mapping: condition 1: shape, color, and size of each element; condition 2: overall area (summation of the areas of all elements depicted in each stimulus); condition 3: overall perimeter (summation of the perimeters of all elements depicted in each stimulus) and density (the mean distance among the elements). Moreover, in condition 3, there was a negative correlation between overall area and number: The overall area of the 8 elements was larger than that of the 32 elements. Furthermore, the elements of each stimulus occupied the same overall spatial frame in conditions 2 and 3. If the overall area, in the presence of the same perimeter, was the crucial factor underlying number-space mapping, chicks would have chosen the right panel in the small number test and the left panel in the large number test. The results showed that in the small number test (8 versus 8), chicks chose the left panel 69.46% and the right panel 30.54% of the times. In the large number test, the left panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times.

30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test the left panel 69.46% and the right panel 30.54% of the times. In the large number test

Therefore, the results of experiment 3 demonstrate that spatial mapping relates to the abstract numerical magnitude, independently of non-numerical cues.

Our results indicate that a disposition to map numerical magnitudes onto a left-to-right-oriented MNL exists independently of cultural factors and can be observed in animals with very little nonsymbolic numerical experience, supporting a nativistic foundation of such orientation. Spatial mapping of numbers from left to right may be a universal cognitive strategy available soon after birth. Experience and, in humans, culture and education (e.g., reading habits and formal mathematics education) may modulate or even be modulated by this innate number sense.

During evolution, the direction of mapping from left to right rather than vice versa, although in principle arbitrary, may have been imposed by brain asymmetry, a common and ancient trait in vertebrates (22), prompted by a right hemisphere dominance in attending visuospatial and/or numerical information. Recent studies have suggested that numerical knowledge constitutes a domain-specific cognitive ability, with a dedicated neural substrate located in the inferior parietal cortices (1,23). Moreover, number-space mapping is implemented in humans through a topographical representation in the right posterior parietal cortex (24). Such topography has not yet been found in neurons responding to number in animals (25,26).

Because nonverbal numerical cognition is shared by many animal classes (1, 27, 28), we suggest that a similar predisposition to map numbers onto space is embodied in the architecture of the animal neural systems.

REFERENCES AND NOTES

1. S. Dehaene, The Number Sense (Oxford Univ. Press, New York, 2011).
2. R. S. Moyer, T. K. Landauer, Nature 215, 1519–1520 (1967).
3. J. F. Cartlton, E. M. Brannon, PLOS Biol. 5, e328 (2007).
4. D. Scarf, H. Hayne, M. Colombo, Science 334, 664 (2011).
5. E. S. Spelke, K. D. Knierer, Dev. Sci. 10, 89–96 (2007).
6. S. Dehaene, S. Bossini, P. Giraux, J. Exp. Psychol. Gen. 122, 371–396 (1993).
7. M. Zorzi, K. Priftis, C. Umiltà, Nature 417, 138–140 (2002).
8. C. Umiltà, K. Priftis, M. Zorzi, Exp. Brain Res. 192, 561–569 (2009).
9. The SNAPP effect refers to the fact that humans respond faster to small numbers when responses are executed on the left side rather than on the right, but they respond faster to large numbers when responses are executed on the right side rather than on the left.
10. E. Zebian, J. Cogn. Cuff, 5, 655–190 (2005).
11. S. Nakai, H. F. Fischer, W. M. Petrusch, Psychon. Bull. Rev. 16, 328–333 (2009).
12. M. D. de Hevia, L. Girelli, M. Addabbo, V. Macchi Cassia, PLOS ONE 9, e96412 (2014).
13. R. Rugani, L. Regolin, G. Vallortigara, J. Exp. Psychol. Anim. Behav. Process. 33, 21–31 (2007).
14. R. Rugani, D. M. Kelly, I. Szelest, L. Regolin, G. Vallortigara, Biol. Lett. 6, 290–292 (2010).
15. C. B. Drucker, E. M. Brannon, Cognition 132, 57–67 (2014).
16. R. Rugani, G. Vallortigara, B. Vallini, L. Regolin, Neuronbiol. Learn. Mem. 95, 231–238 (2011).
17. G. Jewell, M. E. McCourt, Neuropsychologia 38, 93–110 (2000).
18. B. Diekamp, L. Regolin, G. Gantziokin, G. Vallortigara, Curr. Biol. 15, R372–R373 (2005).
19. L. Regolin, Cortex 42, 101–103 (2006).
20. I. Adachi, PLOS ONE 9, e90373 (2014).
21. Materials and methods are available as supplementary materials on Science Online.
Financial data that include noncash and digital payments contain rich metadata on individuals’ behavior. About 60% of payments in the United States are made using credit cards (23), and mobile payments are estimated to soon top $1 billion in the United States (24). A recent survey shows that financial and credit card data sets are considered the most sensitive personal data worldwide (25). Among Americans, 87% consider credit card data as moderately or extremely private, whereas only 68% consider health and genetic information private, and 62% consider location data private. At the same time, financial data sets have been used extensively for credit scoring (26), fraud detection (27), and understanding the predictability of shopping patterns (28). Financial metadata have great potential, but they are also personal and highly sensitive. There are obvious benefits to having metadata data sets broadly available, but this first requires a solid understanding of their privacy.

To provide a quantitative assessment of the likelihood of identification from financial data, we used a data set D of 3 months of credit card transactions for 1.1 million users in 10,000 shops in an Organisation for Economic Co-operation and Development country (Fig. 1). The data set was simply anonymized, which means that it did not contain any names, account numbers, or obvious identifiers. Each transaction was time-stamped with a resolution of 1 day and associated with one shop. Shops are distributed throughout the country, and the number of shops in a district scales with population density ($r^2 = 0.51, P < 0.001$) (fig. S1).

We quantified the risk of reidentification of $D$ by means of unicity $\varepsilon$ (39). Unicity is the risk of reidentification knowing $p$ pieces of outside information about a user (29). We evaluate $\varepsilon_p$ of $D$ as the percentage of its users who are reidentified with $p$ randomly selected points from their financial trace. For each user, we extracted the subset $\mathcal{S}(I_p)$ of traces that match the $p$ known points $(I_p)$. A user was considered reidentified in this correlation attack if $|\mathcal{S}(I_p)| = 1$.

For example, let’s say that we are searching for Scott in a simply anonymized credit card data set (Fig. 1). We know two points about Scott: he went to the bakery on 23 September and to the restaurant on 24 September. Searching through the data set reveals that there is one and only one person in the entire data set who went to these two places on these two days. $|\mathcal{S}(I_p)|$ is thus equal to 1, Scott is reidentified, and we now know all of his other transactions, such as the fact that he spent $79.30 on January 28, 2021.
of increasing size and of our data by aggregating shops in clusters with its resolution. Coarsening the data along any dimension reduces the chances of reidentifying me. In fact, the approximate price of my coffee significantly increased when I was there this morning helps to reidentify me, as Fig. 2 (blue bars) shows that also knowing the price of the transaction from 1 day to up to 15 days. Finally, we increase the size of the bins for price \( a \) from 50 to 75% in practice, this means that the bin in which a transaction falls into will go from $5 to $16 (\( a = 0.50 \)) to $5 to $34 (\( a = 0.75 \)) (table S2).

Figure 3 shows that coarsening the data is not enough to protect the privacy of individuals in financial metadata data sets. Although unicity decreases with the resolution of the data, it only decreases slowly along the spatial (c), temporal (h), and price (a) axes. Furthermore, this decrease is easily overcome by collecting a few more points (table S1). For instance, at a very low resolution of \( h = 15 \) days, \( v = 350 \) shops, and an approximate price \( a = 0.50 \), we have less than an 80% chance of reidentifying an individual knowing four points (\( e_{10} > 0.8 \)) (table S1).

Furthermore, financial traces contain one additional column that can be used to reidentify an individual: the price of a transaction. A piece of outside information, a spatiotemporal tuple can become a triple: space, time, and the approximate price of the transaction. The data set contains the exact price of each transaction, but we assume that we only observe an approximation of this price with a precision \( a \) we call price resolution. Prices are approximated by bins whose size is increasing; that is, the size of a bin containing low prices is smaller than the size of a bin containing high prices. The size of a bin is a function of the price resolution \( a \) and of the median price \( m \) of the bin (29). Although knowing the location of my local coffee shop and the approximate time I was there this morning helps to reidentify me, Fig. 2 (blue bars) shows that also knowing the approximate price of my coffee significantly increases the chances of reidentifying me. In fact, adding the approximate price of the transaction increases, on average, the unicity of the data set by 22% (fig. S2, when \( a = 0.50 \), \( \Delta e_{4} = 0.22 \)).

The unicity \( e \) of the data set naturally decreases with its resolution. Coarsening the data along any or all of the three dimensions makes reidentification harder. We artificially lower the spatial resolution of our data by aggregating shops in clusters of increasing size \( v \) based on their spatial proximity (29). This means that we do not know the exact shop in which the transaction happened, but only that it happened in this geographical area. We also artificially lower the temporal resolution of the data by increasing the time window \( h \) of a transaction from 1 day to up to 15 days. Finally, we increase the size of the bins for price \( a \) from 50 to 75%. In practice, this means that the bin in which a $15.13 transaction falls into will go from $5 to $16 (\( a = 0.50 \)) to $5 to $34 (\( a = 0.75 \)) (table S2).

Figure 3 shows that unicity naturally decreases with the resolution of the data set on any or all of the three dimensions; with four spatiotemporal tuples [(A), no price] and with four spatiotemporal-price triples [(B), \( a = 0.75 \); (C), \( a = 0.50 \)]. Although unicity decreases with the resolution of the data, the decrease is easily overcome by collecting a few more points. Even at very low resolution (\( h = 15 \) days, \( v = 350 \) shops, price \( a = 0.50 \)), we have more than an 80% chance of reidentifying an individual with 10 points (\( e_{10} > 0.8 \)) (table S1).

Figure 4 shows that also knowing the location of the bin (\( v = 1 \), \( h = 1 \)). (A) It is significantly easier to reidentify women (\( e_{4} = 0.93 \)) than men (\( e_{4} = 0.89 \)). (B) The higher a person’s income is, the easier he or she is to reidentify. High-income people (\( e_{4} = 0.93 \)) are significantly easier to reidentify than medium-income people (\( e_{4} = 0.91 \)), and medium-income people are themselves significantly easier to reidentify than low-income people (\( e_{4} = 0.88 \)). Significance levels were tested with a one-tailed \( t \) test (\( P < 0.05 \)). Error bars denote the 95% confidence interval on the mean.

The unicity (\( e_{4} \)) when we lower the resolution of the data set on any or all of the three dimensions; with four spatiotemporal tuples [(A), no price] and with four spatiotemporal-price triples [(B), \( a = 0.75 \); (C), \( a = 0.50 \)]. Although unicity decreases with the resolution of the data, the decrease is easily overcome by collecting a few more points. Even at very low resolution (\( h = 15 \) days, \( v = 350 \) shops, price \( a = 0.50 \)), we have more than an 80% chance of reidentifying an individual with 10 points (\( e_{10} > 0.8 \)) (table S1).
Although future work is needed, it seems likely that most large-scale metadata data sets—for example, browsing history, financial records, and transportation and mobility data—will have a high unicity. Despite technological and behavioral differences (Fig. 5B and fig. S3), we showed credit card records to be as reidentifiable as mobile phone data and their unicity to be robust to coarsening or noise. Like credit card and mobile phone metadata, Web browsing or transportation data sets are generated as side effects of human interaction with technology, are subjected to the same idiosyncrasies of human behavior, and are also sparse and high-dimensional (for example, in the number of Web sites one can visit or the number of possible entry-exit combinations of metro stations). This means that these data can probably be relatively easily reidentified if released in a simply anonymized form and that they can probably not be anonymized by simply coarsening of the data.

Our results render the concept of PII, on which the applicability of U.S. and European Union (EU) privacy laws depend, inadequate for metadata data sets (18). On the one hand, the U.S. specificities approach—for which the lack of names, home addresses, phone numbers, or other listed PII is enough to not be subject to privacy laws—is obviously not sufficient to protect the privacy of individuals in high-unicity metadata data sets. On the other hand, open-ended definitions expanding privacy laws to “any information concerning an identified or identifiable person” (30) in the EU proposed data regulation or “[when the] re-identification to a particular person is not possible” (31) for Deutsche Telekom are probably impossible to prove and could very strongly limit any sharing of the data (32).

From a technical perspective, our results emphasize the need to move, when possible, to more advanced and probably interactive individual (33) or group (34) privacy-conscious technologies, as well as the need for more research in computational privacy. From a policy perspective, our findings highlight the need to reform our data protection mechanisms beyond PII and anonymity and toward a more quantitative assessment of the likelihood of reidentification. Finding the right balance between privacy and utility is absolutely crucial to realizing the great potential of metadata.

REFERENCES AND NOTES
1. S. Higginbotham, “For science, big data is the microscope of the 21st century” (2011); http://gigaom. com/2011/11/08/for-science-big-data-is-the-microscope-of-the-21st-century/.
2. D. Lazer et al., Science 323, 721–723 (2009).
3. J. Giles, Nature 488, 448–450 (2012).
4. D. J. Watts, Winter Issue of The Bridge on Frontiers of Technology, Engineering, and Society (2012).
5. A. Wesolowski et al., Science 338, 267–270 (2012).
6. S. Charaudeau, K. Pakdaman, P.-Y. Boëlle, PLOS ONE 9, e83002 (2014).
7. N. Eagle, M. Macy, R. Claxton, Science 328, 1029–1031 (2010).
8. V. Padmanabhan, R. Ramjee, P. Mohan, U.S. Patent 8,423,255 (2013).
9. G. Boulton, Nature 486, 441 (2012).
10. M. McNutt, Science 346, 679 (2014).
11. T. Bloom, “Data access for the open access literature: PLOS’s data policy” (2013); www.plos.org/data-access-for-the-open-access-literature-plosoa-data-policy.
12. K. Burns, “In US cities, open data is not just nice to have; it’s the norm” The Guardian, 21 October 2013; www.theguardian.com/local-government-network/2013/oct/21/open-data-us-san-francisco.
13. Massachusetts Bay Transportation Authority, “Real-time commuter rail data” (2010); www.mbta.com/ rider_tools/ developers/default.asp?id=21899.
14. Y.-A. de Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, V. D. Blondel, D4D-Senegal: The second mobile phone data for development challenge (2014); http://arxiv.org/abs/1407.4889.
15. V. D. Blondel et al., Data for Development: The D4D challenge on mobile phone data (2012); http://arxiv.org/ abs/1210.0137.
16. P. Matchler, “MetaPhone: The sensitivity of telephone metadata” (2014); http://webpolicy.org/2014/03/12/ metaphone-the-sensitivity-of-telephone-metadata/.
17. Y.-A. de Montjoye, J. Quoidbach, F. Robic, A. Pentland, Predicting personality using novel mobile phone-based metrics in Proc. SBP (Springer, Berlin, Heidelberg, 2013), pp. 48–55.
18. P. M. Schwartz, D. J. Solove, Calif. Law Rev. 102, 877–916 (2014).
19. Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, V. D. Blondel, Sci. Rep. 3, 1376 (2013).
20. A. Narayanan, V. Shmatikov, Robust de-anonymization of large sparse datasets in IEEE Symposium on Security and Privacy, Oakland, CA, 18 to 22 May 2008 (IEEE, New York, 2008), pp. 111–125.
21. A. C. Solomon, R. Hill, E. Janssen, S. A. Sanders, J. R. Heiman, Uniqueness and how it impacts privacy in health-related social science datasets in Proc. IHI (Association for Computing Machinery, New York, 2012), pp. 523–532.
22. L. Sweeney, Int. J. Unc. Fuzz. Knowl. Based Syst. 10, 557–570 (2002).
23. 2013 Federal Reserve payments study (2013); www.frbsereserve.org/files/communications/pdf/research/ 2013, payments_study_summary.pdf.
24. eMarketer, “US mobile payments to top $1 billion in 2013” (2013); www.emarketer.com/commerce/US-Mobile-Payments- Top-1-Billion/2013/1001035.
25. “The trust advantage: How to win with big data” (2013); www.rivispective.com/content/articles/information_technology_strategy_consumer_products_trust_advantage_win_big_data/.
26. C.-L. Huang, M.-C. Chen, C.-J. Wang, Expert Syst. Appl. 33, 847–856 (2007).
27. S. Bhattacharyya, S. Jha, K. Tharakan, J. C. Westland, Decis. Support Syst. 50, 602–613 (2011).
28. C. Krumme, A. Llorente, M. Cebrían, A. S. Pentland, E. Moro, Sci. Rep. 3, 1645 (2013).
29. Materials and methods data are available as supplementary materials on Science Online.
30. European Commission, “General data protection regulation” (2012); http://ec.europa.eu/justice/data-protection/ document/review2012/com_2012_11_en.pdf.
31. Deutsche Telekom, “Guiding principle big data” (2014); www. telekom.de/static/~/209808/1/guiding-principles-big-data-si.
32. Y.-A. de Montjoye, J. Kendall, C. Kerry, Enabling Humanitarian Use of Mobile Phone Data: Brockles Issues in Technology Innovation Proceedings (Brockles Institution, Washington, DC, 2014) vol. 26.
33. Y.-A. de Montjoye, S.S. Wang, A.S. Pentland, IEEE Data Eng. Bull. 35, 5–8 (2012).
34. C. Dwark, in Automata, Languages and Programming (Lecture Notes in Computer Science Series, Springer, Berlin, Heidelberg, 2006), vol. 4052, pp. 1–12.

ACKNOWLEDGMENTS
For contractual and privacy reasons, we unfortunately cannot make the raw data available. Upon request we can, however, make individual-level data of gender, income level, resolution (N, v, a), and unicity (true, false), along with the appropriate documentation, available for replication. This allows the re-creation of Figs. 2 to 4, as well as the GLM model and all of the unicity statistics. A randomly subsampled data set for the four points case can be found at http://web.media.mit.edu/~yva/uniquethehopingmall/ and in the supplementary materials. This work was supported in part by the Gecoway Initial Training Network funded by the European Commission as an FP7-People Marie Curie Action under grant agreement number 264994, and in part by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. Y.-A.d.M. was partially supported by the Belgian American Educational Foundation and Wallonie-Bruxelles International. L. R. did part of this work while visiting the MIT Media Lab. We gratefully acknowledge B. Bokska and a bank that wishes to remain anonymous for access to the data. Views and conclusions in this document are those of the authors and should not be interpreted as representing the policies, either expressed or implied, of the sponsors.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/347/6221/536/suppl/DC1 Materials and Methods Figs. S1 to S5 Tables S1 and S2 Algorithms S1 and S2 Reference (35) Subsampled Data 20 May 2014; accepted 23 December 2014 10.1126/science.1256207.
Unique in the shopping mall: On the reidentifiability of credit card metadata
Yves-Alexandre de Montjoye, Laura Radaelli, Vivek Kumar Singh and Alex “Sandy” Pentland

Science 347 (6221), 536-539,
DOI: 10.1126/science.1256297