Appendix A: Information about the dataset and data cleaning report

Original Number of respondents in the SWG survey

During Wave 1, Swg sent 7563 invitations overall: 1,923 valid interviews were completed and retained according to the predefined quotas (4221 survey recipients did not open the invitation link; 612 participants did not complete the survey, or they were excluded because they did not express their consent to the privacy policy; 1007 interviews were dropped because they exceeded the quotas). The original completion rate for Wave 1 was thus 38,7%. With respect to the longitudinal design, 1,763 individuals filled in the second wave of the study and 1,714 panelists completed Wave 3; 1646 interviewed participated in all the surveys (retention rate 91,7% between Wave 1 and Wave 2; 85,6% for the whole study).

Wave 1 = 1,923 individuals

Wave 2 = 1,763 individuals

Wave 3 = 1,714 individuals

Panelists (wave 1, 2, 3) = 1,646 [Retention Rate 85%]

<u>Data Cleaning process:</u>

1. **<u>Time of completion of the survey.</u>** Following the approach outlined in Baker et al. (2010) and Martini et al. (2021), we dropped from the analysis all those respondents (254) who completed the interview in less than (or equal to) the 50% of the rounded median response time in Wave 1. Therefore, all respondents who took 9 minutes or less to complete the first wave of the panel study were dropped from the dataset. For the other 2 waves, we followed the same criterion and we treated as missing cases all those respondents who completed the survey in less than (or equal to) the 50% of the rounded median response time in Waves 2 and 3. In this way, we treated as missing values 69 and 73 cases, respectively in Wave 2 and Wave 3.

2. **<u>Anomalies with declared gender.</u>** We crossed the information declared by respondents during the 3 waves of the survey with respect to their gender, as well as the information they provided upon enrollment in the panel (SWG variable). When an inconsistency in the whole study was detected, we dropped that respondent from the dataset. Overall, we detected and excluded 33 cases, which added to all those already deleted during the check about the rounded time of completion of the survey in Wave 1 (see Point 1).

3. **<u>Anomalies with declared year of birth</u>**. Similarly, we dropped from the analysis all those respondents who reported two (or more) different years of birth across the survey. Overall, we detected and dropped from the analysis 60 cases, in addition to those we had already excluded by the time of completion of the survey (see Point 1) and their inconsistencies with respect to declared gender (see Point 2). 14 cases had a problem both with declared gender and year of birth.

4. **<u>Anomalies with declared educational level.</u>** In this case, the issue was harder to address, as over-time changes in the educational level is a possible and plausible outcome (for instance, young people who achieve their University degree during the panel study, or who moved from primary to secondary education levels, or who started a University degree to then abandon their studies). Overall, we detected 207 cases of people who had at least one change in their declared educational level across the three waves. Of these:

- 176 cases presented only one change across the study. We kept these cases in the dataset, as the change pertained to individuals who were in high school during Wave 1 but subsequently started a University degree, or to individuals attending their last year of secondary school and subsequently started high school;

- 31 cases presented at least two changes. These occurrences were checked singularly to assess whether these twofold changes took place in the course of a regular educational path or constituted true inconsistencies. Overall, we retained 18 cases and dropped from the dataset 13 people who gave inconsistent responses across the survey (e.g., primary education in one wave and university degree in another wave).

The final sample is as follows:

Wave 1 = 1,563 respondents
Wave 2 = 1,353 respondents
Wave 3 = 1,299 respondents

The number of panelists who participated in the whole panel study (waves 1, 2 and 3) was 1204 [Retention Rate 77%] The completion rate of Wave 1 with respect to the definitive version of the dataset is therefore 34%; the retention rate between Wave 1 and Wave 2 is 86,6%, and the overall retention rate for the whole study amounts to 77%.

The quality of the dataset did improve with the data-cleaning procedure, and Table A2 shows that the cleaned sample in Wave 1 continue to be consistent with predetermined quotas, with a deviation from the theoretical sample that, on average, is equal to or less than +/- 2,5% with respect to gender, age, education, working condition and area of residence. Only age groups 25-34 (-3,4%); 35-44 (- 4,16%), 55-64 (+2,8%) 65-74 (+2,6%) and the lower educated people (-4,4%) were slightly above or below the +/- 2,5% threshold, but still in a very acceptable range.

Table A2: Comparative table – Theoretical quotas (Weighted ISTAT quotas) and effective quotas in the dataset (in percentage with respect to the total). In parentheses the difference (in percentage terms) between the theoretical and the effective quotas in the dataset.

Stratification Variables	Weighted quotas – Italian National Institute of Statistics (ISTAT) 2018[1]	I Wave (Original SWG Version)	I Wave (Post data-cleaning version)	II Wave (Post data-cleaning version)	III Wave (Post data-cleaning version)	I Wave – Only panelists - (Post data-cleaning version)
Gender						
Male	51,16 %	49,95 % (-1,71 %)	48,88 % (-2,28 %)	50,48 % (-0,68 %)	50,42 % (-0,74 %)	51,25 % (+0,09 %)
Female	48,84 %	50,55 % (+1,71 %)	51,12 % (+2,28 %)	49,52 % (+0,68 %)	49,58 % (+0,74 %)	48,75 % (-0,09 %)
Total	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
Age (in categories)						
18-24	11,64 %	11,28 % (-0,36 %)	11,52 % (-0,12 %)	9,68 % (-1,96 %)	9,47 % (-2,17 %)	9,63 % (-2,01 %)
25-34	18,51 %	16,80 % (-1,71 %)	15,10 % (-3,41 %)	15,23 % (-3,28 %)	14,55 % (-3,96 %)	14,78 % (-3,73 %)
35-44	22,27 %	21,42 %	18,11 %	17,89 %	18,24 %	17,69 %

[1] ISTAT Source of data: Weighted microdata – "Multipurpose survey on families: aspects of everyday life 2018", taking as reference of Internet users those who have used Internet over the last three months (age range 18-74). Available on https://www.istat.it/it/archivio/129956

		(-0,85 %)	(-4,16 %)	(-4,38 %)	(-4,03 %)	(-4,58 %)
45-54	24,01 %	24,08 % (+0,07 %)	26,30 % (+2,29 %)	26,16 % (+2,15 %)	25,48 % (+1,47 %)	26,83 % (+2,82 %)
55-64	15,88 %	17,26 % (+1,38 %)	18,68 % (+2,80 %)	19,81 % (+3,93 %)	21,17 % (+5,29 %)	19,85 % (+3,97 %)
65-74	7,69 %	9,15 % (+1,46 %)	10,30 % (+2,61 %)	11,23 % (+3,54 %)	11,09 % (+3,40 %)	11,21 % (+3,52 %)
Total	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
Educational level						
Low level of education	29,93 %	26,42 % (-3,51 %)	25,53 % (-4,40 %)	25,35 % (-4,58 %)	23,71 % (-6,22 %)	24,58 % (-5,35%)
Middle level of education	49,16 %	50,91 % (+1,75 %)	51,38 % (+2,22 %)	50,18 % (+1,02 %)	51,04 % (+1,88 %)	51,16 % (+2%)
High level of education	20,91 %	22,67 % (+1,76 %)	23,10 % (+2,19 %)	24,46 % (+3,55 %)	25,25 % (+4,34 %)	24,25 % (+3,34%)
Total	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
Working status						
Working	61,14 %	59,85 % (-1,29 %)	58,99 % (-2,15 %)	59,65 % (-1,49 %)	61,35 % (+0,21 %)	59,88 % (-1,26%)
No working	38,85 %	40,15 %	41,01 %	40,35 %	38,65 %	40,12 %

		(+1,30 %)	(+2,16 %)	(+1,50 %)	(-0,20 %)	(+1,27%)
Total	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
Region of residence						
Piemonte	7,44 %	7,02 % (-0,42 %)	7,23 % (-0,2 %)	7,46 % (+0,02 %)	7,01 % (-0,43 %)	7,23 % (-0,21%)
Valle d'Aosta	0,21 %	0,26 % (+0,05 %)	0,32 % (+0,11 %)	0,30 % (+0,09 %)	0,31 % (+0,10 %)	0,33 % (+0,12%)
Lombardia	17,84 %	16,90 % (-0,94 %)	17,40 % (-0,44 %)	17,29 % (-0,55 %)	17,24 % (-0,60 %)	17,52 % (-0,32%)
Trentino-Alto Adige	1,86 %	2,44 % (+0,58 %)	1,98 % (+0,12 %)	2,07 % (+0,21 %)	2,31 % (+0,45 %)	2,08 % (+0,22%)
Veneto	8,38 %	8,42 % (+0,04 %)	7,81 % (-0,57 %)	7,17 % (-1,21 %)	7,39 % (-0,99 %)	7,23 % (-1,15%)
Friuli-Venezia Giulia	2,15 %	2,55 % (+0,40%)	2,75 % (+0,60 %)	2,73 % (+0,58 %)	2,85 % (+0,70 %)	2,82 % (+0,67%)
Liguria	2,69 %	2,60 % (-0,09 %)	2,37 % (-0,32 %)	2,66 % (-0,03 %)	2,39 % (-0,30 %)	2,57 % (-0,12%)
Emilia-Romagna	7,93 %	7,49 % (-0,44 %)	7,04 % (-0,89 %)	7,39 % (-0,54 %)	7,08 % (-0,85 %)	6,98 % (-0,95%)
Toscana	6,54 %	6,71 %	7,04 %	6,50 %	7,16 %	6,89 %

		(+0,17 %)	(+0,50 %)	(-0,04 %)	(+0,62 %)	(+0,35%)
Umbria	1,41 %	1,35 % (-0,06 %)	1,28 % (-0,13 %)	1,18 % (-0,23 %)	1,15 % (-0,26 %)	1,25 % (-0,16%)
Marche	2,54 %	2,44 % (-0,10 %)	2,56 % (+0,02 %)	2,44 % (-0,10 %)	2,31 % (-0,23 %)	2,33 % (-0,21%)
Lazio	10,09 %	9,88 % (-.0,21 %)	9,98 % (-0,11 %)	10,50 % (+0,41 %)	10,62 % (+0,53 %)	10,22 % (+0,13%)
Abruzzo	2,14 %	1,98 % (-0,16 %)	2,05 % (-0,09 %)	2,07 % (-0,07 %)	1,92 % (-0,22 %)	1,91 % (-0,23%)
Molise	0,47 %	0,57 % (+0,10 %)	0,51 % (+0,04 %)	0,52 % (+0,05 %)	0,62 % (+0,15 %)	0,42 % (-0,05%)
Campania	8,32 %	8,53 % (+0,21 %)	8,38 % (+0,06 %)	8,35 % (+0,03 %)	8,31 % (-0,01 %)	8,55 % (+0,23%)
Puglia	6,16 %	6,34 % (+0,18 %)	6,46 % (+0,30 %)	6,58 % (+0,42 %)	5,77 % (-0,39 %)	6,40 % (+0,24%)
Basilicata	0,89 %	0,94 % (+0,05 %)	0,9 % (+0,01 %)	0,81 % (-0,08 %)	0,85 % (-0,04 %)	0,91 % (+0,02%)
Calabria	2,82 %	2,96 % (+0,14 %)	3,2 % (+0,38 %)	3,10 % (+0,28 %)	3,31 % (+0,49 %)	3,16 % (+0,34%)
Sicilia	7,26 %	7,59 % (+0,33 %)	7,55 % (+0,29 %)	7,61 % (+0,35 %)	8,01 % (+0,74 %)	7,81 % (+0,55%)
Sardegna	2,85 %	3,02 %	3,2 %	3,25 %	3,39 %	3,41 %

		(+0,17 %)	(+0,35 %)	(+0,40 %)	(+0,54 %)	(+0,56%)
Total	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**
N		**1923**	**1563**	**1353**	**1299**	**1204**

References

Baker R. et al. (2010) Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly* **74**, 711–781.

Martini S. et al. (2021) Paranoid styles and innumeracy: Implications of a conspiracy mindset on Europeans' misperceptions about immigrants, *Italian Political Science Review/Rivista Italiana Di Scienza Politica* 1-17. doi:10.1017/ipo.2021.26