizations over a period of 2 decades (1999–2018; n = 937). Results: Our findings confirm that DivPAR is sufficiently reliable for use in future research. In Study 2, we show that among Dutch organizations, the moral perspective is most prevalent, but the market and innovation perspectives are increasing in popularity. Conclusion: DivPAR can be used to analyze the prevalence and longitudinal development of diversity perspectives in organizational communication. It enables scholars to draw comparisons across different sectors, regions, or countries, to study how diversity perspectives correlate with societal developments, and to uncover the (lack of) relationships between diversity communication and diversity outcomes. Directions for future research are discussed at the end of the article.

Public Significance Statement
In organizational communication, different perspectives can be identified with regard to management of cultural diversity. Examining the prevalence and development of these perspectives across organizations, sectors, and countries, can reveal much about how views on workplace diversity evolve over time. This article describes the development and validation of a digital tool which enables future scholars to conduct such research on a large scale.

Keywords: cultural diversity, diversity perspectives, organizational communication, annual reports, automated content analysis

The management of cultural diversity among employees has emerged as one of the central challenges faced by modern organizations. Previous research shows that successful inclusion of different cultural groups in the workplace is one of the main contributors to equality in society, and can simultaneously provide benefits for organizational productivity (Homan, 2019; Plaut, 2010). Many organizations choose to communicate their diversity initiatives to stakeholders, for example, on company websites, in (job) advertisements, and in periodicals such as annual reports. In these organizational communications, different motivations can be identified for formulating and implementing cultural diversity management practices, often termed diversity perspectives. Based on earlier work by Ely and Thomas (2001), a distinction can be made between a moral perspective (e.g., “We want to eliminate discrimination and provide equal opportunities for all cultural groups”), a market perspective (e.g., “By having a culturally diverse workforce, our organization is better able to understand and adapt to market demands”), or an innovation perspective (e.g., “Cultural diversity is associated with increased learning potential, flexibility, and innovation”).
Over the years, studies within the fields of organizational psychology and human resource management have shown that these perspectives relate to a number of diversity outcomes in organizations, such as diversity climate, inclusion, and employee wellbeing (Hofhuis, van der Zee, et al., 2016; Podsiałłowski et al., 2013; van Knippenberg et al., 2013). However, what remains largely unknown is the prevalence of these diversity perspectives among different types of organizations, how they develop over time, and how they relate to actual policy and management decisions. One of the major reasons for this hiatus is that until now, such research required identification of the perspectives in organizational communication through manual coding, a labor-intensive process which is difficult to do on a large scale.

The present article describes the development and validation of a new digital tool, which can help scholars provide answers to the questions raised above. Specifically, our instrument provides reliable automated coding of Ely and Thomas’ (2001) perspectives in digitalized annual reports, using a Python script. It is designed to process large corpora of documents and create datasets on the prevalence of the diversity perspectives and their longitudinal development. The instrument can be used to draw comparisons across different sectors, regions, or countries, to study how diversity perspectives correlate with societal developments, and to uncover the (lack of) relationships between diversity communication and actual diversity outcomes. This approach answers Plaut’s (2010) call for examining diversity in relation to societal and institutional phenomena. It also contributes to more recent developments in the field, away from perception-based research, and toward examining real-life, non-obtrusive data about diversity policy and its antecedents and consequences (Mor Barak et al., 2016; Reinwald et al., 2019).

Our work combines theory from work- and organizational psychology with methods from digital media and communication science. A brief overview of relevant work from both paradigms will be presented below. Study 1 describes the design and development of the instrument, and provides evidence for its validity and reliability. Study 2 illustrates how the instrument can be used to study longitudinal development of diversity perspectives, by examining their prevalence among a sample of large organizations in the Netherlands, over a period of 2 decades (1999–2018).

The new instrument has been named DivPAR (an acronym of Diversity Perspectives in Annual Reports). The version presented in this article (1.0), as well as future iterations, will be made available to scholars as an open-source tool (see https://github.com/joeophoHuis/DivPAR). The last section of this article provides some directions for future research, and discusses how the instrument can be used to enhance knowledge in the field of diversity science.

Theoretical Background

Cultural Diversity Perspectives in Organizations: A Brief Overview

In organizational psychology and human resource management, the topic of cultural diversity in the workplace has been under investigation since the middle of the last century (see van Knippenberg & Schippers, 2007; Williams & O’Reilly, 1998 for an overview). Specific interest in diversity perspectives, that is, the underlying rationale for organizations to formulate diversity management practices, is a more recent phenomenon. In their widely cited work, Cox and Blake (1991) were among the first to list the potential benefits that diversity may have for organizational performance, including the notion that diversity brings a competitive advantage and could lead to higher group effectiveness. Over the years, this belief in value-in-diversity, has been associated with inclusion, well-being, and job outcomes of cultural minority employees (Hofhuis, van der Rijt, et al., 2016; Homan, 2019; van Knippenberg et al., 2007, 2013). When the value-in-diversity belief is present, organizations are more likely to adopt a multicultural approach, enable minority members to maintain and display their cultural heritage in the workplace, and create a stronger diversity climate (Boehm et al., 2014; Cho et al., 2018; Dwertmann et al., 2016; Hofhuis et al., 2012; McKay et al., 2007).

The most widely used framework for examining diversity perspectives was developed by Thomas and Ely (1996), and verified in a qualitative study among a sample of American organizations (Ely & Thomas, 2001). They identified three specific perspectives, which organizations may adopt in their drive toward diversification. Firstly, the Moral Perspective (originally named Discrimination-and-Fairness) reflects the idea that enhancing cultural diversity in the workplace can be seen as an ethical or moral obligation that organizations have toward society. By promoting cultural diversity, an organization implies it is a socially responsible institution, providing equal opportunities to all cultural groups and aiming to reduce discrimination. The motivations for formulating diversity policy within this perspective may include the normative idea of “doing good,” external legal or regulatory incentives, or the aim of generating a positive impact on the organization’s reputation (Bear et al., 2010; Bird et al., 2007; Podsiałłowski & Reichel, 2014).

Secondly, the Market Perspective (Access-and-Legitimacy), refers to the notion that most organizations operate in a society or market which is inherently culturally diverse. Hence, having a diverse workforce is a valuable tool for gaining knowledge about, and access to, different groups of stakeholders. An example is a supermarket in a culturally diverse neighborhood, which matches the cultural backgrounds of its employees with those of its customers to provide the best customer service. The same principle can be applied to other types of organizations; a governmental organization must have a diverse workforce in order to understand and meet the needs of different groups within the society it serves; a health care provider must be able to communicate effectively about health-related issues in terms that all groups in society are able to comprehend (Ashikali & Groeneveld, 2015; Hofhuis et al., 2018).

Thirdly, the Innovation Perspective (Integration-and-Learning) reflects the idea that cultural diversity may have direct benefits for internal processes within the organization, such as increased learning potential and innovation. According to Ely and Thomas (2001, p. 240) the “insights, skills and experiences employees have developed as members of various cultural identity groups are potentially valuable resources that the work group can use to rethink its primary tasks […] in ways that will advance its mission.” Research in the area of information elaboration and creativity in teams has provided evidence that diversity may indeed lead organizations to become more flexible and innovative in completing their assigned tasks (van Knippenberg et al., 2004). When the opportunity is given to voice different viewpoints, the presence of deviant opinions may increase.
creative thinking (De Dreu & West, 2001). Previous studies suggest that under the right conditions, cultural diversity may increase the effectiveness of idea generation and knowledge sharing in small groups (Hofhuis, van der Rijt, et al., 2016; Nakui et al., 2011). In sum, organizations that employ this perspective believe that cultural diversity may lead to higher flexibility, creativity, and innovation, which in turn could increase the effectiveness of the organization as a whole (Rabl et al., 2020; Shipton et al., 2005).

Effects of Diversity Perspectives on Organizational Outcomes

In the original article in which they present their framework, Ely and Thomas (2001) hypothesized that the three diversity perspectives may influence diversity outcomes in different ways. They approach this issue from the perspective of unequal power distribution, and examine how high- and low-status groups may be affected differently. In doing so, they conclude that the moral perspective may relate to a problematization of diversity, focusing mainly on negative diversity outcomes. In their own words, this perspective assumes that aspects of cultural identity are “relevant only insofar as they trigger others’ negative reactions; they are therefore a potential source of negative intergroup conflict to be avoided in service of the task” (Ely & Thomas, 2001, p. 268). As a result, taking this perspective may enhance existing group boundaries, and may motivate intergroup bias and feelings of hostility, anxiety, and frustration. Contrary to its intended purpose, the moral perspective therefore might be argued to strengthen the existing status-quo between high and low power groups.

The market and innovation perspectives represent a more job-related approach toward diversity, and as such may alleviate some of these issues (cf. Olsen & Martins, 2012). The market perspective, however, presents only a marginal advantage for lower status groups, because organizations with this perspective may value the minority employees’ knowledge about and connection with their own cultural group only as an instrument for making profit, which does not affect their relative status position within the organization. In Ely and Thomas’ view, the innovation perspective, based on the notion of adding value to the organization by giving equal voice to all different cultural groups, is the most promising perspective with regard to inclusion and sustainable diversity management, as well as enhancing cross-cultural learning in the workplace (cf. Podsadlowski et al., 2013; van Knippenberg et al., 2007).

A growing body of research confirms the positive effects of value-in-diversity beliefs, which may include both the market and innovation perspectives (Bader et al., 2019; van Knippenberg et al., 2013). For example, in an experimental study, recruiters who were primed with a value-in-diversity perspective were more likely to hire minority applicants with strong cultural identity (Hofhuis, van der Zee, et al., 2016). However, in a similar experiment, Trawalter et al. (2016) found that emphasizing value-in-diversity may also have unintended consequences: High-status participants who were presented with this perspective were likely to broaden their definitions of diversity, deprivioritizing the hiring of low-status groups. As such, adopting a market or innovation perspective, although well-intentioned, may not always be enough to reach positive diversity outcomes. In similar vein, Ely and Thomas (2020) themselves have recently argued that merely stating a business case for diversity does not contribute to positive outcomes, but that these effects are contingent on the implementation of specific diversity policies, and the presence of a strong diversity climate.

In sum, it appears that diversity perspectives play a role in how organizations deal with cultural differences in the workplace, and can be an indicator of decision making and implementation of policies around minority employees. However, more research is needed to understand which perspectives relate to which types of outcomes, and under which conditions. The tool that is presented in this article may help diversity scholars shed new light on these issues.

Co-Development of Diversity Perspectives

Another point of discussion is how the three diversity perspectives may relate to each other. In their work, Ely and Thomas (2001) state the assumption that “in order for a diversity perspective to produce the results we have observed, a single diversity perspective must prevail in a work group, with no systematic differences along either hierarchical or racial lines” (p. 270). This raises the question whether the perspectives might be contradictory to a certain degree, and whether a perspective must be dominant over the others in order to have a measurable impact on outcomes. More recent work has challenged this assumption, by treating the perspectives as separate and independent variables (e.g., Mehng et al., 2019; Podsadlowski et al., 2013), thus suggesting that they may co-occur within organizations. To date we do not have information on whether the three perspectives are indeed independent, or whether they may correlate either positively or negatively.

Finally, Ely and Thomas have speculated that there may be a pattern in how the perspectives develop over time. For example, organizations may be inclined to adopt the moral perspective first, progressing into a market or innovation perspective as diversity management practices mature. However, to date no longitudinal studies have been conducted to examine these patterns. The new tool presented in this article will enable scholars to initiate such work.

Diversity Perspectives in Organizational Communication

A separate stream of research, mainly rooted in the fields of public relations, marketing, and corporate communication, examines the ways in which organizations communicate their diversity perspectives toward different groups of stakeholders. In this paradigm, diversity management is often viewed as a component of Corporate Social Responsibility (CSR). By broadcasting their diversity policy and the rationale behind it, organizations aim to improve their image and enhance their reputation as an attractive employer (Elving et al., 2013; Hofhuis et al., 2015; Maier & Ravazzani, 2019; Podsadlowski & Reichel, 2014). This practice is also referred to as diversity branding or inclusion branding (Edwards & Kelan, 2011; Jonsen et al., 2019). Bird et al. (2007) report that diversity management is indeed one of the CSR activities that is positively valued by the market. The flipside of the branding approach, however, is that positive communication about diversity does not necessarily imply engaging in actual interventions (cf. Hiemstra et al., 2017). Empty diversity communication, often termed “windowdressing” has been the cause of much skepticism toward diversity management and may discredit existing efforts (Dobbin & Kalev, 2018; Forch & Grier, 2003).

Prior work in the field of organizational communication has analyzed the presence of diversity statements in different types of communication channels. For example, Point and Singh (2003) compared diversity communication on corporate websites in eight
European countries, and conclude that at that time, U.K. companies seemed more likely to promote a value-in-diversity perspective than those from other countries. More recent studies have confirmed the prevalence of this perspective in the U.K. (Guerrier & Wilson, 2011), but show that online diversity branding is gaining in popularity in other countries, such as France, Germany, Spain, the U.S. (Jonsen et al., 2019; Uysal, 2013), Portugal (Barbosa & Cabral-Cardoso, 2010), and South Korea (Mehng et al., 2019).

Other scholars have shown the positive effects of diversity perspectives in management communication (Bader et al., 2019) and job advertising (Casper et al., 2013), or have examined reactions to fictional diversity statements in an experimental setting (Dover et al., 2016; Windscheid et al., 2017). Put together, this body of work has given us some insight in the number and type of diversity statements that are present in different types of corporate media, and how potential employees or customers may react to them. However, it remains difficult to draw generalizable conclusions, because the communications under investigation display a lack of standardization across (or even within) organizations. As a result, more extensive quantitative comparisons of diversity perspectives between countries, sectors or across time, have not been published to date. The instrument presented in this article, is designed to enable such studies to be conducted, through analyzing annual reports.

**Identifying Diversity Perspectives in Annual Reports**

The annual report is one of the most regular and standardized forms of organizational communication. In most countries, it is the norm that private companies publish a yearly report of their financial performance. Over time, annual reports have evolved to include much more information about organizational strategy, management decisions, and long-term perspectives, as such becoming one of the main channels through which organizations communicate with relevant stakeholders (Stanton & Stanton, 2002). More recently, it has become standard practice to also include statements on CSR, which often include the organization’s perspective toward diversity management (Sweeney & Coughlan, 2008). In this way, annual report disclosures are an important channel for reputation management (Bebbington et al., 2008; Neu et al., 1998), and are being used for diversity and inclusion branding purposes (Jonsen et al., 2019). The fact that most large organizations’ annual reports are publicly available, and published at regular intervals, makes them the ideal medium for comparative research on the prevalence and development of diversity perspectives. Furthermore, due to its increased use as a lingua franca in the context of international business, the majority of large organizations now publish their annual reports in English. This allows us to compare organizations from different regions or countries, without the need for translation.

Answering questions about the prevalence, over-time development, and impact of different diversity perspectives in annual reports requires, at a first level, the identification of these perspectives in a body of text. Regardless of the follow-up analysis, we would first need to count their occurrence across a large number of documents. Although it is possible to do so using manual content analysis, this is a time-consuming endeavor, to the point that it is often not feasible to do so with the available means. Modern annual reports are long documents, and diversity management may be mentioned in different, or multiple, sections within each report. Furthermore, large parts of these documents are not about diversity, but a human coder would need to read them anyway—a tiresome and error-prone endeavor. As such, to generate large datasets on the prevalence and development of diversity perspectives over time, automated content analysis is the preferred method. At the same time, a simple keyword search is not sufficient; Simply looking for the word “diversity” does not solve the issue; a more sophisticated approach is needed.

Below, in Study 1, we describe the development and validation of a new instrument, specifically designed for this purpose. Next, Study 2 illustrates how the instrument can be used, by analyzing the development of diversity perspectives among a sample of 55 large organizations in the Netherlands, over a time period of 2 decades (1999–2018). Finally, we will present some directions for future research, which we hope may inspire future scholars to advance the field of diversity science through using the instrument.

**Study 1: Development and Validation of DivPAR**

**Introduction**

The aim of our work was to design and validate a new instrument for automated content analysis of diversity perspectives in annual reports, which was specifically created to identify references to Ely and Thomas’ three perspectives, as described above. In this section, we explain what type of instrument was developed, which steps were taken to validate it, and how it was performed in reliability tests. It is important to note here that the instrument described in this article focuses specifically on perspectives regarding racial/cultural/ethnic diversity in organizational communication. The main advantage of this focus is that we had a large body of existing literature to draw from in the development of our instrument, while still working with a limited number of phrases and manifestations within the documents under investigation. By developing a reliable and accurate tool for identifying these diversity perspectives first, we provide a basis for future scholars to expand our work toward analyzing organizational perspectives toward diversity in gender, age, sexual orientation, etc.

**DivPAR Instrument Design**

In the past few decades, a large number of automated tools for the analysis of texts have become available. These range from deductive, top-down approaches to inductive, bottom-up approaches. Bottom-up approaches, such as unsupervised machine learning, are popular in explorative work where the categories that are to be coded are not known a priori. If, on the other hand, the categories are known in advance, more top-down approaches are appropriate. If it is possible to formulate specific rules (“code as X if word W is present”), then a so-called dictionary-based approach is most suitable; if that is not the case, one has to resort to supervised machine learning (cf. Boumans & Trilling, 2016).

For the aims of the present study, a top-down dictionary-based approach best fits our purpose. We have three pre-defined categories (the moral, market, and innovation perspectives), and we expected to be able to produce a set of rules that distinguish them. Dictionary approaches are especially suitable if the constructs that the researcher is looking for in the texts, manifest themselves through a limited number of indicative words or phrases (Albaugh et al., 2014), which is the case in this study.
Our instrument integrates several tasks, which we implemented using Python scripts. Firstly, a separate script transforms different types of digitalized annual reports (usually in PDF format) into plain text files. These are stored on a cloud platform for data storage and software development (GitHub). In this way, both the annual reports as well as the transformation scripts are available for future projects (see Trilling & Jonkman, 2018).

Next, the main script, our actual tool, consists of a range of search strings. More precisely, it consists of trigger strings and matching search strings, which work in cooperation. In the first stage, the script identifies sections in the annual report that mention trigger words, such as cultural/ethnic/racial diversity, or diversity policy. Next, three matching word lists (hereafter “search strings”) identify the possible presence of the diversity perspectives in that specific section. It has to be noted that the strings can consist of individual words or phrases that need to occur in combination (e.g., “value diversity” <AND> “innovation”) in order for the perspective to be coded as present. The script is programmed to scan each text file line by line, until a trigger is found. Next, it identifies search strings within four lines above or below the trigger. In this way, the script enables us to identify the presence of the three diversity perspectives in the annual report, regardless of whether they occur separately or together in the same paragraphs.

Development of the Trigger and Search String Lists

In the first development phase, a literature-based version of the script was created. Trigger and search strings were chosen based on the definitions and descriptions of the three perspectives in existing publications (Ely & Thomas, 2001; Hofhuis et al., 2015; Podsiadlowski et al., 2013; Thomas & Ely, 1996). The instrument was then used to analyze a sample of digitized annual reports from large organizations (see Study 2 for a more detailed description). As is customary in development of automated scripts, a random subsample (n = 24) was drawn for validation purposes (Albaugh et al., 2014; Haselmayer & Jenny, 2017). The subsample consists of reports from 24 different organizations, each from a random year. Two trained researchers (Coder 1 and Coder 2) independently performed full-text reads of the reports in the subsample, and conducted a manual coding of the diversity perspectives in each report. Intercoder reliability between the coders was sufficient (Krippendorff’s α > .79; Hayes & Krippendorff, 2007).

In line with conventional procedures, the script was then tested, fine-tuned, and validated through an iterative procedure in which the computer coding was compared to the human annotation, and subsequent rounds of improvements were made until sufficient agreement was reached. During the improvement phase, we investigated occurrences of both false positives (the script wrongly counting an instance as relevant) and false negatives (the script not recognizing a relevant instance), and discussed them within the team of researchers. The findings were then used to resolve issues in the coding of the script, with the aim to improve its accuracy.

Not many false positives were identified, but noteworthy examples were phrases such as “… we introduce sustainability clauses covering workers rights and the protection of bio-diversity.” (Shell, 2010, p. 52), which erroneously triggered the script because of the use of the word “diversity.” To overcome such issues, specific phrases such as “bio-diversity” were coded into the script as exceptions that would not trigger the search strings.

False negatives were mostly caused by the limited range of key words in the trigger strings, as well as in the search strings for the respective perspectives. By scanning the documents manually, additional manifestations of the perspectives were identified and subsequently added to the script. An example is the statement “Diversity within companies attracts talent, helps to understand customers better, increases engagement and supports sound business decisions.” (ING, 2017, p. 13), identified by both human coders to contain the market perspective, which was not recognized by the script. By adding the phrase “understand customers” to the search strings for the market perspective, the revised version of the script also identified this correctly. Furthermore, all strings were optimized for use of both American and British spelling, in order to accurately identify phrases using different spelling systems.

Several other important issues surfaced during this procedure. Because of the initial per-paragraph parsing of the documents, the program experienced library-specific difficulties in terms of reading and translating the PDF documents into text documents accurately, and consequently, in terms of labeling perspectives appropriately. To solve that problem, a per-sentence parsing method was used. In order to simulate the paragraph division, a sentence range was used in the revised DIVPAR script. We tested a range of two to five sentences, and found that a range of four sentences (above or below the trigger) yielded the outcome that most closely resembled the manually derived one.

After several rounds of comparison between Coders 1 and 2 and the script, sufficient agreement was achieved. As is customary, the abovementioned procedure was then repeated with a new expert coder (Coder 3), on a second subsample (n = 25) of annual reports from the same dataset. This final comparison revealed sufficient reliability of the search strings for the moral and innovation perspective. However, in this round the human coder identified a higher number of occurrences of the market perspective than the script, prompting us to perform another round of improvements, specifically on this perspective. We again identified which phrases were identified by the coder, but missed by the script, and updated the word lists accordingly until the script reliably identified occurrences of all three perspectives. Table 1 provides examples of trigger and search strings from the final version of the script.

Reliability and Validity of DivPAR 1.0

Table 2 displays the outcomes of reliability tests, comparing the coding of the latest version of the script with coding by all three human coders. Firstly, Krippendorff’s α was assessed for each combination, yielding sufficient intercoder reliability between human coders and the automated coding for all three perspectives in both subsamples (Krippendorff, 2008).

It is important to note here, that reliability indices were calculated in the most conservative way, by taking into account the level of agreement on the presence as well as frequency of occurrences of the three perspectives in each annual report. However, it should be emphasized that frequency is not necessarily a reliable indicator of the emphasis placed on the perspective in the text. For example, some reports mention the same short diversity statement in several places, such as the main text, summary, a framed text-box, and the appendix, thus leading to a high score on a specific perspective.
Other reports include a dedicated paragraph in which a well-argued diversity perspective is explained in-depth, using different phrases and statements. Because all mentions of the perspective are placed within the same section of the report, the script counts this as a single occurrence, even though much more textual emphasis is placed on the perspective. As such, we conclude that at the present stage of development, the frequency of occurrences of a diversity perspective that is coded by DivPAR is not an accurate proxy for the importance placed on each perspective by an organization, and should not be interpreted as such.

In sum, although DivPAR does provide frequencies, its real aim is to provide a reliable indication of whether or not a perspective is present in each report. To assess reliability with respect to this particular aim, we dichotomized the results into “perspective present” (1) and “perspective not present” (0) for each report, and calculated precision and recall.

Precision is the number of reports that—based on the manual coding—indeed did contain a specific perspective, out of all the reports that the script classified as containing the perspective. It is an indicator of the number of false positives in the automated coding process. To give an example: A precision of .80 means that out of 10 documents that were automatically classified as containing X, 8 indeed contained X. Recall, in contrast, indicates how many of the reports that contained X were found; it is an indicator of the number of false negatives in the automated coding. To stay in our example: If we found 8 documents that contained X, but in reality, there were 10 such documents, we have a recall of .80. For reference, precision and recall are often calculated to assess automated news topic classification, where recall values around .80 are common for clearly distinguishable topics, such as sports, entertainment, business, and politics (Vermeer, 2018), but considerably lower ones are often deemed acceptable in more fine-grained topics (Burscher et al., 2015).

Table 2 shows precision and recall for the three perspectives, as compared between the manual coders and the script. The nearly perfect precision scores for the moral and innovation perspective mean that if our script identified these perspectives, we can be confident that the human coders also recognized this perspective in the same report. The precision scores for the market perspective were slightly lower, but still sufficient for further use of the instrument. Future iterations of the script may still improve these scores even further.

Recall of the script is very high when compared to Coders 1 and 2 (.91–1.00), and only slightly lower when compared to Coder 3 (.71–.89). This is an indication that Coder 3 was more likely to identify perspectives in annual reports than Coders 1, 2 and DivPAR. However, these values are within the acceptable range for validating the instrument, and higher scores may still be achieved in future iterations.

Based on the outcomes of the tests, we conclude that overall performance and reliability are sufficient for further use. Therefore, this version of the script was established as DivPAR 1.0, and was subsequently used in the analyses reported in this article.

### Table 1
**Examples of Trigger and Search Strings in DivPAR**

<table>
<thead>
<tr>
<th>Strings</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger strings</td>
<td>Cultural, culture, ethnic, ethnicity, race, racial, national, multinational, etc.</td>
</tr>
<tr>
<td>Search strings—moral perspective</td>
<td>Moral, ethical, fair, fairness, &lt;AND&gt; composition, recruitment, selection, representation; equal opportunities, non-discrimination, against discrimination, social responsibility, socially responsible, moral responsibility, meet diversity standards, achieve diversity targets, etc.</td>
</tr>
<tr>
<td>Search strings—market perspective</td>
<td>Market, markets &lt;AND&gt; local, labor, labour, job, access, accessing; community, society, population, customers; partners; stakeholders; groups; end-users &lt;AND&gt; reflect, reflects, reflecting, serve, serves, serving, mirror, mirrors, mirroring, understand; understanding needs, etc.</td>
</tr>
<tr>
<td>Search strings—innovation perspective</td>
<td>Improve performance, better performance, competitive advantage, competitive edge, flexibility, innovation, creativity, use of human capital, operational excellence, intercultural competence &lt;AND&gt; differences are recognized, valued, learning, inclusion, inclusiveness, problem solving, learning, inspiration, etc.</td>
</tr>
</tbody>
</table>

### Table 2
**Performance of DivPAR 1.0 Compared to Human Coders**

<table>
<thead>
<tr>
<th>Coder 1 and DivPAR</th>
<th>Perspective</th>
<th>Subsample</th>
<th>Kripp. α</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral</td>
<td>#1 (n = 24)</td>
<td>.77</td>
<td>1.00</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>Market</td>
<td></td>
<td>.73</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td></td>
<td>.91</td>
<td>1.00</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Coder 2 and DivPAR</td>
<td>Moral</td>
<td>#1 (n = 24)</td>
<td>.96</td>
<td>1.00</td>
<td>.95</td>
</tr>
<tr>
<td>Market</td>
<td></td>
<td>.78</td>
<td>.71</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td></td>
<td>.91</td>
<td>1.00</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>Coder 3 and DivPAR</td>
<td>Moral</td>
<td>#2 (n = 25)</td>
<td>.73</td>
<td>.92</td>
<td>.86</td>
</tr>
<tr>
<td>Market</td>
<td></td>
<td>.74</td>
<td>.83</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Innovation</td>
<td></td>
<td>.91</td>
<td>1.00</td>
<td>.89</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Precision is the proportion of occurrences identified by DivPAR that were also identified by the human coder. Recall is the proportion of occurrences identified by the human coder that were also identified by DivPAR.

### Study 2: Longitudinal Development of Diversity Perspectives in a Sample of Organizations From the Netherlands

**Study Aim**

Having described the development and validation of the instrument, the aim of Study 2 is to illustrate how DivPAR can be used to generate datasets on the prevalence of the moral, market, and innovation perspectives. Specifically, we will examine whether these perspectives occur in annual reports of large organizations in the Netherlands, over a period of 2 decades (1999–2018).
Method

The sample for this study consisted of 55 Dutch organizations, which were selected from the Volkskrant Top 100 largest employers in the Netherlands (Volkskrant, 2014). It includes for-profit companies in different sectors, such as finance (e.g., ING, ABN AMRO), services (e.g., KPMG, Deloitte), and industry (e.g., Shell, Unilever), as well as non-profit organizations (e.g., Amnesty International, Greenpeace). Most of the corresponding annual reports were downloaded directly from the organizations’ websites. Older reports were also retrieved from existing databases of financial records and company information. Reports from earlier years (1999–2005) were less likely to be digitally available. Out of the total 1100 annual reports that were published by these organizations between 1999 and 2018, we were able to archive 937 (85.2%).

Using DivPAR, we were able to convert all annual reports to usable plain text format, and subsequently coded the prevalence of the moral, market, and innovation perspectives. The resulting dataset provides an overview of the number of times each perspective was mentioned in each annual report. As was explained in Study 1, these data were then converted into dichotomous variables, representing whether the perspective was (1) or was not (0) present in the respective document.

Results

Figure 1 displays the percentage of annual reports that mention each of the three perspectives in each year. Interpreting the graphs, it becomes clear that including statements on cultural diversity and diversity management has become increasingly common over the past 2 decades. Whereas between 1999 and 2003, the perspectives were present in less than 10% of annual reports, we see that since 2010 a majority of organizations in this sample mentions at least one of the perspectives. The moral perspective in particular, displays a sharp increase, being mentioned by more than 95% of organizations in 2018. This is a strong indication that Dutch organizations feel obliged to include statements on cultural diversity in their stakeholder communication, and that diversity branding based around the notion of equality, fairness, and social responsibility (cf. Jonsen et al., 2019) has become the norm in this particular context. It has to be noted, however, that the moral perspective has nearly reached the maximum possible value of 100%, which means that as new data are added in the next few years, the upward trend is expected to flatten out or even turn downward.

The market and innovation perspectives also seem to have gained popularity, being mentioned by 20%–40% of organizations in recent years. This is an interesting finding, particularly in light of divergent effects of such diversity perspectives on organizational outcomes in recent studies (Hofhuis, van der Zee, et al., 2016; Trawalter et al., 2016; van Knippenberg et al., 2013). Using DivPAR, future scholars may be able to examine in greater detail how these perspectives influence the position of high- and low-status groups in organizations. More ideas for future directions are provided below.

Finally, we were interested in examining whether the perspectives are independent, or may occur together. Results of our preliminary analyses show modest positive correlations of the moral perspective with market ($r = .37; p < .001$) and innovation ($r = .34; p < .001$) perspectives, as well as between the market and innovation perspectives ($r = .35; p < .001$). Based on these relationships, we can tentatively conclude that Dutch organizations that mention one of the three perspectives are also slightly more likely to mention the others, indicating that they are neither fully independent, nor incompatible. However, to fully understand how these perspectives relate to each other on the organizational level, as well as across time, more advanced analyses are required, which are beyond the scope of the present article. We encourage future scholars to make
use of DivPAR to engage in such studies, and provide new answers to the questions raised.

Discussion and Future Research Agenda

In Study 1, we have described the development and validation of DivPAR, a new digital tool that allows scholars to conduct automatic content analysis of digitalized annual reports, and create datasets on the presence of diversity perspectives within these documents. Study 2 provides an illustration of how the tool can be used to analyze the prevalence and longitudinal development of these perspectives in a sample of organizations. The current version, DivPAR 1.0, as well as future iterations, will be made available as an open-source digital instrument, that scholars can use free of charge. Below, we will provide some directions for future research, that we hope may inspire scholars to use DivPAR to advance knowledge on the topics of workplace diversity and organizational communication.

Theoretical Understanding of Diversity Perspectives

Firstly, our work may contribute to the understanding of diversity perspectives on a theoretical level. As was mentioned above, Ely and Thomas (2001) reported that in each organization in their study, one of the three perspectives was clearly dominant over the other two. This raises the question whether the three perspectives could occur together, or whether they may be incompatible. Although other scholars have measured the three perspectives as separate variables (Mehng et al., 2019; Podsadlowski et al., 2013), it has so far remained unclear to what degree they are independent, or whether they may be correlated. Based on our results from Study 2, we can provide a first tentative conclusion that the perspectives may indeed correlate positively, meaning that organizations that mention one of the perspectives are also more likely to mention the others. Naturally, this finding needs to be replicated with different samples, from other regions of the world, and within different types of organizations. Our new instrument will allow future scholars to easily conduct such analyses.

Secondly, a major question in the study of diversity perspectives is how they develop over time, and whether there is a distinct pattern in how they evolve. It has been hypothesized that organizations may first adopt the moral perspective, which may progress into a market or innovation perspective as the organization gains experience, and diversity management practices mature (Ely & Thomas, 2001; Hofhuis & Van Drunen, 2019; Mehng et al., 2019). However, no conclusive empirical evidence has been provided for this pattern. As was stated above, it also remains unclear whether one perspective may replace the others, or whether the first perspective may remain present while others develop alongside it. Such questions may be answered by analyzing perspectives in organizational communication through longitudinal modeling. Unfortunately, the sample that was used in Study 2 (consisting of 55 organizations) is too small to reliably conduct such analyses at the present time. However, our findings do show that all three perspectives seem to have gained popularity. Combined with the finding that in recent years nearly all organizations in the sample mention the moral perspective in their annual reports, we come to the preliminary conclusion that the moral perspective indeed seems to develop first, and that the other two appear to develop alongside it, while the moral perspective remains. Again, replication in a larger, more representative sample, as well as using more sophisticated analyses, is needed to be able to draw a definite conclusion.

Apart from understanding the mechanisms behind the development of diversity perspectives, future scholars can also use DivPAR to examine their antecedents and consequences. Prior research has shown that these effects may be related to individual, organizational, and societal level factors.

Individual Level

First, on the individual level, diversity perspectives may be related to a myriad of employee outcomes. A growing body of literature focuses on how different diversity approaches and strategies, such as multiculturalism/colorblindness (Cho et al., 2018; Jansen et al., 2016; Plaut et al., 2009) or value-in-diversity beliefs (Bader et al., 2019; Hofhuis et al., 2012; Hofhuis, van der Zee, et al., 2016; van Knippenberg et al., 2013) affect job satisfaction, inclusion, and career opportunities of cultural majority and minority employees. To date, however, little research has examined Ely and Thomas’ (2001) perspectives in this regard. Using DivPAR, scholars would be able to examine which perspectives are prevalent in organizational communication over a longer time period, and correlate this with existing measures of diversity outcomes on the individual level. This type of study would allow us to verify Ely and Thomas’ (2001, p. 270) claim that the innovation perspective is the most promising of the three, and may overcome some of the issues that may result from organizations adopting the market or moral perspectives. A particularly valuable avenue of exploration would be to compare the experiences of majority and minority employees, to examine how the three perspectives may affect these subgroups differently.

Organizational Level

Second, on the organizational level, our instrument provides opportunities for examining a number of effects that were previously difficult to study. One exciting possibility is to link data on the prevalence of diversity perspectives with existing datasets on organizational characteristics and performance. For example, this would allow us to reveal which perspectives are more prevalent in a certain industry, sector, or region, and to make cross-national comparisons. Furthermore, it would allow us to examine how they relate to organizational innovation, profitability, customer satisfaction, or other relevant performance indicators (e.g., Kochan et al., 2003). Moreover, a logical next step in this research line would be to cross-reference the three diversity perspectives with existing diversity management indices, such as board member diversity or the implementation of successful diversity initiatives (e.g., Brammer et al., 2007). By combining these types of datasets, we would be able to identify whether organizations that communicate certain perspectives are also engaging in actual diversity management, or whether they practice “windowdressing”—that is, merely promoting diversity for the purpose of branding and reputation (e.g., Jonsen et al., 2019). A more detailed analysis would even reveal in which types of organizations windowdressing might be more or less common, and help us recognize it more easily.

Societal Level

Third, DivPAR also provides new opportunities for conducting research on diversity perspectives on the societal level.
Broadening the Scope

Finally, all of the directions for future research that were mentioned above, can be expanded by broadening the scope of the instrument. The current version is designed to specifically identify the moral, market, and innovation perspectives on cultural diversity, but the scripts could be easily adapted to also include measures on other forms of diversity, such as gender, age, or sexual orientation. Furthermore, the current version of the instrument has been optimized to code phrases that are typically used in organizational annual reports. However, expanding the list of trigger words and search strings to include those that are more common in other forms of organizational communication, such as websites, (job) advertisements, or promotional text, would allow for more detailed and comprehensive examination of the use of the perspectives. Extending the instrument to include these uses is relatively straightforward, but requires a new validation procedure, similar to the one described in Study 1.

Conclusion

Above, we have described the development and validation of an innovative new digital tool for automatic content analysis of DivPAR. We have provided a first illustration of how the instrument can be used, by showing the longitudinal development of these perspectives among large organizations in the Netherlands. Finally, we have provided a number of future research directions, that we hope may inspire others to use our instrument in their own work. We believe DivPAR to be a valuable addition to the toolbox of diversity based approaches to topic coding [Conference session]. Proceedings of the 7th Annual Comparative Agendas Project (CAP) Conference, Koblenz, Germany.


Dwertmann, D. J., Nishii, L. H., & van Knippenberg, D. (2016). Disentangling the fairness & discrimination and synergy perspectives on