

Galasso, Vincenzo; Nannicini, Tommaso; Nozza, Debora

**Working Paper**

## We Need to Talk: Audio Surveys and Information Extraction

CESifo Working Paper, No. 11530

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Galasso, Vincenzo; Nannicini, Tommaso; Nozza, Debora (2024) : We Need to Talk: Audio Surveys and Information Extraction, CESifo Working Paper, No. 11530, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/308426>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## We Need to Talk: Audio Surveys and Information Extraction

*Vincenzo Galasso, Tommaso Nannicini, Debora Nozza*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# We Need to Talk: Audio Surveys and Information Extraction

## Abstract

Understanding individuals' beliefs, preferences, and motivations is essential in social sciences. Recent technological advancements—notably, large language models (LLMs) for analyzing open-ended responses and the diffusion of voice messaging—have the potential to significantly enhance our ability to elicit these dimensions. This study investigates the differences between oral and written responses to open-ended survey questions. Through a series of randomized controlled trials across three surveys (focused on AI, public policy, and international relations), we assigned respondents to answer either by audio or text. Respondents who provided audio answers gave longer, though lexically simpler, responses compared to those who typed. By leveraging LLMs, we evaluated answer informativeness and found that oral responses differ in both quantity and quality, offering more information and containing more personal experiences than written responses. These findings suggest that oral responses to open-ended questions can capture richer, more personal insights, presenting a valuable method for understanding individual reasoning.

JEL-Codes: C830, D830.

Keywords: survey design, open-ended questions, Large Language Models, beliefs.

*Vincenzo Galasso\**  
*Department of Social and Political Sciences & IGIER*  
*Bocconi University, Milan / Italy*  
*vincenzo.galasso@unibocconi.it*

*Tommaso Nannicini*  
*European University Institute*  
*San Domenico di Fiesole (FI) / Italy*  
*tommaso.nannicini@eui.eu*

*Debora Nozza*  
*Department of Computing Sciences*  
*Bocconi University, Milan / Italy*  
*debora.nozza@unibocconi.it*

\*corresponding author

November 2024

We thank Adrià Rofes, Nicola Gennaioli, and seminar participants at Bocconi University and CREST Paris for their insightful comments. Alessandro Izzo, Rodolfo Melo, and Bianca Meoli provided excellent research assistance. We received approval from the Bocconi Ethical Review Committee (RA000708). The study is pre-registered at [osf.io/wt3gy](https://osf.io/wt3gy). Any remaining errors are our own.

# 1 Introduction

Understanding individuals’ beliefs, preferences, and motivations has long been a central objective in social sciences (Lau et al., 1991; Bénabou and Tirole, 2016; Bonomi et al., 2021). Despite a wide range of available methodologies, achieving this understanding remains highly challenging. Survey data and survey experiments that gather subjective responses are commonly used to elicit these dimensions, yet they have limitations. Surveys relying exclusively on closed-ended questions are well-suited for structured quantitative analysis but may introduce biases. Framing and priming effects (Ferrario and Stantcheva, 2022), social desirability (Haaland et al., 2024), and response order effects (DeCastellarnau, 2018) are among the cognitive biases that may prevent individuals from accurately reporting their beliefs. Open-ended questions, by contrast, can yield richer, more detailed information, often with a lower risk of biases and a greater capacity to capture how people reason (Stantcheva, 2021). Yet, open-ended questions also present challenges when translating the richness of responses into data for analysis (Haaland et al., 2024).

Two recent technological advancements may prove crucial in enhancing our ability to elicit individuals’ beliefs, preferences, and motivations. First, the use of AI—particularly large language models (LLMs)—in analyzing responses to open-ended, qualitative interviews enables a more quantitative, less subjective analysis compared to traditional methods (Stantcheva, 2021; Dechezlepretre et al., 2022; König and Schmacker, 2022; Liscow and Fox, 2022; Andre et al., 2024; Haaland et al., 2024). In this sense, AI is helping bridge the gap between qualitative and quantitative research. AI is also beginning to be used in conducting qualitative interviews (Chopra and Haaland, 2023). These developments have led to increased interest in open-ended questions, which are particularly effective for eliciting beliefs (Stantcheva, 2021; Jayachandran et al., 2023; Stantcheva, 2023; Andre et al., 2024; Binetti et al., 2024; Graeber et al., 2024), preferences (Zollinger, 2022; Elias et al., 2023; Bursztyn et al., 2024), understanding individuals’ reactions to treatments in experiments (Bergman et al., 2024), and generating new hypotheses to test (Bustos et al., 2022). AI is increasingly recognized for its ability to transform unstructured information from open-ended questions into quantitative data suitable for analysis.

The second technological advancement is the rising popularity of voice messaging. In 2023, a decade after the introduction of this feature on WhatsApp, users sent 7 billion voice messages through the platform. A YouGov survey conducted by Vox (2023) reveals that 62% of Americans have sent a voice message, with 30% using voice messages weekly, daily, or multiple times a day. Among 18- to 29-year-olds, 43% report using this feature at least weekly. Voice messages have already been used to respond to open-ended survey questions, and this usage is expected to grow significantly in the future (Vox, 2023).

In addition to these advancements, the use of smartphones for answering online surveys has significantly increased in recent years (Revilla and Ochoa, 2016; Höhne and Krebs, 2021). Smartphones are equipped with built-in microphones, making it easier to collect oral responses, whereas typing on a smartphone keyboard can be cumbersome. Moreover, research has found that, when responding in English, speaking is three times faster than typing for short responses (Ruan et al., 2017). As a result, surveys with oral responses to open-ended questions have already been used in social sciences (Graeber et al., 2024) and are likely to become very common in the near future. Furthermore, by simplifying the act of responding, this method may help recruit hard-to-reach populations (Shaghaghi et al., 2011; Ellard-Gray et al., 2015). For example, Beam (2023) has used chatbots as a tool for survey recruitment and administration in fieldwork in development economics.

In this paper, we study how oral answers to survey questions differ from written answers. Our aim is to analyze how using voice messages to respond to open-ended questions may improve our understanding of individuals’ beliefs, preferences, and motivations. We conducted a series of randomized control trials across three surveys on distinct topics—namely, artificial intelligence, public policies, and international relations—to compare 7,766 non-skipped responses to open-ended questions provided either by recording an audio message or by typing a text. Exogeneity is ensured by the fact that our 4,749 initial respondents were randomly assigned to one of the two answering methods.

Our initial set of results aligns with previous findings in the literature (Horowitz and Newman, 1964; Gavras et al., 2022; Höhne et al., 2024; Haaland et al., 2024). Respondents provide more words in oral responses than in written ones, though with a lower level of

lexical complexity. This trade-off between word count and lexical complexity raises the question of which response method yields more information from respondents. To address this question, we employ a ChatGPT-enhanced protocol to assess the informativeness of responses to all open-ended questions by calculating both an informativeness dummy and an informativeness score (Gilardi et al., 2023). These measures are designed to capture whether a response is articulated and addresses the complexity of the topic by explaining the reasons behind the stated choices or beliefs, and by providing examples.<sup>1</sup> Our findings based on these measures strongly suggest that oral responses to open-ended questions provide more information than written answers, with this result consistent across the full range of survey topics, question types, and question positions covered by our experiments. Building on this difference in informativeness, we proceed to analyze whether (and how) the content provided by respondents varies according to the response method.

From a cognitive neuropsychological perspective (Gould and Boies, 1978; Ellis, 1988; Cleland and Pickering, 2006), at least in the initial stage of conceptualization (Levitt, 1989), the message a respondent intends to convey in answering a question is not expected to differ by answering method. The two response methods—oral or written—are simply different codes, and the way of encoding the message should remain consistent across modalities. However, written answers require a slower process of typing on a PC or smartphone, which may give respondents time for second thoughts. They may re-read their answers, choosing to delete or rewrite parts. In contrast, oral responses offer less opportunity to revise, as respondents cannot view what they have recorded on a screen, reducing the temptation to alter it. Overall, modifying an oral answer is more cumbersome than modifying a written one. These differences may lead respondents using the oral modality to be more spontaneous and to express more directly the answer initially conceptualized, while written answers are more likely to be thoughtful and considered (Horowitz and Newman, 1964; Redeker, 1984; Chafe, 1985; Tannen, 1985). Consequently, we expect speech recordings to show greater personal involvement and potentially capture different

---

<sup>1</sup>Our definition differs from the one used in signaling theory, according to which informativeness refers to the degree to which a signal affects the likelihood ratio (Holmstrom, 1979). In our context, informativeness is the degree to which a response provides a rich quantity of relevant information and articulates reasoning.

emotions than written text. To test this hypothesis, we construct various lexical measures to evaluate the extent to which responses include references to personal experience. We also use a ChatGPT-enhanced protocol to measure the frequency of these references and the degree of subjectivity in responses to all open-ended questions (Nozza et al., 2017). We then calculate a dummy variable for personal content and a subjectivity index. Our empirical findings show that respondents provide more subjective answers and report more personal content in oral responses, with implications for the rapidly growing literature on the key role of personal experience in shaping values and beliefs (Kubin et al., 2021).

Our findings also suggest that attrition is higher among individuals assigned to record an audio message. Respondents in the audio condition were less likely to complete the survey than those providing a written answer, and they were more likely to skip questions by recording an empty audio (as opposed to typing meaningless characters). We do not overemphasize this difference, however, as it could be partially due to respondents on the commercial survey platform not yet being accustomed to answering open-ended questions by recording an audio message. As this method and the use of smartphones for surveys become more widespread, we expect this difference to diminish.

## 2 Experimental Design

To study the causal effects of the response method on various characteristics of answers to open-ended questions, in collaboration with a commercial survey company, we developed a platform allowing respondents to answer open-ended questions by recording an audio response from their PC or smartphone. We then ran a series of randomized control trials across three surveys on different topics. In each survey, the sample was representative of the adult Italian population in terms of gender, age structure, education, and geographic location, and participants were randomly assigned to respond to open-ended questions by either typing a text or recording an audio. The first survey, conducted in April 2023, included three open-ended questions on artificial intelligence (AI) and common applications, such as ChatGPT. Respondents were asked for their opinion on the spread of artificial intelligence, as well as its advantages and disadvantages. The initial sample of 1,650 invited



respondents resulted in a final sample of 1,176 individuals who completed the survey and 3,167 non-skipped answers to the three open-ended questions.

The second survey, conducted in July 2023, included three open-ended questions on public policies. Respondents were asked for their opinions on the reforms that have altered the Italian pension system over the last few decades, on a well-known retirement scheme introduced in Italy in 2018, and on the age at which people should be allowed to retire. The initial sample of 1,093 invited respondents resulted in a final sample of 827 individuals who completed the survey and 2,171 non-skipped answers to the three open-ended questions.<sup>2</sup>

The third survey, conducted in December 2023, included two open-ended questions on the Israeli-Palestinian conflict. First, respondents were asked whether they had ever felt discriminated against, abused, threatened, or under attack. Then, they were invited to propose ideas for promoting peace in the Israeli-Palestinian conflict. The initial sample of 2,006 invited respondents resulted in a final sample of 1,576 individuals who completed the survey and 2,428 non-skipped answers to the two open-ended questions.

Each survey also collected information on respondents’ characteristics, socio-economic situation, and political preferences. Additionally, each survey included some close-ended questions on its specific topic. All questions from the three surveys are provided in the Appendix. The commercial survey company invited respondents from its community to participate in each survey by email. Individuals were compensated for completing the survey according to the community’s rules and could choose to drop out at any time.

Oral responses to the eight open-ended questions were transcribed into text using OpenAI’s Whisper model.<sup>3</sup> Each answer was then manually checked for possible errors or typos in the transcription process (Ashwin et al., 2023). For both oral and written responses to the open-ended questions, we perform our analysis on the text of each answer. Since respondents were not given the option to skip the open-ended questions, some individuals found alternative ways to avoid answering. For example, in audio-recorded responses,

---

<sup>2</sup>In this survey, unlike the other two, respondents were randomly divided into three groups. Two groups were assigned to either speech recording or text writing, while a third group had the option to choose between the two response modalities. In the following analysis, to ensure comparability with the other surveys, we consider only respondents in the first two groups.

<sup>3</sup>See: <https://openai.com/index/whisper/>.

some respondents recorded an empty audio with only background noise, while in written responses, some respondents typed a few meaningless characters (e.g., “xxx” or “lvjh”). We coded these responses as “skipped” answers. In the analysis of main outcomes associated with responses to the open-ended questions (number of words, lexical structure, informativeness, personal experiences, and subjectivity), we focus on the final sample of 7,766 non-skipped answers, totaling 132,550 words. Specifically, in our main analysis, we pool together responses to all open-ended questions from the three randomized controlled trials and estimate the following OLS model:

$$y_{isq} = \alpha + \beta Audio_{isq} + X_i + \gamma_s + \delta_q + \epsilon_{isq}$$

where  $y_{isq}$  refers to our variables of interest,  $i$  represents the individual respondent,  $s$  the survey,  $q$  the open-ended question,  $X_i$  is a vector of individual-specific characteristics,  $\gamma_s$  and  $\delta_q$  are survey and question fixed effects, respectively, and  $\epsilon_{isq}$  is the error term.<sup>4</sup> Table A.1 in the Appendix presents the descriptive statistics for our main outcomes.

### 3 Empirical Results

Following the existing literature (Gavras et al., 2022; Haaland et al., 2024), we first test whether oral answers to open-ended questions contain more words than written answers. We count the number of words in each answer by converting the text to lowercase and removing any punctuation. Next, we split the text into individual words and count them using the word tokenize function in Python.<sup>5</sup> If the answer is empty or invalid, we return no value. Additionally, as a further indicator, we calculate the number of significant words per answer by first removing (Italian) “stop words” from each answer using the stop-word list in the Python NLTK library, and then counting the remaining significant words.

Table 1 presents the results of this analysis; the units of observation are the 7,766 non-skipped answers to all eight open-ended questions across the three surveys. Oral answers contain nearly fourteen more words than written answers, representing a substantial

---

<sup>4</sup>To corroborate both the internal and external validity of our empirical results, we also estimate separate OLS models in the Appendix, for each survey and each open-ended answer individually.

<sup>5</sup>Specifically, we used the word tokenize function from the NLTK Python library: [https://www.nltk.org/api/nltk.tokenize.word\\_tokenize.html](https://www.nltk.org/api/nltk.tokenize.word_tokenize.html).

increase given that the average number of words in all answers is seventeen. A large difference also appears in the number of significant words, with oral answers containing seven more significant words than written ones. Again, this represents a notable increase, as the average number of significant words in all open-ended answers is almost ten. Interestingly, the point estimates remain nearly unchanged when controlling for individual characteristics ( $X_i$ ) of respondents who provided non-skipped answers. As shown in Tables A.2 to A.4 in the Appendix, these differences are consistent and always statistically significant at the 1% level when the analysis is conducted separately for each survey. Figures A.1 and A.2 in the Appendix display the distribution of total and significant words in oral versus written answers. All distributions are positively skewed, with those of written answers being even more so. Specifically, many written answers are very short, whereas several oral answers have a high word count.

Next, we analyze the lexical structure of these answers. The existing literature suggests that the lexical structure in spoken responses is generally simpler than in written ones (Gavras et al., 2022). To assess lexical complexity, we use two standard measures (Yule, 1944; Templin, 1957): Yule’s K and the Type-Token Ratio (TTR). Yule’s K is a statistical measure that quantifies vocabulary richness and lexical diversity in a text, particularly useful for comparing lexical complexity across different texts or sections of the same text. This metric is based on word frequency, with higher values indicating lower lexical diversity. The TTR is another measure of lexical diversity, calculated by dividing the number of unique words (types) by the total number of words (tokens) in the text. Unlike Yule’s K, a higher value of TTR indicates a more diverse and complex vocabulary.

Table 2 presents the results on lexical complexity. Oral answers display higher values of Yule’s K and lower values of TTR compared to written answers.<sup>6</sup> These results indicate that oral answers are characterized by a lower level of lexical diversity and a less complex vocabulary. Once again, point estimates remain remarkably stable when controlling for respondent characteristics. Tables A.5 to A.7 in the Appendix show that these differences are consistent and statistically significant at the 1% level across all surveys.

---

<sup>6</sup>The Yule’s K is not defined for single-word answers. Consequently, in Table 2, the number of observations in the first and second columns is 27 fewer than in the third and fourth columns.

These findings suggest that oral answers to open-ended questions contain more significant words than written answers but are also characterized by more repetitions and a simpler lexical structure. Thus, these results alone do not allow us to conclude which response method provides more information. To address this, we employ an AI-enhanced protocol to analyze all open-ended responses from the three surveys, using ChatGPT (GPT-3.5-turbo-0125) to compute an informativeness score and an informativeness dummy for each answer. ChatGPT has been shown to outperform crowd workers and research assistants in text annotation tasks (Gilardi et al., 2023). For replicability, we set the temperature to zero. In our prompt, we instructed ChatGPT to act as “an independent judge in a competition to determine how informative a response is.” We further instructed ChatGPT to consider a response highly informative if it “explains its reasoning or provides examples,” is “articulated,” and “attempts to address the complexity of the topic at hand.” To ensure that evaluation is not based on text length, we specified: “The evaluation should not be tied to the number of words or phrases contained in the response, but rather to the quantity and complexity of the information.” For each response, we provided ChatGPT with the question and the answer, instructing it to “choose the correct option and explain the reasons for your choice.” For the dummy variable, ChatGPT was instructed to report *H* for highly informative responses and *L* for less informative ones, and to assign a score from 1 to 10 for the informativeness score. The Appendix provides the full prompt and examples of ChatGPT’s assessments, including explanations for the choice made.

Table 3 shows that, according to these measures, oral answers to open-ended questions are more informative than written answers. The percentage-point increase in the informativeness score for oral answers is 0.67 (significant at the 1% level), representing an increase of almost 20% over the average score (3.39). The difference captured by the informativeness dummy is even larger, with a marginal increase of 0.09 percentage points for oral answers (significant at the 1% level), representing a 57% increase over the average (0.16). Tables A.8 to A.10 in the Appendix show that these differences are consistent and statistically significant at the 1% level across all surveys.

Differences in the informativeness level of responses may also mask possible variations in

the type of information provided. The neuropsychology literature, as mentioned, suggests that the cognitive process of conceptualization occurring before answering a question does not vary with the response method—speaking or writing (Gould and Boies, 1978; Ellis, 1988; Cleland and Pickering, 2006). However, some distinctions may emerge during the act of typing or speaking (Pickering and Ferreira, 2008). Typing a response on a PC or smartphone may allow enough time for the respondent to reconsider and potentially revise their answer. Additionally, respondents can easily delete and rewrite parts of their response. In oral responses, once the record button is clicked, individuals begin recording and cannot pause or rewind. While respondents in our experiments could cancel and re-record their audio, this process is clearly more cumbersome than retyping a text, as the entire audio would need to be re-recorded rather than selectively edited. These differences may lead to different types of information being provided. Specifically, oral responses might be more spontaneous and introspective, aligning more closely with the initial answer formulated, whereas written responses might undergo additional cognitive filtering.

To test whether oral responses are more personal than written ones, we use four measures. First, we determine whether respondents use the pronoun “I” (in Italian, “Io”) in their answers. Unlike in some other languages, such as English, in Italian, the pronoun is not strictly required and is typically used to emphasize a personal perspective. Second, we check if respondents use verbs in the first-person singular, which similarly implies a personal viewpoint. Third, we use a ChatGPT-enhanced protocol (ChatGPT-4o-mini-2024-07-18) to assess whether the answer includes a reference to the respondent’s personal experience. In our prompt, we instructed ChatGPT to act as “an expert in text analysis” and to “determine whether the respondent’s answer refers to their own personal experience.” We also specified that ChatGPT should “ensure that its evaluation is fair and focused exclusively on the presence of personal experiences (e.g., I, me, my family).” For each response, we provided ChatGPT with the question and the answer and instructed it to report “True” or “False” for the presence of a personal experience, along with a brief explanation of the rationale for its choice. Fourth, following Nozza et al. (2017), we apply a ChatGPT-driven analysis (ChatGPT-4o-mini-2024-07-18) to assess the subjectivity of

an answer. A subjective answer expresses personal feelings, views, or beliefs, while an objective answer conveys factual information. In our prompt, we provided this definition of subjectivity and objectivity and asked ChatGPT to classify answers accordingly. For each response, we supplied ChatGPT with the question and the answer and asked it to report “0” for objective or “1” for subjective, along with a brief explanation of its choice. For both the dummies on personal experience and subjectivity, we set the temperature to zero to ensure replicability. The Appendix provides the full prompt and examples of ChatGPT’s assessments, including explanations for the choice made.

Table 4 reports the results of the analysis using these four measures. Respondents use the pronoun “I” (in Italian “Io”) much more frequently in oral than in written answers. The estimated coefficient of this difference, significant at the 1% level, is about 0.05, while the average usage of the pronoun is 0.03. Similarly, the use of verbs in the first-person singular is more common in oral than in written responses. This difference is substantial too: the estimated coefficient, also significant at the 1% level, is 0.13, while the average usage is 0.27. The analysis using the ChatGPT-generated dummy for personal experience indicates that respondents refer more to their personal experiences in oral than in written answers, with an estimated coefficient of 0.03 (significant at the 1% level) compared to an average usage of 0.08. Consistent with these results, the analysis using the subjectivity dummy suggests that oral answers are more subjective than written ones, with an estimated coefficient of 0.05 (significant at the 1% level) compared to an average subjectivity level of 0.83. Tables A.11 to A.13 in the Appendix show that these differences are consistent and statistically significant at the 1% level across all surveys.<sup>7</sup> On the whole, speech recordings provide more informative and more personal answers to open-ended survey questions.

---

<sup>7</sup>To further demonstrate the robustness of our results, for each set of indicators used above (number of words, lexical complexity, informativeness, and personal experience), we repeat the analysis by question in Appendix Tables A.14 to A.17. Furthermore, to complement the informativeness analysis, we apply both human-coded and AI-driven topic analysis models to responses from an open-ended question in the second survey, which was particularly suited for this analysis, as respondents were asked the age at which people should be allowed to retire. The findings confirm that written responses are less informative than oral ones. Specifically, written answers are significantly more likely to propose a single retirement rule (e.g., the same age or years of contribution for all workers), whereas oral responses are more likely to include complex answers that relate retirement age to workers’ occupation or socio-economic status. A detailed description of this analysis and its results is provided in the Appendix (see Table A.18).

These findings do not appear to be affected by the individual characteristics of survey respondents. Even if they were, however, this would not pose a problem, as our focus is on the treatment effect of using one answering modality over the other—a treatment effect that may, in turn, be a combination of cognitive and selection effects. Nonetheless, it is interesting to investigate how individual selection into surveys is influenced by the response method: oral versus written. Among the panelists invited from the community, 4,749 individuals started the survey, with their initial allocation to oral or written responses perfectly balanced on observable characteristics (see Table A.19 in the Appendix). However, the attrition rate varies by response method: 98% of respondents assigned to the written format complete the survey, compared to 65% of those assigned to the oral format. This differential attrition rate correlates with certain characteristics; for example, men, non-retirees, individuals with higher (self-declared) technological knowledge, and respondents using a smartphone were relatively more likely to complete the survey when assigned to an oral answer, though the resulting differences are only significant at the 10% level in the sample of 3,579 individuals who completed the surveys (see Table A.20). Additional selection effects may arise, as individuals may opt to skip one or more open-ended questions by recording an empty audio or by typing only a few meaningless characters instead of dropping from the survey. To account for this, Table A.21 repeats the balance tests on this sample of 2,946 individuals who provided non-skipped responses, again showing that men, non-retirees, individuals with higher technological knowledge, and those responding via smartphone are slightly overrepresented in the oral modality.

## 4 Discussion

Our empirical evidence suggests that recent technological advancements in the open-ended measurement of thoughts may help uncover the roots of individuals’ beliefs and motivations, while also enabling the collection of more detailed information on personal experiences. Speech recordings enable individuals to provide more informative answers than written texts. These oral responses tend to be longer, albeit with a simpler lexical structure, and our analysis suggests that their overall informational content is greater. Perhaps more

importantly, respondents are more likely to report personal experiences in oral answers, which more closely relate to their process of belief formation. Thus, the oral modality of responding to open-ended questions may serve as a more effective tool for eliciting individual beliefs.<sup>8</sup> This evidence speaks to emerging perspectives in social sciences.

In recent decades, behavioral scientists have increasingly emphasized the role of emotions and personal experiences in shaping preferences, beliefs, and ultimately behavior. While rationality would prioritize hard evidence and logic in forming beliefs—suggesting that individuals should favor facts, objective data, and statistics over subjective personal experiences (Stahl et al., 2016)—recent empirical evidence points otherwise (Graeber et al., 2024). In the political arena (Kubin et al., 2021), people tend to believe both facts and personal experiences in cases of non-moral disagreement; however, in cases of moral disagreement, subjective experiences are doubted less than objective facts, implying that personal experiences may have a greater impact on beliefs than facts. Similar findings emerge in economics: individuals may perceive past outcomes they have experienced as more likely to occur again, even when presented with information about the actual likelihood of these events. Knowledge of personal experiences has been shown to improve understanding of belief formation and decision-making in various economic areas, such as inflation, home purchases, mortgage choices, and consumption expenditures (Koudijs and Voth, 2016; Malmendier, 2021). The extent to which past experiences influence belief formation varies across individuals, who may assign different weights to these experiences—as opposed to objective data—in their learning process (Bordalo et al., 2020). Furthermore, the influence of experiences may be mediated by memory retrieval, which is itself affected by the salience of these experiences (Bordalo et al., 2012).

Our results also suggest promising avenues for future research on persuasion (Falk and Scholz, 2018; Kamenica, 2019; Druckman, 2022). Survey experiments could be designed to increase the salience of issues related to individuals’ personal experiences (Galasso, 2024). Oral responses to open-ended questions could be used to gather information on these experiences and, ultimately, on beliefs. Follow-up survey experiments using narratives related

---

<sup>8</sup>In addition, speech recordings may put respondents at ease, increasing satisfaction and enhancing the integrity and quality of survey research (Jamieson et al., 2023; Berenbaum and Thorp, 2024).



to personal experiences could then test the effectiveness of these narratives in shaping individuals’ (posterior) beliefs. Oral responses in follow-up surveys could also provide insights into the mental process of belief updating. Such analyses are particularly relevant for public policy, which often relies on information provision. We leave to future research the integration of speech emotion recognition into this perspective. Indeed, oral responses to open-ended questions may become even more valuable as technological advancements enable the evaluation of emotions, tone, and emphasis in voice recordings.

## References

- P. Andre, I. Haaland, C. Roth, M. Wiederholt, J. Wohlfart. Narratives About the Macroeconomy. *SAFE Working Paper*, (N. 426), 2024.
- J. Ashwin, A. Chhabra, V. Rao. Using Large Language Models for Qualitative Analysis Can Introduce Serious Bias. *World Bank Working Paper*, (N. 10597), 2023.
- E. Beam. Social Media as a Recruitment and Data Collection Tool: Experimental Evidence on the Relative Effectiveness of Web Surveys and Chatbots. *Journal of Development Economics*, 162:103069, 2023.
- R. Bénabou, J. Tirole. Mindful Economics: The Production, Consumption, and Value of Beliefs. *Journal of Economic Perspectives*, 30(3):141–164, 2016.
- M. R. Berenbaum, H.H. Thorp. Publishing survey research. *Proceedings of the National Academy of Sciences (PNAS)*, 121(18):e2406826121, 2024.
- P. Bergman, R. Chetty, S. DeLuca, N. Hendren, L.F. Katz, C. Palmer. Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice. *American Economic Review*, 114(5):1281–1337, 2024.
- A. Binetti, F. Nuzzi, S. Stantcheva. People’s Understanding of Inflation. *NBER Working Paper*, (N. 32497), 2024.

- G. Bonomi, N. Gennaioli, G. Tabellini. Identity, Beliefs, and Political Conflict. *Quarterly Journal of Economics*, 136(4):2371–2411, 2021.
- P. Bordalo, N. Gennaioli, A. Shleifer. Salience Theory of Choice under Risk. *Quarterly Journal of Economics*, 127(3):1243–1285, 2012.
- P. Bordalo, N. Gennaioli, A. Shleifer. Memory, Attention, and Choice. *Quarterly Journal of Economics*, 135(3):1399–1442, 2020.
- L. Bursztyn, B.R. Handel, R. Jimenez, C. Roth. When Product Markets Become Collective Traps: The Case of Social Media. *NBER Working Paper*, (N. 31771), 2024.
- S. Bustos, D. Pomeranz, J.C. Suárez Serrato, J. Vila–Belda, G. Zucman. The Race Between Tax Enforcement and Tax Planning: Evidence from a Natural Experiment in Chile. *NBER Working Paper*, (N. 30114), 2022.
- W. Chafe. Linguistic Differences Produced by Differences Between Speaking and Writing. In N. Torrance D. R. Olson and A. Hildyard, editors, *Literacy, Language, and Learning: The Nature and Consequences of Reading and Writing*, pages 105–123. Cambridge University Press, 1985.
- F. Chopra, I. Haaland. Conducting Qualitative Interviews with AI. *CESifo Working Paper*, (N. 10666), 2023.
- A.A. Cleland, M.J. Pickering. Do Writing and Speaking Employ the Same Syntactic Representations? *Journal of Memory and Language*, 54(2):185–198, 2006.
- A. DeCastellarnau. A Classification of Response Scale Characteristics That Affect Data Quality: A Literature Review. *Quality Quantity*, 52:1523–1559, 2018.
- A. Dechezleprêtre, A. Fabre, T. Kruse, B. Planterose, A. Sanchez Chico, S. Stantcheva. Fighting Climate Change: International Attitudes Toward Climate Policies. *NBER Working Paper*, (N. 30265), 2022.

- J.N. Druckman. A Framework for the Study of Persuasion. *Annual Review of Political Science*, 25(1):65–88, 2022.
- J. Elias, N. Lacetera, M. Macis. Is the Price Right? The Role of Morals, Ideology, and Tradeoff Thinking in Explaining Reactions to Price Surges. *CESifo Working Paper*, (N. 9712), 2023.
- A. Ellard-Gray, N.K. Jeffrey, M. Choubak, S.E. Crann. Finding the Hidden Participant: Solutions for Recruiting Hidden, Hard-to-Reach, and Vulnerable Populations. *International Journal of Qualitative Methods*, 14(5), 2015.
- A.W. Ellis. Modelling the Writing Process. In G. Denes, G. Semenza, P. Bisiacchi, editor, *Perspectives on Cognitive Neuropsychology*, page 24. Routledge, 1988.
- E. Falk, C. Scholz. Persuasion, Influence, and Value: Perspectives from Communication and Social Neuroscience. *Annual Review of Psychology*, 69(1):329–356, 2018.
- B. Ferrario, S. Stantcheva. Eliciting People’s First-Order Concerns: Text Analysis of Open-Ended Survey Questions. *AEA Papers and Proceedings*, 112(5):163–169, 2022.
- V. Galasso. The Role of Salience and Memory in Fertility Decisions: Experimental Evidence. *Population Research and Policy Review*, 43(51), 2024.
- K. Gavras, J. Karem Höhne, A.G. Blom, H. Schoen. Innovating the Collection of Open-Ended Answers: The Linguistic and Content Characteristics of Written and Oral Answers to Political Attitude Questions. *Journal of the Royal Statistical Society Series A*, 185(2):772–808, 2021.
- F. Gilardi, M. Alizadeh, M. Kubli. ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks. *Proceedings of the National Academy of Sciences (PNAS)*, 120(30): e2305016120, 2023.
- J.D. Gould, S.J. Boies. Writing, Dictating, and Speaking Letters. *Science*, 201:1145–1147, 1978.

- T. Graeber, C. Roth, F. Zimmermann. Stories, Statistics, and Memory. *Quarterly Journal of Economics*, 139(4):2181–2225, 2024.
- I. K. Haaland, C. Roth, S. Stantcheva, J. Wohlfart. Measuring What Is Top of Mind. *NBER Working Paper*, (N. 32421), 2024.
- J. K. Höhne, D. Krebs. Mismatching Middle Options: Consequences for Attitude Measurement in Smartphone Surveys. *International Journal of Social Research Methodology*, 24(3):381–386, 2021.
- J. K. Höhne, K. Gavras, J. Claassen. Typing or Speaking? Comparing Text and Voice Answers to Open Questions on Sensitive Topics in Smartphone Surveys. *Social Science Computer Review*, 42(4):1066–1085, 2024.
- B. R. Holmstrom. Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1):74–91, 1979.
- M. W. Horowitz, J. B. Newman. Spoken and Written Expression: An Experimental Analysis. *Journal of Abnormal and Social Psychology*, 68(6):640–647, 1964.
- S. Jayachandran, M. Biradavolu, J. Cooper. Using Machine Learning and Qualitative Interviews to Design a Five-Question Survey Module for Women’s Agency. *World Development*, 161:106076, 2023.
- K.H. Jamieson, A. Lupia, A. Amaya, H.E. Brady, R. Bautista, J. D. Clinton, J.A. Dever, D. Dutwin, D.L. Goroff, D.S. Hillygus, C. Kennedy, G. Langer, J.S. Lapinski, M. Link, T. Philpot, K. Prewitt, D.Rivers, L. Vavreck, D.C. Wilson, M.K. McNutt. Protecting the Integrity of Survey Research. *PNAS Nexus*, 2(3):1–10, 2023.
- E. Kamenica. Bayesian Persuasion and Information Design. *Annual Review of Economics*, 11(1):249–272, 2019.
- T. König, R. Schmacker. Preferences for Sin Taxes. *CESifo Working Paper*, (N. 10046), 2022.

- P. Koudijs, H.J. Voth. Leverage and Beliefs: Personal Experience and Risk-Taking in Margin Lending. *American Economic Review*, 106(11):3367–3400, 2016.
- E. Kubin, C. Puryear, C. Schein, K. Gray. Personal Experiences Bridge Moral and Political Divides Better than Facts. *Proceedings of the National Academy of Sciences (PNAS)*, 118(6):e2008389118, 2021.
- R.R. Lau, R.A. Smith, S.T. Fiske. Political Beliefs, Policy Interpretations, and Political Persuasion. *Journal of Politics*, 53(3):644–675, 1991.
- W.J.M. Levelt. Working Models of Perception: Five General Issues. In *Working Models of Perception*, pages 489–503. Academic Press, 1989.
- Z. Liscow, E. Fox. The Psychology of Taxing Capital Income: Evidence from a Survey Experiment on the Realization Rule. *Journal of Public Economics*, 213:104714, 2022.
- U. Malmendier. Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics. *Journal of the European Economic Association*, 2021.
- D. Nozza, E. Fersini, E. Messina. A Multi-View Sentiment Corpus. In A. Koller M. Lapata, P. Blunsom, editor, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 273–280, 2017.
- M.J. Pickering, V.S. Ferreira. Structural Priming: A Critical Review. *Psychological Bulletin*, 134(3):427, 2008.
- G. Redeker. On Differences between Spoken and Written Language. *Discourse Processes*, 7(1):43–55, 1984.
- M. Revilla, C. Ochoa. Open Narrative Questions in PC and Smartphones: Is the Device Playing a Role? *Quality & Quantity*, 50:2495–2513, 2016.
- S. Ruan, J.O. Wobbrock, K. Liou, A. Ng, J.A. Landay. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings*

- of the *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–23, 2017.
- A. Shaghghi, R.S. Bhopal, A. Sheikh. Approaches to Recruiting “Hard-To-Reach” Populations into Research: A Review of the Literature. *Health Promotion Perspectives*, 1(2):86–94, 2011.
- T. Stahl, M.P. Zaal, L.J. Skitka. Moralized Rationality: Relying on Logic and Evidence in the Formation and Evaluation of Belief Can Be Seen as a Moral Issue. *PLOS ONE*, 11(11):e0166332, 2016.
- S. Stantcheva. Understanding Tax Policy: How do People Reason? *Quarterly Journal of Economics*, 136(4):2309–2369, 2021.
- S. Stantcheva. How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible. *Annual Review of Economics*, 15:205–234, 2023.
- D. Tannen. Relative Focus on Involvement in Oral and Written Discourse. In N. Torrance D. R. Olson and A. Hildyard, editors, *Literacy, Language, and Learning: The Nature and Consequences of Reading and Writing*, pages 124–147. Cambridge University Press, 1985.
- M.C. Templin. *Certain Language Skills in Children: Their Development and Interrelationships*. University of Minnesota Press, Minneapolis, MN, USA, 1957.
- Vox. Why All Your Friends Are Sending You Voice Notes, 2023. Available at: <https://www.vox.com/technology/23665101/voice-message-whatsapp-apple-text>.
- G.U. Yule. *The Statistical Study of Literary Vocabulary*. Cambridge University Press, Cambridge, UK, 1944.
- D. Zollinger. Cleavage Identities in Voters’ Own Words: Harnessing Open-Ended Survey Responses. *American Journal of Political Science*, pages 1–48, 2022.

# Tables

Table 1. Number of Words (All Surveys)

	Words		Significant words	
Audio	13.880*** (0.609)	13.861*** (0.615)	7.366*** (0.335)	7.347*** (0.339)
Constant	11.020*** (0.190)	6.237*** (1.649)	6.505*** (0.107)	3.614*** (0.896)
$\mathbb{E}(y)$	17.068 (25.693)	17.068 (25.693)	9.715 (14.170)	9.715 (14.170)
R-squared	0.114	0.125	0.116	0.129
Obs	7,766	7,766	7,766	7,766
Question fixed effects	Yes	Yes	Yes	Yes
Respondent-level controls	No	Yes	No	Yes

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table 2. Lexical Complexity Metrics (All Surveys)

	Yule's K		TTR	
Audio	85.849*** (10.355)	84.347*** (10.194)	-0.057*** (0.002)	-0.057*** (0.002)
Constant	52.137*** (2.801)	78.276* (43.231)	0.970*** (0.001)	0.979*** (0.007)
$\mathbb{E}(y)$	91.384 (403.818)	91.384 (403.818)	0.945 (0.094)	0.945 (0.094)
R-squared	0.015	0.018	0.125	0.133
Obs	7,039	7,039	7,766	7,766
Question fixed effects	Yes	Yes	Yes	Yes
Respondent-level controls	No	Yes	No	Yes

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. One-word answers are excluded from the calculation of the Yule's K by construction. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table 3. Informativeness Metrics (All Surveys)

	Score		Dummy	
Audio	0.666*** (0.055)	0.645*** (0.054)	0.093*** (0.008)	0.090*** (0.008)
Constant	3.102*** (0.032)	2.141*** (0.161)	0.122*** (0.005)	0.000 (0.024)
$\mathbb{E}(y)$	3.392 (2.460)	3.392 (2.460)	0.162 (0.368)	0.162 (0.368)
R-squared	0.110	0.135	0.076	0.094
Obs	7,766	7,766	7,766	7,766
Question fixed effects	Yes	Yes	Yes	Yes
Respondent-level controls	No	Yes	No	Yes

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table 4. Personal Experience Metrics (All Surveys)

	1st Person (pronoun)		1st Person (verb)		Personal Experience		Subjectivity	
Audio	0.045*** (0.004)	0.046*** (0.004)	0.127*** (0.010)	0.130*** (0.010)	0.027*** (0.006)	0.028*** (0.006)	0.052*** (0.008)	0.053*** (0.008)
Constant	0.012*** (0.002)	0.008 (0.012)	0.217*** (0.006)	0.245*** (0.033)	0.070*** (0.004)	0.072*** (0.020)	0.803*** (0.006)	0.788*** (0.029)
$\mathbb{E}(y)$	0.032 (0.175)	0.032 (0.175)	0.272 (0.445)	0.272 (0.445)	0.081 (0.273)	0.081 (0.273)	0.825 (0.380)	0.825 (0.380)
R-squared	0.022	0.025	0.023	0.032	0.020	0.023	0.032	0.035
Obs	7,766	7,766	7,766	7,766	7,766	7,766	7,766	7,766
Question fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Respondent-level controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.



# Appendix

## A Surveys' Design

### A.1 Recruitment and Completion

In a collaboration with a commercial survey company, we developed a platform for respondents to answer open-ended questions by recording audio from their PC or smartphone. We then ran a randomized control trial in three surveys on different topics with respondents assigned to either the oral or the written response mode. In each survey, potential respondents were recruited by the commercial survey company, which sent emails to their community sample made of around 60 thousands individuals. The commercial survey company monitored the flow of respondents to ensure that the initial sample of individuals who started the survey was representative of the Italian adult population in terms of gender, age structure, macro-areas of residence and education. The balance tests on individual characteristics for the initial sample of respondents are reported in Table A.19. In Table A.19, and in the entire analysis, the individuals in survey 2 that could choose between text and audio were excluded. The sample of individuals who started the survey is made of 4,749 respondents, of whom 1,650 in survey 1, 1,093 in survey 2 (the individuals that had the choice between oral and written answers are not considered in this sample) and 2,006 in survey 3.

Not all individuals who started a survey completed it. Table A.20 reports the balance tests on individual characteristics for the sample of respondents that completed a survey. This sample is made of 3,579 individuals (out of the 4,749 who started the survey), of whom 1,176 in survey 1, 827 in survey 2 and 1,576 in survey 3. This is the sample used in our analysis. The descriptive statistics are reported in Table A.1. These balance tests in Table A.19 suggest that in the final sample of respondents there are slightly more men, more high tech individuals, fewer retired people and more individuals answering from a smartphone in the oral than in the written answers.

Finally, individuals who completed the survey could decide to skip single open-ended questions by recording an empty audio or by typing meaningless characters (such as "xxx").

In Table A.21, we report the balance tests on individual characteristics for the sample of respondents, who did not skip any questions in a survey. This sample is made of 2,946 individuals (out of the 4,749 who started the survey and of the 3,579 who completed it), of whom 1,017 in survey 1, 699 in survey 2 and 1,230 in survey 3. These balance tests are in line with those in Table A.20 on the sample of individuals who completed the survey.

## A.2 Closed-Ended Questions

We run a randomized control trial in three surveys on different topics. Each survey contained the following close-ended questions on demographic, socio-economic and political background:

- Your gender is Male/Female/Else
- Your date of birth is....
- Which municipality do you live in? What is your ZIP code?
- What is your level of education? [Close-ended question with the following possible answers: no schooling/ high school dropout/ high school degree/ some college/ college graduate/ master’s degree or more]<sup>9</sup>
- What is your main occupational status? [Close-ended question with the following possible answers permanently employed/ occasionally employed/ looking for your first job/ unemployed/ retired/ student/ housewife/ not working nor studying/ else]
- Where would you place yourself on a political scale [Close-ended question with the following possible answers: right/ center-right/ center/ center-left/ left/ nowhere/ prefer not to answer]
- In the national elections held on close-ended September 25, 2022 for which of the following parties did you vote? [Close-ended question with the following possible answers: Azione-Italia Viva/ Fratelli d’Italia/ Lega/ Forza Italia/ Noi Moderati/ Par-

---

<sup>9</sup>This question was not asked in the first survey.

tito Democratico/ Alleanza Verdi-Sinistra/ +Europa/ Impegno Civico-Centro Democratico/ Movimento 5 Stelle/ Italexit per l'Italia/ Unione Popolare/ Italia Sovrana e Popolare/ Sud chiama Nord/ Vita/ SVP - PATT/ Another party/ I did not vote/ I returned a blank ballot/ Prefer not to answer]

- If the national elections were to take place today, for which of the following parties would you be more likely to vote? [Closed question with the following possible answers: Per l'Italia con Paragone/ +Europa/ Unione Popolare di de Magistris/ Movimento 5 stelle/ Fratelli d'Italia/ Azione-Italia Viva/ Lega/ Forza Italia/ Noi Moderati/ Partito Democratico/ Alleanza Verdi-Sinistra/ Another party/ I would not vote/ I would return a blank ballot/ Prefer not to answer]
- How do you relate to technology in general and to digital technology? [Close-ended question with the following possible answers: I am passionate about it/ I am good at it/ I am an average user/ I am able to tolerate technology/ if possible I avoid technology]

The following close-ended question was asked only in the first survey:

- Did you ever use ChatGPT or other applications based on AI? [Close-ended question with the following possible answers: Yes, regularly/ Yes, until ChatGPT was blocked/ No, I never used them but I heard about them/ No, I never used them and I do not know what they are for]

The following close-ended question was asked only in the second survey:

- According to the Quota 100 pension policy, introduced in 2019 for three years, what are the minimum age and the minimum number of years of contributions needed to retire? [Closed question with two counters that individuals could use to indicate age and number of years of contributions]

The following close-ended question was asked only in the third survey:

- How much do you think it is right to support Ukraine in the war against Russia? [Close-ended question with the following answers: We should never have supported

it; It was right to do so, but now we have done enough; We should support them a little longer until a peace is reached; We should support them indefinitely; I do not know.]

- In the First Republic, which of these parties did you or would you have felt closest to? [Close-ended question with the following answers: Christian Democracy (DC); Italian Communist Party (PCI); Italian Socialist Party - Italian Democratic Socialist Party (PSI-PSDI); Italian Social Movement (MSI); Italian Republican Party - Italian Liberal Party (PRI-PLI); Other parties; I wouldn't know]
- How concerned are you that the current international situation could negatively impact your and your family's standard of living? [Answers from 1 (not at all) to 10 (a lot)]
- In your opinion, how many people out of 10,000 in Italy have red hair?
- Let's now talk about the longstanding conflict between Israelis and Hamas, which has further intensified since October 7, when Hamas launched a surprise multi-front attack on Israel, followed by the Israeli army's response. In your opinion, which of the two factions is more responsible for the escalation of the conflict? [Close-ended question with the following answers: Israelis; Hamas; Both factions are equally responsible; Neither of them; I don't know]
- How confident do you feel about your answer to the previous question? [Answers from 1 (not confident at all) to 10 (extremely confident)]
- Personally, at this moment, do you feel closer to... [Close-ended question with the following answers: the Israeli community; the Palestinian community; to both communities; to neither one; I do not know]
- How confident do you feel about your answer to the previous question? [Answers from 1 (not confident at all) to 10 (extremely confident)]

- We ask if personally... 1) Have you ever been subject to discrimination or abuse due to economic or social reasons by someone stronger or more powerful than you? 2) Have you ever felt that yourself, your family, your property, or people close to you were threatened or under attack by other people or groups? 3) Have you ever experienced highly conflicting situations that were out of your control? 4) Have you ever experienced the death of a friend or family member due to an accident or sudden illness? 5) Have you ever personally experienced a dramatic event such as an earthquake or a flood? 6) Have you ever experienced a situation where you feared for your life? [Possible answers Yes or No]
- Do you think these experiences are similar to what the Palestinian community is going through? 1) Being subject to discrimination or abuse due to economic or social reasons by someone stronger or more powerful than you? 2) Feeling that yourself, your family, your property, or people close to you are threatened or under attack by other people or groups? 3) Experiencing highly conflicting situations that were out of your control? 4) Experiencing the death of a friend or family member due to an accident or sudden illness? 5) Personally experiencing a dramatic event such as an earthquake or a flood? 6) Experiencing a situation where you feared for your life? [Answers from 1 (not at all) to 10 (very much)]
- Do you think these experiences are similar to what the Israeli community is going through? 1) Being subject to discrimination or abuse due to economic or social reasons by someone stronger or more powerful than you? 2) Feeling that yourself, your family, your property, or people close to you are threatened or under attack by other people or groups? 3) Experiencing highly conflicting situations that were out of your control? 4) Experiencing the death of a friend or family member due to an accident or sudden illness? 5) Personally experiencing a dramatic event such as an earthquake or a flood? 6) Experiencing a situation where you feared for your life? [Answers from 1 (not at all) to 10 (very much)]
- Could you tell us if... 1) You have donated or would donate in favor of the Palestinian

community? 2) You have donated or would donate in favor of the Israeli community? 3) You have signed or would sign a petition in favor of the Palestinian community? 4) You have signed or would sign a petition in favor of the Israeli community? 5) You have boycotted or would boycott products, universities, or initiatives of the Palestinian community? 6) You have boycotted or would boycott products, universities, or initiatives of the Israeli community? [Answers: Yes or No]

### A.3 Open-Ended Questions

The first survey, run in April 2023, contained three open-ended questions on artificial intelligence (AI) and common Apps, such as ChatGPT. Here is the text of the three questions.

*Question 1:* Artificial intelligence and other technological innovations are spreading rapidly, also thanks to new apps like ChatGPT. What do you think about this phenomenon?

*Question 2:* In your opinion, what opportunities are linked to the spread of artificial intelligence? What will be the advantages, and who will benefit the most?

*Question 3:* In your opinion, what are the risks associated with the spread of artificial intelligence? What will be the disadvantages, and who will bear the highest cost?

The second survey, run in July 2023, had three open-ended questions on public policies. Respondents were asked their opinions about the reforms that modified the Italian pension system in the last few decades, about a well-known retirement scheme introduced in Italy in 2018 and about the age at which people should be allowed to retire. Here is the text of the three questions.

*Question 1:* Over the past three decades, the retirement age in Italy has been modified many times. What do you think about these reforms and generally about the pension system?

*Question 2:* What do you think about the Quota 100 policy?

*Question 3:* According to you, what should be the age at which people should be allowed to retire? Alternatively, or in addition, what should be the number of years of

contributions, upon reaching which people should be allowed to retire regardless of their age? Give a number and explain why it seems to you to be the right age (or the right number of years of contributions).

The third survey, run in December 2023, had two open-ended questions on the Israeli-Palestine conflict. Initially, respondents were asked whether they ever felt discriminated, abused, threatened or under attack. Then, they were asked to put forward proposals to promote peace in the Israeli-Palestinian conflict. Here is the text of the two questions.

*Question 1:* Have you ever felt subject to discrimination or abuse by someone stronger or more powerful than you? OR Have you ever felt as if: yourself, your family, your belongings, or people close to you were threatened or under attack by other individuals or groups?

*Question 2:* There are many proposals to promote peace in the Israeli-Palestinian conflict, what do you think about them? Which proposals would you put forward?

## B Outcome Measures

### B.1 Words

We consider the number of words and of significant words, namely excluding stop-words. To count the number of words in each answer, we convert the text to lowercase and remove any punctuation, splitting the text into individual words and simply count them using the Word Tokenize function in Python.<sup>10</sup> If the answer is empty or invalid, we return no value. We calculated the number of significant words per answer by removing from each answer the (Italian) stop-words contained in the list provided by python NLTK library ( `nltk.corpus.stopwords.words('italian')` ) and counting the remaining (significant) words.

### B.2 Lexical Complexity

To analyze the lexical complexity of the text, we use two off-the-shelf measures: Yule's K and Type Token Ratio (TTR) (Yule, 1944; Templin, 1957).

Yule's K is a statistical measure used in text analysis to quantify the vocabulary richness and lexical diversity of a given text. It is particularly useful in comparing the lexical complexity of different texts or different sections of a same text. The metric is based on the frequency of word occurrences. Higher values of Yule's K indicate a lower level of lexical diversity.

Its mathematical expression,  $K$  can be obtained as follows. Let  $N$  be the total number of words in a text,  $V(N)$  be the number of distinct words,  $V(m, N)$  be the number of words appearing  $m$  times in the text, and  $mmax$  be the largest frequency of a word. Yule's K is then defined as follows, through the first and second moments of the vocabulary population distribution of  $V(m, N)$ :

$$K = 10,000[(S_2 - S_1)/S_1^2]$$

where  $S_1 = N = \sum_m V(m, N)$  and  $S_2 = N = \sum_m m^2 V(m, N)$

The Type-Token Ratio (TTR) is a linguistic metric that measures lexical diversity within a given text. It is calculated by dividing the number of different words (types) by

---

<sup>10</sup>We used the `word_tokenize` function from NLTK python library ([https://www.nltk.org/api/nltk.tokenize.word\\_tokenize.html](https://www.nltk.org/api/nltk.tokenize.word_tokenize.html))



the total number of words (tokens) in the text. A higher TTR indicates a greater diversity of words used, implying a more complex or varied vocabulary.

### B.3 Informativeness

To evaluate the informativeness of the answers to the open-ended questions, we used unsupervised AI for all the answers to open-ended questions in the three surveys by asking ChatGPT (GPT-3.5-turbo-0125) to compute an information score and an information dummy for each answer. For replicability purposes, we set the temperature to zero.

Below is the **prompt** used to instruct ChatGPT.

You are an independent judge in a competition to determine how informative a response is. You will be presented with responses to a survey question and you must determine whether it is informative or not informative. You only need to respond with the output in JSON format. There are three keys: The first key is *detailed reasons*, which explains the reasons for the decision in string format. The second key is *informative dummy*, which can be *H* or *L*. The third key is *informative score*, which is an integer from 1 to 10. We would like assistance in analyzing the responses given to a survey that was administered to about a thousand people. The goal is to gather individual opinions about [description of the survey]. This question asks respondents [description of the question] We will provide you with the responses. Specifically, you should return a JSON file with three keys: *detailed reasons*, *informative dummy* and *informative score*. *detailed reasons* should contain a string explaining the reasons for your decision. *informative dummy* can be H or L. *informative score* can be any integer from 1 to 10. We consider a question to be very informative if it explains its reasons or provides examples. A very informative answer is articulate and attempts to address the complexity of the topic at hand. The evaluation we ask of you is not tied to the number of words or phrases contained in the response, but rather to the quantity and complexity of the information contained. Below is the response for analysis: Question - [text of the question. in original language (italian)] Answer - [text of the answer. in original language (italian)] Choose the correct option and explain the reasons for your choice. For informative dummy: H) The response is very informative. L)

The response is not informative. For informative score: Assign a value from 1 to 10 to each response and explain your decision.

Below we provide three examples of the question and answer fed into ChatGPT and of its evaluation of informativeness.

*Example from Survey 1*

Question 1: Artificial intelligence and other technological innovations are spreading rapidly, also thanks to new apps like ChatGPT. What do you think about this phenomenon?

Answer: AI is too underestimated and represents an extremely dangerous phenomenon.

ChatGPT Assessment: The response provided a brief opinion without elaborating on the reasons behind considering the rapid diffusion of AI as underestimated and extremely dangerous. It lacks depth, examples, or further explanation to support the viewpoint. Informativeness. Dummy: 0. Score: 3.

*Example from Survey 2*

Question 3: According to you, what should be the age at which people should be allowed to retire? Alternatively, or in addition, what should be the number of years of contributions, upon reaching which people should be allowed to retire regardless of their age? Give a number and explain why it seems to you to be the right age (or the right number of years of contributions).

Answer: retirement ages should be differentiated based on the type of work: heavy manual labor jobs like masonry and construction should retire at variable ages whereas office workers could retire at 70 years.

ChatGPT Assessment: The response provides a specific suggestion on how to differentiate retirement ages based on the type of work, giving examples of heavy manual labor jobs like masonry and construction versus office work. This shows an attempt to address the complexity of the pension system and reforms by considering different job requirements. Informativeness. Dummy: 1. Score: 8.

*Example from Survey 3*

Question 2: There are many proposals to promote peace in the Israeli-Palestinian conflict, what do you think about them? Which proposals would you put forward?

Answer: eliminating terrorist factions from both sides, achieving a two-state solution, and removing Jewish settlements

ChatGPT Assessment: The response provides specific proposals to promote peace in the Israeli-Palestinian conflict, including eliminating terrorist factions from both sides, achieving a two-state solution, and removing Jewish settlements. These proposals address key issues in the conflict and demonstrate an understanding of the complexity of the situation. Informativeness. Dummy: 1. Score: 9.

## B.4 Personal Experience

To test whether oral responses are more personal than written ones, we use four measures. First, we determine whether respondents use the pronoun "I" (in Italian "io") in their answers. Second, we determine whether respondents use the verb in its first singular person. Third, we used unsupervised AI by asking ChatGPT (ChatGPT-4o-mini-2024-07-18) to assess whether the answer includes a reference to a personal experience of the respondent. Fourth, we used unsupervised AI by asking ChatGPT (ChatGPT-4o-mini-2024-07-18) to assess whether the answer is subjective or objective. In both cases, we set the temperature to zero.

To calculate our dummy for personal experience, we used the following prompt to instruct ChatGPT.

You are an expert in analyzing texts. Given the question [text of the question. in original language (Italian)] and the answer [text of the answer. in original language (Italian)], please determine whether the respondent is referring to a personal experience. Your output should be a JSON file, with "personal experience" = True or False and justification: "a brief explanation for the classification" Please make sure that your assessment is impartial and concentrates exclusively on the existence of personal experiences (for instance, "I", "me", "my family").

To calculate our dummy for subjective answer, we used the following prompt to instruct ChatGPT.

I need your help to classify an answer based on its subjectiveness or objectiveness. Subjectiveness refers to answers expressing opinions, personal feeling and beliefs. Objectiveness refers to answers presenting factual information. ADD QUESTION The answer to be analyzed is [text of the answer. in original language (Italian)] Your output should be a JSON file, that includes the classification and an explanation of your choice. Your output should be subjectivity: 1 or 0, explanation: "explanation for the classification."

Below we provide three examples of the question and answer fed into ChatGPT and of its evaluation of personal experience and of subjectiveness.

*Example from Survey 1*

Question 3: In your opinion, what are the risks associated with the spread of artificial intelligence? What will be the disadvantages, and who will bear the highest cost?

Answer: People with more limited digital skills or digital literacy.

ChatGPT Assessment on *Personal Experience*: The answer refers to a group of individuals, 'People with more limited digital skills or digital literacy,' without mentioning any direct personal experience of the respondent. . Personal Experience Dummy = 0

ChatGPT Assessment on *Subjectivity*: The sentence 'People with more limited digital skills or digital literacy.' presents factual information about a group of people and their digital competence. It does not express personal opinions or feelings, but simply describes an observable condition. Subjectivity Dummy = 0

*Example from Survey 2*

Question 2: What do you think about the Quota 100 policy?

Answer: A reform designed solely and exclusively for electoral purposes, which resulted in the expenditure of substantial public financial resources. In this case as well, it is a short-term, non-structural policy, without considering the long-term effects and focusing only on narrow groups of workers.

ChatGPT Assessment on *Personal Experience*: The response focuses on a critical as-

assessment of the Quota 100 policy, with no reference to any direct or indirect personal experience of the respondent. Personal Experience Dummy = 1

ChatGPT Assessment on *Subjectivity*: The response expresses opinions and personal evaluations regarding a reform, describing it as "designed solely and exclusively for electoral purposes" and criticizing its "short-term, non-structural" nature. These statements reflect a subjective judgment rather than factual information. Subjectivity Dummy = 1

### *Example from Survey 3*

Question 1: Have you ever felt subject to discrimination or abuse by someone stronger or more powerful than you? OR Have you ever felt as if: yourself, your family, your belongings, or people close to you were threatened or under attack by other individuals or groups?

Answer: When I was in middle school, I was bullied for being the smallest in the class and also for being sensitive. For several years, I felt very, very uncomfortable.

ChatGPT Assessment on *Personal Experience*: The respondent describes a personal experience of bullying during middle school, using personal pronouns like 'I was' and 'I felt,' clearly indicating that they are talking about a direct experience. Personal Experience Dummy = 1

ChatGPT Assessment on *Subjectivity*: The response expresses personal experiences and feelings related to the bullying endured during middle school. The use of phrases like 'I felt very, very uncomfortable for several years' clearly indicates an emotional reaction and a personal reflection, making the response subjective. Subjectivity Dummy = 1

## **B.5 Addition Measure on Information (Second Survey)**

The third question in the (second) survey on public policies – reported below – allows for an alternative analysis of the informativeness of the response.

Open Question (Pension Survey, Q3): According to you, what should be the age at which people should be allowed to retire? Alternatively, or in addition, what should be the number of years of contributions, upon reaching which people should be allowed to retire regardless of their age? Give a number and explain why it seems to you to be the right

age (or the right number of years of contributions).

We used human coding to define different informational categories. First, we consider whether the answer provides a "Single Retirement Rule", namely whether in the answer the same age or number of years of contributions is used for all workers (regardless of their individual characteristics). We then construct a Single Retirement Rule taking value 1 if the answer provides a single retirement rule and 0 otherwise. Second, we consider the alternative case, in which different retirement rules are provided. We categorized three different topics, according to which different rules can be provided: (i) Job Categories, if the answer suggests that retirement rules should be different based on types of jobs or on socio-economic categories; (ii) Flexibility, if the answer suggests that retirement rules should embed some degree of flexibility for the timing of retirement; and (iii) Turnover Narrative, if the answer suggests that lower minimum retirement age fosters the hiring of young workers. For each topic, we construct a dummy variable taking value 1 if the topic applies to the answer and 0 otherwise. Finally, we construct a "Complexity Score" variable by summing the three topic dummies, to account for the possibility that an answer discusses more topics.

Once these five variables are defined, we proceed to classify the answers in three different ways. First, we asked two research assistants to classify each answer according to the five variables. Second, we feed the categories in the Open AI propt and ask ChatGPT to classify the answers. We report the results using GPT 3.5 turbo and GPT 4 turbo.

Table A.18 reports the coefficients of the t-test of the difference between oral and written responses to the third open-ended question in the second survey for these five variables (Single Retirement Rule, Job Categories, Flexibility, Turnover Narrative, Complexity Score), as classified by Human Coding, Chat GPT 3.5 turbo and Chat GPT 4 turbo.

According to all three classification methods, oral answers are much less likely to provide a single retirement rule and more likely to mention some topics. In particular, the role of job categories in retirement rules are more likely to be discussed in the oral answers. These results confirm our previous findings suggesting that oral answers are more informative than written answers.

## C Additional Empirical Results

Table A.1. Descriptive Statistics

	Obs	Mean	S.D.	Min	Max
<b>Panel A: Outcomes</b>					
Words	7766	17.068	25.693	1	411
Significant words	7766	9.715	14.170	0	224
Yule's K	7039	91.384	403.818	0	8000
TTR	7766	0.945	0.094	0.027	1
Information Score	7766	3.392	2.460	1	9
Information Dummy	7766	0.162	0.368	0	1
1st Person (pronoun)	7766	0.032	0.175	0	1
1st Person (Verb)	7766	0.272	0.445	0	1
Personal Experience	7766	0.081	0.273	0	1
Subjectivity	7766	0.825	0.380	0	1
<b>Panel B: Respondent-Level Controls</b>					
Male	3579	0.515	0.500	0	1
High Education	3579	0.373	0.484	0	1
Babyboomers	3579	0.373	0.484	0	1
GenX	3579	0.319	0.466	0	1
GenZ	3579	0.058	0.234	0	1
Millennials	3579	0.205	0.404	0	1
Retired	3579	0.227	0.419	0	1
Employed	3579	0.566	0.496	0	1
Main cities	3579	0.486	0.500	0	1
Right	3579	0.265	0.441	0	1
Left	3579	0.421	0.494	0	1
Other parties	3579	0.232	0.422	0	1
Hightec	3579	0.606	0.489	0	1
Mobile phone	3579	0.428	0.495	0	1

Notes. One-word answers have been excluded from the Yule's K measure, as they do not provide enough data for the measure to function properly. Only individuals who completed the survey are selected into the sample.

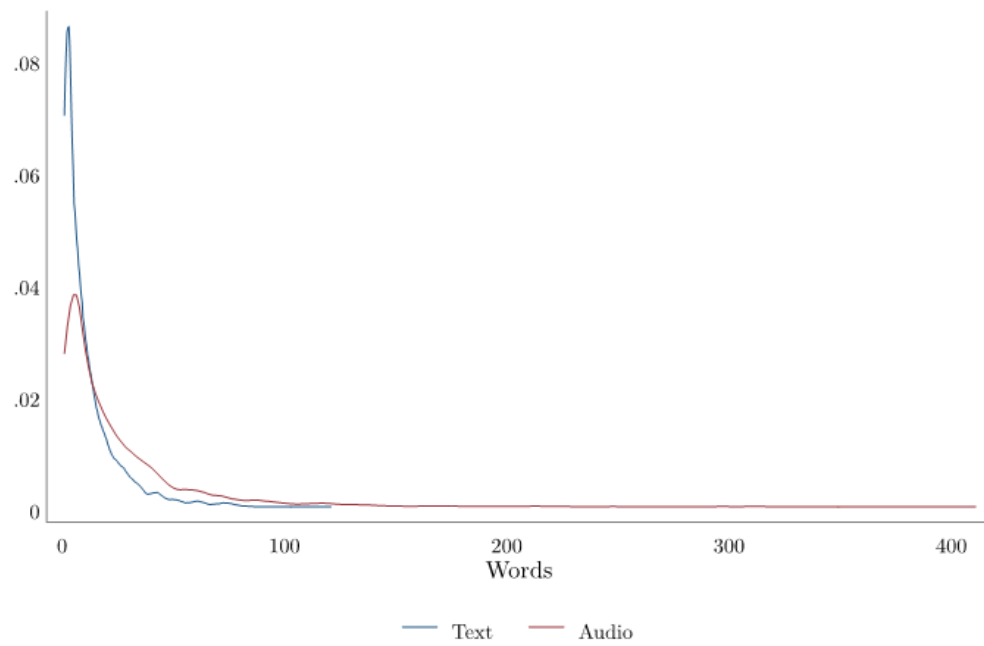


Figure A.1. Number of Words

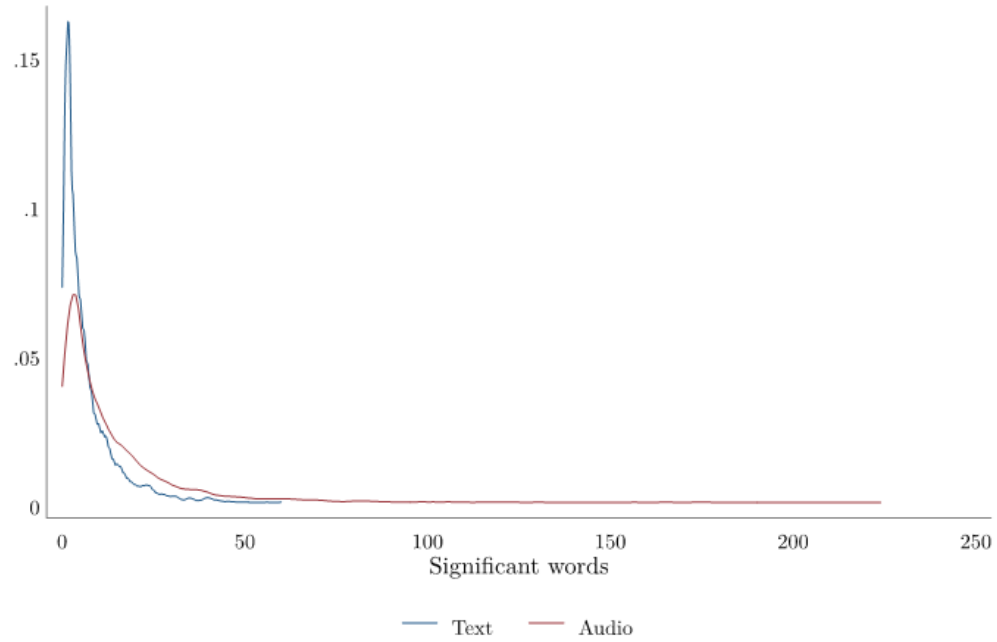


Figure A.2. Number of Significant Words



Table A.2. Number of Words (Survey 1)

	Words		Significant words	
Audio	15.335*** (0.921)	15.271*** (0.939)	8.246*** (0.521)	8.190*** (0.531)
Constant	10.125*** (0.243)	2.958* (1.655)	5.909*** (0.140)	1.846** (0.933)
$\mathbb{E}(y)$	16.609 (24.026)	16.609 (24.026)	9.396 (13.534)	9.396 (13.534)
R-squared	0.114	0.151	0.104	0.144
Obs	3,167	3,167	3,167	3,167
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.3. Number of Words (Survey 2)

	Words		Significant words	
Audio	16.168*** (1.301)	16.136*** (1.336)	8.653*** (0.725)	8.632*** (0.744)
Constant	15.301*** (0.455)	8.493** (3.614)	9.140*** (0.261)	5.107** (2.013)
$\mathbb{E}(y)$	22.666 (29.985)	22.666 (29.985)	13.082 (16.745)	13.082 (16.745)
R-squared	0.107	0.133	0.108	0.136
Obs	2,171	2,171	2,171	2,171
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals who could choose between text and audio were excluded from the analysis. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.4. Number of Words (Survey 3)

	Words		Significant words	
Audio	9.929*** (0.989)	9.847*** (0.959)	5.065*** (0.509)	5.032*** (0.500)
Constant	8.343*** (0.318)	10.557** (4.122)	4.919*** (0.174)	5.860*** (2.223)
$\mathbb{E}(y)$	12.662 (22.523)	12.662 (22.523)	7.121 (11.649)	7.121 (11.649)
R-squared	0.052	0.061	0.055	0.066
Obs	2,428	2,428	2,428	2,428
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Randomization for the two open questions were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.5. Lexical Complexity Metrics (Survey 1)

	Yule's K		TTR	
Audio	99.743*** (17.651)	97.900*** (17.278)	-0.061*** (0.003)	-0.060*** (0.003)
Constant	37.596*** (2.905)	111.500 (73.307)	0.976*** (0.001)	0.984*** (0.010)
$\mathbb{E}(y)$	80.931 (423.282)	80.931 (423.282)	0.950 (0.088)	0.950 (0.088)
R-squared	0.017	0.023	0.122	0.142
Obs	3,006	3,006	3,167	3,167
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). One word answers have been excluded from Yule's K, as they do not provide enough data for the measure to function properly. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.6. Lexical Complexity Metrics (Survey 2)

	Yule's K		TTR	
Audio	31.029** (14.730)	27.575** (13.289)	-0.059*** (0.004)	-0.058*** (0.004)
Constant	77.242*** (7.190)	58.005 (40.011)	0.954*** (0.002)	0.981*** (0.013)
$\mathbb{E}(y)$	92.014 (321.757)	92.014 (321.757)	0.927 (0.101)	0.927 (0.101)
R-squared	0.011	0.017	0.131	0.147
Obs	2,008	2,008	2,171	2,171
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals who could choose between text and audio were excluded from the analysis. One word answers have been excluded from Yules'K, as they do not provide enough data for the measure to function properly. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.7. Lexical Complexity Metrics (Survey 3)

	Yule's K		TTR	
Audio	120.079*** (19.945)	120.353*** (20.497)	-0.052*** (0.004)	-0.051*** (0.004)
Constant	49.588*** (5.029)	48.551 (102.911)	0.976*** (0.001)	0.968*** (0.015)
$\mathbb{E}(y)$	106.277 (444.955)	106.277 (444.955)	0.954 (0.093)	0.954 (0.093)
R-squared	0.018	0.025	0.084	0.096
Obs	2,025	2,025	2,428	2,428
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Randomization for the two open questions were independent, we include each question into its corresponding group. One word answers have been excluded from Yules'K, as they do not provide enough data for the measure to function properly. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.8. Informativeness Metrics (Survey 1)

	Score		Dummy	
Audio	0.831*** (0.081)	0.797*** (0.080)	0.110*** (0.012)	0.107*** (0.012)
Constant	2.766*** (0.042)	1.768*** (0.203)	0.068*** (0.006)	-0.042 (0.029)
$\mathbb{E}(y)$	3.117 (2.200)	3.117 (2.200)	0.115 (0.319)	0.115 (0.319)
R-squared	0.045	0.095	0.039	0.066
Obs	3,167	3,167	3,167	3,167
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.9. Informativeness Metrics (Survey 2)

	Score		Dummy	
Audio	0.713*** (0.117)	0.678*** (0.117)	0.120*** (0.019)	0.116*** (0.019)
Constant	4.009*** (0.073)	2.496*** (0.367)	0.226*** (0.012)	-0.003 (0.058)
$\mathbb{E}(y)$	4.333 (2.817)	4.333 (2.817)	0.281 (0.450)	0.281 (0.450)
R-squared	0.096	0.135	0.056	0.089
Obs	2,171	2,171	2,171	2,171
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals who could choose between text and audio were excluded from the analysis. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.10. Informativeness Metrics (Survey 3)

	Score		Dummy	
Audio	0.409*** (0.091)	0.402*** (0.090)	0.046*** (0.013)	0.045*** (0.013)
Constant	2.732*** (0.054)	2.493*** (0.310)	0.098*** (0.008)	0.084* (0.047)
$\mathbb{E}(y)$	2.910 (2.197)	2.910 (2.197)	0.117 (0.322)	0.117 (0.322)
R-squared	0.013	0.035	0.010	0.027
Obs	2,428	2,428	2,428	2,428
Question FE	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Randomization for the two open questions were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.11. Personal Experience Metrics (Survey 1)

	1st Person (pronoun)		1st Person (verb)		Personal Experience		Subjectivity	
Audio	0.034*** (0.006)	0.036*** (0.006)	0.150*** (0.016)	0.159*** (0.016)	0.026*** (0.008)	0.031*** (0.008)	0.042*** (0.012)	0.044*** (0.012)
Constant	0.005*** (0.002)	-0.002 (0.013)	0.224*** (0.010)	0.275*** (0.052)	0.039*** (0.005)	0.057** (0.027)	0.852*** (0.008)	0.870*** (0.040)
$\mathbb{E}(y)$	0.020 (0.140)	0.020 (0.140)	0.288 (0.453)	0.288 (0.453)	0.050 (0.218)	0.050 (0.218)	0.870 (0.336)	0.870 (0.336)
R-squared	0.016	0.020	0.027	0.041	0.004	0.018	0.004	0.007
Obs	3,167	3,167	3,167	3,167	3,167	3,167	3,167	3,167
Question FE	YES	YES	YES	YES	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.12. Personal Experience Metrics (Survey 2)

	1st Person (pronoun)		1st Person (verb)		Personal Experience		Subjectivity	
Audio	0.070*** (0.010)	0.069*** (0.010)	0.197*** (0.019)	0.195*** (0.019)	0.031*** (0.011)	0.030*** (0.011)	0.076*** (0.015)	0.074*** (0.015)
Constant	0.017*** (0.004)	0.015 (0.031)	0.200*** (0.012)	0.160** (0.064)	0.054*** (0.007)	0.043 (0.034)	0.823*** (0.011)	0.806*** (0.052)
$\mathbb{E}(y)$	0.049 (0.216)	0.049 (0.216)	0.289 (0.454)	0.289 (0.454)	0.068 (0.252)	0.068 (0.252)	0.858 (0.349)	0.858 (0.349)
R-squared	0.027	0.036	0.047	0.067	0.004	0.009	0.012	0.019
Obs	2,171	2,171	2,171	2,171	2,171	2,171	2,171	2,171
Question FE	YES	YES	YES	YES	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Individuals who could choose between text and audio were excluded from the analysis. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.13. Personal Experience Metrics (Survey 3)

	1st Person (pronoun)		1st Person (verb)		Personal Experience		Subjectivity	
Audio	0.037*** (0.008)	0.036*** (0.008)	0.034** (0.018)	0.037** (0.017)	0.024* (0.014)	0.024* (0.014)	0.043** (0.018)	0.043** (0.018)
Constant	0.016*** (0.003)	0.012 (0.022)	0.222*** (0.011)	0.263*** (0.059)	0.123*** (0.009)	0.107** (0.049)	0.720*** (0.012)	0.636*** (0.069)
$\mathbb{E}(y)$	0.032 (0.176)	0.032 (0.176)	0.237 (0.425)	0.237 (0.425)	0.134 (0.341)	0.134 (0.341)	0.738 (0.440)	0.738 (0.440)
R-squared	0.012	0.022	0.002	0.029	0.002	0.011	0.009	0.024
Obs	2,428	2,428	2,428	2,428	2,428	2,428	2,428	2,428
Question FE	YES	YES	YES	YES	YES	YES	YES	YES
Respondent-level controls	NO	YES	NO	YES	NO	YES	NO	YES

Notes. Respondent-level controls include dummies for: gender, high education (college or more), cohort (Babyboomer, GenX, GenZ, Millennial), occupation status (Retired, Employed), size of the municipality (Main cities), political attitudes (Right, Left, Other parties), relation to technology and to digital technology (Hightech), device used for the survey (Mobile phone). Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.14. Number of Words (By Question)

	Words	Significant words	Obs.
<b>Audio</b>			
S1Q1	10.289*** (1.303)	5.653*** (0.734)	1043
S1Q2	16.282*** (1.519)	8.780*** (0.866)	1058
S1Q3	19.282*** (1.878)	10.229*** (1.062)	1066
S2Q1	13.730*** (2.411)	6.816*** (1.336)	717
S2Q2	13.177*** (1.573)	7.140*** (0.875)	721
S2Q3	21.483*** (2.616)	11.932*** (1.465)	733
S3Q1	10.015*** (1.339)	5.353*** (0.744)	1322
S3Q2	9.827*** (1.468)	4.721*** (0.677)	1106

Notes. Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.15. Lexical Complexity Metrics (By Question)

	Yule's K	Obs.	TTR	Obs.
<b>Audio</b>				
S1Q1	213.023*** (48.212)	976	-0.059*** (0.006)	1043
S1Q2	53.483*** (16.572)	1010	-0.062*** (0.005)	1058
S1Q3	38.007** (17.594)	1020	-0.062*** (0.005)	1066
S2Q1	48.705*** (18.091)	691	-0.050*** (0.007)	717
S2Q2	75.582** (36.291)	659	-0.059*** (0.007)	721
S2Q3	-31.973* (18.126)	658	-0.068*** (0.008)	733
S3Q1	88.969*** (21.597)	1241	-0.050*** (0.005)	1322
S3Q2	168.741*** (38.353)	784	-0.053*** (0.006)	1106

Notes. Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.



Table A.16. Informativeness Metrics (By Question)

	Score	Dummy	Obs.
<b>Audio</b>			
S1Q1	0.579*** (0.123)	0.085*** (0.018)	1043
S1Q2	0.903*** (0.145)	0.105*** (0.021)	1058
S1Q3	1.004*** (0.152)	0.138*** (0.023)	1066
S2Q1	0.364* (0.209)	0.056* (0.033)	717
S2Q2	1.087*** (0.188)	0.150*** (0.030)	721
S2Q3	0.684*** (0.209)	0.153*** (0.036)	733
S3Q1	0.454*** (0.136)	0.057*** (0.020)	1322
S3Q2	0.355*** (0.118)	0.032* (0.018)	1106

Notes. Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.17. Personal Experience Metrics (By Question)

	1st Person (pronoun)	1st Person (verb)	Personal Experience	Subjectivity	Obs.
<b>Audio</b>					
S1Q1	0.044*** (0.011)	0.151*** (0.029)	0.013 (0.014)	0.041** (0.021)	1043
S1Q2	0.034*** (0.010)	0.140*** (0.029)	0.030** (0.014)	0.031 (0.021)	1058
S1Q3	0.025*** (0.008)	0.159*** (0.028)	0.035** (0.014)	0.052** (0.020)	1066
S2Q1	0.067*** (0.018)	0.198*** (0.034)	0.047** (0.020)	0.075*** (0.026)	717
S2Q2	0.071*** (0.017)	0.214*** (0.034)	-0.007 (0.019)	0.076*** (0.025)	721
S2Q3	0.073*** (0.015)	0.179*** (0.033)	0.052*** (0.019)	0.076*** (0.025)	733
S3Q1	0.055*** (0.011)	0.055** (0.024)	0.005 (0.018)	0.025 (0.025)	1322
S3Q2	0.016 (0.010)	0.010 (0.025)	0.048** (0.022)	0.063** (0.025)	1106

Notes. Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include each question into its corresponding group. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.18. Human-Coded and AI-Driven Topic Analysis (Survey 2)

	Obs	Text	s.d.	Audio	s.d.	Coeff	p-value
<b>Panel A: Question 3 - Human Coding</b>							
Single ret rule	734	0.771	0.421	0.617	0.487	-0.154***	0.000
Turnover narrative	734	0.028	0.164	0.053	0.225	0.026*	0.082
Job categories	734	0.136	0.343	0.228	0.420	0.092***	0.001
Flexible schemes	734	0.088	0.284	0.104	0.306	0.016	0.474
Complexity score	734	0.252	0.519	0.386	0.567	0.134***	0.001
<b>Panel B: Question 3 - GPT 4 Turbo</b>							
Single ret rule (corr 0.4873)	734	0.506	0.501	0.374	0.485	-0.132***	0.000
Turnover narrative (corr 0.8393)	734	0.048	0.214	0.062	0.242	0.014	0.395
Job categories (corr 0.8446)	734	0.169	0.375	0.220	0.415	0.051*	0.084
Flexible schemes (corr 0.6789)	734	0.116	0.320	0.145	0.353	0.030	0.239
Complexity score (corr 0.8006)	734	0.332	0.555	0.427	0.552	0.095**	0.021

Notes. Individuals who could choose between text and audio were excluded from the analysis. Question FE and respondent-level controls are not included. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*.

Table A.19. Balance Tests (Survey Started)

	Obs	Text	s.d.	Audio	s.d.	Coeff	p-value
Male	4,749	0.493	0.500	0.503	0.500	0.006	0.695
High education	4,749	0.368	0.482	0.357	0.479	-0.013	0.374
Babyboomer	4,749	0.386	0.487	0.372	0.483	-0.014	0.358
GenX	4,749	0.312	0.464	0.312	0.463	-0.002	0.873
GenZ	4,749	0.054	0.226	0.063	0.242	0.011	0.117
Millennial	4,749	0.198	0.399	0.199	0.399	-0.001	0.958
Retired	4,749	0.244	0.430	0.239	0.426	-0.004	0.748
Employed	4,749	0.549	0.498	0.549	0.498	-0.003	0.825
Main cities	4,749	0.468	0.499	0.490	0.500	0.021	0.191
Right	4,749	0.257	0.437	0.253	0.435	-0.006	0.684
Left	4,749	0.420	0.494	0.431	0.495	0.012	0.439
Other parties	4,749	0.241	0.428	0.233	0.423	-0.008	0.572
Hightech	4,749	0.589	0.492	0.578	0.494	-0.009	0.562
Mobile phone	4,749	0.409	0.492	0.414	0.493	0.005	0.734

Notes. Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include in the group *audio* individuals who face an audio-type question at least once. Coefficients are estimated including survey FE. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*. Survey 1 includes 1,650 observations, Survey 2 includes 1,093 observations, and Survey 3 includes 2,006 observations.

Table A.20. Balance Tests (Survey Completed)

	Obs	Text	s.d.	Audio	s.d.	Coeff	p-value
Male	3,579	0.493	0.500	0.531	0.499	0.034*	0.050
High education	3,579	0.367	0.482	0.377	0.485	0.008	0.653
Babyboomer	3,579	0.387	0.487	0.364	0.481	-0.025	0.143
GenX	3,579	0.313	0.464	0.323	0.468	0.008	0.603
GenZ	3,579	0.053	0.224	0.062	0.241	0.011	0.170
Millennial	3,579	0.199	0.399	0.209	0.407	0.010	0.488
Retired	3,579	0.242	0.429	0.215	0.411	-0.027*	0.067
Employed	3,579	0.552	0.497	0.576	0.494	0.021	0.227
Main cities	3,579	0.472	0.499	0.496	0.500	0.021	0.225
Right	3,579	0.256	0.437	0.271	0.444	0.012	0.420
Left	3,579	0.419	0.494	0.423	0.494	0.008	0.646
Other parties	3,579	0.241	0.428	0.225	0.417	-0.019	0.194
Hightech	3,579	0.589	0.492	0.618	0.486	0.033*	0.054
Mobile phone	3,579	0.411	0.492	0.440	0.497	0.029*	0.091

Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include in the group *audio* individuals who face an audio-type question at least once. Coefficients are estimated including survey FE. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*. Survey 1 includes 1,176 observations, Survey 2 includes 827 observations, and Survey 3 includes 1,576 observations.

Table A.21. Balance Tests (No Open-Ended Question Skipped)

	Obs	Text	s.d.	Audio	s.d.	Coeff	p-value
Male	2,946	0.492	0.500	0.534	0.499	0.037*	0.051
High education	2,946	0.370	0.483	0.383	0.486	0.013	0.479
Babyboomer	2,946	0.388	0.488	0.368	0.482	-0.025	0.175
GenX	2,946	0.312	0.464	0.338	0.473	0.023	0.185
GenZ	2,946	0.053	0.223	0.057	0.231	0.007	0.405
Millennial	2,946	0.196	0.397	0.200	0.400	0.004	0.789
Retired	2,946	0.244	0.430	0.213	0.409	-0.035**	0.028
Employed	2,946	0.551	0.498	0.584	0.493	0.033*	0.074
Main cities	2,946	0.473	0.499	0.504	0.500	0.026	0.176
Right	2,946	0.256	0.436	0.265	0.441	0.006	0.707
Left	2,946	0.426	0.495	0.450	0.498	0.029	0.125
Other parties	2,946	0.242	0.428	0.218	0.413	-0.026*	0.097
Hightech	2,946	0.594	0.491	0.632	0.482	0.043**	0.018
Mobile phone	2,946	0.409	0.492	0.464	0.499	0.055***	0.004

Individuals in survey 2 who could choose between text and audio were excluded from the analysis. Randomization for the two open questions in survey 3 were independent, we include in the group *audio* individuals who face an audio-type question at least once. Coefficients are estimated including survey FE. Significance at the 10% level is represented by \*, at the 5% by \*\*, and at the 1% by \*\*\*. Survey 1 includes 1,017 observations, Survey 2 includes 699 observations, and Survey 3 includes 1,230 observations.