



Number 19: Another Victim of the COVID-19 Pandemic?

Patrick Roger¹ · Catherine D'Hondt² · Daria Plotkina³ · Arvid Hoffmann⁴

Accepted: 5 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Conscious selection is the mental process by which lottery players select numbers nonrandomly. In this paper, we show that the number 19, which has been heard, read, seen, and googled countless times since March 2020, has become significantly less popular among Belgian lottery players after the World Health Organization named the disease caused by the coronavirus SARS-CoV-2 “COVID-19”. We argue that the reduced popularity of the number 19 is due to its negative association with the COVID-19 pandemic. Our study triangulates evidence from field data from the Belgian National Lottery and survey data from a nationally representative sample of 500 Belgian individuals. The field data indicate that the number 19 has been played significantly less frequently since March 2020. However, a potential limitation of the field data is that an unknown proportion of players selects numbers randomly through the “Quick Pick” computer system. The survey data do not suffer from this limitation and reinforce our previous findings by showing that priming an increase in the salience of COVID-19 prior to the players’ selection of lottery numbers reduces their preference for the number 19. The effect of priming is concentrated amongst those with high superstitious beliefs, further supporting our explanation for the reduced popularity of the number 19 during the COVID-19 pandemic.

Keywords Games of chance · Lotteries · Conscious selection · COVID-19 · Availability heuristic

Introduction

On February 11, 2020, the World Health Organization (WHO) published Situation Report 22.¹ The first page of this report states the following: “Following WHO best practices for naming of new human infectious diseases, which were developed in consultation and collaboration with the World Organisation for Animal Health (OIE) and the Food and Agriculture Organization of the United Nations (FAO), WHO has named the disease Covid-19, short for coronavirus disease 2019.” Since then, almost every person in the world has

¹ https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2.

✉ Catherine D'Hondt
catherine.dhondt@uclouvain.be

Extended author information available on the last page of the article

heard/read/seen/googled the number 19 several times a day, associated with the word Covid. As an illustration of this intense exposition, Fig. 1 shows the evolution of Google searches in the news for different random sets of numbers during the year 2020.² Not surprisingly, the number 19 exhibits a completely different behavior compared to other numbers, with a tremendous increase in searches starting at the end of February 2020.

In everyday life, many decisions involve integer numbers, either as quantities, rankings, or codes. A number of papers relate economic decisions to personal preferences for numbers. For example, in China, the number eight is considered lucky, and the number four is considered unlucky (Hirshleifer et al., 2018; Bhattacharya et al., 2018). As a consequence, when license plates are auctioned in Hong Kong, those including the number eight fetch a higher price than those containing the number four (Woo et al., 2008; Chong and Du, 2008). The same kind of distortion is observed in the Hong Kong housing market, where houses with a number ending in eight are sold at a higher price than those with a number ending in four (Fortin et al., 2014; Shum et al., 2014).

For almost thirty years, the literature on lottery gambling has shown that lottery players do not choose their numbers at random, although winning numbers are drawn randomly (Baker and McHale, 2009, 2011; Cook and Clotfelter, 1993; Farrell et al., 2000; Papachristou and Karamanis, 1998; Roger and Broihanne, 2007; Simon, 1998; Turner, 2010; Wang et al., 2016). Since the seminal work of Cook and Clotfelter (1993), the nonrandom choice of lottery numbers is called *conscious selection*. There are multiple reasons for such non-random choice. Lottery players may prefer some specific combinations of numbers (e.g., meaningful dates such as birth dates or random-looking sequences) or “lucky” numbers (e.g., 7 in most Western countries (Roger, 2011), 9 in Nigeria (Vandewiele et al., 1986), and 8 in China (Shum et al., 2014; Brown and Mitchell, 2008; Hirshleifer et al., 2018)).³

Gamblers may also avoid numbers perceived as unlucky. Typically, as mentioned above, in China, the number 4 is viewed as unlucky (Chung et al., 2014) and the number 17 is viewed as unlucky in Italy (De Paola et al., 2014). In Western countries, the status of the number 13 remains unclear. Although it was historically considered unlucky, the number 13 is overplayed by Lotto players in countries such as France (Roger and Broihanne, 2007) or the Netherlands (Wang et al., 2016). The gambler’s fallacy and the hot hand fallacy may also cause conscious selection. For example, Suetens et al. (2016) show the gambler’s fallacy among regular players, whereby the amount bet on numbers drawn in the preceding draw is 2% lower than the amount bet on numbers that were not previously drawn.⁴

In a standard expected utility framework, gambling on state lotteries is difficult to explain, and choosing popular numbers when playing a Lotto game seems even more irrational. The reason is intuitive: Most state lotteries, such as Lotto games, are Parimutuel games, which means that a given amount devoted to winners at a given rank is

² Google Trends allows searching for up to 5 numbers at a time. For comparison, we included 19 in each set of 5 numbers.

³ For example, Baker and McHale (2011) refer to a draw of the Canadian lottery (March 19th, 2008), when the six winning numbers were 23, 40, 41, 42, 44 and 45, with a bonus number of 43. Nobody won the jackpot, but an unusually high number of winners (239) won second prize (5 correct numbers plus the bonus number) because many players bet on sequences, in this case 40-41-42-43-44-45.

⁴ See also Clotfelter and Cook (1993), Dek (1994), and Suetens and Tyran (2012) for other illustrations of these two fallacies in lottery gambling.

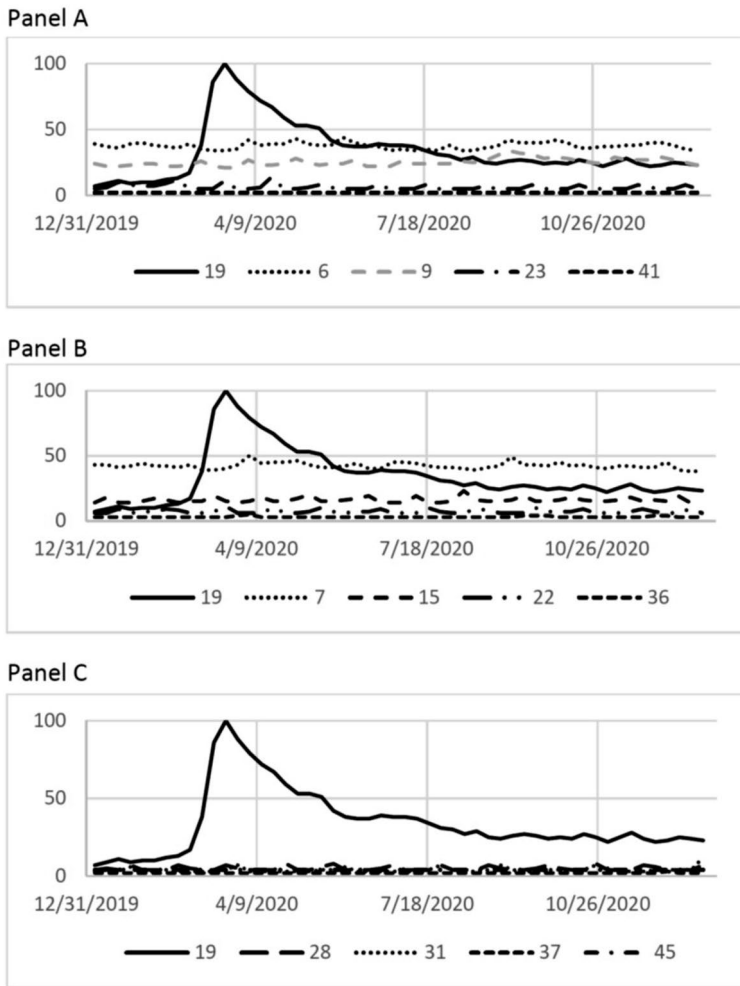


Fig. 1 Google searches for random numbers in the news during 2020 (including the number 19). This figure shows the evolution of the worldwide search results from Google Trends for different sets of numbers during 2020. Each set includes the number 19 for comparison purposes. Panel A shows the predominance of searches for the number 19 in the set including 6, 9, 23, and 41. Panel B shows the predominance of searches for the number 19 in the set including 7, 15, 22, and 36. Panel C shows the predominance of global searches for the number 19 in the set including 28, 31, 37, and 45

shared among all those winners. As a consequence, betting on popular numbers reduces the expected value of the gain. In fact, individual prizes are lower when numbers that are popular among players appear in the official draw.⁵

⁵ Although conscious selection seems to refer to deliberate choices, preferences for numbers can be unconscious. For example, in the criminal context, Dhimi et al. (2020) show that sentencers prefer certain numbers when meting out the sentence length (for custody and community service) and penalty amount (for fines/compensation).

The intense exposition of the number 19 in the media following the outbreak of the COVID-19 pandemic may influence the decision-making process of lottery players in several ways. The availability heuristic (Tversky and Kahneman, 1973) tells us that since the onset of COVID-19, the number 19 should come to mind more readily when people select numbers on their lottery ticket. Indeed, Wang et al. (2016) investigate preferences for specific numbers in large proprietary data sets from two different lottery games. They find that number selection does not obey uniform distributions. Instead, players select personally meaningful and situationally available numbers. If the availability heuristic affects lottery players, one would expect the number 19 to be played more frequently (perhaps unconsciously) since March 2020, when it became pervasive due to its association with COVID-19. However, because the association with COVID-19 is most likely negative, there may instead be a reduced preference for the number 19, despite its increased availability.

Indeed, preferences for specific lottery numbers may be affected by the interaction between emotions, preferences, and heuristic decision-making. In their review of the influence of cognitive, emotional, and hormonal factors on decision-making under risk, Kusev et al. (2017) refer to Loewenstein and Lerner (2003), who distinguish between immediate and anticipated emotions. Anticipated emotions are the emotions people *expect* to feel as a consequence of their decisions. In contrast, immediate emotions are the affective states that people *actually* experience when they make a decision. Under this approach, the association of the number 19 with the word COVID is expected to generate negative emotions (through bad feelings and/or memories). Since March 2020, such negative emotions may have led people to perceive the number 19 as unlucky or, at least, less lucky than before. Hence, the immediate negative emotions related to COVID-19 might influence gamblers when they select numbers on a Lotto ticket, affecting their preferences.

We performed an empirical study using Belgian data to investigate whether the popularity of the number 19 has been affected by the COVID-19 pandemic and, if so, whether its popularity varies depending on how salient COVID-19 is in people's minds (as that affects reliance on the availability heuristic). The motivation for relying on Belgian data is two-fold. First, Belgium is one of the most severely hit European countries in terms of deaths per capita linked to COVID-19 (see Fig. 6 in Appendix A.1). It thus provides an ideal setting to examine research questions on the impact of COVID-19 on lottery players' number preferences. Second, the Belgian National Lottery provides the general public with unique and clean data, concerning not only the results of lottery games but also the number of global sales per draw and the number of players per draw. To the best of our knowledge, the latter information is not publicly available in other countries. To provide a comprehensive investigation whether the number 19 has been another victim of the COVID-19 pandemic, we used two empirical approaches and triangulated evidence from field and survey data.

The first approach can be viewed as a field experiment based on the aforementioned data from the Belgian National Lottery. Among lottery games, Lotto (a 6/45 game with one bonus number) and Euromillions (a 5/50 game with two bonus numbers) are the most popular, which is the reason this study focused on these two games. An average of 635,000 (590,000) players participate in each of the two weekly draws for Euromillions (Lotto) out of a national population of 11 million people. We performed an econometric analysis of a sample of 836 lottery draws between March 2017 and February 2021 (418 draws for each game). The first three years define our benchmark period in terms of conscious selection and number preference. The fourth year (March 2020 to February 2021) defines the COVID period. The number 19 was drawn only 31 times during the COVID period. To address the small size of the sample of draws including the number 19 during the COVID

period, we used a permutation test to reject our null hypothesis that the popularity of the number 19 remained unchanged; nothing comparable occurred for the 44 other numbers. As conscious selection is a well-documented phenomenon, we know that birth dates are especially popular among lottery players (see Wang et al. (2016) or Roger and Broihanne (2007)). The former result obtained for the number 19 could therefore be a statistical artifact. We solved this issue with a multivariate analysis controlling for a set of standard explanatory variables in lottery studies, specifically, the average value of the numbers appearing in each draw and the per-player amount of the bets in a given draw. The results remain unchanged, indicating a significant decrease in the popularity of the number 19 among lottery players during the COVID-19 pandemic.

The second approach relied on survey data. In particular, we complemented the aforementioned field data with a survey of 500 individuals representative of the Belgian population in terms of age, gender, income, and education. In the middle of the survey, participants were asked to fill in two Lotto tickets. Several questions related to the COVID-19 pandemic and participants' associated experiences during the pandemic also appeared in the survey. To vary the salience of COVID-19 and therefore the availability of the pandemic and its associations in individuals' mind, we designed the survey so that for half the sample, COVID-19-related questions were asked before completing the Lotto tickets, while for the other half, the COVID-19-related questions came after. The survey study reinforces the findings of the field study by showing that when the salience of COVID-19 is increased through question ordering, individuals choose the number 19 significantly less frequently. Furthermore, we find that the priming effect of COVID-19 salience on number preference is concentrated among individuals with more pronounced superstitious beliefs.

The remainder of this paper is organized as follows. Section 2 describes the functioning of the Lotto games under investigation, including the Euromillions lottery and the Belgian Lotto game. Section 3 presents our field study using secondary data. Section 4 presents our survey study using primary data. Section 5 discusses the results. Section 6 concludes the paper.

Lotto Games

General Presentation

In a typical Lotto game, a player picks 5 or 6 numbers out of N , where N varies between 40 and 70, depending on the design of the game (the odds are chosen as a function of the population of the country). In most games, one or two bonus numbers are added to the principal draw. These bonus numbers can be drawn from an independent set of numbers or picked in the principal set of N numbers.

Lotto games are based on the Parimutuel principle, whereby the sponsor of the game applies a takeout rate (usually close to 50%) to the global amount bet by the players. The remainder goes to the players, according to prespecified sharing rules. The amount A to be redistributed is shared across ranks of gain by means of a percentage of A devoted to each rank. Therefore, more winners at a given rank mean lower individual prizes which range from the jackpot when a player has selected all of the right (i.e., drawn) numbers to consolation prizes for choosing 2 or 3 right numbers (with or without bonus numbers).

As an investment, a Lotto ticket exhibits a large negative expected return (because of the takeout rate) and a large positive skewness (because of a huge jackpot compared to the ticket cost). The typical Lotto player is therefore a skewness seeker, as are many individual investors (Mitton and Vorkink, 2007; Kumar, 2009; Hoffmann and Shefrin, 2014; Kumar et al., 2016; Broihanne et al., 2016). Moreover, the existence of rollovers may dramatically increase the skewness of the distribution of prizes.⁶

The Euromillions Lottery and the Belgian Lotto

The Euromillions Lottery

Euromillions is a Lotto game played in nine countries: Austria, Belgium, France, Ireland, Luxembourg, Portugal, Spain, Switzerland and the United Kingdom. A ticket consists of 5 different numbers selected from the range 1 to 50, and 2 “stars” that are numbers between 1 and 12, which are drawn from an independent set of numbers.⁷ There are 2 draws per week, on Tuesday and Friday. The probability of hitting the jackpot is approximately one in 140 million. From the ticket price of €2.50, €0.30 is devoted to paying the prize money for an included, independent game called MyBonus. Of the remaining €2.20, 50% is redistributed to the players, namely, €1.10. At the start of 2021, the minimum jackpot was €17 million. Over our study period, the maximum jackpot increased from €190 million to €250 million. Table 7 in Appendix A.2 gives the winning ranks and the corresponding probabilities.

The Belgian Lotto

The Belgian Lotto is a 6/45 Lotto game with a bonus number drawn in the set of 39 numbers remaining after the principal draw. The 6/45 refers to the fact that players have to select 6 numbers from a range of 1 to 45. There are two draws per week, on Wednesday and Saturday. Individual prizes for ranks 1 to 6 are variable from one draw to another. The amount devoted to a given rank is a percentage of the amount bet by the players; it is equally shared among the winners at this rank.⁸ In contrast, the individual prizes for the last three ranks are fixed (respectively €6.25, €3.75, and €1.25).⁹ At each draw, the minimum jackpot was €1 million, and the jackpot increases when there are rollovers. Table 8 in Appendix A.2 gives the winning ranks and the corresponding winning probabilities.

⁶ When the jackpot is not hit on a given draw, it is added to the next jackpot (rollover), thereby increasing the demand for Lotto tickets.

⁷ Details can be found in the legal document (Arrêté royal, March 21, 2018) available at https://www.etaamb.be/fr/arrete-royal-du-01-avril-2016_n2016003257.html. Before May 26, 2018, the fixed gains for ranks 7 and 8 were €5 and €3, respectively.

⁸ In the current version of the game, the percentages for ranks 1 to 6 are 17.50%, 2.95%, 2.80%, 1.40%, 2.59%, and 1.38%, respectively.

⁹ Details can be found in the legal document (Arrêté royal, March 21, 2018) available at https://www.etaamb.be/fr/arrete-royal-du-21-mars-2018_n2018011371.html. Before May 26, 2018, the fixed gains for ranks 7 and 8 were €5 and €3, respectively.

Field Study

Data and Descriptive Statistics

The Data from the Belgian National Lottery

We downloaded the data from the website of the Belgian National Lottery. The data files include the total amount of each bet and the number of players in each draw, which is unique to Belgium compared to other countries. In most countries, sales are estimated, and to the best of our knowledge, the number of players is not publicly available.¹⁰

We collected the data over a four-year period starting on March 3, 2017, and ending on March 2, 2021. The first three years are considered the “benchmark” period (March 2017 to February 2020), and the last year is considered the “COVID” period. We selected March 2017 as the starting point to align the beginning of the benchmark period (March 2017) with that of the COVID period (March 2020) in order to deal with entire years in our analysis.¹¹ As Euromillions draws occur on Tuesdays and Fridays and Lotto draws on Wednesdays and Saturdays, we have 4 draws per week over 209 weeks, leading to a total of 836 draws. Of these, 627 draws belong to the benchmark period and 209 belong to the COVID period.

Descriptive Statistics

Table 1 provides descriptive statistics for the number of players, amounts bet, and number of tickets for each draw. Panel A (B) gives the results for Euromillions (Lotto). To illustrate the stability of these variables over time, the two subperiods are separated. For averages, the two periods are almost identical, and we only notice a slightly higher average amount bet on Lotto during the COVID period. However, no conclusion related to the role of COVID-related lockdowns can be made with respect to such differences. Moreover, a difference between the average amounts bet on Lotto during the two subperiods might arise from differences in the history of jackpots. Table 1 also shows that despite periods of lockdown in Belgium, average demand did not fall during the COVID period. Unfortunately, the data do not reveal whether players moved from onsite to online betting during the pandemic.

¹⁰ For each game, the data are stored in two different files. The first file contains the official draws, with the date of the draw, and the numbers drawn, including two stars for Euromillions and a bonus number for Lotto. The second file contains the financial results of the game. There are 29 columns for Euromillions (21 for Lotto) with the date in the first column, the number of players in the second column, the global amount of bets in the third column, and 13 (9 for Lotto) pairs of columns for each of the 13 (9) winning ranks. Each pair of columns gives the number of winners and the individual prize for the rank under consideration. The first file name for Euromillions is EuroMillionsGameData-YYYYMMDD, where YYYYMMDD denotes the date the data was downloaded, and the name of the second file is EuroMillionsFinancialData-YYYYMMDD. The same rules are used for the files related to the Belgian Lotto. Examples of the data files for Euromillions are available in Appendix A.3.

¹¹ An important change in the Euromillions game rules occurred in September 2016. Before then, only eleven stars were available.

Over the entire sample period, the correlation between the number of players (tickets) and the number of winners is 97.5% (97.9%) for Euromillions and 92.3% (93.1%) for the Lotto.¹² If we focus on the COVID period, these correlations are slightly higher at 98.4% (98.7%) for Euromillions and 95.05% (94.61%) for the Lotto. If players were selecting numbers randomly, correlations would be very close to 1 because the ratio of the number of winners to the number of tickets would be almost constant and equal to the objective probability of winning. As mentioned earlier, small numbers are more popular because many people bet on birth dates. As a result, the ratio of the number of winners to the number of tickets deviates significantly from the probability of winning when small numbers appear in the official draw. Conscious selection therefore explains the imperfect correlation between the number of tickets and the number of winners.

A simple illustration of conscious selection in our sample is the correlation between the aggregate proportion of winners (i.e., number of winners divided by number of tickets) in each draw and the average of the numbers drawn. For Euromillions (Lotto), this correlation is -0.685 (-0.618) over the entire period and -0.745 (-0.543) over the COVID period. When small (large) numbers appear in the official draw, there is a large (small) number of winners because small numbers are preferred; many people play birth dates, that is, numbers lower than 12 (months) and 31 (days). For example, on March 6, 2020, the Euromillions draw was 15-38-43-45-46. Therefore, the aggregate proportion of winners was equal to 6.09%, far below the expected proportion of winners under random choice (7.71%). In contrast, on March 17, 2020, the draw was 5-7-8-16-20, and the aggregate proportion of winners jumped to 8.89%. These examples illustrate the popular preference for small numbers. The first draw contained mainly large “unpopular” numbers, while the second draw contained mainly small “popular” numbers (also see Wang et al. (2016), Roger and Broihanne (2007)). This anecdotal evidence, and the strong correlation between the proportion of winners and the average of numbers drawn, is the reason why we controlled for the average value of drawn numbers in our multivariate analysis reported in Sect. 3.3.

Table 2 provides detailed information on the number of winners per rank. The first key point explaining our methodological choices relates to the median number of winners for the first ranks. For Euromillions, more than 50% of draws did not have a winner with 5 correct numbers, corresponding to ranks 1 to 3. For Lotto, the median number of winners with 6 correct numbers was also zero. While a meaningful popularity index must account for winners at lower ranks, a Euromillions (Lotto) winner with 4 (5) correct numbers might not have played the number we are interested in. Because our focus is on the popularity of specific numbers, not combinations, we did not consider winners among all ranks (as in Turner (2010)). Because consolation prizes are given to players with combinations of only two or three correct numbers, the inclusion of these winners would not provide any meaningful information pertaining to the popularity of a specific individual number. We present a rigorous definition of our popularity index in Sect. 3.2.

The second key point related to Table 2 is the large variation among the ratios of maximum to minimum number of winners at a given rank. For intermediate ranks, this ratio is approximately 10. There are two main explanations for these variations. The first is the number of tickets sold; sales are strongly driven by the amount of the jackpot. Conscious selection—primarily, players’ preference for small numbers—provides the second explanation, as discussed above.

¹² Due to the addition of a winning rank to the Lotto game in the middle of the benchmark period (May 2018), the number of winners was calculated using the 8 first ranks.

Table 1 Descriptive statistics on the number of players, the amount of cash bet by players, and the number of tickets on each draw

	Mean	Median	Std dev	Min	Max
Panel A: Euromillions					
Benchmark period (03/2017-02/2020), 313 draws					
Number of players	637,352	620,357	203,703	358,100	1,588,165
Amount of cash	5,061,063	4,779,488	2,078,028	2,554,743	15,657,730
Number of tickets	2,024,425	1,911,795	831,211	1,021,897	6,263,092
Covid period (03/2020-02/2021), 105 draws					
Number of players	634,471	580,563	243,971	346,236	1,653,185
Amount of cash	5,055,992	4,445,806	2,580,285	2,401,880	16,278,395
Number of tickets	2,022,397	1,778,322	1,032,114	960,752	6,511,358
Panel B: Lotto					
Benchmark period (03/2017-02/2020), 314 draws					
Number of players	588,243	610,915	150,224	384,361	902,941
Amount of cash	3,900,857	4,195,847	1,266,069	2,158,160	6,240,587
Number of tickets	3,418,252	3,640,673	1,111,733	1,943,379	6,023,201
Covid period (03/2020-02/2021), 104 draws					
Number of players	591,686	599,512	143,302	361,014	972,739
Amount of cash	4,201,444	4,413,718	1,332,580	2,257,593	8,324,598
Number of tickets	3,361,155	3,530,975	1,066,064	1,806,075	6,659,679

This table reports descriptive statistics on the number of players, the amount bet by players, and the number of tickets in each draw. Panel A (B) refers to Euromillions (Lotto)

Each panel of Table 2 ends with a row indicating the aggregate proportion of winners over the sample period. Tables 7 and 8 in Appendix A.2 show that the expected values for the aggregate proportion of winners are 7.71% for Euromillions and 3.94% for Lotto (8 first ranks). The actual averages are 7.74% (Euromillions) and 3.92% (Lotto). The role of conscious selection appears more clearly in the *proportion* of winners compared to the number of winners because the effect of sales is neutralized when using proportions.

Table 3 provides, for each number $n = 1, \dots, 45$, the frequency of n in the official draws over the entire four-year period and the one-year COVID period. For convenience, Euromillions and Lotto are aggregated in this table. For example, the number 1 showed up 100 times over the four-year period but only 22 times during the COVID period. More generally, during the COVID period, most frequencies lie between 20 and 30, meaning a relatively small number of draws included a given number. As expected, the frequencies during the benchmark period are approximately three times larger than the frequencies during the COVID period. These two elements—a small sample size during the COVID period and a much larger sample size during the benchmark period—justify the use of permutation tests to compare the popularity of numbers in the two successive subperiods. We explain this methodology in Sect. 3.2.

Table 2 Descriptive statistics on the number of winners per rank

	Mean	Median	Std dev	Min	Max
Panel A: Euromillions winners					
Rank1 (5+2)	0.01	0	0.12	0	1
Rank2 (5+1)	0.27	0	0.52	0	3
Rank3 (5+0)	0.64	0	0.97	0	7
Rank4 (4+2)	3.39	3	3.22	0	20
Rank5 (4+1)	66.12	58	36.30	15	220
Rank6 (3+2)	147.87	124	87.63	36	682
Rank7 (4+0)	148.50	132	77.86	48	491
Rank8 (2+2)	2106.22	1819	1151.05	640	9491
Rank9 (3+1)	2904.04	2530	1394.97	1071	9319
Rank10 (3+0)	6514.63	5855	3054.92	2487	22,417
Rank11 (1+2)	10,960.98	9556	5656.25	3416	45,557
Rank12 (2+1)	41,309.46	37,300	18,805.05	17,113	124,479
Rank13 (2+0)	92,593.03	84,502	41,265.52	40,476	312,464
% winners	7.74%	7.71%	0.70%	6.09%	9.92%
Panel B: Lotto winners					
Rank1 (6+0)	0.49	0	0.80	0	5
Rank2 (5+1)	2.48	2	2.44	0	21
Rank3 (5+0)	95.09	84	47.21	22	272
Rank4 (4+1)	233.49	216	102.72	77	776
Rank5 (4+0)	4424.06	4097	1769.76	1603	10,656
Rank6 (3+1)	5778.63	5514	2182.63	2446	13,083
Rank7 (3+0)	70,585.44	69,056	24,891.56	30,188	144,974
Rank8 (2+1)	52,150.48	49,688	18,352.24	25,034	109,504
Rank9 (1+1)	175,611.15	165,675	56,617.28	93,143	345,694
% winners	3.92%	3.87%	0.46%	2.87%	5.32%

This table reports statistics on the number of winners per rank. Panel A (B) refers to Euromillions (Lotto). The expected proportion of winners (i.e., the aggregate probability of winning) is 7.71% (3.94%) for Euromillions (Lotto)

Methodology

The Popularity Index from the Belgian National Lottery Data

Appendix A.4 provides the details of our methodology to compute a popularity index of numbers to use in the empirical analysis. For brevity, we summarize the key points here. A given number $n = 1, 2, 3, \dots, 45$ appears in a subset of draws during each of the two subperiods under scrutiny. When a popular number appears in a draw, the proportion of winners deviates from the theoretical probability of winning. Hence, over a given subperiod, averaging the proportion of winners of draws in which n appeared provides a good estimate of the popularity of number n .

As noted earlier, the median number of winners with all winning numbers is 0 for the two games. Therefore, we must include tickets with 4 winning numbers in our estimation of the proportion of winners for Euromillions, that is, ranks 4, 5, and 7. For the Lotto

Table 3 Frequencies for each number: Euromillions and Lotto draws aggregated (Entire period and Covid period, 836 draws)

Number	Entire period	Covid period	Number	Entire period	Covid period
1	100	22	24	100	25
2	92	24	25	96	23
3	94	16	26	85	14
4	104	29	27	109	31
5	103	28	28	98	29
6	83	17	29	105	28
7	89	20	30	85	21
8	101	27	31	95	21
9	97	28	32	101	28
10	97	29	33	94	23
11	89	24	34	109	25
12	107	26	35	93	26
13	80	16	36	83	22
14	99	29	37	85	22
15	105	30	38	103	31
16	112	37	39	105	24
17	99	22	40	92	25
18	97	21	41	102	28
19	112	31	42	115	31
20	104	25	43	105	22
21	102	26	44	101	24
22	83	16	45	98	25
23	97	26			

In this table, columns 1 and 4 list the numbers from 1 to 45. Columns 2 and 5 give the number of times each number showed up in official draws over the entire four-year period, while columns 3 and 6 indicate the number of times each number showed up in official draws over the one-year Covid period

game, we must include winners with 5 winning numbers. However, Euromillions and Lotto strongly differ with respect to bonus number(s). Euromillions players select bonus numbers in an independent set of 12 numbers. In contrast, Lotto players select the bonus number in the same set as the principal draw (numbers 1 to 45). Consequently, measuring the popularity of a given number by including winners from rank 2 in the Lotto game would raise interpretive difficulties due to the special role of the bonus number. For the sake of simplicity, we aggregated only winners at ranks 1 and 3, excluding rank 2, which contains the bonus number. Importantly, the median number of Lotto winners at rank 2 is 2, compared to 84 at rank 3. Measuring popularity without any comparison to the theoretical probability of winning is not a limitation here, provided that the subset of ranks remains the same over the two subperiods, which is the case in our analysis.

Permutation Tests

To be relevant, a test of popularity change should apply to all 45 numbers, not only the number 19. Even when aggregating Euromillions and Lotto games, some numbers did not

show up frequently during the COVID period (e.g., the number 26 appeared only 14 times, compared to 71 times during the benchmark period). Of special interest, the number 19 appeared 31 times during the COVID period and 81 times during the benchmark period, as reported in Table 3.

To address the small sample problem and the disequilibrium of frequencies across sub-periods, we used a nonparametric permutation test.¹³ Consider the case of the number 19: It was drawn 81 times during the benchmark period and 31 times during the Covid period. As explained in Appendix A.4, the popularity index for the benchmark (COVID) period is the average of the 81 (31) standardized proportions of winners, which we call popularity scores for the sake of simplicity. The permutation test involved performing a large number of simulations (5,000 in our analysis), which were carried out according to the following steps: We drew two random samples (with replacements) of 81 and 31 popularity scores within the global sample of 112 popularity scores. We then calculated the difference in average popularity scores between these two subsamples. We repeated the draws 5,000 times to obtain a distribution of random differences that takes into account the difference in sample sizes (81 in the benchmark period and 31 in the COVID period). We calculated the observed difference in popularity indices (which are simply the average popularity scores) between the benchmark and COVID periods. Finally, we positioned the difference of popularity indices on the distribution of 5,000 values. The proportion of the 5,000 differences above the actual difference (benchmark period vs. COVID period) is the p-value of the permutation test. The histogram exhibited in Fig. 2 displays the distribution of random differences over 5,000 random samples. Of utmost importance, the vertical line on the right shows the difference between the benchmark and COVID periods. Only 3.04% of the random differences exceed the observed value. This means that a null assumption of unchanged popularity across subperiods (benchmark and COVID periods) can be rejected at the 5% significance level (since the p-value of the test is 3.04%).

While Fig. 2 allows us to highlight that the popularity index of the number 19 decreases significantly during the COVID period compared to the benchmark period, it does not provide us with any insights about what happens for the other 44 numbers. To compare the number 19 to the 44 other numbers, we need to repeat the permutation test for all 45 numbers. Figure 3 shows the p-values for the permutation test for all numbers. It appears that the only number obtaining a p-value below 5% is the number 19, meaning that 19 is the only number with a significant decrease in popularity during the COVID period.¹⁴

Multivariate Analysis

Our previous analysis based on permutation tests provides significant results but in a univariate setting that does not control for potentially confounding effects. As mentioned in the Introduction, the literature on lotteries reveals that conscious selection influences the proportion of winners (e.g., Papachristou and Karamanis (1998), Roger and Broihanne (2007)). First, an important observation from the study of conscious selection is that people prefer small numbers (e.g., Figure 1 in Wang et al. (2016) or Figure 3 in Roger and

¹³ See Ludbrook and Dudley (1998) for a detailed description of the permutation tests.

¹⁴ This result does not mean that the popularities for other numbers have not changed. As our test is one-sided, p-values higher than 95% identify numbers that are significantly more popular during the COVID period. These are the numbers 2, 17, and 40. We will come back more specifically regarding the number 17 result when discussing our findings in Sect. 5.

Broihanne (2007)). This well-documented phenomenon indicates that draws with small numbers generate larger proportions of winners than draws with large numbers. In a multivariate setting, it is therefore necessary to control for the average magnitude of numbers drawn. Second, we combined the Euromillions and Lotto data to increase the sample size, especially during the COVID period. However, because the two games have a somewhat different structure, we needed to control for the possibility that our results could be driven by only one of the two games. Third, there are two draws per week for each game, and the draws occurring on Fridays and Saturdays attract more players than those on Tuesdays and Wednesdays. Although there is no obvious reason to think that this “seasonality” in sales would influence gamblers’ preferences for subsets of numbers, a dummy variable may help to control for any difference in preference for the two parts of the week. Finally, the number of tickets played by any given player influences that player’s number choice because, under the Parimutuel principle, it is not optimal to play the same ticket several times. We therefore expect that two tickets played by two different players are less “different” in terms of influences on number selection than two tickets played by the same player.

To account for these potentially confounding effects, we ran the following regression model:

$$\begin{aligned} PropWinners_i = & a_0 + a_1 Covid_i + a_2 Dummy19_i + a_3 Covid_i \times Dummy19_i \\ & + a_4 MeanDraw_i + a_5 Amount_i + a_6 DummyLotto_i + a_7 DummyDay_i + \epsilon_i \end{aligned} \quad (1)$$

where $PropWinners_i$ is the standardized proportion of winners for draw i , $Covid_i$ is a dummy variable equal to 1 if draw i is from the Covid period, $Dummy19_i$ is another dummy variable set to 1 if 19 appears in draw i , $MeanDraw_i$ is the average value of numbers drawn at draw i , $Amount_i$ is the average per-player amount bet at draw i , $DummyLotto_i$ is equal to 1 if draw i is a Lotto draw, and $DummyDay_i$ is set to 1 for draws on either Tuesdays or Wednesdays. We cluster standard errors by week to address potential issues related to cross-sectional correlation.

Table 4 reports the results of estimating the regression model, in Panel A without control variables, and in Panel B with control variables. In each panel, we present three versions of the model. The first version includes the entire sample (836 draws), the second focuses on the COVID period (209 draws), and the third relates to the benchmark period (627 draws). The results provide several insights. First, 19 is a popular number since the coefficient estimate of $Dummy19$ is positive and significant (at the 1% or 5% level, depending on the panel) when considering both the benchmark and entire periods. However, this coefficient estimate is no longer significant for the COVID period. Second, the popularity of the number 19 significantly decreased during the COVID period, as shown by the coefficient estimate of the interaction dummy $Covid \times Dummy19$, which is negative and significant (at the 1% or 5% level, depending on the panel).

Furthermore, Panel B shows that only the coefficient estimate of the $MeanDraw$ variable is statistically significant, which indicates that the average value of numbers in a draw is negatively related to the proportion of winners in that draw, as expected. This is consistent with the most well-known example of conscious selection, the preference for small numbers.¹⁵ We did not find any significant relationship between the average amount bet

¹⁵ For the 418 Lotto draws, the correlation between the proportion of winners at ranks 1 and 3 and the average value of the numbers drawn is -0.392. For the 418 Euromillions draws, the corresponding correlation is -0.477 for ranks 1, 2, 3, 4, 5 and 7.

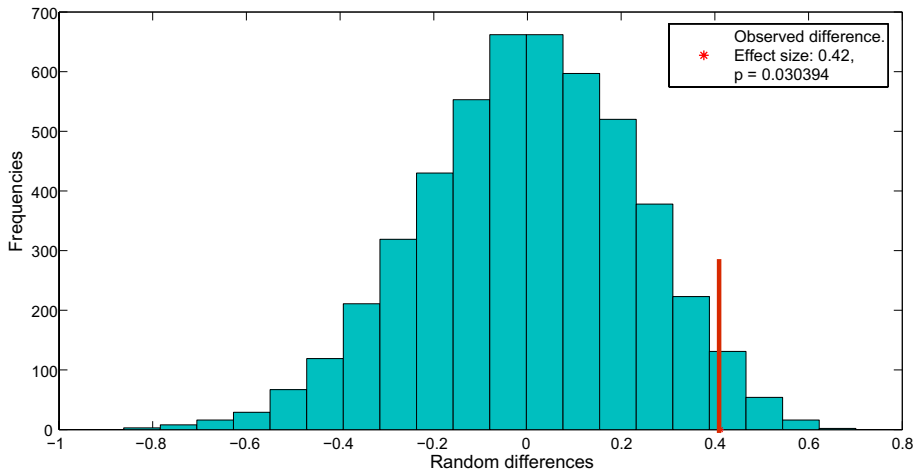


Fig. 2 Number 19: p-value of the permutation test (5,000 simulations of the combined Lotto and Euromillions samples). This figure shows a histogram of random differences (benchmark period - COVID period) over 5,000 simulations. The vertical line on the right shows the p-value of the permutation test, which is 3.04% for this set of simulations. Number 19 appeared in 81 (31) draws during the benchmark (COVID) period

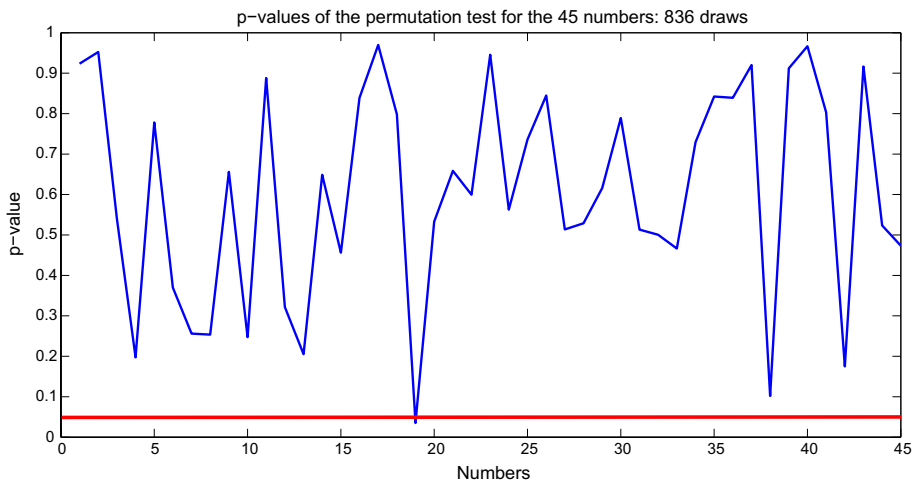


Fig. 3 All 45 numbers: p-values of the permutation test (5,000 simulations of the combined Lotto and Euromillions samples). The graph provides the p-values of the permutation test for the 45 numbers. For each number, the test is based on 5,000 simulations from the combined Euromillions and Lotto samples. The horizontal line corresponds to a 5% significance level

per player and the type of game; nor does the day of the week have any effect. Overall, our regression results point to a significant decline in the popularity of the number 19 in games of chance since the start of the COVID-19 pandemic. Our multivariate findings thus fully confirm the previous univariate results reported in Sect. 3.2.

Table 4 Evolution of the popularity of number 19: Benchmark period vs. Covid period

	Entire period	Covid period	Benchmark period
Panel A: Without control variables			
Intercept	– 0.0669	0.0612	– 0.0669
Covid	0.1281		
Dummy19	0.4094***	– 0.1291	0.4094***
Covid.Dummy19	– 0.5385***		
N	836	209	627
R ²	0.0153	0.0024	0.0183
Panel B: With control variables			
Intercept	– 0.0137	0.0473	0.0050
Covid	0.1012		
Dummy19	0.3010**	– 0.1369	0.3028**
Covidx Dummy19	– 0.4220**		
Mean Draw	– 0.4038***	– 0.3650***	– 0.4194***
Amount	– 0.0060	0.0608	– 0.0261
DummyLotto	– 0.0126	0.1431	– 0.0626
DummyDay	– 0.0600	– 0.1076	– 0.0454
N	836	209	627
R ²	0.1774	0.1587	0.1867

This table provides the results for the regression model depicted in Eq. (1), in Panel A without control variables and in Panel B with control variables. The dependent variable is the standardized proportion of winners for draw i , $Covid_i$ is a dummy variable equal to 1 if draw i is in the Covid period, $Dummy19_i$ is a dummy variable set to 1 if 19 shows up in draw i , $MeanDraw_i$ is the average value of numbers drawn at draw i , $Amount_i$ is the average per player amount bet at draw i , $DummyLotto_i$ is equal to 1 if draw i is a lotto draw, and $DummyDay_i$ is set to 1 for draws on either Tuesdays or Wednesdays. N gives the number of observations, that is, draws. Standard errors are clustered by week. ** and *** indicate statistical significance at the level of 5% and 1%, respectively

Survey Study

Previously, we used a field study to demonstrate that the popularity of the number 19 in games of chance declined during the COVID-19 pandemic, and our hypothesis was that this was the result of the negative associations created by the pandemic with the number 19. In our survey study, we addressed a potential shortcoming of the field study, which was the inability to control the extent to which individuals had been exposed to information related to COVID-19. There is no control group that has not been exposed to media coverage of the COVID-19 pandemic since the first mention of the virus in the media in late 2019. But in the survey study, we nevertheless manipulated players' exposure to COVID-19-related information to either increase or decrease negative associations with the number 19. Before answering the gambling-related questions and selecting Lotto numbers, we exposed one treatment group to an experimental manipulation intended to increase the saliency of the COVID-19 pandemic, thereby increasing the availability of negative associations with the number 19, aiming to replicate and better understand the field study results. We also measured the strength of the players' superstitious beliefs, as we expected that

only the most superstitious would form negative conscious or subconscious associations with the number 19 (e.g., Hirshleifer et al. (2018); Hoffmann et al. (2022)). In general, superstitious beliefs are an important driver of individuals' gambling activities and reflect a failure to appreciate the lack of a cause-and-effect relationship between their gambling behavior and subsequent events in the world around them (e.g., Rogers (1998)). More specifically, superstition is considered to be an important factor influencing lottery players' selection of specific numbers (Chou et al., 2009). Hence, we expect an interaction effect between the strength of players' superstitious beliefs and the extent to which our experimental manipulation affects their number preferences.

Data and Method

Data Collection

To test our hypothesis and match the country setting of the field study, we recruited 500 participants for our survey study from an online panel of Belgians managed by the data provider Qualtrics, using quotas to reflect, as much as possible, national statistics on the socio-demographics of the Belgian population (i.e., age, gender, income, and education). Because Belgium consists of two different language-speaking regions, the sample was equally distributed between French-speaking Walloons and Dutch-speaking Flemings. We excluded 11 participants with incomplete or invalid responses. The remaining 489 participants were randomly assigned to one of two different experimental conditions, only one of which primed the players with the salience of the COVID-19 pandemic. In the primed condition ($n = 241$), players first answered questions regarding their attitudes and experiences relating to COVID-19; while in the unprimed condition ($n = 248$), participants first answered gambling-related questions and selected Lotto numbers before answering questions regarding their attitudes and experiences with COVID-19.

Sample Description

The final sample of $N = 489$ participants was equally distributed between men ($n = 244$) and women ($n = 245$) and players from the Flemish ($n = 253$) and Walloon ($n = 236$) regions. The average age of the players was 44.9 years, and most players held university degrees (29.2% had a bachelor's degree and 15.3% had a master's degree), while a large fraction of the remaining sample had completed secondary school (37.8% of the global sample). Furthermore, 34.4% were married, 27.8% were single, and 14.3% were divorced. On average, the players had 1.52 underage children. In addition, the sample appeared to be representative of the national population of Belgium in terms of COVID experience. In response to the statement, "I had a confirmed COVID-19 infection," 7.57% of players completely agreed, marking 7 on a 7-point scale where 1 = completely disagree and 7 = completely agree, and 3.5% answered with a 5 or 6. These results suggest that approximately 10% of participants have been infected with COVID-19, which reflects the 1.2 million positive cases in Belgium at the time of the writing of this paper (approximately 10% of the Belgian population (September 2021)).¹⁶ Table 5 below provides descriptive statistics

¹⁶ See <https://datastudio.google.com/embed/reporting/c14a5cfc-cab7-4812-848c-0369173148ab/page/ZwmOB> for official statistics.

of the survey participants across the experimental conditions (i.e., primed vs. unprimed), which shows that they are similar according to the main socio-demographic variables. Furthermore, we found no significant differences among the experimental conditions in terms of any of the socio-demographic variables (all F-tests, $p > .10$), indicating that the random assignment of players to the two experimental conditions was successful, which is also evidenced by the similar number of participants in the primed ($n = 241$) and unprimed ($n = 248$) conditions.

Experimental Design

The experimental design of the survey study followed the recommendations within the psychology research literature (Spencer et al., 2005) and began with measuring sociodemographic variables and the moderator variable (i.e., the strength of participants' superstitious beliefs). Next, half of the participants were randomly selected to receive the experimental manipulation in order to increase the saliency of COVID-19 before (after) answering a series of other questions. In particular, participants in the primed condition were asked about their attitudes towards COVID-19 with the 10-item Brief Illness Perception Questionnaire (Moss-Morris et al., 2002). These questions related to COVID-19-induced anxiety and the players' perceived impact of COVID-19, including statements such as "I am worried about being infected by Covid-19", "I think Covid-19 is dangerous", and "Covid-19 affects my life a lot". Participants also answered questions pertaining to their actual experience with COVID-19 (e.g., whether they had been previously diagnosed with it or had experienced symptoms that made them suspect they had a COVID-19 infection). The number of players with a direct COVID-19 experience did not differ significantly between the primed and unprimed conditions. After the participants answered the COVID-19-related questions, we measured the dependent variable (i.e., their choice of numbers on the Lotto tickets). For participants in the unprimed condition, the order of the groups of questions was counterbalanced, in that we first measured the players' number selection on the lottery tickets before they answered the COVID-19-related questions.

Measurement Scales

To measure the strength of participants' superstitious beliefs, we used an established 4-item scale from Westjohn et al. (2017), which includes statements such as "I sometimes carry good luck charms with me" and "People who know me would say that I am superstitious". Cronbach's alpha value (a measure of a scale's reliability) is high at 0.84 (Netemeyer et al., 2003). To measure the dependent variable, participants had to select 6 numbers from 1 to 45 on two consecutive Lotto tickets. The ticket template corresponds to the national Belgian Lotto 6/45.

Common Method Variance

Given that the experiment was conducted through an online survey, we examined the potential of common method variance (CMV) to influence our results, relying on statistical approaches from the literature (Craighead et al., 2011). First, we performed Harman's single-factor test with an exploratory factor analysis based on the recommendations of Podsakoff et al. (2003) and confirmed that the different variables that we measured did not all load onto a single factor. Second, we included in our experiment a theoretically unrelated

Table 5 Sample description of participants across the experimental conditions (i.e., primed vs. unprimed)

	Primed	Unprimed
n	241	248
Age [mean (std.dev.), min-max]	45.08 (14.75), 18–82	44.8 (15.54), 18–82
Gender (male)	49%	50.8%
Language	51.9% Flemish 47.7% Walloon	51.6% Flemish 48.4% Walloon
Relationship status	53.6% married/in a relationship 24.5% single 19.1% divorced or separated 2.9% widowed	51.8% married/in a relationship 31% single 14.1% divorced or separated 2% widowed
Number of children [mean (std.dev.), min-max]	1.53 (1.03), 0–7	1.51 (0.88), 0–6
Education (highest completed)	8.7% primary school 42.8% secondary school 29.9% Bachelor or equivalent 16.2% Master or equivalent 0.4% PhD or equivalent	10.1% primary school 44.8% secondary school 28.6% Bachelor or equivalent 14.5% Master or equivalent 2% PhD or equivalent
Use of Quick Pick [mean (std.dev.), min-max]	4.63 (2.15), 1–7	4.52 (2.14), 1–7
Residential situation	40.2% rural 17% suburban 28.6% urban (not city-center) 13.7% city-center	39.9% rural 15.3% suburban 30.2% urban (not city-center) 14.5% city-center
Income (net per month) [mean (std. dev.), min-max]	€2618.90 (€1493.09), €0–€7000	€2701.40 (€1455.17), €0–€7000

The table provides descriptive statistics of the sample of the survey study. The primed group answered COVID-19 questions in the beginning of the survey. The sample is described according to the distribution among the categories or as a mean (standard deviation) and minimal and maximal value of the answer. There are no significant differences in any of the socio-demographic variables between the two experimental groups (all F-tests, $p > 0.10$)

question relating to the participants' attitude towards the color blue, a so-called "marker variable" as per Simmering et al. (2015) and perform a marker variable test (Lindell and Whitney, 2001). Participants' scores on the marker variable test did not correlate significantly with the key variables relating to their selection of the number 19 or their superstitious beliefs. Overall, these test results indicate that CMV bias did not affect the results of the survey study.

Results

We used ordinal logistic regression to test the extent to which priming of the players with the salience of COVID-19 impacted their selection of the number 19 in two consecutive lottery tickets. In particular, we estimated the following regression model:

$$Chose19_i = a_0 + a_1 Priming_i + \epsilon_i \quad (2)$$

where $Chose19_i$ is a categorical variable for participant i , coded as “0” if the number 19 is not selected in any of the two Lotto grids (357 out of 489 participants), “1” if the number 19 is selected in one of the two grids (109 out of 489 participants), and “2” if the number 19 is selected in both grids (23 out of 489 participants). $Priming_i$ is a dummy variable equal to 1 when participant i is in the COVID-19 primed condition. To test the robustness of this effect, we subsequently estimated the following regression in which we controlled for a set of individual characteristics among the participants:

$$Chose19_i = a_0 + a_1Priming_i + a_2SDI_i + a_3Covid19_i + \epsilon_i \quad (3)$$

where SDI_i is a vector of sociodemographic variables (i.e., age, gender, language, relationship status, number of minor children, education, residence, income) plus a dummy variable set to one for nonusers of the Quick Pick system. And $Covid19_i$ is another dummy variable set to one for participants who have had COVID-19.

Finally, we expected that players’ number selection (“conscious selection”) depends upon the strength of their superstitious beliefs (He et al., 2020; Hirshleifer et al., 2018) and that the COVID-19 pandemic therefore influenced their beliefs and perceptions of the number 19 (Dutta, 2020). To test this hypothesis, we measured the levels of participants’ superstitious beliefs by coding the level of participants’ superstitious beliefs as a dummy variable *Superstition* equaling 1 for participants above the median score on this variable (i.e., above 3.25):

$$Chose19_i = a_0 + a_1Priming_i + a_2Superstition_i + a_3Priming_i \times Superstition_i + a_4SDI_i + a_5Covid19_i + \epsilon_i \quad (4)$$

Table 6 presents the regression results. Model 1 shows the effect of priming alone; Model 2 shows the effect of priming, while accommodating for the effect of the control variables; and Model 3 additionally shows the effect of participants’ superstitious beliefs.

The results show that COVID-19 priming significantly reduces the probability that the number 19 will be selected by participants in their Lotto grids (at the 1% level), confirming the results of the field study. Furthermore, we found that COVID-19 priming maintains its significant effect (at the 1% level) while accounting for control variables.

Finally, according to the regression results, the interaction of the experimental manipulation and the superstitious beliefs dummy is significant (at the 5% level) while the main effect of priming is no longer significant. The fact that the main effect goes away when including the interaction term suggests that the effect of priming only holds for a certain group of individuals, namely, those with high superstitious beliefs. Building on Mize (2019), we performed a supplementary analysis to detect non-linear effects of the priming condition for different levels of superstition (see Appendix A.5). This procedure relies on marginal effects that allow one to estimate the distinct role of every independent variable on the model’s predictions (Mize, 2019). According to the observed average marginal effects (AMEs),¹⁷ there is a significant effect of priming among individuals with high superstitious beliefs (at 1% level). Considering second differences in AMEs indicating whether there is a significant difference in the reaction to the priming between individuals with low and high superstitious beliefs for every outcome level of the dependent variable

¹⁷ AMEs are calculated as the average of marginal effects computed for every observation and every independent variable. These results are extracted from Stata14 output based on Mize (2019)’s prescribed procedure (.do file with syntax).

(i.e., not choosing 19, choosing 19 in one draw, or choosing 19 in both draws), we confirm that priming was significantly more effective on the sub-group of individuals with high superstitious beliefs (at 5% level). Hence, there is an interaction effect between priming the salience of COVID-19 and superstitious beliefs on individuals' number preference in terms of selecting (or not) the number 19 in lotteries.¹⁸

Discussion

The results based on the primary data from the survey study and the secondary data from the field study fully converge. The consistency among the findings of these studies is highlighted in Figs. 4 and 5. Figure 4 depicts the overall preferences for specific numbers based on both field and survey data. Lottery numbers from 1 to 45 are found on the horizontal axis. The dashed curve gives the standardized popularity index measured over the entire four-year period of the field study, while the solid curve provides the standardized choice frequency of the entire sample of the survey study. Thanks to the standardization of each of the two sets of 45 popularity indices/choice frequencies, a direct comparison is meaningful. On the one hand, there is a substantial correlation of 0.798 between the two vectors. On the other hand, the two curves are arguably close to each other and display several similarities. First, the number 7 is an obvious outlier on each curve, revealing a much higher popularity/choice frequency than the other numbers. This is consistent with research reporting that the number 7 has been the most popular Lotto number in many countries for decades (e.g., Farrell et al. (2000), Roger and Broihanne (2007), Polin et al. (2021)). It is also noteworthy that the general shape of the two curves in Fig. 4 is close to what is typically found in the aforementioned work. In particular, numbers higher than 30 are unpopular, primarily due to the “birthday date effect”, resulting in an overall decrease in the curve from the numbers 30 to 45.

The strong correlation between the two vectors and the similar shapes of the two curves also reveal that the popularity index used in the field study is a reliable estimate of choice frequency, despite the unknown proportion of players who do not choose the lottery numbers themselves, but used the Quick Pick system. Moreover, we note that the survey study did not suffer from this shortcoming and contained two questions related to the Quick Pick system to further examine this issue. The first question asked regular players ($n = 151$) to grade their level of agreement with the statement, “When I play Lotto or Euromillions ...”. The available answers consisted of a seven-point scale ranging from -3 (“I always use Quick Pick”) to +3 (“I always select numbers myself”). Splitting the participants into two groups based on the superstition dummy,¹⁹ we found that superstitious players had a significantly higher tendency to select the numbers themselves (the mean difference between the groups of superstitious and non-superstitious players is 0.966, significant at the 1% level). Nonplayers were asked a similar question (with the same response scale) that began with

¹⁸ We performed a robustness check to test the influence of priming and superstitious beliefs on players' selection of the number 19 on at least one of the two Lotto tickets. The results are presented in Appendix A.6 and are consistent with those of the main analysis. Again, we found that the priming of players resulted in the number 19 being selected significantly less frequently than among players who had not been primed. Additionally, superstition interacts with priming so that superstitious individuals are less prone to choose the number 19 when primed than non-superstitious individuals.

¹⁹ This dummy variable is used in Eq. (4) and is set to 1 for participants above the median score of superstitious beliefs (i.e., 3.25).

Table 6 Ordinal logistic regressions on the effect of priming on selection of the number 19

	Model 1 (Eq. 2)	Model 2 (Eq. 3)	Model 3 (Eq. 4)
Priming	– 0.571***	– 0.552***	– 0.103
Superstition			0.456
Priming x Superstition			– 0.849**
Age		– 0.004	– 0.003
Gender		– 0.152	– 0.206
Language		0.219	0.196
Relationship		– 0.048	– 0.037
Minor Children		– 0.070	– 0.075
Education		– 0.054	– 0.047
No use of Quick Pick		0.005	0.015
Residence		0.034	0.029
Income		0.000	0.000
Covid19		0.038	0.031
N	489	489	489
Pseudo-R ²	0.011	0.018	0.024
Log Likelihood	– 342.336	– 339.890	– 337.815
Chi-Square	7.81	12.70	16.85

The table provides the results for ordinal logistic regressions (Eqs. 2 to 4) on the effect of the priming of the availability of the number 19 in the experiment on the selection of number 19 in two consecutive Lotto grids. In each model, the dependent variable encompasses the number of times the number 19 was chosen in the two grids from 1 to 45: none, in one of the grids, or in both grids. *Priming* is a dummy variable equal to 1 for participants in the Covid-19 primed condition. *Superstition* is a dummy variable coded as 0 for participants below the median level of superstitiousness, and 1 for those above the median level of superstitiousness. Both Models 2 and 3 include a set of participants' individual characteristics (including age, gender, language, relationship status, minor children, education, no use of Quick Pick system, residence, income) as well as *Covid19* which is a dummy variable set to one for participants having (had) Covid-19. *N* gives the number of participants, i.e., survey participants. The results are presented as coefficient estimates. ** and *** indicate statistical significance at the level of 5% and 1%, respectively

“If I had to play Lotto or Euromillions ...”, instead of “When I play Lotto or Euromillions ...”. Among the subset of people who never play the lottery ($n = 92$), we also found a significant difference between superstitious and non-superstitious individuals (the mean difference between the groups is 0.783, significant at the 5% level). These results provide further evidence of the complementarity of our two studies and the primary and secondary data we relied upon. Building upon the findings from the survey study, we determined that the difference in choice frequency for the number 19 between primed and unprimed lottery players was essentially driven by the choices of superstitious players. Therefore, we expected that the decline in the popularity of the number 19 during the COVID period, as reported in the field study, was also driven by preference shifts among superstitious people, who tend to select the numbers themselves when playing Euromillions/Lotto.

Figure 5 depicts the overall preferences of unprimed individuals for specific numbers when comparing data on unprimed individuals in the survey study to the field study over the benchmark period (Fig. 5A) or when comparing data on primed individuals in the survey study to field data over the COVID period (Fig. 5B). Again, the correlations between the two vectors are high (0.781 and 0.644, respectively), and in both cases, the general shape of the two curves displays obvious similarities. Moreover, as illustrated in Fig. 5B,

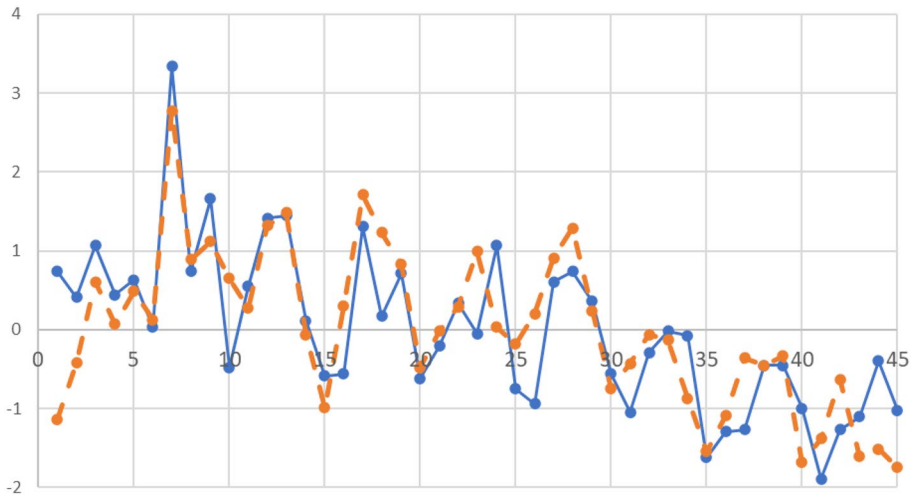


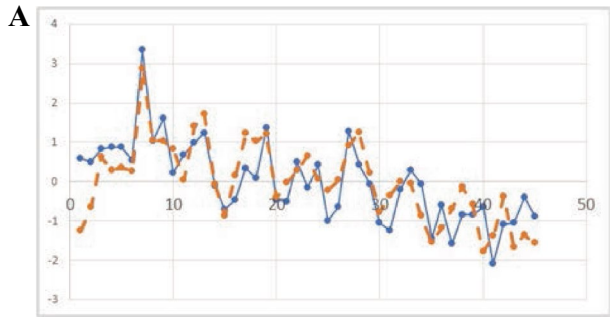
Fig. 4 Field and survey (standardized) data: Popularity index (entire period of field study) vs. frequency of choices (survey study, full sample) The solid curve represents the (standardized) number of times each number (on the horizontal axis) has been chosen by participants in the survey study when aggregating choices over the two Lotto tickets. The dashed curve represents the standardized popularity index for the 45 numbers in the field study over the entire four-year period. The correlation between the two variables is 0.798

the number 19 exhibits the strongest decline in popularity among the 45 numbers in both studies, that is, a decrease of 1.73 based on the popularity index (field data) and a decrease of 1.55 based on choice frequency (survey data). For comparison, the second highest decline is observed for the number 13 in the field study (with a decrease of 1.29) and the number 27 in the survey study (with a decrease of 1.45). In contrast, the number 17 exhibits the strongest gain in popularity in both studies (i.e., 1.65 in the field data and 1.83 in the survey data). The number 37 comes in second in the field study (with an increase of 1.40), while the number 24 exhibits the second highest popularity gain in the survey study (with an increase of 1.16). Unfortunately, the data do not allow us to test whether there was any (possibly unconscious) substitution effect on players' number choices (e.g., between 19 and 17).

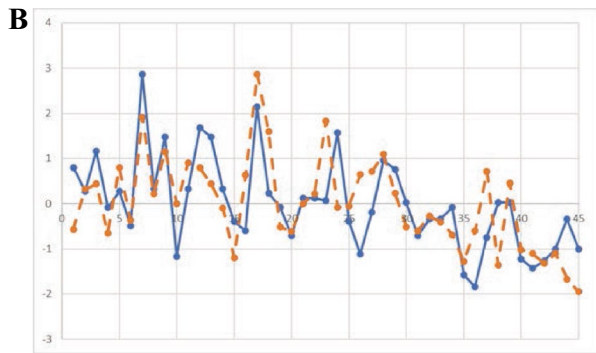
Finally, we analyzed participants' textual responses to an open question regarding their associations with the number 19 that was included at the end of the survey. The results show that the vast majority of participants (80.2%) maintain a neutral attitude toward the number 19 (e.g., by responding, "cannot think of anything" or "1 less than 20"), and that 15.3% maintain a positive association (e.g., by associating 19 with youth or the birthday of someone they are close to). Only 4.5% declared a negative association (e.g., by associating 19 with bad luck or directly with the COVID-19 pandemic). In total, when asked about their associations with the number 19, only 9 out of 489 participants explicitly mentioned COVID-19. Hence, we conclude that the COVID-19 pandemic mainly led to a *subconscious* aversion to the number 19 through the increased salience of COVID-19, reinforcing the availability of negative feelings/memories associated with this particular number. This subconscious aversion to the number 19 influences players to select it less frequently in games of chance.

Fig. 5 Field and survey (standardized) data. The solid curve represents the (standardized) number of times each number (on the horizontal axis) has been chosen by unprimed individuals in the survey study. The dashed curve represents the standardized popularity index for the 45 numbers in the field study over the three-year benchmark period. The correlation between the two variables is 0.781.

The solid curve represents the (standardized) number of times each number (on the horizontal axis) has been chosen by primed individuals in the survey study. The dashed curve represents the standardized popularity index for the 45 numbers in the field study over the one-year COVID period. The correlation between the two variables is 0.644



The solid curve represents the (standardized) number of times each number (on the horizontal axis) has been chosen by unprimed individuals in the survey study. The dashed curve represents the standardized popularity index for the 45 numbers in the field study over the three-year benchmark period. The correlation between the two variables is 0.781.



The solid curve represents the (standardized) number of times each number (on the horizontal axis) has been chosen by primed individuals in the survey study. The dashed curve represents the standardized popularity index for the 45 numbers in the field study over the one-year COVID period. The correlation between the two variables is 0.644.

Conclusion

The number 19 has been heard/read/seen/googled countless times since the beginning of the COVID-19 pandemic in March 2020. This study used Belgium as the testing ground for a natural experiment investigating whether the global pandemic has influenced the popularity of the number 19 among lottery players. To obtain a comprehensive understanding of the impact of the COVID-19 pandemic on the popularity of the number 19 in the context of lotteries, we triangulated evidence from field and survey data. Specifically, using Belgian National Lottery data, we analyzed a sample of 836 draws over the period March 2017 to February 2021. This sample included data on both Euromillions and the Lotto and allowed us to define the first three years as the benchmark period and the last year as the COVID period. To accurately measure the popularity of a given number in the lottery data, we built a popularity index that was inferred from the actual proportion of winners among a subset of ranks. The main advantage of this methodology was that it enabled us to identify any potential shift in conscious selection between the two periods.

Relying on nonparametric permutation tests, our univariate results led us to reject the null hypothesis of unchanged popularity of the number 19 between the two periods. Specifically, we found that 19 is the only number with a significant decline in popularity during the COVID period. Next, we ran regression models to control for potentially confounding effects that could impact the proportion of winners at a given draw. These multivariate findings confirmed a significant decline in the popularity of the number 19 since the start of the COVID-19 pandemic.

To complement our results from the field study that used data obtained from the Belgian lottery games, we performed a survey study in the same country. The advantage of this approach was twofold. First, the survey study allowed us to vary the extent to which participants were exposed to information about COVID-19. Second, the survey study allowed us to measure the strength of participants' superstitious beliefs, which is recognized as an important factor explaining lottery players' choice of numbers (e.g., Simon (1998), Farrell et al. (2000), Wang et al. (2016), Polin et al. (2021)). The survey data thus allowed us to check for an interaction effect between the strength of players' superstitious beliefs and the extent to which our experimental manipulation affected their number preferences.

Using ordinal logistic regression to determine the extent to which priming with the salience of COVID-19 impacts the survey participants' selection of the number 19 in the two consecutive Lotto tickets, we found evidence that COVID-19 priming significantly reduces the probability that players will select the number 19, which fully supports the results from the field study. Importantly, the effect of COVID-19 priming is still significant when controlling for participants' socio-demographic variables. Of particular interest, the findings reveal that the effect of COVID-19 priming is moderated by participants' superstitious beliefs, being concentrated amongst individuals with high levels of superstitious beliefs.

Overall, although there could be unobserved variables affecting the results, the results of both studies provide converging evidence that is consistent with our theoretical explanation that an increased salience of the number 19 due to the COVID-19 pandemic is associated with an increase in the availability of negative feelings or memories associated with this number, which is associated with a decline in the popularity of the number 19 in games of chance. The strong consistency between both studies is also evident from the substantial correlation between the overall preferences for specific numbers in the survey data and the field data. Furthermore, building upon the findings from the survey study showing that the difference in choice frequency for the number 19 between primed and unprimed players is mainly driven by the choices of superstitious players, we hypothesize that the actual decline in the popularity of the number 19 during the Covid period as observed in the field study is likewise most likely driven by the preferences of superstitious people, who are more likely to select numbers themselves when playing Euromillions/Lotto (as opposed to using the Quick Pick system to randomly select numbers for them). Finally, based on a textual analysis of the responses provided by the survey participants regarding their associations with the number 19 which only incidentally mentioned COVID-19, we conclude that the COVID-19 pandemic might mainly generate a subconscious aversion to the number 19, which could lead people to select it less frequently when playing games of chance.

Our results have various implications for public policy, raising several questions. First, the existence of conscious selection in lottery games indicates that people can take decisions that are not based on economic objectives. In particular, playing popular numbers

leads to paying an overvalued effective price.²⁰ However, the distribution of numbers actually chosen by lottery players remains unknown to the general public. Researchers have used various methods to estimate this distribution (e.g., Farrell et al. 2000; Roger and Broihanne 2007) or obtained real data for some draws (e.g., Simon 1998; Polin et al. 2021). At a time where open data policies develop in various countries,²¹ an important policy question is transparency around games of chance. Should players be able to know, after each draw, the aggregate distribution of choices (like Fig. 7)? Should players have free access to the entire set of combinations chosen by players at the preceding draw (like Fig. 8)? The well-being of players who bet unconsciously on popular numbers (for reasons that are not linked to religious considerations, superstitious beliefs, etc.) would be enhanced by such measures if they change their number selection based on this information.

Second, our results raise another, and even more important, question linked to education about probability theory. Economic decisions under risk and uncertainty require estimations of probability distributions. For at least four decades, it is well-known that people have a distorted understanding of probabilities (Kahneman and Tversky, 1979). This distortion seems to be related to intelligence/cognitive ability (Choi et al., 2022). Moreover, since probability distortion is already present at a young age (Steelandt et al., 2013), learning elementary probability theory through games of chance in the first years of school might be beneficial for improving people's future decision processes. In extreme cases, understanding what a uniform distribution is and what independent events are could potentially save lives. For example, in 2005, the number 53 did not show up in the Italian lottery for almost two years. As a result, some people had completely unreasonable reactions about this number, falling prey to the gambler's fallacy (e.g., Suetens et al. 2016). In an article published just after the number 53 was finally drawn,²² newspaper The Guardian reported that "Four died in 53-related incidents. A woman drowned herself in the sea off Tuscany leaving a note admitting that she had spent her family's savings on the number. A man from Signa near Florence shot his wife and son before killing himself. A man was arrested in Sicily this week for beating his wife out of frustration at debts incurred by his 53 habit."

Another insightful illustration of the need for a better understanding of probability theory in the general population is the case of saliva Covid tests for children. The analysis of a number of different tests by Kivelä et al. (2021) shows that the average specificity of these tests is 99%. In France, these tests were used on a systematic basis in schools when the prevalence of the disease was close to 0.5% in children. Bayes' theorem tells us that getting a positive test means a probability of being really positive of only one third.²³ In other words, two thirds of positive tests were false positive that may have contributed to potentially unjustified school closures.

To a broader extent, illustrating conscious selection for children would be easy and might also make them develop their critical thinking skills, a crucial skill for real-problem solving that is not included in standard IQ tests (Halpern and Dunn, 2021). For example,

²⁰ In the literature on state lotteries, "effective price" is the usual name for the difference between the cost of a ticket and the expected return (Forrest et al., 2002).

²¹ See <https://digital-strategy.ec.europa.eu/en/policies/open-data> for Europe and <https://www.archives.gov/data> for the U.S..

²² <https://www.theguardian.com/world/2005/feb/11/italy.sophiearie>.

²³ Explanation: test 10,000 children, 0.5% (prevalence) are positive, that is 50. However, a specificity of 99% means that 1% of tests will be declared positive when they are negative, that is roughly 100 tests. Overall, out of 150 positive tests, only 50 are really positive, that is one third.

following Kahneman and Tversky (1974), a classroom could be divided in two groups, each group being exposed to a draw of a wheel of fortune (numbered from 1 to 10). Then, the two groups are asked a given question (e.g., estimating the weight of a pet shown on a picture). The comparison of the resulting estimates in the two groups would show that these estimates are related to the number drawn on the wheel of fortune, which is completely irrelevant information to the task at hand.

Our results add to the literature on number preferences and on the interaction between emotions, preferences, and decision-making. One limitation of this study was that it does not provide insights into whether the observed change in the popularity of the number 19 is permanent or transitory. There are several arguments supporting the hypothesis that the effect may be transitory. First, the COVID-19 pandemic should eventually come to an end and/or be replaced by another worldwide topic of concern. Chun and Turk-Browne (2007) show that the limited capacity of human memory indicates that attention determines what is encoded by memory. Moreover, Roy et al. (2005) and Kress and Aue (2017) report that biased memories are the result of an optimism bias, making it likely that the negative association between the number 19 and bad memories will weaken over time. At the time of this writing (September 2021), there is also anecdotal evidence that the media and people in everyday life have begun to shorten “COVID-19” to “COVID”. In the future, this change in terminology in the popular press could lead to a (partial) disconnect in the public’s mind between the number 19 and any bad memories or feelings associated with COVID-19. Google searches for the number 19 already provide tangible evidence of such an evolution (<https://trends.google.com>). For example, Google searches for the number 19 halved from May to October 2021, after a sharp increase in March 2020 (followed by a decrease until June 2020 and a period of stability for almost one year, until April 2021).

This study paves the way for further research on the impact of the COVID-19 pandemic on number preferences, emotions, and associated decision-making. First, future research should address the aforementioned question whether the decline in the popularity of the number 19 is temporary or permanent. Second, it would also be of interest to investigate whether other games of chance (e.g., sport betting) or other types of decision-making involving a selection of numbers have been similarly affected by the COVID-19 pandemic. Finally, research using data from other countries that were impacted by the COVID pandemic in different respects than Belgium (e.g., because they are more geographically isolated or have less developed health systems) would provide additional insights into the interaction between emotions and decision-making.

A Appendix

A.1 Covid deaths in Belgium

See Fig. 6.

Cumulated number of COVID-19-related deaths in five countries

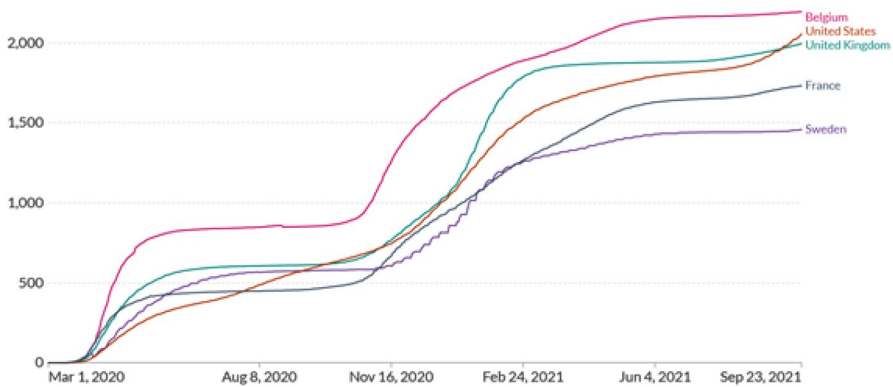


Fig. 6 Cumulated number of COVID-19-related deaths in five countries. This figure shows the evolution of the cumulative number of deaths per million inhabitants in five countries: Belgium, the United States, the United Kingdom, France, and Sweden. Source: Johns Hopkins University CSSE COVID-19 Data, through Our World in Data (<https://ourworldindata.org/>)

A.1 Winning Ranks and Winning Probabilities for the Belgian Lotto and the Euromillions Lottery

See Tables 7 and 8.

Table 7 Euromillions winning ranks and winning probabilities

Winning Ranks	Correct numbers	Correct stars	Combinations	Probability
1	5	2	1	7.15×10^{-9}
2	5	1	20	1.43×10^{-7}
3	5	0	45	3.22×10^{-7}
4	4	2	225	1.61×10^{-6}
5	4	1	4500	3.22×10^{-5}
6	3	2	9900	7.08×10^{-5}
7	4	0	10,125	7.24×10^{-5}
8	2	2	141,900	1.01×10^{-3}
9	3	1	198,000	1.42×10^{-3}
10	3	0	445,500	3.00×10^{-3}
11	1	2	744,750	5.33×10^{-3}
12	2	1	2,838,000	2.03×10^{-2}
13	2	0	6,385,500	4.57×10^{-2}

This table reports the 13 winning ranks of the Euromillions game and their associated odds and probabilities. Columns two and three indicate the number of correct numbers/stars of the winning rank indicated in the first column. Columns four and five give the number of winning combinations and the winning probability of the corresponding rank. The total number of combinations is 139,838,160. Let $N_p = 50$ and $N_s = 12$ denote the cardinals of the two sets of numbers (principal set and stars) used to draw the winning combination. Additionally, $n_p = 5$ ($n_s = 2$) is denoted as the number of numbers drawn in the principal set (set of stars), and finally k_p and k_s the number of correct numbers and stars in a given ticket. The probability of winning with k_p and k_s correct numbers and stars is equal to: $P(k_p, k_s) = \frac{C(n_p, k_p) C(N_p - n_p, n_p - k_p) C(n_s, k_s) C(N_s - n_s, n_s - k_s)}{C(N_p, n_p) C(N_s, n_s)}$ (5)

where $C(M, j)$ is the number of combinations of j numbers out of M without replacement. The cumulative probability of winning a strictly positive amount is approximately 7.71%

Table 8 Lotto winning ranks and winning probabilities

Winning Ranks	Correct numbers	Correct Bonus	Combinations	Probability
1	6	0	1	1.22×10^{-7}
2	5	1	6	7.37×10^{-7}
3	5	0	228	2.80×10^{-5}
4	4	1	570	7.00×10^{-5}
5	4	0	10,545	1.29×10^{-3}
6	3	1	14,060	1.73×10^{-3}
7	3	0	168,720	2.07×10^{-2}
8	2	1	126,540	1.55×10^{-2}
9	1	1	442,890	5.44×10^{-2}

This table reports the 9 winning ranks of the Lotto game and their associated odds and probabilities. Columns two and three indicate the number of correct numbers/bonus of the winning rank indicated in the first column. Columns four and five give the number of winning combinations and the winning probability of the corresponding rank. The total number of combinations is 8,145,060. Belgian Lotto was launched in 1978 and has experienced changes since then. The last important modification, the introduction of a 9th rank of gain, occurred on May 26, 2018. This new rank (1 correct number plus the bonus number) pays a fixed amount equal to the ticket price, which is currently €1.25 (€1 before May 26, 2018). Our methodology, developed in Sect. 3.2, is not impacted by this change because this new rank of gain does not play any role in the definition of our popularity index, which is based on the proportion of winners at ranks 1 and 3 only. Moreover, we do not use the amounts won by players. Therefore, the change in ticket price is used only to calculate the number of tickets sold. With the same notations as in Table 7, the probability of winning with k_p correct numbers and k_s bonus numbers ($k_s = 0$ or 1 in the Lotto game) is equal to $P(k_p, k_s)$ defined as:

$$P(k_p, k_s) = \frac{C(n_p, k_p) C(N_p - n_p - n_s, n_p - k_p - k_s)}{C(N_p, n_p)} \quad (6)$$

A.3 Examples of Data Files for Euromillions

See Figs. 7 and 8.

Date	Nombre 1	Nombre 2	Nombre 3	Nombre 4	Nombre 5	Etoile 1	Etoile 2
05/02/2021	12	19	37	44	45	2	6
02/02/2021	18	20	35	38	48	9	12
29/01/2021	1	5	27	36	42	1	6
26/01/2021	5	7	25	37	40	2	8
22/01/2021	8	16	42	44	47	6	7
19/01/2021	25	27	31	35	43	5	6
15/01/2021	4	10	27	38	40	3	11
12/01/2021	2	10	18	34	35	6	9
08/01/2021	18	23	37	41	42	4	6
05/01/2021	12	34	37	38	42	1	3
01/01/2021	16	28	32	44	48	1	9
29/12/2020	7	32	39	40	45	3	8
25/12/2020	16	21	27	30	32	3	5
22/12/2020	2	10	16	42	50	4	11

Fig. 7 Screenshot of the Euromillions game data file. Example of Euromillions game data (in French) provided by www.lotto.be

Date	Joueurs	Cash	P-R1	W-R1	P-R2	W-R2	P-R3	W-R3	P-R4	W-R4	P-R5	W-R5	P-R6	W-R6	P-R7	W-R7
05/02/2021	1130152	10217440,00	0	0,00	0	151022,60	1	26472,30	17	875,60	166	113,20	454	50,50	329	39,20
02/02/2021	490403	3466720,00	0	0,00	1	576287,00	0	13468,70	0	1498,20	19	147,10	63	82,90	97	45,80
29/01/2021	662492	5021062,50	0	0,00	1	238052,20	0	41727,50	2	1732,90	45	143,50	149	59,40	116	45,90
26/01/2021	432795	3012935,00	0	0,00	0	0,00	1	122065,60	2	1440,50	46	111,40	134	41,90	108	35,80
22/01/2021	691808	5388607,50	0	80566353,00	0	797450,90	0	93188,70	6	1658,60	54	148,10	135	67,50	136	49,60
19/01/2021	489278	3531905,00	0	0,00	1	99196,60	0	34775,80	1	1883,70	57	146,10	85	65,80	98	49,10
15/01/2021	642575	4903772,50	0	0,00	0	384315,10	0	59880,50	3	1271,60	52	124,60	121	63,00	140	38,80
12/01/2021	474108	3339202,50	0	0,00	0	562879,70	0	32888,50	0	1639,00	52	124,30	93	57,60	104	40,10
08/01/2021	635667	4781757,50	0	0,00	0	180403,70	0	28108,80	2	1117,60	43	133,80	163	59,50	105	48,80
05/01/2021	448310	3130980,00	0	0,00	0	131718,10	0	0,00	0	10093,30	40	160,20	82	70,80	61	54,00
01/01/2021	684978	5310420,00	0	44958140,00	0	109750,60	1	20520,40	4	1359,90	72	116,80	139	66,20	145	35,90
29/12/2020	512290	3666252,50	0	0,00	0	581409,40	0	33971,20	1	1511,60	40	154,30	92	61,90	66	57,20
25/12/2020	598926	4529777,50	0	0,00	0	0,00	0	207160,10	2	888,90	61	110,90	143	40,30	104	41,40
22/12/2020	533087	3859320,00	0	43704037,00	0	156220,30	0	48681,60	0	3032,60	40	141,30	110	69,50	89	43,30
18/12/2020	701801	5356090,00	0	0,00	0	192267,30	1	17974,40	2	699,80	113	61,90	164	43,60	197	27,30

Fig. 8 Screenshot of the Euromillions financial data file. Example of Euromillions financial data (in French) provided by www.lotto.be. Columns starting with “P” give the number of winners, and columns starting with “W” give the individual prizes. The entire file includes the results for the 13 ranks

A.4 Building the Popularity Index

For the sake of simplicity, we define the popularity index of a given number $n = 1, \dots, 45$, starting from a sample of N draws of the Lotto game. A direct transposition gives the index for the Euromillions game. Quantities will be possibly indexed by i corresponding to the draw, $T = B$ or C , corresponding to the Benchmark and COVID subperiods. The notations and definitions are the following:

- G_i is the number of tickets sold for draw i ;
- $\Theta(n, T)$ is the set of draws in which n appeared during subperiod T ;
- R is the relevant set of ranks used to calculate the popularity index. As mentioned in the text, $R = \{1, 3\}$ for the lotto game ($R = \{1, 2, 3, 4, 5, 7\}$ for the Euromillions game);
- $W_i(R)$ is the cumulated number of winners at draw i for the set of ranks R ;
- For $i \in \Theta(n, T)$, $\Gamma(n, i) = \frac{W_i(R)}{G_i}$ is the cumulated percentage of winners at ranks in R at draw i . This is an estimate of the popularity of n for a single draw in which n shows up.

- $\Gamma(n, T) = \frac{1}{\#\Theta(n, T)} \sum_{i \in \Theta(n, T)} \Gamma(n, i)$ is the average popularity score for number n over draws in subperiod T .

Applying the above formulas for the two subperiods B and C provides a time series of popularity scores for the Lotto game $S_L = \{\Gamma(n, B), \Gamma(n, C)\}$. The same process produces a second time series, S_E of popularity scores for the Euromillions game. To mix the scores of the two games, we need to standardize scores for each game because the probability of winning at ranks $R = \{1, 3\}$ for the lotto game is different from the probability of winning at ranks $R = \{1, 2, 3, 4, 5, 7\}$ for the Euromillions game. We therefore standardize each of the time series S_L and S_E in such a way that each of the time series has a mean equal to 0 and a standard deviation of 1.

Remember that Euromillions (Lotto) draws occur on Tuesdays and Fridays (Wednesdays and Saturdays). As a consequence, we can mix the two standardized time series and keep a standardized time series. In the end, each number n is characterized by a set of $\Theta(n, B \cup C)$ popularity scores. The popularity index of a number n on a subperiod $T = B, C$ is simply defined as the average of the standardized popularity scores over the subperiod under consideration.

A.5 Average Marginal Effects for Priming Condition and Level of Superstitious Beliefs

See Table 9.

Table 9 Results for how priming condition is associated with superstition: tests of average marginal effects (AMEs), average adjusted predictions (AAPs), and second differences (N = 489)

Effect of priming	$AME_{LowSuperstition}$	$AME_{HighSuperstition}$	Second diff.	$AAP_{LowSuperstition}$	$AAP_{HighSuperstition}$
Panel A: 19 is not selected					
Priming	0.024 (0.058)	0.186*** (0.054)	-0.163** (-0.080)	0.753*** (0.039)	0.818*** (0.035)
Control				0.729*** (0.041)	0.632*** (0.041)
Panel B: 19 is selected once					
Priming	-0.019 (0.045)	-0.144*** (0.042)	0.125** (0.062)	0.205*** (0.032)	0.153*** (0.029)
Control				0.224*** (0.034)	0.297*** (0.033)
Panel C: 19 is selected twice					
Priming	-0.005 (0.013)	-0.042*** (0.015)	0.037** (0.019)	0.040*** (0.011)	0.027*** (0.008)
Control				0.045*** (0.012)	0.070*** (0.016)

The table provides average marginal effects (AME), average adjusted predictions (AAP), and second differences (Second diff.) estimated for the ordinal logistic regression on the effect of number 19 availability primed in the experiment on the selection of the number 19 in two consecutive Lotto grids and the full interaction effect with Superstition (see Eq. (4)). Priming is a dummy variable equal to 1 for participants in the Covid-19 primed condition. Superstition is a dummy variable coded as 0 for participants below the median level of superstitiousness, and 1 for those above the median level of superstitiousness. AME factor level refers to the discrete change from the base level (i.e., control condition). Second difference shows the non-linear interaction effect between individuals with a low and high superstition level. Standard errors are reported in parentheses. ** and *** indicate statistical difference at the level of 5% and 1%, respectively

A.6 Robustness Check with Binary Logistic Regressions on the Effect of Priming on Selection of the Number 19 at least Once in Two Tickets

See Table 10.

Table 10 Binary logistic regressions on the effect of priming on selection of the number 19

	Model 1 (Eq. 2)	Model 2 (Eq. 3)	Model 3 (Eq. 4)
Priming	−0.591***	−0.572***	0.100
Superstition			0.419
Priming x Superstition			−0.891**
Age		−0.007	−0.007
Gender		−0.144	−0.189
Language		0.156	0.130
Relationship		−0.065	−0.054
Minor Children		−0.115	−0.121
Education		−0.062	−0.055
No use of Quick Pick		0.000	0.010
Residence		0.026	0.022
Income		0.000	0.000
Covid19		0.033	0.027
N	489	489	489
Pseudo-R ²	0.014	0.023	0.031
Log Likelihood	−281.053	−278.521	−276.348
Chi-Square	8.26	13.32	17.67

The table provides the results for binary logistic regressions (Eqs. 2 to 4) on the effect of number 19 availability primed in the experiment on the selection of number 19 in two consecutive Lotto grids. In each model, the dependent variable is a dummy equal to 1 if the player chose the number 19 at least once in two tickets. *Priming* is a dummy variable equal to 1 for participants in the Covid-19 primed condition. *Superstition* is a dummy variable coded as 0 for participants below the median level of superstitiousness, and 1 for those above the median level of superstitiousness. Both Models 2 and 3 include a set of participants' individual characteristics (including age, gender, language, relationship status, minor children, education, no use of Quick Pick system, residence, income) as well as *Covid19* which is a dummy variable set to one for participants having (had) Covid-19. *N* gives the number of participants, i.e., survey participants. The results are presented as coefficient estimates. **, and *** indicate statistical difference at the level of 5% and 1%, respectively

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis of Study 1 were performed by P. Roger and C. D'Hondt; data collection and analysis of Study 2 were performed by D. Plotkina and A. Hoffmann. All authors participated in drafting the manuscript and commented on previous versions of the manuscript. All authors approved the final manuscript.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. No funding was received for conducting this study.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the research institutions and did not require a separate ethics committee approval.

Informed consent This study was conducted from publicly available data and survey data with informed consent.

References

- Baker, R. D., & McHale, I. G. (2009). Modelling the probability distribution of prize winnings in the UK national lottery: Consequences of conscious selection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 813–834.
- Baker, R. D., & McHale, I. G. (2011). Investigating the behavioural characteristics of lottery players by using a combination preference model for conscious selection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 1071–1086.
- Bhattacharya, U., Kuo, W.-Y., Lin, T.-C., & Zhao, J. (2018). Do superstitious traders lose money? *Management Science*, 64, 3772–3791.
- Broihanne, M., Merli, M., & Roger, P. (2016). Diversification, gambling and market forces. *Review of Quantitative Finance and Accounting*, 47, 129–157.
- Brown, P., & Mitchell, J. (2008). Culture and stock price clustering: Evidence from the peoples' Republic of China. *Pacific-Basin Finance Journal*, 16, 95–120.
- Choi, S., Kim, J., Lee, E., & Lee, J. (2022). Probability weighting and cognitive ability. *Management Science* 0, null.
- Chong, T.T.-L., & Du, X. (2008). Hedonic pricing models for vehicle registration marks. *Pacific Economic Review*, 13, 259–276.
- Chou, M. C., Kong, Q., Teo, C., Wang, Z., & Zheng, H. (2009). Benford's law and number selection in fixed-odds numbers game. *Journal of gambling studies*, 25, 503–521.
- Chung, R., Darrat, A. F., & Li, B. (2014). Chinese superstition in U.S commodity trading. *Applied Economics Letters*, 21, 171–175.
- Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, 17, 177–184.
- Clotfelter, C. T., & Cook, P. J. (1993). Notes: The gambler's fallacy in lottery play. *Management Science*, 39, 1521–1525.
- Cook, P. J., & Clotfelter, C. T. (1993). The peculiar scale economies of Lotto. *The American Economic Review*, 83, 634–643.
- Craighead, C. W., Ketchen, D. J., Dunn, K. S., & Hult, G. T. M. (2011). Addressing common method variance: Guidelines for survey research on information technology, operations, and supply chain management. *IEEE Transactions on Engineering Management*, 58, 578–588.
- De Paola, M., Gioia, F., & Scoppa, V. (2014). Overconfidence, omens and gender heterogeneity: Results from a field experiment. *Journal of Economic Psychology*, 45, 237–252.
- Dek, T. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty*, 8, 309–317.
- Dhami, M. K., Belton, I. K., Merrall, E., McGrath, A., & Bird, S. M. (2020). Criminal sentencing by preferred numbers. *Journal of Empirical Legal Studies*, 17, 139–163.

- Dutta, A. (2020). Beliefs and superstitions about COVID-19: Observations in India under lockdown. *Journal of Scientific Temper* pp. 38–52.
- Farrell, L., Hartley, R., Lanot, G., & Walker, I. (2000). The demand for Lotto: The role of conscious selection. *Journal of Business & Economic Statistics*, 18, 228–241.
- Forrest, D., Simmons, R., & Chesters, N. (2002). Buying a dream: Alternative models of demand for lotto. *Economic Inquiry*, 40, 485–496.
- Fortin, N. M., Hill, A. J., & Huang, J. (2014). Superstition in the housing market. *Economic Inquiry*, 52, 974–993.
- Halpern, D. F., & Dunn, D. S. (2021). Critical thinking: A model of intelligence for solving real-world problems. *Journal of Intelligence*, 9, 22.
- He, J., Liu, H., Sing, T. F., Song, C., & Wong, W.-K. (2020). Superstition, conspicuous spending, and housing market: Evidence from Singapore. *Management Science*, 66, 783–804.
- Hirshleifer, D., Jian, M., & Zhang, H. (2018). Superstition and financial decision making. *Management Science*, 64, 235–252.
- Hoffmann, A. O., & Shefrin, H. (2014). Technical analysis and individual investors. *Journal of Economic Behavior & Organization*, 107, 487–511.
- Hoffmann, A. O. I., Plotkina, D., Roger P., & D'Hondt, C. (2022). Superstitious beliefs, locus of control, and feeling at risk in the face of Covid-19. *Personality and Individual Differences*, 196, 111718. <https://doi.org/10.1016/j.paid.2022.111718>
- Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kivelä, J. M., Jarva, H., Lappalainen, M., & Kurkela, S. (2021). Saliva-based testing for diagnosis of sars-cov-2 infection: A meta-analysis. *Journal of Medical Virology*, 93, 1256–1258.
- Kress, L., & Aue, T. (2017). The link between optimism bias and attention bias: A neurocognitive perspective. *Neuroscience & Biobehavioral Reviews*, 80, 688–702.
- Kumar, A. (2009). Who gambles in the stock market. *Journal of Finance*, 64, 1889–1933.
- Kumar, A., Page, J. K., & Spalt, O. (2016). Gambling and comovement. *Journal of Financial and Quantitative Analysis*, 51, 85–111.
- Kusev, P., Purser, H., Heilman, R., Cooke, A. J., Van Schaik, P., Baranova, V., et al. (2017). Understanding risky behavior: The influence of cognitive, emotional and hormonal factors on decision-making under risk. *Frontiers in Psychology*, 8, 1–10.
- Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86, 114–121.
- Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In Davidson, R. J., Scherer, K., H. H. Goldsmith, E. (eds.), *Series in affective science. Handbook of affective sciences*, Oxford University Press, Oxford, pp. 619–642.
- Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician*, 52, 127–132.
- Mitton, T., & Vorkink, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies*, 20, 1255–1288.
- Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, 6, 81–117.
- Moss-Morris, R., Weinman, J., Petrie, K., Horne, R., Cameron, L., & Buick, D. (2002). The revised illness perception questionnaire (IPQ-R). *Psychology and Health*, 17, 1–16.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks: Sage Publications.
- Papachristou, G., & Karamanis, D. (1998). Investigating efficiency in betting markets: Evidence from the Greek 6/49 Lotto. *Journal of Banking & Finance*, 22, 1597–1615.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Polin, B. A., Isaac, E. B., & Aharon, I. (2021). Patterns in manually selected numbers in the Israeli lottery. *Judgment and Decision Making*, 16, 1039–1059.
- Roger, P. (2011). Testing alternative theories of financial decision making: A survey study with lottery bonds. *Journal of Behavioral Finance*, 12, 219–232.
- Roger, P., & Broihamne, M.-H. (2007). Efficiency of betting markets and rationality of players: Evidence from the French 6/49 Lotto. *Journal of Applied Statistics*, 34, 645–662.

- Rogers, P. (1998). The cognitive psychology of lottery gambling: A theoretical review. *Journal of Gambling Studies*, 14, 111–134.
- Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005). Underestimating the duration of future events: memory incorrectly used or memory bias? *Psychological Bulletin*, 131, 738–756.
- Shum, M., Sun, W., & Ye, G. (2014). Superstition and lucky apartments: Evidence from transaction-level data. *Journal of Comparative Economics*, 42, 109–117.
- Simmering, M. J., Fuller, C. M., Richardson, H. A., Ocal, Y., & Atinc, G. M. (2015). Marker variable choice, reporting, and interpretation in the detection of common method variance: A review and demonstration. *Organizational Research Methods*, 18, 473–511.
- Simon, J. (1998). An analysis of the distribution of combinations chosen by UK national lottery players. *Journal of Risk and Uncertainty*, 17, 243–277.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89, 845–851.
- Stelandt, S., Broihanne, M., Romain, A., Thierry, B., & Dufour, V. (2013). Decision-making under risk of loss in children. *PLoS ONE*, 8, e52316.
- Suetens, S., Galbo-Jorgensen, C. B., & Tyran, J.-R. (2016). Predicting lotto numbers: A natural experiment on the gambler's fallacy and the hot-hand fallacy. *Journal of the European Economic Association*, 14, 584–607.
- Suetens, S., & Tyran, J.-R. (2012). The gambler's fallacy and gender. *Journal of Economic Behavior & Organization*, 83, 118–124.
- Turner, N. E. (2010). Lottery ticket preferences as indicated by the variation in the number of winners. *Journal of Gambling Studies*, 26, 421–439.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232.
- Vandewiele, M., d'Hondt, W., Didillon, H., Iwawaki, S., & Mwamwenda, T. S. (1986). Number and color preferences in four countries. *Perceptual and Motor Skills*, 63, 945–946.
- Wang, T. V., van Loon, R. J. P., van den Assem, M. J., & van Dolder, D. (2016). Number preferences in lotteries. *Judgment and Decision Making*, 11, 243–259.
- Westjohn, S. A., Roschk, H., & Magnusson, P. (2017). Eastern versus western culture pricing strategy: Superstition, lucky numbers, and localization. *Journal of International Marketing*, 25, 72–90.
- Woo, C.-K., Horowitz, I., Luk, S., & Lai, A. (2008). Willingness to pay and nuanced cultural cues: Evidence from Hong-Kong's license-plate auction market. *Journal of Economic Psychology*, 29, 35–53.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Patrick Roger¹ · Catherine D'Hondt²  · Daria Plotkina³ · Arvid Hoffmann⁴

Patrick Roger
proger@unistra.fr

Daria Plotkina
daria.plotkina@em-strasbourg.eu

Arvid Hoffmann
arvid.hoffmann@adelaide.edu.au

¹ LaRGE Research Center, EM Strasbourg Business School, University of Strasbourg, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France

² Louvain Finance (LIDAM), Louvain School of Management, UCLouvain, Chaussée de Binche 151, 7000 Mons, Belgium

³ HuManiS Research Center, EM Strasbourg Business School, 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France

⁴ University of Adelaide Business School, 10 Pulteney Street, Adelaide, South Australia 5005, Australia