Automatic Summarization of Corporate Disclosures

Eddy Cardinaels
Department of Accountancy, Finance and Insurance, KU Leuven
Department of Accountancy, Tilburg University
eddy.cardinaels@kuleuven.be

Stephan Hollander
Department of Accountancy
Tilburg University
s.hollander@tilburguniversity.edu

Brian J. White
Department of Accounting
The University of Texas at Austin
brian.white@mccombs.utexas.edu

July 14, 2017

We thank Sarah Zechman (discussant), David Veenman, and workshop participants at the London Business School Accounting Symposium, Radboud University Nijmegen, the University of Texas at Austin, the University of Melbourne, the University of Pittsburgh, the University of Amsterdam, the University of New South Wales, Jinan University, and Tilburg University for helpful comments. We also thank Arnaud Nicolas for generating expert summaries of earnings releases, and Ties de Kok and Shannon Garavaglia for their capable research assistance.

Automatic Summarization of Corporate Disclosures

ABSTRACT

In practice, key disclosures such as earnings releases and MD&A often include summaries. These manager-provided summaries, however, may be prone to strategic tone and content management compared to the underlying disclosures they summarize. In contrast, automatic, algorithm-based summaries have the potential to provide useful summary information with less bias than management summaries. We provide archival evidence on the use of summaries in practice and conduct three experiments to investigate how management and automatic summaries compare on several dimensions (e.g., bias, usefulness), and how summaries affect investor information processing, beliefs about firm fundamentals (e.g., performance, risk), and valuation judgments. Our results suggest that automatic summaries compare favorably to management summaries for earnings releases, but fare less well for MD&A. Importantly, investors who receive an earnings release accompanied by an automatic summary arrive at more conservative (i.e., lower) valuation judgments, and are more confident in those judgments, compared to investors who receive the same earnings release with a management summary. Our findings provide important input to recent discussions by policy makers on the use of summaries for corporate disclosures.

Keywords: *Management summary; automatic summary; disclosure; investor judgment*

Data availability: *Contact the authors*

1. Introduction

Public companies disclose more information than ever before (e.g., KPMG 2011, Loughran and McDonald 2014). Given the large volume of disclosure and evidence that investors are boundedly rational (Hirshleifer and Teoh 2003; Elliott, Hobson and White 2015), managers often provide summaries of key disclosures, such as earnings releases and management discussion and analysis (MD&A) (Ernst & Young 2014, 14). However, rather than presenting a balanced picture of the information disclosed in the underlying document, managers may selectively highlight information that is favorable to the company (Henry 2008; Guillamon-Saorin, Osma and Jones 2012; Huang, Nekrasov and Teoh 2013; Huang, Teoh and Zhang 2014). Against this backdrop, there may be a role for automatic, algorithm-based summarization of corporate disclosures. Summarization algorithms rely on statistical heuristics for sentence extraction, and can summarize large amounts of text without human intervention. As such, automatic summaries have the potential to reduce both information overload and bias. In this study, we conduct three experiments to investigate how automatic summaries compare to management summaries, and how automatic and management summaries affect investors' judgments.

Investigating automatic summarization of corporate disclosures is important for several reasons. First, because disclosures have become lengthier and more redundant (Dyer, Lang and Stice-Lawrence 2016), regulators and standard setters are starting to explore ways of simplifying financial reports (SEC 2013; FASB 2015), including summaries (SEC 2016). These efforts have led to calls for research on summarization to aid investors and others (Barth 2015). Thus, investigating automatic summarization has the potential to provide new insights to financial reporting regulators and accounting standard setters. Second, management-generated summaries are already part of the financial reporting landscape. Our review of S&P 100 firms' disclosure practices indicates that 72% provided summaries of MD&A in their fiscal 2015 10-K filings, and 81% provided summaries of their fourth quarter 2015 earnings releases. However, there is little evidence about how summaries affect investors' judgments. Third, the technology underlying algorithm-based summarization has advanced considerably in recent years and is now recognized as viable and useful across a number of disciplines, for example in law (Nenkova and McKeown 2011), medicine (Garcia-Gathright et al. 2016) and journalism (Blankespoor, deHaan and Zhu 2017; Holmes 2016). This suggests that it may be appropriate to assess the viability of automatic summarization for corporate disclosures. However, summarization algorithms take a number of different approaches to summarization. Empirical evidence is thus needed regarding the usefulness of the various tools in the financial reporting domain.

We conduct three experimental studies. Studies One and Two provide evidence regarding the viability of automatic summarization of corporate disclosures by comparing attributes of management summaries with those of several algorithm-based automatic summaries for several earnings releases and MD&As.[1] Study Three tests the effect of automatic and management summaries on investors' valuation and other investment-related judgments. Following other accounting studies in which the experimental task does not require specialized accounting knowledge, we recruit participants for each of our studies from Amazon's Mechanical Turk platform (e.g., Rennekamp 2012; Farrell, Grenier and Leiby 2017).

Study One compares automatic and management summaries for two types of disclosure—earnings releases and MD&A—that differ in important respects. For example, managers are more strategic in their use of optimistic versus pessimistic language in earnings releases than in

[1] Studies One and Two provide some information on the usefulness of automatic summaries relative to management summaries for the set of companies and disclosures we use in these two studies. Insights of these studies also help us to shape the design of Study Three. We do not intend to deliver a comprehensive test of the general usefulness of any summarization algorithm nor for any corporate disclosure in particular.

MD&A (Davis and Tama-Sweet 2012).[2] In addition, MD&A is more influenced by auditors, which increases homogeneity across firms that share auditors and may reduce the informativeness of MD&A (DeFranco, Fogel-Yaari and Li 2016). We find that automatic summaries compare favorably to management summaries for earnings releases, but fare less well for MD&A. On average, participants rate automatic summaries of earnings releases as *more* informative and credible than management summaries, and *equal* to management summaries in terms of readability and overall usefulness. In contrast, participants rate automatic summaries of MD&A lower than management summaries on each of these dimensions.

Building on results of Study One, Study Two focuses on earnings releases. Participants compare automatic and management summaries to the underlying text of two earnings releases on a number of dimensions. The key result from Study Two is that automatic summaries reflect the underlying text of the earnings release with less bias (i.e., present a more balanced picture) than management summaries. Participants also rate automatic summaries as no different from management summaries in capturing the important information in the earnings releases, and participants are equally likely to rely on automatic and management summaries. Both Study One and Two also show that one algorithm, known as LexRank, produces summaries that are consistently rated as superior to management summaries.[3] When we compare the summaries against a summary generated by an experienced Investment Relations Officer, the LexRank summary again outperforms the management summary as LexRank *better* captures elements of the earnings release that the experienced Investment Relations Officer deems important.

[2] While this suggests that earnings releases are more biased than MD&A, ex ante the implications for automatic versus management summaries are not clear. For example, automatic summarization is based on the underlying text, so greater bias in the body of the earnings release may also translate into greater bias in the automatic summary of the earnings release. On the other hand, managers may be more strategic in summarizing MD&A because the MD&A section tends to be longer and thus offers greater choice of content to highlight.

[3] We discuss the different summarization algorithms used in this study in the Online Appendix. For further details on the LexRank algorithm, we refer the reader to Section 2.2.

Based on Studies One and Two, we select the algorithm (LexRank) and type of disclosure (earnings release) that together produce the most useful and least biased automatic summary. Study Three then tests the effects of automatic and management summaries of an earnings release on investors' judgments of firm value and other investment-related judgments. While Studies One and Two use actual disclosures of real companies, for Study Three we follow prior literature (e.g., Rennnekamp 2012) and anonymize an earnings release from a real company so that participants' familiarity with the real company does not affect their judgments. Participants in Study Three receive earnings releases that include either a management summary, an automatic summary or no summary. All participants receive linked access to the full text of the earnings release and the accompanying tables, which they can search for additional information.

Results from Study Three indicate that participants who receive the earnings release with the management summary value the company's common stock more highly than those who receive the automatic summary or no summary. Importantly, participants who receive the automatic summary are more confident in their (lower) valuation judgments than those who receive the management summary. Our results further show that judgments of future earnings growth potential and perceptions of the favorability of the earnings release explain the effect of summary type (i.e., management vs. automatic) on common stock valuation. Analyses of information search data reveal no differences in overall search time across summary types. However, summaries, regardless of type, increase search efficiency. Specifically, both automatic and management summaries direct participants' information search toward more complex and economically relevant sections of the earnings release compared to providing no summary.

This paper contributes to the literature in several ways. First, our paper is the first to examine characteristics of automatic summaries of corporate disclosures and the effect of automatic summaries on investors' judgments. In so doing, our paper contributes to research on the impact of computer-based textual analysis and linguistic processing technology on investor behavior (see Loughran and McDonald 2016 for a review). Our results indicate both that managers tend to bias their summaries beyond any bias present in the underlying disclosure, and that the use of technology to generate automatic summaries can potentially undo this bias. For regulators and standard setters interested in summary information, our study provides evidence that encouraging management-generated summaries would not necessarily lead to the most value-relevant information being highlighted.

Second, we contribute to the broad literature, spanning accounting, economics and finance, on investors' bounded rationality. Analytical and empirical studies find that bounded rationality—i.e., investors displaying limited attention and processing power—affects market price efficiency (e.g., Daniel, Hirshleifer and Teoh 2002; Hirshleifer and Teoh 2003; Hirshleifer, Lim and Teoh 2009; Elliott et al. 2015; Umar 2017). Further, regulators have expressed concerns that in the presence of bounded rationality, information overload can exacerbate inefficiency (Paredes 2003, 2013). Providing summarized information is often considered an easy fix for this problem. However, we provide evidence that investors' reactions to summary information depend on whether the summary is generated automatically or by management.

Third, while information intermediaries (e.g., analysts, business journalists) contribute to the efficient allocation of capital, research also consistently shows that conflicts of interest (e.g., *quid pro quo* relations between journalists and their sources, analysts' incentives to collude with management; Desai, Rajgopal and Yu 2016, Dyck and Zingales 2003) and behavioral biases tend to stand in the way of information intermediaries fulfilling their full potential. Our paper contributes to this literature by highlighting the possibility of automating one aspect of information dissemination in capital markets, thereby removing one opportunity for conflicts of interest and behavioral biases to negatively affect information.

The paper proceeds as follows. Section 2 provides background for our study. Section 3 presents the design and results of Studies One and Two. Section 4 develops hypotheses and presents the design and results of Study Three. Section 5 concludes the paper.

2. Background

As corporate disclosures get longer (Dyer et al. 2016; Francis, Schipper and Vincent 2002), investors with limited attention may find it difficult to process all information contained in company disclosures. This suggests that it may be useful to study how summaries help (or hinder) investors in their information search and investment-related judgments, an issue that policy makers also deem to be relevant (SEC 2013; SEC 2016). The question of how summaries affect investors' judgments is even more important when one considers the flexibility that the Securities and Exchange Commission (SEC) offers to its registrants, as the SEC does not offer any guidance on summary length or the items that the summary should cover (SEC 2016, 3-4).[4]

However, despite the recognition that summarization can be useful in the financial reporting domain, there has been no systematic research on how summarization affects investors' judgments (Barth 2015). Our studies provide evidence on this issue in two important ways. First, we systematically compare investors' assessments of automatic and management summaries on several dimensions, including informativeness, readability, credibility and bias (section 3). Second, we test the impact of summarization on investors' valuation and other investment-related judgments (section 4).

[4] In June 2016, the SEC adopted an interim final rule that allows issuers to include, at their option, a summary page in their Form 10-Ks. As noted therein, summary information must be presented "fairly and accurately."

In the remainder of this section, we provide background on management and automatic summaries. In section 2.1, we provide evidence on the prevalence and characteristics of management summaries in practice. In section 2.2, we provide background on the technology underlying automatic summarization. In section 2.3, we discuss how management and automatic summaries are generated and how they may differ in information content and tone.

2.1 The Use of Management Summaries in Practice

Table 1 provides evidence on the use of management summaries in practice. We collected fourth-quarter earnings releases and MD&A for S&P 100 firms in fiscal 2015. After excluding firms for which disclosures were unavailable or which provided the disclosure in a format (e.g., a picture file) that was not suitable for analysis, we analyzed the remaining disclosures to determine: (1) how many included summaries, (2) whether the summaries differed in tone from the underlying text of the disclosures, and (3) whether summaries of earnings releases and MD&A differed in length, format and/or tone.

Insert Table 1 about here

As shown in Panel A of Table 1, a majority of firms included summaries of their earnings release (78 of 96, or 81%) and MD&A (64 of 89, or 72%). Forty-eight firms included summaries with both their earnings release and their MD&A. Panels B and C of Table 1 compare the tone of these summaries to the tone of the underlying text of the document being summarized. For this comparison, we identify positive and negative tone words using Henry's (2008) context-specific word list, and divide the number of tone words by the total number of words in the summary or the full text (excluding the summary).[5] As shown in Panel B of Table 1, compared to the underlying text, earnings release summaries include a smaller percentage of negative words, and a larger percentage of positive words (both *p*-values < 0.02). Panel C of Table 1 shows that, compared to the underlying text, summaries of MD&A also include a larger percentage of positive words ($t_{63} = 6.19$, $p < 0.01$); however, the percentage of negative words does not differ significantly between the summary and the underlying text of the MD&A ($t_{63} = 1.57$, $p = 0.12$).

Comparing summaries across type of disclosure, Panel A of Table 1 shows that summaries are considerably shorter for earnings releases (mean = 127 words) than for MD&A (mean = 764 words). As shown in Panel D of Table 1, summary tone also differs by disclosure type. For the 48 firms that provide summaries of both earnings releases and MD&A, the earnings release summaries include a smaller percentage of negative words, and a larger percentage of positive words, than the MD&A summaries (both *p*-values < 0.01). Untabulated results also reveal that MD&A summaries exhibit greater variation in format compared to earnings release summaries. For example, of the 78 earnings releases that include summaries, all but one use bullet points, only occasionally supplemented by a short paragraph or table. In contrast, of the 64 MD&A summaries, only 55% include bullet points, while 41% include a table.

Overall, this analysis indicates that summaries are commonly used by large public companies both for earnings releases as well as for MD&A. Further, summaries tend to exhibit more positive tone than the documents they summarize. Finally, observed differences in length, tone, and format suggest that summaries play a different role for earnings releases compared to MD&A.

---

[5] To facilitate this analysis, we remove tables that contain less than 50% alphabetic characters from the body of earnings releases and sentences that contain less than 50% alphabetic characters or consist of fewer than 50 alphabetic characters from the body of the MD&A.

2.2 Extraction-Based Automatic Summarization

The Online Appendix provides a primer on extraction-based automatic summarization, including details on the six algorithms we use to generate the automatic summaries in Study One. These algorithms differ principally in the statistical heuristics applied to identify the most salient sentences of a text. The algorithm that performs best in the context of earnings releases, according to the results from Studies One and Two, is LexRank (Erkan and Radev 2004). To give a sense of how summarization algorithms work in general, and how LexRank works specifically, we briefly describe the LexRank algorithm below.

The LexRank algorithm first generates a graph, composed of all sentences in the underlying document. In this graph, each sentence represents a node, with similarity between nodes as edges.[6] A similarity matrix is then constructed, wherein each entry is the similarity between a sentence pair. The LexRank algorithm then computes sentence importance by considering each sentence's relative importance to its neighboring sentences. Next, following Erkan and Radev (2004), in our implementation LexRank computes sentence importance based on eigenvector centrality. Finally, a summary is generated by combining the top-ranked sentences, using a threshold and length cutoff to limit the size of the summary. In our implementation, we set the number of sentences per summary equal to the bullet points in management's summary of the disclosure.

2.3 Human versus Algorithm-Based Summarization

Human-generated summaries are typically based on text understanding (i.e., summarization by abstraction). A typical process for a person generating a summary would involve (1) getting

[6] To define similarity, each sentence is represented as a "bag-of-words" model, meaning that grammar and the order of words in a sentence are disregarded. The similarity between two sentences is computed by the frequency of word occurrence (specifically, tf*idf cosine similarity) in a sentence.

an understanding of the content of the source document, (2) identifying the most relevant content contained in the document, and (3) writing up this information, usually in the person's own words (Brandow et. al., 1995). Importantly, steps 1 and 3 are beyond the capability of most automatic summarization techniques (Brandow et. al., 1995; Salton, Singhal, Mitra and Buckley 1997). For step 2, as explained above, automatic summaries rely on statistical heuristics that attempt to identify the most important lexical units (typically, sentences) in a document (i.e., summarization by extraction).

Given that management should have a good understanding of the information content of the underlying source document (step 1), one would expect that with regard to content selection (step 2), management should be able to highlight information that investors deem relevant. However, ample evidence suggests that managers tend to bias their disclosures when they have discretion to do so.[7] Given that the set of news items from which to select—at least for companies with reasonably complex operations—tends to be large in the underlying document (Henry 2008), we expect managers to select items that depict a more favorable view of the company's performance when they write up a summary. That is, we expect *incremental* bias in management summaries compared to any bias already present in the underlying document. Henry (2008, 371) describes selective inclusion of information in bulleted introductory points of earnings releases (e.g., a firm disclosing an increase in operating margin for one of its divisions, while overall profit margin of the company declined). Prior studies also show evidence of news management by managers (Ahern and Sosyura, 2014; Guillamon-Saorin et al. 2012). In contrast

[7] Prior research on management bias in disclosures tends to focus on underlying source documents, rather than summaries. Such biases arise more often in earnings releases and conference calls (Henry 2008) and to a lesser extent in MD&A (Davis and Tama-Sweet 2012). However, even in the presence of bias, both positive and negative news items are likely to occur in the underlying document and thus have the potential to be included or excluded in a summary. Thus, even when the underlying document contains a degree of bias, item selection may differ between a manager-generated summary and an automatic summary.

to a management summary, an automatic summarization algorithm selects lexical units based on statistical heuristics. If a sentence is deemed important according to the statistical heuristics, the automatic summary will include it regardless of whether it is good or bad news.

Regarding write-up (step 3), extraction-based summarization algorithms cannot change the language that appears in the summary given that they extract sentences directly from the source document. Thus, in addition to content selection, managers can also rewrite and manager the tone of the content that they include in a summary. Prior research shows that managers use tone to depict a more positive view of the company, for example by using positive words and vivid language (e.g., Davis et. al. 2012; Hales, Kuang and Venkataraman 2011; Henry and Leone 2016; Henry 2008; Huang, Nekrasov, and Teoh 2013; Loughran and McDonald 2016). Thus, we expect that managers manage tone when preparing summaries, given that they have incentives and discretion to do so (Arslan-Ayaydin, Boudt and Thewissen 2016). The next section explicitly tests this presumption by comparing attributes of algorithm-based and management summaries.

3. Studies One and Two: Attributes of Automatic and Management Summaries

3.1 Study One: Participants

In common with other recent studies in which the experimental task does not require specialized accounting knowledge, we recruited participants for each of our studies from Amazon's Mechanical Turk (MTurk) platform (e.g., Asay, Elliott and Rennekamp 2016; Farrell et al. 2017). The MTurk platform then directed potential participants to an online survey designed in Qualtrics. Because we used the same procedures to recruit and screen participants for all three studies, we report procedures and aggregate demographics in this section, and do not repeat them in full for Studies Two and Three.

As in other financial accounting studies using MTurk participants (e.g., Bonner, Clor-Proell and Koonce 2014), participants were required to pass certain screening questions in order to proceed with the study. Specifically, they were required to be over 18 years of age, to be native English speakers, to have previous investing experience, and to be at least moderately familiar with financial disclosures (indicated by reported familiarity of 60 or higher out of 100). A large majority of our participants (88%) had taken college courses, and 71% held a bachelor's degree or higher. Participants had taken an average of 3.8 accounting or finance classes, and had an average of 14.5 years of full-time work experience. According to Elliott, Hodge, Kennedy and Pronk (2007), nonprofessional investors on average have taken 3.5 accounting or finance courses, and 97 percent have experience with financial statements. Our participants thus had similar profiles to those in Elliott et al. (2007), suggesting that they were appropriate proxies for nonprofessional investors. This is in line with recent evidence from Krische (2015), who shows that MTurk participants can be used with confidence to proxy for nonprofessional investors.

3.2 Study One: Method

In Study One, participants rated the informativeness, readability, credibility and overall usefulness of six automatic summaries and one management summary for one of six real company disclosures.[8] The automatic summaries were generated by algorithms known as KLSum (KL), LexRank (LEX), Latent Semantic Analysis (LSA), Luhn (LUHN), SumBasic (SB), and TextRank (TR); the Online Appendix contains additional detail on each of these algorithms.[9] Summary type was manipulated within participants, and disclosures were

[8] The corporate disclosures examined in this study vary widely in terms of writing style (e.g., sentence length), vocabulary, and structure, all of which may influence the performance of the algorithms. Accordingly, we do not make a prediction as to which algorithm will perform best.

[9] We employ a variety of frequently used sentence-extraction based approaches for generic summarization, applicable for which no additional information or prior knowledge (e.g., about user need) is needed. We exclude from our analysis genre-specific (e.g., academic journal articles) and domain-specific (e.g., medical) approaches; see Nenkova and McKeown 2011 for an overview.

manipulated between participants. Thus, each participant viewed all six automatic summaries plus the management summary for one disclosure.[10] Participants were not told that any of the summaries were generated automatically. While we kept the number of bullet points constant across the summaries, text length could differ. Participants therefore rated the length of each summary before making the other judgments in order to reduce any subconscious effects of length on subsequent judgments (e.g., Schwarz 2004), and to test whether differences in perceived length explained differences in other judgments.

In total, 863 people volunteered to take part in Study One, and 303 (35.1%) met the qualification requirements and completed the study. Qualtrics randomly assigned participants to one of the six disclosure conditions, and the order in which participants viewed the summaries was also randomized. On completion, participants were paid $1.50 via MTurk. A mean (median) completion time of 12 (8) minutes resulted in a mean (median) hourly rate of $7.50 ($11.25).[11]

3.3 Study One: Results

Results of Study One are presented in Figure 1 and Table 2. All responses were made on 101-point scales with appropriately labeled endpoints; in each case, higher values indicated higher levels of the measured variable. The key insight from Study One is that automatic summaries compare favorably to management summaries for earnings releases, but fare less well for MD&A. On average, automatic summaries of earnings releases are rated as *more* informative and *more* credible than management summaries (both $p < 0.01$), and do not differ from management summaries in terms of readability or overall usefulness (both $p > 0.10$). In contrast, automatic summaries of MD&A are rated *less* favorably than management summaries on all

[10] The final sample included earnings releases for Alibaba (Q1 2016), Boeing (Q2 2008), and Target (Q4 2013), and MD&A for Macy's (2014), Mattel (2014), and Target (2013).
[11] Results from all three studies are inferentially identical when controlling for time spent completing materials.

dimensions (all $p < 0.05$). This pattern of results is also supported by the significant interactions—reported in Panel C of Table 2—between disclosure type (earnings release versus MD&A) and summary type (automatic versus management). For each of the four measures—informativeness, readability, credibility, and usefulness—the interaction of disclosure type and summary type is significant at $p < 0.02$.

Insert Figure 1 and Table 2 about here

Consistent with using more strategic language in earnings releases (Davis and Tama-Sweet 2012) and with the differences in summary tone reported in section 2.1, we speculate that the different results for earnings releases and MD&A result from managers being more strategic in their summaries of earnings releases.[12] In untabulated analyses, we also confirm that the superior (inferior) performance of automatic summaries for earnings releases (MD&A) is robust to controlling for firm and time-period, as we observe the same interactive pattern of results for Target's Q4 2013 earnings release and the MD&A section of Target's 2013 annual report.

We also observe variation among different algorithms to generate automatic summaries, with LexRank, Luhn and TextRank generally getting the highest ratings for summaries of earnings releases. Of these, Luhn and TextRank produce by far the longest summaries. While results generally remain significant when controlling for summary length (both actual word count and perceived length), significance levels decrease for Luhn and TextRank summaries when controlling for length. This suggests that at least part of their outperformance is explained by greater length. For this reason, we exclude these two summaries from Study Two, and focus instead on summaries that are more similar in length to management summaries.

[12] Differences may also result from more boilerplate language in MD&A (e.g., DeFranco et al. 2016), which the algorithms incorrectly identify as important.

Overall, results from Study One suggest that automatic summarization is a viable tool for summarizing earnings releases. Automatic summaries increase informativeness and credibility compared to management summaries, without sacrificing readability or overall usefulness. Results, however, also suggest that automatic summarization is less viable for MD&A.

3.4 Study Two: Participants and Method

Study Two builds on the results of Study One by having participants complete a more in-depth analysis of the relation between summaries of earnings releases and the underlying source document. Two earnings releases—Boeing for Q2 of 2008 and Target for Q4 of 2013—were selected for Study Two based on the strength of results for these two disclosures in Study One.

Participants in Study Two first read one of the two earnings releases in full.[13] To facilitate judgments, participants could take notes while reading the earnings release, and these notes were reproduced for reference when participants rated the summaries. After reading the full earnings release, participants evaluated four automatic summaries and one management summary (holding constant the number of bullet points) of the earnings release. Participants responded to the following measures: (1) "Capture"—the extent to which the summary captured important information in the earnings release, (2) "Reliance"—the extent to which participants would rely on the summary in judging the company's performance, (3) "Bias"—the extent to which the summary made the company's performance look better or worse than the full document, and (4) "Should be included"—participants' overall preference for whether the summary should or should not be included with the earnings release. "Capture" and "Reliance" judgments were

[13] We excluded the following sections from Boeing's Q2-2008 earnings release: "Non-GAAP measure disclosure" (containing definitions) and "Forward-looking statements" (containing a disclaimer regarding forward-looking information). From Target's Q4-2013 earnings release, we excluded "Miscellaneous" (e.g., the company announcing the date for its next quarter earnings conference call) and "About Target" sections. The earnings releases contained 1,564 (Boeing) and 1,662 (Target) words, respectively.

made on 101-point scales with endpoints of 0 and 100 (both endpoints appropriately labeled). "Bias" judgments were made on a 101-point scale with endpoints of –50 ("Summary makes [Company] look worse") and +50 ("Summary makes [Company] look better"). For the "Should be included" judgments, participants selected either "Yes, the summary should be included" or "No, the summary should not be included." Participants also indicated what important information (if any) was missing from each summary, and whether any information included in each summary should not have been included (e.g., because it was redundant or irrelevant).

Participants for Study Two were again recruited from MTurk, following the same procedures described for Study One. In total, 334 people volunteered to take part and 98 (29.3%) met the qualification requirements and completed the study. Qualtrics randomly assigned participants to one of the two earnings releases, and the order in which participants viewed the summaries was also randomized. Participants were not told the source of any of the summaries. On completion, participants were paid $4.00 via MTurk. A mean (median) completion time of 39 (24) minutes resulted in a mean (median) hourly rate of $6.15 ($10).

3.5 Study Two: Results

3.5.1 Main Results

Table 3 presents the results of Study Two. Panel A of Table 3 presents results for the full sample, and Panels B and C break down results by company. A key insight of Study two is that participants judge the automatic summaries to be less biased in the company's favor than the management summaries. On average, participants further believe automatic summaries capture the important information from earnings releases as well as the management summaries. They are also just as likely to rely on the automatic summaries as the management summaries, and to believe that the automatic summaries should be included in the earnings release.

Insert Table 3 about here

Consistent with the results of Study One, LexRank appears to be the top overall performer among automatic summarization tools, in that it is rated high for "Capture," "Reliance," and "Should be included," but is also rated low for "Bias." LexRank also compares favorably to management summaries for both earnings releases on these measures.[14] Further supporting this conclusion, in an untabulated analysis, we compare the number of "fundamental" terms (e.g., "sales", "expenses", "margins") discussed in the body of the earnings release that are also included in the LexRank and management summaries.[15] For Target, of the 16 fundamental terms thus identified, the LexRank summary includes 10 of these terms, compared to seven in the management summary. For Boeing, the LexRank summary includes six of the 12 fundamental terms, compared to four in the management summary.

Overall, Study Two extends and reinforces our conclusions from Study One. Specifically, automatic summaries have the potential to capture the important information from earnings releases as well as management summaries, but with less bias. With respect to the specific automatic summarization tools, LexRank appears to perform particularly well.[16]

---

[14] For each earnings release, we find that LexRank produces summaries that are consistently rated as superior to the summary contained in the earnings release. Statistically, however, only participants assigned to Target rated bias as significantly lower for the LexRank summary compared to the management summary ($F_{1,56} = 12.45$, $p < 0.01$). A probable explanation for this difference is that, in the case of Target, management gave too little attention to two major events in the management summary: the credit-card breach and the struggling Canadian segment. Three raters (two co-authors and an independent rater) coded how often participants indicated that important information related to these events was missing from each summary. In untabulated analysis, for the LexRank (management) summary, we find that 19.30% (35.09%) of the participants indicated important information was missing related to these events. This difference is statistically significant ($F_{1,56} = 4.53$, $p = 0.04$) and significantly related to the difference in bias ($F_{1,56} = 5.81$, $p = 0.02$). This evidence suggests that Target's management avoids highlighting important negative events in its summary.

[15] For each earnings release, a list of fundamentals was agreed upon by two of the authors, who independently identified fundamental terms mentioned in the underlying text.

[16] LexRank's superior performance is consistent with findings in previous studies (e.g., Verma and Om 2016) and may be attributable to its use of a "reranker." According to Erkan et al. (2004), the reranker "penalizes the sentences that are similar to the sentences already included in the summary so that a better information coverage is achieved."

3.5.2 Additional Analysis: Intrinsic Evaluation

Summaries may be evaluated extrinsically or intrinsically (Nenkova and McKeown 2011). Extrinsic evaluation uses a criterion external to the summary to evaluate it (e.g., a summary's usefulness in carrying out a task). The results reported above reflect this approach. Intrinsic evaluation, on the other hand, considers the content of the summary relative to a benchmark or reference summary. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is the most commonly used intrinsic evaluation system (Nenkova and McKeown 2011). Under the ROUGE approach, the distribution of words in the reference summary, Sum_{Ref}, is compared to the distribution of words in candidate summaries, Sum_{Can} (Lin 2004), with a higher ROUGE score indicating a better match between the reference and candidate summaries. Specifically,

$$\text{ROUGE} = \frac{\sum_{\text{Words}} \in \{\text{Sum}_{\text{Can}} \cap \text{Sum}_{\text{Ref}}\}}{\sum_{\text{Words}} \in \text{Sum}_{\text{Ref}}}.$$

To construct a reference summary, we rely on an experienced Investor Relations Officer (IRO). The IRO received electronic copies of the two earnings releases used in Study Two (without summaries). We asked the IRO to read each earnings release and produce a summary consisting of five sentences, each presented as a bullet point. We asked that the summary should focus on the important information that best captured the content of the earnings release.

Insert Table 4 about here

Table 4 reports the ROUGE scores for Target (Q4 2013) and Boeing (Q2 2008) respectively. After correcting for stop words and allowing the software to use synonyms, we find that in all cases except one (when using bigrams, or word pairs, in the case of Boeing's Q2-2008 earnings release), LexRank summaries have higher ROUGE scores relative to management summaries. That is, LexRank summaries contain a larger number of words from the set of relevant words identified by the IRO. This intrinsic analysis suggests that, compared to the management summary, the LexRank summary better captures elements of the earnings release that are deemed important by the experienced IRO.

4. Study Three: The Effect of Summaries on Investors' Judgments

Whereas Studies One and Two focused on characteristics of summaries, Study Three tests the effects of automatic and management summaries on investors' judgments. We begin this section by developing hypotheses regarding the effect of automatic and management summaries on investors' judgments.

4.1 Hypothesis Development

If investors are rational and capable of extracting value relevant information from the underlying disclosure to which they have access, summaries preceding an underlying source document should have little impact on investors' judgments.[17] However, based on theories of bounded rationality and limited attention (Hirshleifer and Teoh 2003; Elliott et al. 2015), we expect summarization to affect investors' judgments.

4.1.1 Management Summary versus Automatic Summary

We first consider the impact of a management summary compared to an automatic summary. As shown in Studies One and Two, management summaries tend to be positively biased, depicting a more favorable image of the company compared to both automatic summaries (Study One) and the underlying document (Study Two). We posit that this positive bias affects investors' judgments via two psychological mechanisms: primacy effects and tone effects.

Research in psychology suggests that decision makers are prone to primacy effects, in that the order in which information is presented affects information processing (Asch 1946; Hogarth

[17] All participants in Study Three could access the full earnings release through hyperlinks of different sections of the earnings release.

and Einhorn 1992; Nisbett and Ross 1980). This theory suggests that encoding initial positive information tends to result in more positive global impressions of a target, relative to when participants first encode initial negative information (Sinclair 1988, 25). For example, Sinclair (1988) explicitly manipulates information order, and shows that participants initially receiving positive information (compared to those who initially review negative information), tend to make more favorable judgments when reviewing the performance of employees. Further, information retrieval seems to be directionally consistent with information order manipulation, with participants retrieving more positive information when seeing positive information first.

Building on this research in psychology, we argue that a summary is a piece of narrative information that is likely to be read first, and hence more likely to be remembered when constructing a problem representation due to primacy effects (Pennington and Hastie 1986). This is likely to then bias the way that participants acquire and/or interpret the information they review in the underlying earnings release. Thus, even when the full underlying text is available, investors' judgments may differ when management depicts a more favorable picture of the company in their summary compared to an automatic summary.

Tone management can also affect investors' impressions of a target company. Archival research suggests that investors react to opportunistic use of tone (Huang, Nekrasov, and Teoh 2013; Davis and Tama-Sweet 2012). Prior experimental research in accounting suggests that less sophisticated investors are influenced by the opportunistic use of tone in earnings releases, indicating that judgments about the firm's future earnings performance are more favorable when the earnings release is positively written (Tan, Wang and Zhou 2014). Further, as we document in our analysis of management summaries in practice (Table 1), manager-provided summaries are often even more favorable in tone than the underlying documents they summarize. In contrast, an automatic summary is likely to reflect the tone of the underlying document because it is based on sentence extraction.[18]

In sum, whereas the automatic summaries rely on sentence extraction, management gives information in the earnings release a positive spin by selectively emphasizing positive words and news items when they generate a summary. When we consider this tone management together with the selective content selection and primacy effects discussed earlier, we predict in H1 that a management summary will have a more favorable effect on investors' valuation judgments compared to an automatic summary generated using an algorithm.

H1: Investors' valuation judgments will be more favorable when a management summary accompanies an earnings release compared to when an automatic summary accompanies the earnings release.

4.1.2 Management Summary versus No Summary

We also consider the effect of a management summary compared to cases where investors do not receive a summary preceding the earnings release. In these cases, we also predict that a management summary will affect investors' valuation judgments positively, given that management summaries are likely to depict a more favorable picture of the company than the underlying document. Thus, H2 is based on the same theory as H1.

H2: Investors' valuation judgments will be more favorable when a management summary accompanies an earnings release compared to when no summary is provided.

4.1.3 Automatic Summary versus No Summary

Ex ante, neither theory nor previous research provide a clear directional prediction for how an automatic summary will affect investors' judgments compared to when no summary is

[18] For our study it is important to mention that we do not manipulate the tone of the underlying earnings release, but instead look at potential differences in tone management within a summary that accompanies the release.

provided. In the absence of a summary, investors must navigate and process a large amount of information, which can be more difficult when no initial guidance is provided (McDonald and Stevenson 1998). A summary arguably offers some initial guidance. Further, research suggests that a query-based summary—which extracts important content from the whole document, rather than simply displaying the first sentences of an article—positively affects information search (Tombros and Sanderson 1998).

However, while an automatic summary has potential to guide investors in retrieving relevant information, it may not be completely free from bias, given that it relies on sentence extraction of the underlying earnings release, which can itself be biased. Further, in our study, as in many real-world disclosures, all participants get a hyperlinked overview of the different sections of the earnings release and such a road map can already be beneficial for identifying relevant information (Arnold, Bedard, Phillips and Sutton 2012; McDonald and Stevenson, 1998). Given the uncertainty about the effects of an automatic summary, we leave the effect of an automatic summary compared to no summary as an empirical question.

4.2 Research Design and Materials

Study Three had a 1 × 3 between-subjects design, with summary type (automatic summary, management summary, or no summary) as the manipulated variable. For this third study, we created an earnings release for a hypothetical retail company based on Target Corporation's earnings release for the first quarter of 2016. Following prior literature (e.g., Rennnekamp 2012), we disguised the company's identity so that participants' familiarity with a real company would not influence their valuation or other judgments.[19] However, to preserve the external validity of

[19] Familiarity was not a concern in Studies One or Two because we were simply asking participants to rate characteristics of summaries and disclosures. In contrast, Study Three asked participants to form judgments about value, risk, and other characteristics of the firm.

the earnings release and the summaries, we changed only information that would clearly identify Target or other companies named in the earnings release: the firm name and logo, the location of its headquarters, names and contact details for company employees, and the name of another firm identified in the earnings release. In addition, to be consistent with the time of year in which the study was administered (late October), the earnings release for the hypothetical company reported results from the third quarter of fiscal year 2016 rather than the first quarter. Other than these changes, the earnings release and the management summary used in the study were identical to Target's actual earnings release and summary. The Appendix contains further details of the materials used in Study Three, including the management and automatic summaries.[20]

We then generated the automatic summary, with the same number of bullet points (six) as the management summary. Because of its superior performance in Studies One and Two, we used LexRank to generate the automatic summary. In addition, to ensure that the summaries exhibited similar characteristics as the summaries tested in Studies One and Two, we compared the automatic and management summaries for two potential sources of bias: content management (i.e., managers highlighting certain news items, while withholding others), and tone management (i.e., managing the tone of words in the management summary). With respect to content management, the company experienced a 5.4% sales decrease during the quarter compared to the same quarter in the previous year. Management does not mention this sales decrease in its summary, while the automatic summary does include a sentence from the underlying earnings release on the sales decrease. Consistent with the results of Studies One and Two, this suggests that management avoided mentioning an important negative news item in its summary.[21]

[20] The full earnings release contained 1,509 words, excluding tables. The management summary contained 106 words, and the automatic summary contained 164 words.
[21] Media reports on Target's Q1 2016 performance also indicate that the sales decrease was interpreted as both significant and negative (e.g., CNBC 2016, Oyedele 2016, Zacks 2016).

To compare tone management between the management and automatic summaries, we measure the frequency of negative/positive words (as a % of total words used) using the context-specific wordlist from Henry (2008). In addition, following Allee and DeAngelis (2015), we compute a measure of linguistic dispersion from the computational linguistics literature—(average) reduced frequency, or (A)RF—to measure the degree to which tone words are evenly distributed throughout the document. A higher RF (closer to 1) indicates that words are more "evenly" distributed throughout the document, while smaller values of RF indicate a "chunkier" distribution. A more even distribution of tone throughout the narrative reflects a portrayal of good or bad news as pervasive, while a less even distribution isolates the news to fewer components of performance. Table 5 presents the results of this tone analysis.

Insert Table 5 about here

The frequency of negative (positive) words in the underlying text of the earnings release—excluding the headline, management summary, and the tables—is a modest one percent (two percent). Both tone dispersion measures are higher for negative words. Considering that our algorithm extracts sentences from the underlying source document, it is reassuring that Table 5 documents tone frequencies and dispersion scores for the automatic summary similar to those documented for the earnings release. However, the frequency of positive words (nine percent) is considerably higher in the management summary than in the underlying earnings release, and (A)RF dispersion measures are also relatively high. Combined with the content management, this analysis indicates that management positively biases information in its summary, making this a powerful and representative setting in which to compare investors' reactions to management and automatic summaries.

4.3 Participants and Procedures

Participants for Study Three were again recruited from MTurk, using the same procedure as in Studies One and Two. In total, 308 people volunteered to take part and 90 (29.2%) met the qualification requirements and completed the study. Qualtrics randomly assigned participants to one of the three summary conditions (automatic, management, or no summary). Participants who viewed a summary were not told whether the summary was generated automatically or by management. On completion, participants were paid $2.00 via MTurk. A mean (median) completion time of approximately 16 (10) minutes resulted in a mean (median) hourly rate of $7.49 ($11.63).

Participants in Study Three first read background information about the hypothetical firm (called "Home Square Stores" or "HSQ"), and then provided an initial valuation judgment for the company's common stock on a 101-point scale with endpoints of 0 ("Very low value") and 100 ("Very high value"). Consistent with prior work (e.g., Asay et al. 2016), eliciting an initial valuation judgment allows us to more precisely measure the effect of our manipulation by measuring the difference between valuation judgments before and after observing the earnings release and accompanying summary (or lack of summary) according to each participant's assigned condition. This procedure thus controls for individual differences, such as participants' views about the retail industry as an investment and their use of the scale.

After making the initial valuation judgment, participants received HSQ's earnings release for the third quarter of 2016. In the automatic and management summary conditions, participants were asked to first read the summary provided before clicking a button that revealed hyperlinks to the sections and tables of the earnings release. Participants in the no summary condition were asked to click the button when they were ready to proceed. Clicking on any of the hyperlinks opened a new window containing the section or table. After reviewing the earnings release information, but before moving to the next screen, participants made a final valuation judgment for HSQ's common stock on a 101-point scale that was identical to the scale used for the initial valuation judgment.[22] On the next screen, participants made several additional judgments. First, participants indicated how confident they felt when making their final valuation judgment on a 101-point scale with appropriately labeled endpoints. Participants then indicated—via a free response—which factor was most important to their final valuation judgment, and also indicated up to four additional factors that were important. Next, participants rated HSQ's future earnings growth potential, the risk of investing in HSQ's common stock, and the favorability and credibility of HSQ's earnings release, all on 101-point scales with appropriately labeled endpoints (for a similar approach, see Frederickson and Miller 2004).

4.4 Results

Table 6 presents the results of Study Three. Panel A presents descriptive statistics by summary condition. Panel B shows planned comparisons between summary conditions. Based on our hypotheses, we expect higher valuation judgments when participants receive the earnings release with management's summary compared to the automatic summary (H1) or no summary (H2).

Insert Table 6 about here

Results support this prediction. With respect to our main dependent measure—the change in participants' valuation judgments (i.e., the final valuation judgment minus the initial valuation judgment)—participants who received the earnings release with management's summary

[22] We presented the final evaluation separate from the additional judgments to ensure that the final judgment is not confounded by any of these other judgments.

increased their valuation judgments by 7.70 points, compared to an increase of only 0.41 points for those who received the automatic summary (t_{60} = 1.85, p = 0.03, one-tailed), and a decrease of 0.14 points for those who received no summary (t_{59} = 2.06, p = 0.02, one-tailed). Valuation judgments of participants who received automatic summaries do not differ significantly from those of participants who received no summary (t_{55} = 0.13, p = 0.90, two-tailed).[23]

Results for the other investment-related judgments are also reported in Table 6. Consistent with our prediction for the valuation judgment, we would expect higher judgments for earnings growth potential and earnings release favorability, and lower risk judgments, when participants receive the earnings release with management's summary compared to the automatic summary or no summary. For the earnings growth potential and earnings release favorability measures, the judgments of participants who received the management summary are at least marginally higher than those of participants who received the automatic summary or no summary (all $p < 0.10$, one-tailed). For the risk measure, participants who received the automatic summary judge the risk of investing in HSQ's common stock to be higher compared to participants who received the management summary, and this difference is marginally significant (t_{60} = 1.35, p = 0.09, one-tailed). However, we observe no difference in the risk judgments of those who received the management summary compared to those who received no summary (t_{59} = 0.03, p = 0.52, one-tailed). The risk judgments of participants who received the automatic summary are directionally, but not significantly, higher than the risk judgments of participants who received no summary (t_{55} = 1.43, p = 0.16, two-tailed).

We also elicited two additional judgments from participants: the credibility of the earnings release, and the confidence they felt when making their final valuation judgments. We observe

[23] Results for valuation judgments are inferentially identical if we instead compare final valuation judgments across summary conditions, controlling for initial valuation judgments.

no differences in the credibility of the earnings release. Interestingly, however, participants who receive the automatic summary are more confident in their (lower) final valuation judgments than participants who receive the management summary (t_{60} = 1.84, p = 0.07, two-tailed). This difference in confidence is important to interpreting the valuation results, as it suggests participants who receive automatic summaries do not simply ignore the earnings release information in forming their final valuation judgments.

4.4 Mediation Analysis

The analysis reported above indicates that summary type had a significant effect on four of the additional investment-related measures: earnings growth potential, earnings release favorability, risk, and confidence. We next conduct a mediation analysis using Structural Equation Modeling (SEM) to determine which, if any, of these four measures explained investors' valuation judgments.[24] The results of this analysis are presented in Figure 2. Panel A presents results for the effect of the management summary compared to the automatic summary. Panel B presents results for the effect of the management summary compared to no summary.

Insert Figure 2 about here

We start by testing the overall goodness of fit for each model. For the management versus automatic model in Panel A, the Tucker-Lewis Index, which measures the improvement in fit compared to a null model, is 1.05, indicating that the model is a good fit for the data. The goodness of fit is confirmed by various other measures, including an Incremental Fit Index of 1.00, and an insignificant χ^2 test ($\chi^2_{(1)}$ = 0.62, p = 0.43) (Iacobucci 2010, Kline 2011). The

[24] SEM has several advantages over regression in testing for mediation, especially in cases that deviate from the simple X→M→Y relationship, as is the case in our models with their multiple potential mediators (e.g., Iacobucci, Saldanha and Deng 2007).

management versus no summary model in Panel B is also a good fit for the data, as confirmed by a Tucker-Lewis Index of 1.08, an Incremental Fit Index of 1.00, and an insignificant χ^2 test ($\chi^2_{(1)} = 0.35$, $p = 0.55$).

We next turn to the sign and significance of the path coefficients. Each model includes paths from summary type (the independent variable) to each of the four potential mediators, and paths from each of the four mediators to the change in valuation judgments (the dependent variable).[25] Full mediation is indicated if the following conditions hold: (1) the path from summary type to the mediator is significant, (2) the path from the mediator to the valuation judgment is significant, and (3) the path from summary type to the change in valuation judgment is insignificantly different from zero with the mediator included in the model (Baron and Kenny 1986; Iacobucci et al. 2007).

Results reveal that, for both models, these conditions are met for two of the potential mediators: earnings growth potential and earnings release favorability. Specifically, the path coefficients for the effect of summary type on earnings growth potential and earnings release favorability are significantly positive. In addition, the path coefficients from these two measures to common stock valuation judgments are significantly positive. Finally, with the potential mediators included in the models, the path coefficients from summary type to common stock valuation are no longer significant. These results indicate that the effect of the management summary on participants' valuation judgments is fully explained by their judgments of earnings growth potential and earnings release favorability.

Two points about these results are in order. First, the mediating effect of potential future earnings growth indicates that participants are sufficiently knowledgeable about the determinants

[25] We also allow error terms for the mediators to covary (these covariance paths are omitted from Figure 2 to simplify the presentation).

of equity value that their judgments of common stock value are closely related to their judgments of the company's potential for future earnings growth. Second, the mediating effect of earnings release favorability suggests that the management summary changes participants' overall impression of the earnings release, despite receiving the same underlying information, consistent with the theory underlying our hypotheses.

4.5 Information Search and Processing

As noted, all participants had access to the full text of the earnings release, and accompanying tables, via hyperlinks. This design allowed us to measure the time participants spent searching specific sections of the earnings release. Further, participants listed up to five factors that were important to their valuation judgment, providing insights into participants' processing of the earnings release information, including the summaries. Below, we discuss several insights from this information search and processing data.

We first compare, across summary conditions, the time that participants spent searching for and processing information. We measured time in two different ways: total time spent on the study and time spent on the earnings release information. We detect no differences in these time measures across summary conditions (all p-values > 0.10, two-tailed). Next, we compare time spent on each of the earnings release sections and the five tables across summary conditions. Results indicate that participants who received summaries, regardless of summary type, spent significantly more time searching for information in the following four sections: Capital Returned to Shareholders, Discontinued Operations, Reconciliation of Non-GAAP Financial Measures (table), and Segment Results (table). Compared to participants who did not receive a summary, search time for these sections was significantly higher for participants who received automatic summaries and for participants who received management summaries (all p-values <

0.10, two-tailed). Notably, these sections and tables of the earnings release provide detail on the more complex and economically meaningful aspects of the company's performance during the quarter.[26] Interestingly, we do not observe any significant effects of summary *type* (i.e., automatic versus management) on search time related to these sections (all p-values > 0.10, two tailed). These results suggest that providing a summary, regardless of whether generated automatically or by management, improves the efficiency of participants' information search, directing their focus toward more important or complex sections of the earnings release.

We also coded the influential factors listed by participants for mentions of "sales" or "revenue." As previously noted, the company reported a year-over-year sales decline in the quarter, which was widely interpreted as significant and negative news. The sales decline was explicitly mentioned in the automatic summary, but not in the management summary. Perhaps surprisingly, then, we do not observe differences in mentions of "sales" or "revenue" between summary conditions. Specifically, 34.4% of participants who received an automatic summary (10 of 29), 36.4% of participants who received a management summary (12 of 33), and 46.4% of participants who received no summary (13 of 28) explicitly mentioned sales or revenue as a significant factor that influenced their valuation judgments ($\chi^2_{(2)} = 1.00$, $p = 0.61$).[27] Nevertheless, combined with the fact that we do observe differences in valuation and other investment-related judgments, this result suggests that summary type does not seem to affect the acquisition of information from the underlying document, but rather affects the processing and interpretation of the information.

[26] For example, the company distributed an unusually large amount of cash—more than $1.2 billion, representing more than twice net income from continuing operations from the quarter—to shareholders during the quarter, either as dividends or as share repurchases. This information was included in the Capital Returned to Shareholders section.

[27] We note, however, that these results should be interpreted with caution, as we do not look at qualifiers that accompany these words, and some participants mention sales in an ambiguous way (i.e., without stating explicitly whether sales influenced their judgments positively or negatively).

5. Conclusion

Automatic summarization technology is today recognized as a useful tool in various disciplines. In this paper, we assess the viability of automatic summarization in the domain of corporate disclosures. Specifically, we present archival evidence on the use of summaries in practice and conduct three experiments to investigate how automatic summaries compare to management summaries on several dimensions, and how summaries affect investor information processing, beliefs about firm fundamentals, and valuation judgments. Our study thus responds to the call by Barth (2015, 506) for research on summarization "to aid investors and other outside providers of capital in their decision making."

Our archival evidence shows that summaries are widely used in practice. However, our analysis also shows that management summaries introduce incremental bias compared to the underlying documents they summarize. Summaries related to earnings releases seem to depict a stronger favorable bias compared to MD&As. Results of our experiments suggest important advantages of automatic summaries (e.g., less bias, more investor confidence) over summaries written by management in particular for earnings releases. The key result from Study Three is that investors who receive an earnings release accompanied by an automatic summary arrive at more conservative (i.e., lower) valuation judgments, and are more confident in those judgments, compared to investors who receive the same earnings release with a management summary. As such, our study also informs policy makers, including the SEC, which is considering a rule that explicitly allows Form 10-K filers to provide summary information (SEC 2016). Based on our results, we argue that automatic summarization algorithms have potential in a corporate disclosure context. Because they rely on sentence extraction, automatic summaries have the potential to present a more balanced picture than management summaries.

This study opens up avenues for future research on the role played by summarization in capital markets. For example, in practice, investors could generate an automatic summary and use it alongside a summary provided by management. When automatic summaries differ from management summaries, management bias (i.e., tone management and/or content management) might become evident, and could potentially affect investors' interpretation of the underlying information. At the same time, given individual investors' tendency to disregard the content of earnings releases (Blankespoor, deHaan, Wertz and Zhu, 2017), automatic summaries may be better at enhancing "the ability of investors and other users to process relevant information and/or [reduce] their processing time and search costs" (SEC Release No. 34-77969).

Research could also examine how the existence or widespread use of automatic summarization affects management summaries or indeed the underlying source documents. If managers are aware that their disclosures will be summarized automatically and investors tend to trade on this information, they might alter their own summaries to be less biased and/or change the disclosure itself so that more positive information is identified by the automatic summarization algorithm. Finally, different types of investors have different information needs (e.g., Hales et al. 2011). As automatic summarization technology matures, research could investigate automatic summaries that are customized based on investors' preferences and/or for other types of disclosures (e.g., conference calls, prospectuses) for which investors may find summaries useful.

APPENDIX
Summaries used in Study Three

Panel A: Earnings Release Header

HOME SQUARE STORES

FOR IMMEDIATE RELEASE

Contacts: James Connelly, Investors, (726) 616-7216
Emily Hubble, Media, (726) 829-5167
Home Square Media Hotline, (726) 696-4300

Home Square Reports Third Quarter 2016 Earnings

Panel B: Management Summary

- *Third quarter Adjusted EPS of $1.29 was above the company's guidance of $1.15 to $1.25.*
- *Third quarter comparable sales increased 1.2 percent, driven by growth in both traffic and basket.*
- *Comparable digital channel sales increased 23 percent, on top of 38 percent growth in third quarter 2015.*
- *Third quarter comparable sales in signature categories (Style, Baby, Kids and Wellness) grew more than three times as fast as the company average.*
- *The third quarter marked Home Square's sixth consecutive quarter of traffic growth, reflecting increases in both stores and digital channels.*
- *Home Square returned $1.2 billion to shareholders in the third quarter through dividends and share repurchases.*

Panel C: Automatic Summary

- *Third quarter GAAP earnings per share (EPS) from continuing operations were $1.02, compared with $1.01 in third quarter 2015.*
- *Third quarter 2016 GAAP EPS from continuing operations reflects $261 million of pre-tax early debt retirement losses, costs related to the sale of the pharmacy and clinic businesses to DA Pharma and the resolution of income tax matters.*
- *Third quarter 2016 sales decreased 5.4 percent to $16.2 billion from $17.1 billion last year, as a 1.2 percent increase in comparable sales was more than offset by the impact of the sale of the pharmacy and clinic businesses.*
- *The Company's third quarter 2016 net interest expense was $415 million, compared with $155 million last year, driven by a $261 million charge related to the early retirement of debt.*
- *Third quarter 2016 effective income tax rate from continuing operations was 31.6 percent, compared with 34.8 percent last year.*
- *Third quarter net earnings from discontinued operations were $18 million, compared with after-tax losses of ($16) million last year.*

Panel D: Earnings Release Sections

Sections

- Preface
- Fiscal 2016 Earnings Guidance
- Segment Results
- Interest Expense and Taxes from Continuing Operations
- Capital Returned to Shareholders
- Discontinued Operations
- Conference Call Details
- Miscellaneous
- About Home Square

Tables

- Consolidated Financial Statements:
  - Consolidated Statements of Operations
  - Consolidated Statements of Financial Position
  - Consolidated Statements of Cash Flows
- Segment Results
- Reconciliation of Non-GAAP Financial Measures

Note: After viewing the earnings release header and summary (header only for no summary condition), participants clicked a button labeled ‘access earnings release sections’, which displayed hyperlinks to the different parts of the company’s full earnings release, as shown in Panel D.

REFERENCES

Ahern, K.R. and D. Sosyura. 2014. Who Writes the News? Corporate Press Releases during Merger Negotiations. *The Journal of Finance* 69 (1): 241-291

Allee, K., and M. DeAngelis. 2015. The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research* 53(2): 241-274.

Arnold, V., J. C. Bedard, J. R. Phillips, and S.G. Sutton. 2014. The Impact of Tagging Qualitative Financial Information on Investor Decision Making: Implications for XBRL. *International Journal of Accounting Information Systems* 13: 2-20.

Arslan-Ayaydin, O., K. Boudt and J. Thewissen. 2016. Managers Set the Tone: Equity Incentives and the Tone of Earnings Press Releases. *Journal of Banking & Finance* S132-S147.

Asay, H.S., W.B. Elliott, and K.R. Rennekamp. 2016. Disclosure Readability and the Sensitivity of Investors' Valuation Judgments to Outside Information. *The Accounting Review*, in press.

Asch, S. E. 1946. Forming impressions of personality. *The Journal of Abnormal and Social Psychology*. 41(3): 258-290.

Baron, R.M., and D. Kenny. 1986. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology* 51(6): 1173-1182.

Barth, M. E. 2015. Financial Accounting Research, Practice, and Financial Accountability. *ABACUS: A Journal of Accounting, Finance and Business Studies* 51(4): 499-510.

Blankespoor, E., E. deHaan, and C. Zhu. 2017. Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-Journalism. *Review of Accounting Studies*, forthcoming.

Blankespoor, E., E. deHaan, J. Wertz, and C. Zhu. 2017. Why do Some Investors Disregard Earnings News? Awareness Costs versus Processing Costs. *Working paper*. Stanford University.

Bonner, S. E., S. M. Clor-Proell, S. M., and L. Koonce. 2014. Mental Accounting and Disaggregation Based on the Sign and Relative Magnitude of Income Statement Items. *The Accounting Review* 89(6): 2087-2114.

Brandow, R., K. Mitze, and L. E. Rau. 1995. Automatic Condensation of Electronic Publications by Sentence Selection. *Information Processing & Management*, 31(5): 675-685

CNBC. 2016. Target Earnings Top Expectations, But Revenue Is Light. Retrieved from http://www.cnbc.com/2016/05/18/target-q1-earnings-report.html.

Daniel, K., D. Hershleifer, and S. H. Teoh. 2002. Investor Psychology in Capital Markets: Evidence and Policy Implications. *Journal of Monetary Economics* 49: 139-209.

Davis, A., and I. Tama-Sweet. 2012. Managers Use of Language Across Alternative Disclosure Outlets: Earnings Press Releases Versus MD&A. *Contemporary Accounting Research* 29: 804–837.

DeFranco, G., H. Fogel-Yaari, and H. Li. 2016. MD&A Textual Similarity and Auditors. Working paper, University of Toronto, Tulane University, and Nanyang Technological University.

Desai, H., S. Rajgopal, and J. J. Yu. 2016. Were Information Intermediaries Sensitive to the Financial Statement-Based Leading Indicators of Bank Distress Prior to the Financial Crisis. *Contemporary Accounting Research* 33(2): 576-606.

Dyck, A., and L. Zingales. 2003. The Media and Asset Prices. Harvard Business School Working Paper Series, No. 04-003.

Dyer, T., Lang, M., and Stice-lawrence, L. 2016. The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. Available at SSRN: https://ssrn.com/abstract=2741682.

Elliott, W. B., J. L. Hobson, J. L., and B. J. White. 2015. Earnings Metrics, Information Processing, and Price Efficiency in Laboratory Markets. *Journal of Accounting Research* 55(3): 555-592.

Elliott, W. B., F. D. Hodge, J. J. Kennedy, and M. Pronk. 2007. Are MBA Students a Good Proxy for Nonprofessional Investors? *The Accounting Review* 82(1): 139-168.

Erkan, G., and D. R. Radev. 2004. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research* 22: 457-479.

Ernst & Young. 2014. *Disclosure Effectiveness: What Companies Can Do Now*. Retrieved from http://www.ey.com/Publication/vwLUAssets/EY-disclosure-effectiveness-what-companies-can-do-now/$FILE/EY-disclosure-effectiveness-what-companies-can-do-now.pdf.

Farrell, A., J. H. Grenier, and J. Leiby. 2017. Scoundrels or Stars? Theory and Evidence on the Quality of Workers in Online Labor Markets. *The Accounting Review* 92(1): 93-114.

FASB (Financial Accounting Standards Board). 2015. FASB's Simplification Initiative: An Update. Retrieved at http://www.fasb.org/jsp/FASB/Page/SectionPage&cid=1176165963019

Frederickson, J., and J. Miller. 2004. The Effects of Pro Forma Earnings Disclosures on Analysts' and Nonprofessional Investors' Valuation Judgments. *The Accounting Review* 79(3): 667-686.

Francis, J., K. Schipper, and L. Vincent. 2002. Expanded Disclosures and the Increased Usefulness of Earnings Announcements. *The Accounting Review* 77(3): 515-546.

Garcia-Gathright, J., N. Matiasz, E. Garon, D. Aberle, R. Taira, and A. Bui. 2016. Toward Patient-Tailored Summarization of Lung Cancer Literature. *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 449-452.

Guillarmon-Saorin, E., B. Garcia Osma, and M. J. Jones. 2012. Opportunistic Disclosure in Press Release Headlines. *Accounting and Business Research* 42(2): 143-168.

Hales, J., X. Kuang, and S. Venkataraman. 2011. Who Believes the Hype? An Experimental Investigation of How Language Affects Investor Judgments. *Journal of Accounting Research* 49(1): 223-255.

Henry, E. 2008. Are Investors Influenced by How Earnings Press Releases are Written? *Journal of Business Communications* 45: 363–407.

Henry, E., and A. Leone. 2016. Measuring Qualitative Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *The Accounting Review* 91(1): 153-178.

Hirshleifer, D., and S. H. Teoh. 2003. Limited Attention, Information Disclosure, and Financial Reporting. *Journal of Accounting and Economics* 36: 337-386.

Hirshleifer, D., S. S. Lim, and S. H. Teoh. 2009. Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *The Journal of Finance* 64(5): 2289-2325.

Hogarth, R.M. and H. J. Einhorn. 1992. Order Effects in Belief Updating: The Belief-Adjustment Model. *Cognitive Psychology* 24: l-55.

Holmes, J. 2016. AI is Already Making Inroads Into Journalism But Could It Win a Pulitzer. *The Guardian*. http://www.theguardian.com/media/2016/apr/03/artificla-intelligence-robot-reporter-pulitzer-prize.

Huang, X., A. Nekrasov, and S. Teoh. 2013. Headline Salience and Over- and Underreactions to Earnings. Working paper. University of California, Irvine.

Huang, X., S. Teoh, and Y. Zhang. 2014. Tone Management. *The Accounting Review* 89(3): 1083-1113.

Iacobucci, D., N. Saldanha, and X. Deng. 2007. A Meditation on Mediation: Evidence That Structural Equations Models Perform Better than Regression. *Journal of Consumer Psychology* 17(2): 139-153.

Iacobucci, D. 2010. Structural Equations Modeling: Fit Indices, Sample Size, and Advanced Topics. *Journal of Consumer Psychology* 20: 90-98.

Kline, R. B. 2011. *Principles and Practice of Structural Equation Modeling*. 3rd edition. New York: The Guilford Press.

KPMG. 2011. *Disclosure Overload and Complexity: Hidden in Plain Sight.* Retrieved from https://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/disclosure-overload-complexity.pdf.

Krische, S.D. 2015. The Impact of Individual Investors' Financial Literacy on Assessments of Conflicts of Interest. Working paper American University.

Loughran, T., and B. McDonald. 2014. Measuring Readability in Financial Disclosures. *The Journal of Finance* 69(4): 1643-1671.

Loughran T. and B. McDonald. 2016. Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research* 54(4): 1187-1230.

McDonald, S., and R. J. Stevenson. 1998. Navigation in Hyperspace: An Evaluation of the Effects of Navigational Tools and Subject Matter Expertise on Browsing and Information Retrieval in Hypertext. *Interacting with Computers* 10: 129-142.

Nenkova, A. and K. McKeown. 2011. Automatic Summarization. *Foundations and Trends in Information Retrieval* 5(2-3): 103-233.

Nisbett, R., and L. Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs*, NJ: Prentice Hall.

Oyedele, A. 2016. Target Says Its Sales Will Stink. *Business Insider*. Retrieved from http://www.businessinsider.com/target-earnings-2016-5.

Paredes, T. 2003. Blinded by the Light: Information Overload and Its Consequences for Securities Regulation. *Washington University Law Quarterly* 81: 417-485.

Paredes, T. 2013. Remarks at the SEC speaks in 2013. Retrieved from www.sec.gov/News/Speech/Detail/Speech/1365171492408.

Pennington, N., and R. Hastie. 1986. Evidence Evaluation in Complex Decision Making. *Journal of Personality and Social Psychology* 51: 242-258.

Rennekamp, K.R. 2012. Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research* 50: 1319-1354.

Salton, G., A. Sighhal, M. Mitra., and C. Buckley. 1997. Automatic Text Structuring and Summarization. *Information Processing & Management* 33(2): 193-207.

Schwarz, N. 2004. Metacognitive Experiences in Consumer Judgment and Decision Making. *Journal of Consumer Psychology* 14(4): 332-438.

SEC (Securities and Exchange Commission). 2013. *Report on Review of Disclosure Requirements in Regulation S-K.* Retrieved from https://www.sec.gov/news/studies/2013/reg-sk-disclosure-requirements-review.pdf.

SEC (Securities and Exchange Commission). 2016. *SEC Adopts Amendment to Form 10-K Implementing FAST Act Provision.* Retrieved from https://www.sec.gov/news/pressrelease/2016-101.html.

Sinclair, R.C. 1988. Mood, Categorization Breadth, and Performance Appraisal: The Effects of Order of Information Acquisition and Affective State on Halo, Accuracy, Information Retrieval, and Evaluations. *Organizational Behavior and Human Decision Processes* 42: 22-46.

Tan, H-T., E. Y. Wang and B. Zhou. 2014. When the Use of Positive Language Backfires: The Joint Effect of Tone, Readability, and Investor Sophistication on Earnings Judgments. *Journal of Accounting Research* 52 (1): 273-302.

Tombros, A. and M. Sanderson. 1998. Advantages of Query Biased Summaries in Information Retrieval. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 2-10.

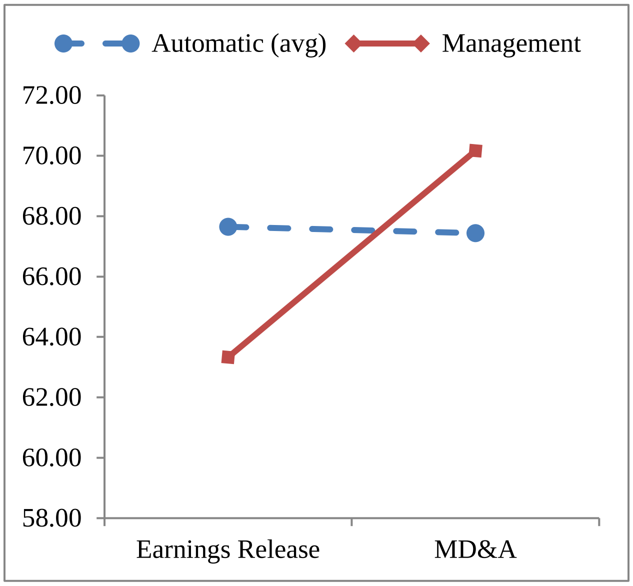
Umar, T. 2017. Complexity Aversion when *Seeking Alpha*. Working paper. University of Chicago, Booth School of Business.

Verma, P., and H. Om. 2016. Extraction Based Text Summarization Methods on User's Review Data: A Comparative Study. In: Unal A., et al. (eds) *Smart Trends in Information Technology and Computer Communications*. Communications in Computer and Information Science, Vol 628: 346-354.

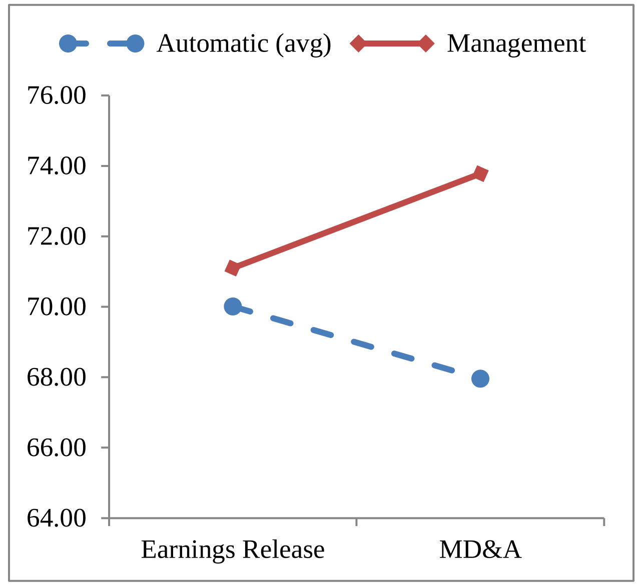
Zacks. 2016. Target (TGT) Q1 Earnings Beat, Sales Miss; Stock Plunges. Retrieved from https://www.zacks.com/stock/news/217811/target-tgt-q1-earnings-beat-sales-miss-stock-plunges.

FIGURE 1
Study One Results

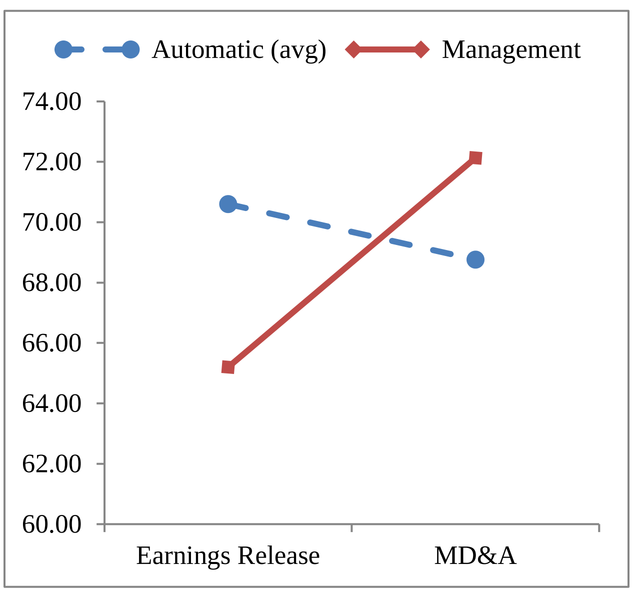
Panel A: Informativeness



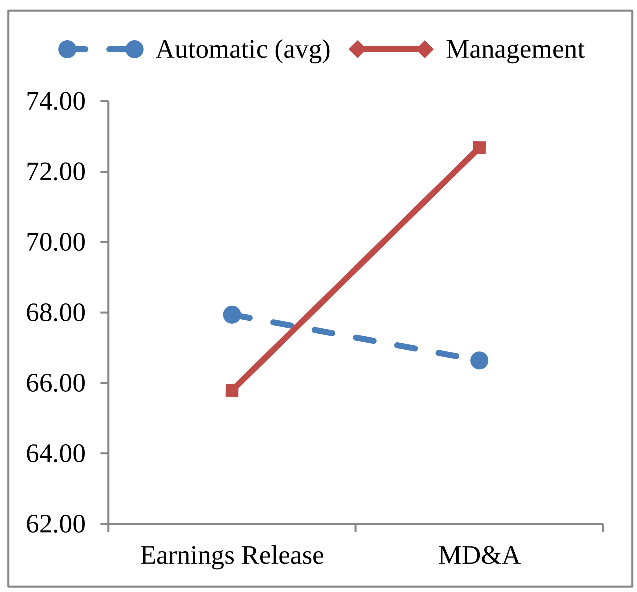
Panel B: Readability



Panel C: Credibility



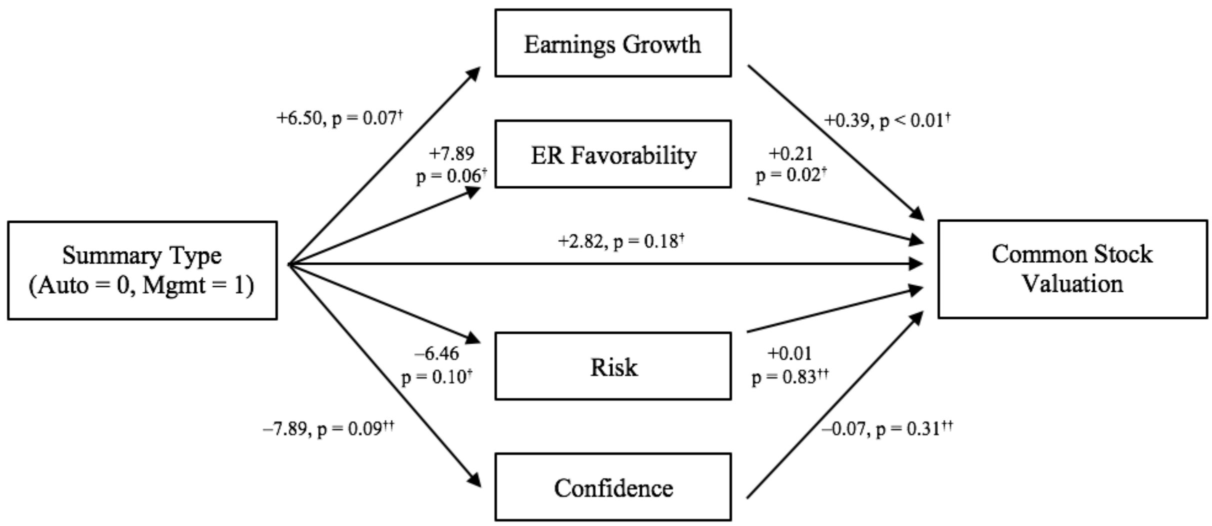
Panel D: Overall Usefulness



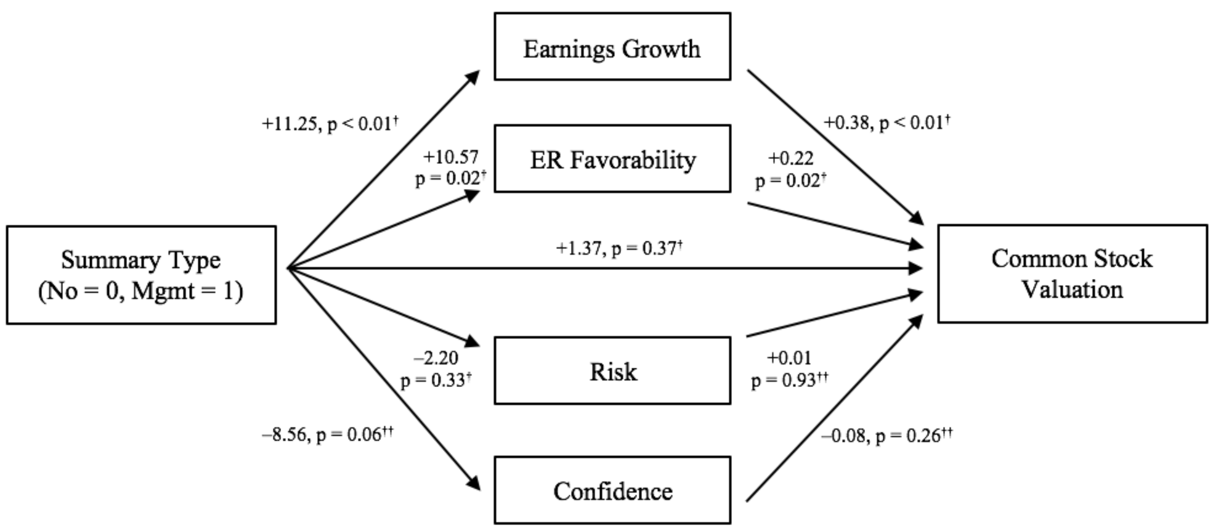
Note: In Study One, participants reviewed (between-subjects) summaries for one of six company disclosures (three earnings releases and three MD&A). Summary type was administered within-subjects (six automatic summaries and one management summary). Participants were informed about the company and disclosure type for which the summaries were generated. All questions were answered on 101-point scales. Means presented are averaged across the three earnings release conditions and the three MD&A conditions. See Table 2 for descriptive and inferential statistics.

FIGURE 2
Study Three Mediation Analysis

Panel A: Management versus Automatic Summary



Panel B: Management versus No Summary



Note: Panel A (B) presents results of a structural equation analysis that tests potential mediators of the effect of a management summary compared to an automatic summary (no summary) on participants' judgments of common stock value. Next to each arrow are path coefficients and *p*-values (with † and †† indicating one-tailed and two-tailed tests, respectively). Overall goodness of fit is high for both models, as measured by the following measures. Panel A: Tucker-Lewis Index (1.05), Incremental Fit Index (1.00), χ^2 test ($\chi^2_{(1)} = 0.62$, $p = 0.43$). Panel B: Tucker-Lewis Index (1.08), Incremental Fit Index (1.00), χ^2 test ($\chi^2_{(1)} = 0.35$, $p = 0.55$).

TABLE 1
Management Summaries in S&P 100 Corporate Disclosures

Panel A: Sample Selection

	Q4-2015 earnings release	Annual 2015 MD&A
All S&P 100 firms	100	100
Document format unsuitable for analysis	(4)	-
MD&A not included in 10-K	-	(11)
Subtotal: disclosures available for analysis	96	89
Disclosures not including summaries	(18)	(25)
Disclosures including summaries available for analysis[a]	78	64
Full text, excluding summary: average word count	2,867	19,458
Summary: average word count	127	764

[a] *Forty-eight firms provided summaries for both earnings releases and MD&A*

Panel B: Earnings Release Summary vs. Full Text (N = 78)

	Summary	Full text	Difference = (1) – (2)	Two-sided t-test of difference
	(1)	(2)	(3)	(4)
Negative tone	0.73%	1.18%	-0.45%	-2.35
				(*p* = 0.02)
Positive tone	5.45%	2.32%	3.13%	7.03
				(*p* < 0.01)

Panel C: MD&A Summary vs. Full Text (N = 64)

	Summary	Full text	Difference = (1) – (2)	Two-sided t-test of difference
	(1)	(2)	(3)	(4)
Negative tone	1.60%	1.30%	0.29%	1.57
				(*p* = 0.12)
Positive tone	3.01%	1.74%	1.27%	6.19
				(*p* < 0.01)

Panel D: Earnings Release vs. MD&A Summaries (N = 48)

	Earnings release summary	MD&A summary	Difference = (1) – (2)	Two-sided t-test difference
	(1)	(2)	(3)	(4)
Negative tone	0.79%	1.69%	-0.90%	-2.79
				(*p* < 0.01)
Positive tone	5.33%	3.07%	2.26%	3.57
				(*p* < 0.01)

TABLE 2
Results of Study One

Panel A: Mean Judgments for Earnings Releases (Alibaba Q1 2016, Boeing Q2 2008, Target Q4 2013), N = 153

	(1) KL	(2) LEX	(3) LSA	(4) LUHN	(5) SB	(6) TR	**Average Auto**	**Average Management**
Length	33.46**	48.99***	45.41***	70.03***	34.35*	69.48***	**50.30*** **	**36.93**
Informativeness	63.25	67.42**	62.00	78.42***	58.10**	76.70***	**67.65*** **	**63.33**
Readability	71.03	73.25	70.97	69.60	69.10	66.14**	**70.01**	**71.10**
Credibility	66.12	71.37***	69.24**	76.69***	65.71	74.47***	**70.60*** **	**65.20**
Usefulness	64.55	69.95*	64.18	75.33**	59.54***	74.46**	**67.94**	**65.79**
Written by management	56.86	69.92***	64.39**	68.61***	53.39**	68.98***	**63.72** **	**58.87**

Panel B: Mean Judgments for MD&A (Macy's 2014, Mattel 2014, Target 2013), N = 150

	(1) KL	(2) LEX	(3) LSA	(4) LUHN	(5) SB	(6) TR	**Average Auto**	**Average Management**
Length	42.02	38.68***	44.52	77.89***	40.91**	68.13***	**52.03*** **	**44.81**
Informativeness	66.95**	63.77***	63.32***	74.53**	66.35**	69.70	**67.44** **	**70.17**
Readability	73.00	73.85	70.81*	57.55***	72.80	59.75***	**67.96*** **	**73.78**
Credibility	69.11**	68.11***	66.10***	70.83	70.58	67.81***	**68.76*** **	**72.13**
Usefulness	68.70***	66.81***	65.13***	66.27***	67.58***	65.36***	**66.64*** **	**72.68**
Written by management	63.58**	63.36**	64.91*	67.69	64.17**	62.60***	**64.39*** **	**68.47**

Panel C: Repeated Measures ANOVA for Selected Participant Judgments

	Informativeness	Readability	Credibility	Usefulness
Between subjects:				
Disclosure type (ER vs. MD&A)	$F = 3.46$, $p = 0.06$	$F = 0.03$, $p = 0.86$	$F = 1.89$, $p = 0.17$	$F = 2.26$, $p = 0.13$
Within subjects:				
Summary type (Auto vs. Management)	$F = 0.60$, $p = 0.44$	$F = 12.09$, $p < 0.01$	$F = 1.25$, $p = 0.27$	$F = 3.31$, $p = 0.07$
Summary type × Disclosure type	$F = 11.80$, $p < 0.01$	$F = 5.68$, $p = 0.02$	$F = 23.47$, $p < 0.01$	$F = 14.69$, $p < 0.01$

Panels A and B: *,**,*** indicate different from management summary at $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively (all two-tailed).

TABLE 3
Results of Study Two

Panel A: Mean Judgments (Overall), N = 98

	(1) KL	(2) LEX	(3) LSA	(4) SB	**Avg Auto**	**Avg Mgmt**
Capture	57.70	64.39	56.77^{*}	52.77^{***}	**57.91**	**61.70**
Reliance	55.39	61.27	55.00	50.50^{***}	**55.54**	**59.11**
Bias	6.08	4.01^{***}	8.83	5.81^{*}	**6.18^{**}**	**11.14**
Should be included (% yes)	65.3%	78.6%	61.2%*	56.1%**	**65.33%**	**73.5%**

Panel B: Mean Judgments (Boeing Q2 2008 Earnings Release), N = 41

	(1) KL	(2) LEX	(3) LSA	(4) SB	**Avg Auto**	**Avg Mgmt**
Capture	61.07	63.71	60.80	59.61	**61.30**	**58.44**
Reliance	58.66	58.00	58.32	56.05	**57.76**	**56.22**
Bias	15.27	5.15	1.29	13.76	**8.87**	**8.44**
Should be included (% yes)	73.2%	80.5%	68.3%	65.9%	**72.0%**	**73.2%**

Panel C: Mean Judgments (Target Q4 2013 Earnings Release), N = 57

	(1) KL	(2) LEX	(3) LSA	(4) SB	**Avg Auto**	**Avg Mgmt**
Capture	55.28^{**}	64.88	53.86^{***}	47.84^{***}	**55.46^{***}**	**64.05**
Reliance	53.04^{**}	63.65	52.61^{***}	46.51^{***}	**53.95^{***}**	**61.19**
Bias	-0.53^{***}	3.19^{***}	14.25	0.09^{***}	**4.25^{***}**	**13.09**
Should be included (% yes)	59.6%	77.2%	56.1%**	49.1%**	**60.6%**	**73.7%**

*,**,*** indicate different from management summary at $p < 0.10$, $p < 0.05$ and $p < 0.01$, respectively (all two-tailed).

TABLE 4
Study Two – Intrinsic Evaluation

		ROUGE score			
		Unigrams			Bigrams
Earnings release:	*Summary type:*	NN	YN	YY	YY
Target Q4 2013	Automatic	0.3759	0.2933	0.4359	0.0958
	Management	0.3007	0.2666	0.3974	0.0684
Boeing Q2 2008	Automatic	0.4393	0.4230	0.4821	0.2352
	Management	0.3939	0.3846	0.4285	0.2352

ROUGE scores are obtained using ROUGE 2.0, a Java package developed by Kavita Ganesan (http://kavita-ganesan.com/content/rouge-2.0). The automatic summary was generated using LexRank. NN = no stop words correction, no synonyms allowed; YN = stop words correction, no synonyms allowed; YY = stop words correction, synonyms allowed. We obtained the latest version of WordNet from wordnet.princeton.edu/wordnet/download/current-version/.

TABLE 5
Study Three – Tone Analysis

	Earnings release full text	Management summary	Automatic summary
	(1)	(2)	(3)
Panel A: Negative Tone Words			
Frequency	5	0	1
(as % of total words)	(1%)	(0%)	(1%)
Negative RF	0.60	N/A	1.00
Negative ARF	0.74	N/A	1.00
Panel B: Positive Tone Words			
Frequency	21	9	2
(as % of total words)	(2%)	(9%)	(1%)
Positive RF	0.52	0.78	0.50
Positive ARF	0.54	0.69	0.53

Positive and negative words are identified using the context-specific wordlist developed by Henry (2008).

TABLE 6
Study Three – Results

Panel A: Mean (Standard Deviation) of Participants' Judgments

	Summary type		
Judgment	Automatic N = 29	Management N = 33	None N = 28
	(1)	(2)	(3)
Initial valuation (pre-manipulation)	55.72	53.61	54.46
	(12.81)	(14.73)	(10.36)
Final valuation (post-manipulation)	56.14	61.30	54.32
	(18.59)	(18.49)	(18.77)
Change in valuation (final minus initial)	0.41	7.70	–0.14
	(16.87)	(14.09)	(15.67)
Earnings growth potential	58.21	64.45	53.71
	(18.11)	(18.51)	(17.18)
Earnings release favorability	56.90	64.67	54.89
	(24.90)	(18.79)	(18.21)
Risk	59.55	52.21	52.04
	(21.29)	(21.50)	(18.33)
Earnings release credibility	72.76	68.94	71.96
	(15.31)	(16.64)	(14.89)
Confidence	73.31	63.45	69.89
	(20.88)	(21.27)	(12.65)

Panel B: Comparisons

Judgment	Contrast	Expectation	*t*-stat	*p*-value
Change in valuation	Mgmt vs. Auto	Mgmt > Auto	1.85	0.03†
	Mgmt vs. None	Mgmt > None	2.06	0.02†
	Auto vs. None	?	0.13	0.90††
Earnings growth potential	Mgmt vs. Auto	Mgmt > Auto	1.34	0.09†
	Mgmt vs. None	Mgmt > None	2.33	0.01†
	Auto vs. None	?	0.96	0.34††
Earnings release favorability	Mgmt vs. Auto	Mgmt > Auto	1.40	0.09†
	Mgmt vs. None	Mgmt > None	2.05	0.02†
	Auto vs. None	?	0.35	0.73††
Risk	Mgmt vs. Auto	Mgmt < Auto	1.35	0.09†
	Mgmt vs. None	Mgmt < None	0.03	0.52†
	Auto vs. None	?	1.43	0.16††
Earnings release credibility	Mgmt vs. Auto	?	0.94	0.35††
	Mgmt vs. None	?	0.74	0.46††
	Auto vs. None	?	0.20	0.84††
Confidence	Mgmt vs. Auto	?	1.84	0.07††
	Mgmt vs. None	?	1.40	0.17††
	Auto vs. None	?	0.74	0.46††

†, †† designate one-tailed and two-tailed *p*-values, respectively.

Online Appendix "Automatic Summarization of Corporate Disclosures"

This online appendix contains a primer on automatic summarization. This primer is not intended to be exhaustive. For further discussion, we refer the reader to the original papers referenced herein, and textbooks such as Juan-Manuel Torres-Moreno's *Automatic Text Summarization* (Wiley 2014), Inderjeet Mani's *Automatic Summarization* (John Benjamins Publishing Company 2001), or Inderjeet Mani and Mark Maybury's *Advances in Automatic Text Summarization* (MIT Press 1999). For a review of the literature, see Nenkova and McKeown 2011.

OA.1 Text Summarization

A summary is “a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that.” (Radev, Hovy and McKeown 2002). To summarize a text implies taking “an information source, extract[ing] content from it, and present[ing] the most important content to the user in a condensed form and in a manner sensitive to the user’s or application’s needs” (Mani 2001).

OA.2 Types of Text Summarization

Summarization techniques can be classified into two types: summarization by abstraction and summarization by extraction.

OA.2.1 Summarization by Abstraction

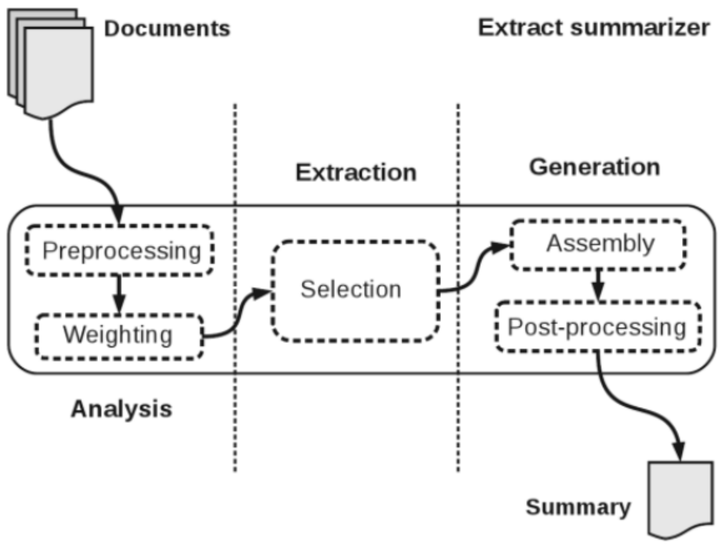
Based on semantic understanding, abstraction-based summaries convey the main information in the input, may reuse phrases or clauses from it, expressed in the words of the summarizer (Nenkova and McKeown 2011). In contrast to extraction-based summarization there has been limited research on summarization by abstraction, probably because abstraction-based summarization is beyond the capability of even state-of-the-art automatic summarization techniques. “Very few abstract summarization systems have been created (…). We are (…) a long way off achieving genuine automatic text understanding” (Torres-Moreno 2014, 35).

OA.2.2 Summarization by Extraction

The essence of extraction-based summarization is to select lexical units containing a document’s essential information (i.e., informative content), concatenated into an extractive summary, aiming to give an overview of the original text’s content. “Currently, extraction algorithms dominate the landscape and are at the center of countless automatic summarization systems. The ease with which these methods can be implemented and their good performance are the key to their success" (Torres-Moreno 2014, 271).

Figure OA.1 summarizes the summarization-by-extraction process (figure taken from Torres-Moreno 2014).

FIGURE OA.1
Summarization-by-Extraction Process



The basic idea is to first split a document into lexical units (i.e., sentences). After weighting those using statistical heuristics, the algorithm extracts the units with the highest scores, and assembles them to create a summary.

OA.3 Extraction-Based Summarization Algorithms

Below we provide some detail on the six algorithms we used to generate the automatic summaries in Study One. In each case, given a text, the summarization task consists in extracting sentences to be included in the summary such that they cover important information with minimal redundancy, while satisfying a length constraint. The algorithms differ in the statistical heuristics (see Figure OA.1 above) applied.

OA.3.1 Luhn

The Luhn algorithm—named after its creator, H.P. Luhn: IBM Research Center—collects the frequencies of words in the text and identifies a subset of significant words, excluding the most frequent and the least frequent. The algorithm, then, treats all significant words as having equal weight and computes the weight of a sentence as a function of the concentration of significant words in the sentence (Luhn 1958).

OA.3.2 SumBasic

In contrast to Luhn, the SumBasic algorithm relies only on word probability to calculate importance; it uses true initial probabilities and computes the weight of a sentence as equal to the average probability of the words in a sentence (Vanderwende, Suzuki, Brockett and Nenkova 2007). Specifically, for each sentence S_j in the input, the algorithm assigns a weight equal to the average probability $p(w_i)$ of the content words in the sentence, estimated from the input for summarization:

$$\text{Weight}(S_j) = \frac{\sum_{w_i \in S_j} p(w_i)}{|\{w_i | w_i \in S_j\}|}$$

Then SumBasic picks the best scoring sentence that contains the word that currently has the highest probability. This selection strategy assumes that at each point when a sentence is selected, a single word—that with highest probability—represents the most important topic in the document and the goal is to select the best sentence that covers this word. After the best sentence is selected, the probability of each word that appears in the chosen sentence is adjusted. It is set to a smaller value, equal to the square of the probability of the word at the beginning of the current selection step, to reflect the fact that the probability of a word occurring twice in a summary is lower than the probability of the word occurring only once. This selection loop is repeated until the desired summary length is achieved.

OA.3.3 TextRank, LexRank

In graph-based summarization research, TextRank (Mihalcea and Tarau 2004) and LexRank (Erkan and Radev 2004) are the most well-known and often cited. These methods model text as a graph with sentences as nodes and edges based on word overlap. A sentence node is then ranked according to its similarity with other nodes. Specifically, if a sentence S_i is represented as a set of words:

$$S_i = w_1^i, w_2^i, ...w_{|S_i|}^i$$

then the similarity between two sentences S_i and S_j is defined as:

$$Sim(S_i, S_j) = \frac{|\{w_k : w_k \in S_i \wedge w_k \in S_j\}|}{log(|S_i|) + log(|S_j|)}$$

An edge based on similarity can be seen as a process of "recommendation": a sentence that addresses certain concepts, gives the reader a "recommendation" to refer to other sentences that address the same concepts. The underlying assumption for calculating relevance is that the sentences which are similar to a large number of other important sentences are "central." Finally, PageRank (Brin and Page 1998) is used to calculate a relevance score for each sentence based on the relevance score of its similar sentences. Top ranked sentences are selected for the summary such that their total length satisfies the summary length constraint.

OA.3.4 Latent Semantic Analysis

At the heart of the Latent Semantic Analysis (LSA) approach is the representation of the input documents as a word by sentence matrix A: each row corresponds to a word that appears in the input and each column corresponds to a sentence in the input. Each entry a_{ij} of the matrix corresponds to the weight of word *i* in sentence *j*. If the sentence does not contain the word, the weight is zero, otherwise the weight is equal to the tf*idf weight of the word. Standard techniques for singular value decomposition (SVD) from linear algebra are applied to the matrix *A*, to represent it as the product of three matrices:

$$A = U\Sigma V^{\mathrm{T}}$$

The rows of V^T can be regarded as mutually independent topics discussed in the input, while each column represents a sentence from the document. In order to produce an extractive summary, the algorithm consecutively considers each row of V^T, and selects the sentence with the highest value, until the desired summary length is reached (Gong and Liu 2001; Steinberger and Jezek 2004).

OA.3.5 KLSum

The KLSum algorithm selects a set of sentences from the source document, *D*, such that the distribution of words in the selected sentences—i.e., the "candidate summary," *S*—is as close as possible to distribution of words in document *D*. Specifically, the algorithm introduces the following selection criterion:

$$\mathbf{S}^* = \min_{\mathbf{S}:words(\mathbf{S}) \leq L} KL(P_{\mathcal{D}} \| P_{\mathbf{S}})$$

where P_S (P_D) is the word (i.e., unigram) distribution of candidate summary *S* (document *D*). To measure similarity across the word distributions, P_S and P_D, the Kullback-Lieber (KL) divergence measure is used (Haghighi and Vanderwende 2009).

REFERENCES

Brin, S., and L. Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Seventh International World-Wide Web Conference,* April 14-18, 1998, Brisbane, Australia.

Erkan, G., and D. Radev. 2004. LexRank: Graph-Based Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence Research* 22: 1-23.

Gong Y., and X. Liu. 2001. Generic Text Summarization using Relevance Measure and Latent Semantic Analysis. *Proceedings of SIGIR:* 19-25.

Haghighi, A., and L. Vanderwende. 2009. Exploring Content Models for Multi-Document Summarization. *Proceedings of Human Language Technologies*: 362-370.

Luhn, H.P. 1958. The Automatic Creation of Literature Abstracts, *IBM Journal* April: 159-165.

Mani, I. 2001. *Automatic Summarization*, Natural Language Processing. Amsterdam, Philadelphia, John Benjamins Publishing Co.

Mani I. and M. Maybury. 1999. *Advances in Automatic Text Summarization.* Cambridge, MIT Press

Mihalcea, R., and P. Tarau. 2004. TextRank: Bringing Order Into Texts. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc30962/. Accessed December 19, 2016.

Nenkova, A. and K. McKeown. 2011. Automatic Summarization. *Foundations and Trends in Information Retrieval* 5(2-3): 103-233.

Radev, D., E. Hovy, and K. McKeown. 2002. Introduction to Special Issue on Summarization. *Computational Linguistics* 28(4): 399-408.

Steinberger, J., and K. Jezek. 2004. Using Latent Semantic Analysis in Text Summarization and Summary Evaluation. *Proceedings of ISIM*: 93-100

Torres-Moreno, J-M. 2014. *Automatic Text Summarization*. Hoboken, NJ: Wiley.

Vanderwende, L., H. Suzuki, C. Brockett, and A. Nenkova. 2007. Beyond SumBasic: Task-Focused Summarization with Sentence Simplification and Lexical Expansion. *Information Processing and Management*, Special Issue on Summarization, 43(6): 1606-1618.