RESEARCH ARTICLE

Human behavioral complexity peaks at age 25

Nicolas Gauvrit[1,2], **Hector Zenil**[1,3,4]*, **Fernando Soler-Toscano**[1,5], **Jean-Paul Delahaye**[1,6], **Peter Brugger**[1,7]

1 Algorithmic Nature Group, Laboratoire de Recherche Scientifique LABORES For the Natural and Digital Sciences, Paris, France, **2** Human and Artificial Cognition Lab, EPHE, Paris, France, **3** Department of Computer Science, University of Oxford, Oxford, United Kingdom, **4** Information Dynamics Lab, Unit of Computational Medicine, Department of Medicine Solna, Centre for Molecular Medicine and Science for Life Laboratory (SciLifeLab), Karolinska Institute, Stockholm, Sweden, **5** Grupo de Lógica, Lenguaje e Información. Universidad de Sevilla, Sevilla, Spain, **6** Centre de Recherche en Informatique, Signal et Automatique de Lille (CRISTAL), UMR CNRS 9189, University of Lille 1, Lille, France, **7** Department of Neurology, Neuropsychology Unit, University Hospital, Zurich, Switzerland

* hector.zenil@algorithmicnaturelab.org

Abstract

Random Item Generation tasks (RIG) are commonly used to assess high cognitive abilities such as inhibition or sustained attention. They also draw upon our approximate sense of complexity. A detrimental effect of aging on pseudo-random productions has been demonstrated for some tasks, but little is as yet known about the developmental curve of cognitive complexity over the lifespan. We investigate the complexity trajectory across the lifespan of human responses to five common RIG tasks, using a large sample (n = 3429). Our main finding is that the developmental curve of the estimated algorithmic complexity of responses is similar to what may be expected of a measure of higher cognitive abilities, with a performance peak around 25 and a decline starting around 60, suggesting that RIG tasks yield good estimates of such cognitive abilities. Our study illustrates that very short strings of, i.e., 10 items, are sufficient to have their complexity reliably estimated and to allow the documentation of an age-dependent decline in the approximate sense of complexity.

Author summary

It has been unclear how this ability evolves over a person's lifetime and it had not been possible to be assessed with previous classical tools for statistical randomness. To better understand how age impacts behavior, we have assessed more than 3,400 people aged 4 to 91 years old. Each participant performed a series of online tasks that assessed their ability to behave randomly. The five tasks included listing the hypothetical results of a series of 12 coin flips so that they would "look random to somebody else," guessing which card would appear when selected from a randomly shuffled deck, and listing the hypothetical results of 10 rolls of a die. We analyzed the participants' choices according to their algorithmic randomness, which is based on the idea that patterns that are more random are harder to encode in a short computer program. After controlling for characteristics such as gender, language, and education. We have found that age was the only factor that affected the ability to behave randomly. This ability peaked at age 25, on average, and

declined from then on. We also demonstrate that a relatively short list of choices, say 10 hypothetical coin flips, can be used to reliably gauge randomness of human behavior. A similar approach could be then used to study potential connections between the ability to behave randomly, cognitive decline, neurodegenerative diseases and abilities such as human creativity.

Introduction

Knowledge forming the content of several academic fields, including mathematics, follows from precocious core knowledge [1–3], and then follows a specific developmental course along the lifespan [4]. Numerosity (our approximate implicit sense of quantity) has been a privileged target of recent research, because numbers form one of the main pillars of elementary mathematical knowledge [5], but the study of randomness perception and statistical reasoning has also yielded striking results in the field of probability: adults with no formal education [6] as well as 8 to 12 month-old children [7, 8] have the wherewithal for simple implicit probabilistic reasoning. One of the toughest problems when it comes to Bayesian reasoning, however, is the detection of randomness, i.e., the ability to decide whether an observed sequence of events originates from a random source as opposed to produced by a deterministic origin [9].

Formally, the algorithmic (Kolmogorov-Chaitin) complexity of a string is the length of the shortest program that, running on a universal Turing machine (an abstract general-purpose computer), produces the string and halts. The algorithmic complexity of a string is a measure of how likely it is to have been produced deterministically by a computer program rather than by chance. In this way, a random string is a string that cannot be compressed by any means, neither statistically or algorithmically, that is a string for which no computer program shorter than the string itself exists. Humans, adults and infants [10, 11], have a keen ability to detect structure, both of statistic and algorithmic nature (e.g. 0101. . . and 1234. . .) that only algorithmic complexity can intrinsically capture (as opposed to e.g. entropy rate).

Within the field of study devoted to our sense of complexity, the task of randomly arranging a set of alternatives is of special interest, as it poses almost insurmountable problems to any cognitive system. The complexity of a subject-produced pseudorandom sequence may serve as a direct measure of cognitive functioning, one that is surprisingly resistant to practice effects [12] and largely independent of the kind of alternatives to be randomized, e.g., dots [13], digits [14], words [15], tones [16] or heads-or-tails [17]. Although random item generation (RIG) tasks usually demand vocalization of selections, motor versions have comparable validity and reliability [18, 19]. RIG tasks are believed to tap our approximate sense of complexity (ASC), while also drawing heavily on focused attention, sustained attention, updating and inhibition [20, 21]. Indeed to produce a random sequence of symbols, one has to avoid any routine and inhibit prepotent responses. The ability to inhibit such responses is a sign of efficient cognitive processing, notably a flexibility assumed to be mediated by the prefrontal cortex.

Instructions may require responding at various speeds [22], or else the generation of responses may be unpaced [27]. Participants are sometimes asked to guess a forthcoming symbol in a series ("implicit randomization", [51]), or vaguely instructed to "create a random-looking string" [23]. The consensus is that, beyond their diversity, all RIG tasks rely heavily on an ASC, akin to a probabilistic core knowledge [24, 25].

Theoretical accounts of the reasons why RIG tasks are relevant tests of prefrontal functions are profuse, but pieces of experimental evidence are sparse. Sparse empirical factors indirectly

validate the status of RIG tasks as measures of controlled processing, such as the detrimental effect of cognitive load or sleep deprivation [26] or the fact that they have proved useful in the monitoring of several neuropsychological disorders [27–31].

As a rule, the development of cognitive abilities across the lifespan follows an inverse U-shaped curve, with differences in the age at which the peak is reached [4, 32]. The decrease rate following the peak also differs from one case to another, moving between two extremes. "Fluid" components tend to decrease at a steady pace until stabilization, while "crystalized" components tend to remain high after the peak, significantly decreasing only in or beyond the 60s [33]. Other evolutions may be thought of as a combination of these two extremes.

Two studies have addressed the evolution of complexity in adulthood, but with limited age ranges and, more importantly, limited 'complexity' measures. The first [22] compared young and older adults' responses and found a slight decrease in several indices of randomness. The second [15] found a detrimental effect of aging on inhibition processes, but also an increase of the cycling bias (a tendency to postpone the re-use of an item until all possible items have been used once), which tends to make the participants' productions more uniform. In both studies, authors used controversial indices of complexity that only capture particular statistical aspects, such as repetition rate or first-order entropy. Such measures have proved some usefulness in gauging the diversity and type of long sequences (with e.g., thousands of data points) such as those appearing in the study of physiological complexity in [34–36], but are inadequate when in comes to short strings (e.g., of less than a few tens of symbols), such as the strings typically examined in the study of behavioral complexity. Moreover, such indexes are only capable of detecting statistical properties. Authors have called upon algorithmic complexity to overcome these difficulties [37, 38]. However, because algorithmic complexity is uncomputable, it was believed to have no practical interest or application. In the last years, however, methods were introduced related to algorithmic complexity that are particularly suitable for short strings [39, 40], and native n-dimensional data [41]. These methods are based on a massive computation to find short computer programs producing short strings and have been made publicly available [42] and have been successfully applied in a range of different applications [41, 43, 44].

The main objective of the present study is to provide the first fine-grained description of the evolution over the lifespan of the (algorithmic) complexity of human pseudo-random productions. Secondary objectives are to demonstrate that, across a variety of different tasks of random generation, the novel measure of behavioral complexity does not rely on the collection of tediously long response sequences as hitherto required. The playful instructions to produce brief response sequences by randomizing a given set of alternatives are suitable for children and elderly people alike, can be applied in work with various patient groups and are convenient for individual testing as well as Internet-based data collection.

Participants with ages ranging from 4 to 91 performed a series of RIG tasks online. Completion time (CT) serves as an index of speed in a repeated multiple choice framework. An estimate of the algorithmic complexity of (normalized) responses was used to assess randomization performance (e.g., response quality). The testing hypothesis is that the different RIG tasks are correlated, since they all rely on similar core cognitive mechanisms, despite their differences. To ensure a broad range of RIG measurements, five different RIG tasks were selected from the most commonly used in psychology.

The experiment is some sort of reversed Turing test where humans are asked to produce configurations of high algorithmic randomness that are then compared to the occurrence of what computers can produce by chance according to the theory of algorithmic probability [39, 40].

Table 1. Description of the 5 RIG tasks used in the experiment. Order was fixed across participants.

Task	Description
Tossing a coin	Participants had to create a series of 12 head-or-tails that would "look random to somebody else" by clicking on one of the two sides of a coin appearing on the screen. The resulting series was not visible on the screen (the participant could only see the last choice made).
Guessing a card	Participants had to select one of 5 types of cards (Zener cards; see e.g. [45]), ten times. In contrast to the other tasks, they were not asked to make the result look random. Instead, they were asked to guess which card will appear after a random shuffle.
Rolling a die	Participants had to generate a string of 10 numbers between 1 and 6, as random as possible ("the kind of sequence you'd get if you really rolled a die"). In contrast to the preceding cases, they could here see all previous choices, but could not change any of them.
Pointing to circles	Participants had to point 10 times at one out of 9 circles displayed simultaneously on the screen. They could not see their previous choices. This task is an adaptation of the classical Mittenecker pointing test [13].
Filling a grid	Participants had to blacken cells in a 3x3 grid such that the result would look randomly patterned, starting from a white grid. In contrast to the other tasks, they could see their choice and click as many times as they wished. Clicking on a white cell made it black, and vice versa.

https://doi.org/10.1371/journal.pcbi.1005408.t001

Methods

Ethics statement

This study was approved by the University of Zurich Institutional Review Board (Req00583).

The five tasks used, described in Table 1, are purposely different in ways that may affect the precise cognitive ability that they estimate. For instance, some tasks draw on short-term memory because participants cannot see their previous choices (e.g., "pointing to circles"), whereas in other tasks memory requirements are negligible, because the participant's previous choices remain visible ("rolling a die"). Completion times across the various tasks showed a satisfactory correlation (Cohen's α = .79), suggesting that participants did not systematically differ in the cognitive effort they devoted to the different tasks. Any difference between task-related complexities is thus unlikely to be attributable to differences in time-on-task.

Complexities were weakly to moderately positively correlated across the different tasks (Cohen's α = .45), mostly as a consequence of the "filling the grid" task being almost uncorrelated with the other tasks (for more details, see *SI Principal Component Analysis*). Despite this moderate link, however, all trajectories showed a similar pattern across the lifespan, with a peak around 25, a slow, steady decline between 25 and 60, followed by accelerated decline after 60 as shown in Fig 1.

The hypothesis to test that the different tasks are positively related to each other was partially supported by the data, especially in view of the results obtained on the "filling the grid" task. At the same time, CTs showed correlations supporting the testing hypothesis together with the developmental complexity curves in agreement pointing in the same direction. This suggests that all tasks do tap into our ASC as well as into other cognitive components with similar developmental trajectories, but that different tasks actually require different supplementary abilities or else they weight the components of these abilities differently.

The "filling the grid" task appeared unique in that it was loosely correlated with all the other tasks. The fact that it required binary responses cannot account for this lack of association, since the "tossing a coin" task yielded results uncorrelated with the "filling the grid" responses. Bi-dimensionality could possibly have had an effect, but the "pointing to circles" task was also unrelated to the grid task. On the other hand, one factor distinguished the grid task from all
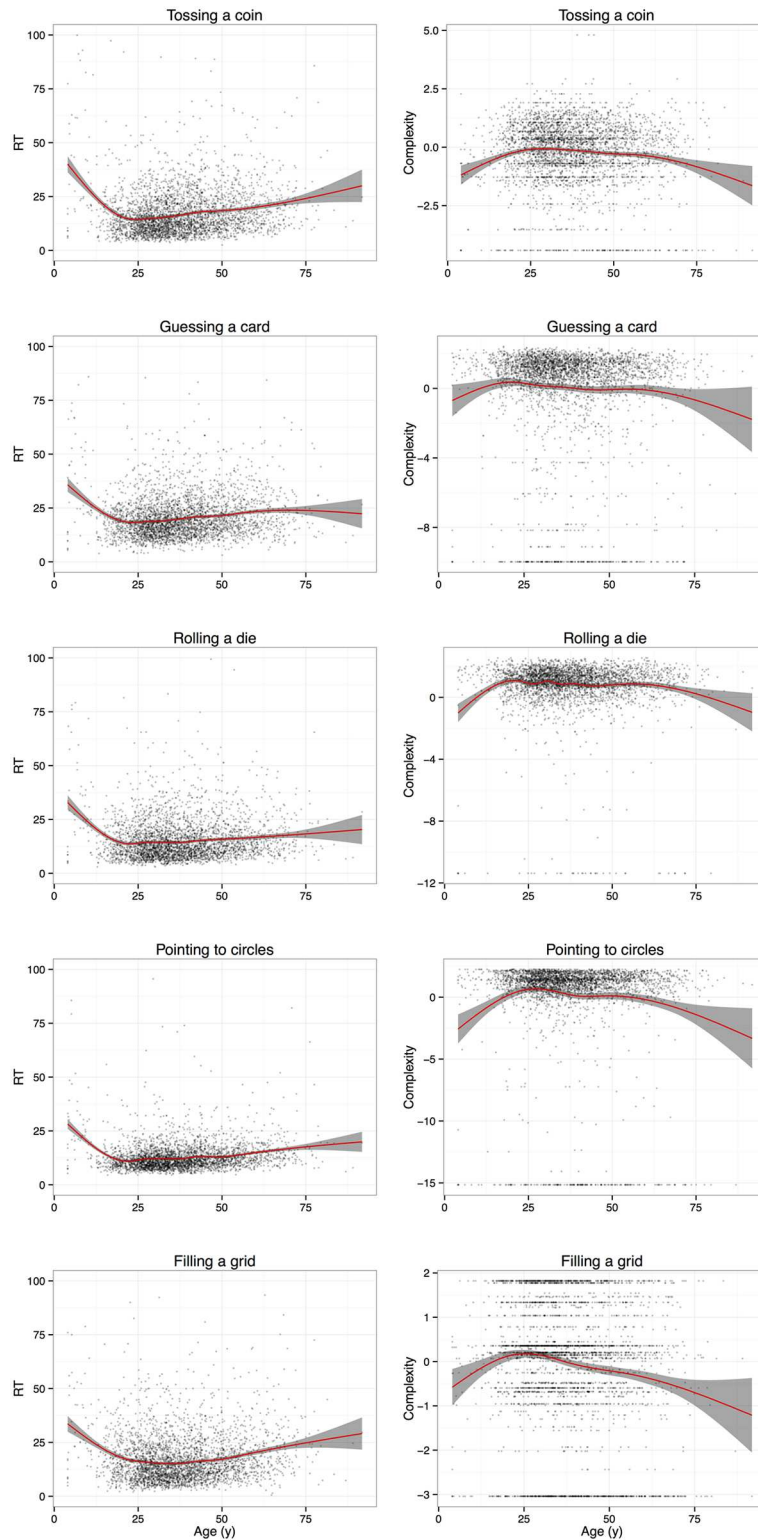
Fig 1. Developmental curves of completion time and complexity, split by task, with trend curves and 95% confidence regions (shaded).

others in the set: the option offered to the participants to change their previous choices as many times as they wished. For that reason, the grid task may in fact have relied more on randomness perception, and less on inhibition and randomness production heuristics. Indeed, participants could change the grid until they felt it was random, relying more on their ASC than on any high order cognitive ability serving output structure. This hypothesis is supported by the fact that participants did indeed change their minds. There were only 9 cells (that could turn white or black) on the grids and participants' end responses had a mean of 4.08 (SD = 1.8) selected (black) cells thus generally favoring whiter configurations (possibly as a result of the all-white initial configuration). However, the number of clicks used by participants during this task was far larger (M = 10.16, SD = 9.86), with values ranging from 5 to 134 (this latter trying almost a fifth of all possible configurations). Thus the option to change previous choices in a RIG task may have been an important factor, and should accordingly be considered a novel variable in future explorations of randomization tasks (and balanced with an all-black configuration). In this view, the "filling the grid" task would reflect our ASC in a more reliable fashion than other tasks, while being less dependent on inhibition processes.

Limitations

The present findings are based on data collected online. One could argue that this might bias the sample toward participants more comfortable with technology. Although direct control over a participant's behaviour online is certainly limited, as compared to the laboratory environment, there is an increasing number of studies demonstrating the convergence of laboratory-based and web-based data collection [4]. This is the case in very particular procedural situations, such as lateralized [46] or very brief [47] stimulation, presentation and the measurement of reaction times [48], and it also holds for the assessment of cognitive changes across the lifespan [49]. Compared to these special situations, our research procedure was simple, the tasks were entertaining, and response times did not have to be collected with millisecond precision [50]. We thus think that any disadvantages of online-testing have been more than compensated for by the advantages of enrolling a large number of participants across a wide age range.

Modulating factors

To investigate possible modulating factors (besides age), we used general linear models with complexity and CT as DV, and age (squared), sex, education, field and paranormal beliefs as IV.

The variable *Sex* was chosen in order to test in a large sample whether the absence of differences in laboratory RIG experiments could be replicated in an online test. Similarly, *Education* was important to test given previous claims in the RIG literature that human randomization quality may be independent of educational level [51]. Paradoxically, participants with a scientific background may perform worse at producing random sequences, thanks to a common belief among them that the occurrence of any string is as statistically likely as any other (a bias deriving from their knowledge of classical probability theory), which further justifies controlling for *Field of education*, simplified as humanities v. science. Finally, the variable Paranormal Belief was included as it has been related to RIG performance in previous studies [52].

The variables *field* and *paranormal belief* were, however, only asked in a subset of the 3313 participants that were above the age of 15 and we ignored the responses of younger participants as they were not considered to have a differentiated education background nor a fixed belief concerning paranormality. The analysis was performed on a task-wise basis. As we report, neither field or education level had no significant effect on any of the complexity or CT scores.

Experiment

A sample of 3429 participants took part in the experiment (age range: 4–91y, M = 37.72, SD = 13.38). Participants were recruited through social networks, radio broadcasts and journal calls during a 10 month period. Basic demographic characteristics of the sample are displayed in Table 2, the experiment is to this date still available online for people to test (URL available in the next section). Each of the five (self-paced) RIG tasks consisted in the production, as fast and accurately as possible, of a short (with length range 9–12) pseudo-random series of symbols, with various numbers of possible symbols and variations among other instructional features (Table 1).

Procedure

A specific web application was designed to implement the experiment online. Participants freely logged on to the website (http://complexitycalculator.com/hrng/). The experiment was available in English, French, German, and Spanish (all translated by native speakers for each language). In the case of young children as yet unable to read or use a computer, an adult was instructed to read the instructions out loud, make sure they were understood, and enter the child's responses without giving any advice or feedback. Participants were informed that they would be taking part in an experiment on randomness. They then performed a series of tasks (all participants performed the tasks in this order, see *SI* for screen shots) before entering demographic information such as sex, age, mother tongue, educational level, and main field of education (science, humanities, or other) if relevant. Before each task, participants (or a parent, in the case of youngsters) read the specific instructions of the task and only press "start" key after full agreement that they have understood the requirements of the task. Only then, that action initiated the measurement of the completion time (CT) for each task, which was recorded alongside the responses. Practice trials were not allowed in order to minimize boredom effects leading to drop-out rates and bias and to maximize spontaneity.

Table 2. Sample descriptive statistics (*n*).

Sex	Male	2333
	Female	1085
	Unknown	11
Mother tongue	English	274
	French	1448
	German	1303
	Spanish	220
	Other	184
Education level	Kindergarten or below	38
	Primary school	83
	Secondary school	387
	High School	621
	Undergraduate	347
	Graduate	1364
	Post graduate	538
	Unknown	51
Field	Humanities	609
	Science	1684
	Other	550
	Irrelevant	586

https://doi.org/10.1371/journal.pcbi.1005408.t002

One last item served as an index of paranormal beliefs and was included since probabilistic reasoning is among the factors associated with the formation of such beliefs [53, 54]. Participants had to rate on a 6-point Likert scale how well the following statement applied to them: "Some 'coincidences' I have experienced can best be explained by extrasensory perception or similar paranormal forces."

Measures

For each task, CT (in seconds) was recorded. The sum of CTs (total CT) was also used in the analyses. An estimate of the algorithmic complexity of participants' responses was computed using the acss function included in the freely publicly available acss R-package [42] that implements the complexity methods used in this project. Complexities were then normalized, using the mean and standard deviation of all possible strings with the given length and number of possible symbols, so that a complexity of 0 corresponds to the mean complexity of all possible strings. For each participant, the mean normalized complexity (averaged over the five tasks) was also computed, serving as a global measure of complexity.

Results and discussion

Sex had no effect on any of the complexity scores, but a significant one on two CT scores, with male participants performing faster in the first two tasks: "tossing a coin" ($p = 6.26 \times 10^{-10}$, $\eta_p^2 = .012$) and "guessing a card" ($p = 2.3 \times 10^{-10}$, $\eta_p^2 = .012$). A general linear model analysis of the total CT scores as a function of sex, age (squared), field, education level and paranormal belief was performed on the same subset and revealed a strongly significant effect of sex ($p = 9.41 \times 10^{-8}$, $\eta_p^2 = .009$), with male participants performing faster than female participants. A simpler model, including only sex and age (squared) as IV, still showed an effect of sex ($p = 6.35 \times 10^{-13}$, $\eta_p^2 = .016$), with male participants needing less time. The sex difference in CT, mostly appearing in adulthood (during the 60s; *SI, Fig. 4*), was in line with previous findings that in adults, choice CT is lower in men than in women [55].

Paranormal belief scores were unrelated to CTs for all tasks. However, they were negatively linked with the complexity score in the "filling a grid" task ($p = .0006$, $\eta_p^2 = .004$), though not with any other task.

Paranormal beliefs have been previously reported to be negatively linked to various RIG performances [52, 53]. Our results replicated these findings but only on the "filling a grid" task. One possible explanation is that the grid task actually measures randomness perception rather than randomness production, and that a paranormal bent is more strongly linked with a biased perception of randomness than with a set of biased procedures used by participants to mimic chance. This hypothesis is supported by the finding that believers in extrasensory perception are more prone to see "meaningful" patterns in random dot displays [52]. Another complementary hypothesis is that the type of biases linked to beliefs in the paranormal only usually appear over the long haul, and are here preempted by the fact that we asked participants to produce short sequences (of 12 items at most). Indeed, when it comes to investigating the effects of paranormal belief on pure randomness production, rather long strings are needed, as the critical measure is the number of repetitions produced [53, 56].

To get a better sense of the effect of paranormal belief, we performed a general linear model analysis of the mean complexity score as a function of sex, age (squared), field, education level and paranormal belief. For this analysis, we again used the subset of 3,313 participants over the age of 15. Paranormal belief no longer had an effect.

Mean complexity and total CT trajectories

Our main objective was to describe the evolution over the lifespan of mean complexity, which is achieved here using an approximation of algorithmic complexity for short strings (but see *SI Entropy* for a discussion of entropy (under)performance). Following Craik and Bialystok's [33] view, the developmental curve of complexity found in Fig 2, suggests that RIG tasks measure a combination of fluid mechanics (reflected in a dramatic performance drop with aging) and more crystallized processes (represented by a stable performance from 25 to 65 years of age). This trajectory indirectly confirms a previous hypothesis [15]: attention and inhibition decrease in adulthood, but an increased sense of complexity based on crystallized efficient heuristics counters the overall decline in performance.

Plotting complexity and CT trends on a single two-dimensional diagram allowed a finer representation of these developmental changes (Fig 3). It confirmed the entanglement of complexity (accuracy) and CT (speed). In the first period of life (<25), accuracy and speed increased together in a linear manner. The adult years were remarkable in that complexity remained at a high level for a protracted period, in spite of a slow decrease of speed during the same period. This suggest that during the adult period, people tend to invest more and more computational time to achieve a stable level of output complexity. Later in life (>70), however, speed stabilizes, while complexity drops in a dramatic way.

These speed-accuracy trade-offs were evident in the adult years, including the turn toward old age. During childhood, however, no similar pattern is discernible. This suggests that aging cannot simply be considered a "regression", and that CT and complexity provide different complementary information. This is again supported by the fact that in the 25–60 year range,
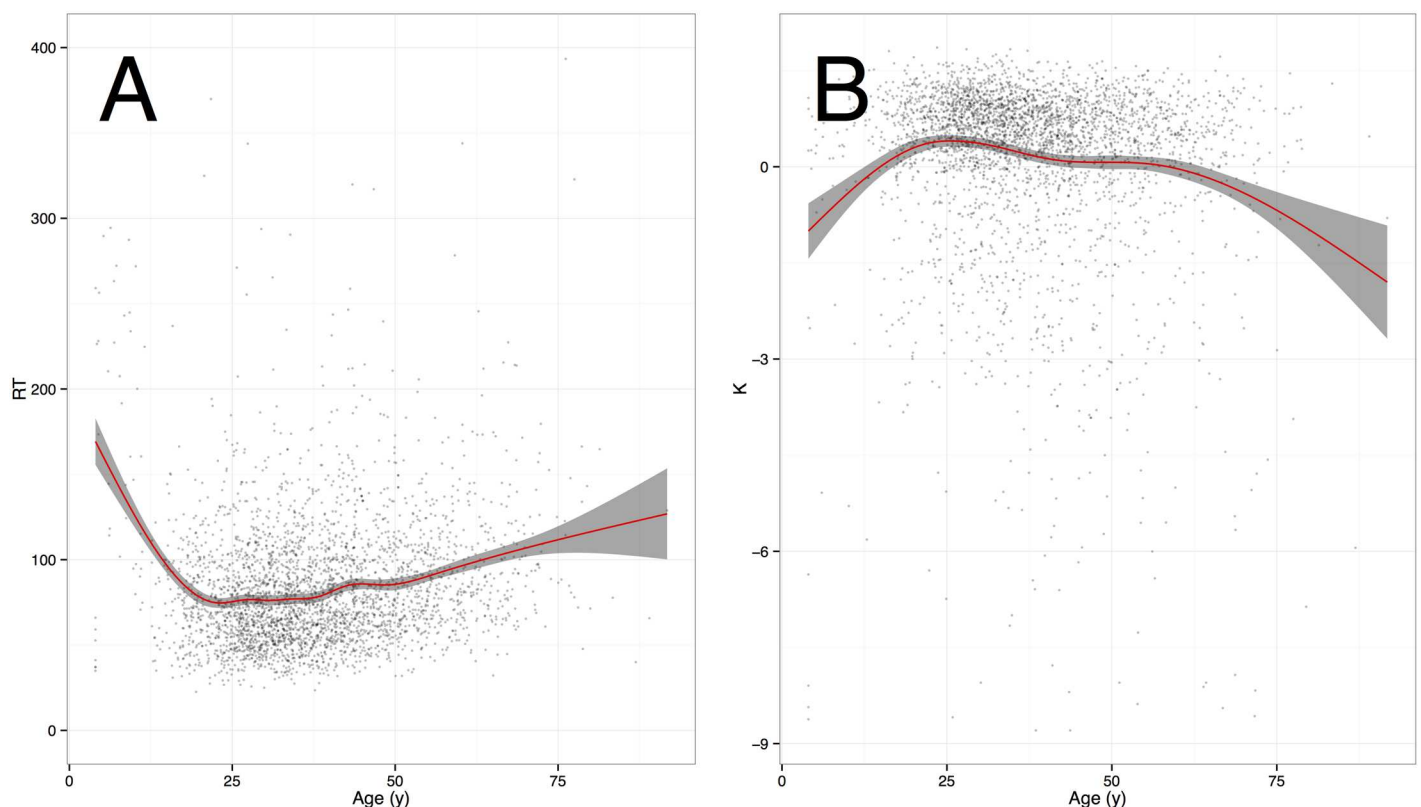

Fig 2. (A) Total completion time (CT) and (B) mean complexity as a function of age, with trend curve and 95% confidence region (shaded area).
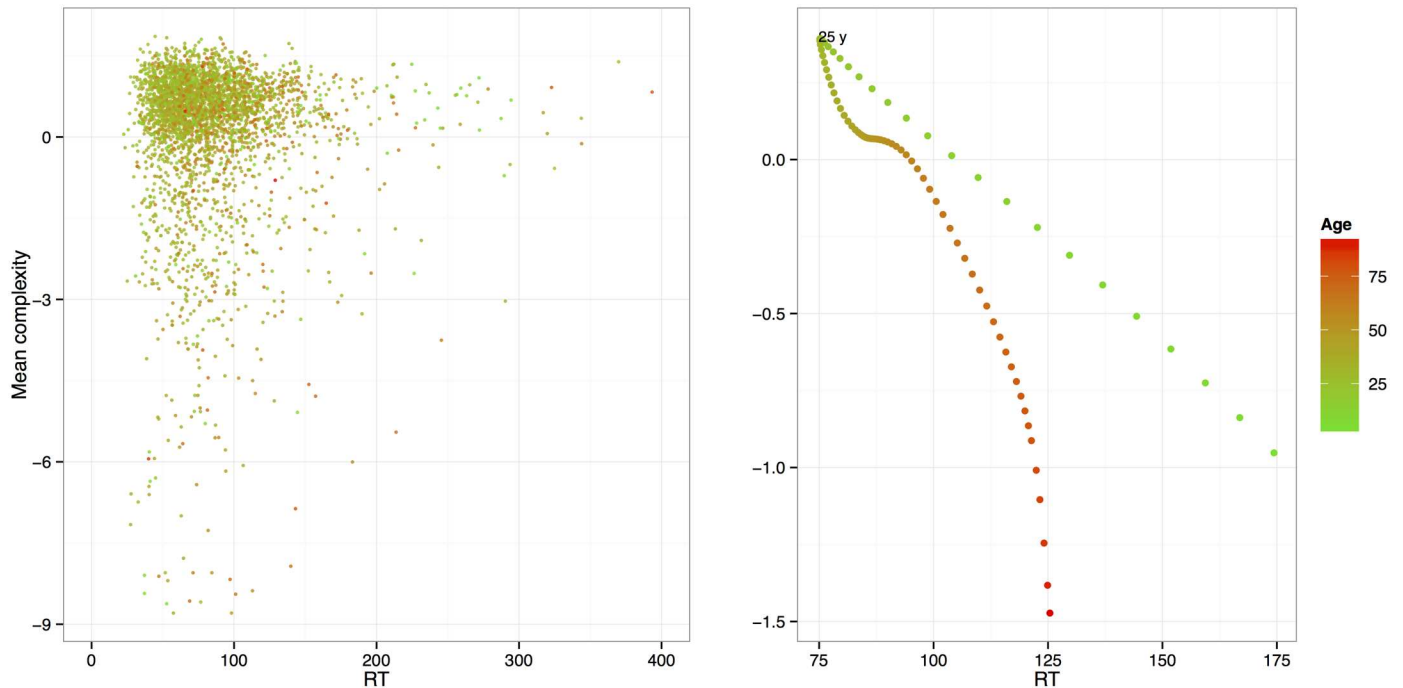
Fig 3. Scatterplot and developmental change trend of the CT and complexity combined. The trend is obtained by use of smooth splines of CT and complexity (df = 7).

where the effect of age is reduced, CT and complexity are uncorrelated ($r = -.012$, $p = .53$). These findings add to a rapidly growing literature that views RIG tasks as good measures of complex cognitive abilities [21, for a review].

We have gone further here in several respects than any previous literature. First, we present a set of data collected in RIG tasks with a broad variety of instructions as to what and how to randomize: our participants playfully solved binary randomization tasks along with multiple-alternative variants; they explicitly attempted to generate random sequences, but also distributed their responses in a guessing task, typically considered "implicit random generation". The expected outcome was unidimensional in some tasks and two-dimensional in others; constraints imposed by working memory capacity were high in some tasks, but almost absent in others. In the cognitive science literature, such diverse tasks have never been compared directly. We do not deny that the various tasks we used may tap into slightly different subcategories of prefrontal functioning, with some relying more on working memory and others on inhibitory control. Yet, we set out to illustrate the commonalities among the different tasks leaving a more fine-grained analysis to future studies.

Cross-sectional studies should try to relate behavioural complexity to the degree of maturation or degeneration of specific prefrontal cortex regions. Neuropsychological investigations could use the tasks and measures employed here with selected patient groups to perform lesion-symptom mappings, as has been done recently [57], but preferably in patients with focal vascular lesions. In parallel with such investigations, Internet-based work such as the project presented here may still play a powerful role. They may complement RIG tasks with brief behavioural tasks having a known neurocognitive basis and well-studied developmental trajectories. Thus, laboratory testing and web-based approaches may conjointly help pinpoint the cognitive mechanisms underlying the age-related modulation of behavioural complexity.

A second extension of the existing literature on subject-generated random sequences is the number of participants tested and their age-range. To date, only two studies have investigated age-related changes in RIG tasks with a range comparable to the one investigated here [15, 22]. They both compared groups of young adults and older adults and were thus unable to describe the continuous evolution of complexity across the lifespan.

Finally, one of the most exciting novel aspects of this research is that we have presented an estimate of algorithmic complexity that relies on sequences shorter than any that research on RIG reported in the psychological literature would have dared to use because of the limitations of other complexity indexes.

Conclusion

RIG tasks require a sense of randomness or complexity, as well as cognitive functions such as attention, inhibition and working memory. The evolution of algorithmic complexity over the lifespan is compatible with the idea that RIG tasks, even in a computerized and shortened format, reflect such abilities. The developmental curve reveals an evolution compatible with the concept of a combination of fluid and crystallized components in cognition, with the latter predominating.

Beyond the similarity of complexity trajectories, we found that the variety of RIG tasks offered different and probably complementary information about a participant's cognitive abilities. The exact component of cognition that is assessed by RIG tasks, and which factors differentiate the tasks, are still open questions.

Our findings shed light on the developmental change in ASC, on which inductive reasoning about randomness is built. They will hopefully further our understanding of human probabilistic core knowledge. Like other complex cognitive abilities, the trend in evidence here must not occlude important intra- and inter-subject variations. Age (squared) explains about 2% of the variance in mean complexity, and 4% of the variance in CT. Although age is the predominant variable, CT and complexity are also affected, in the case of some tasks, by sex, statistical intuition and paranormal belief. Future research should investigate the impact of other variables on RIG performance. Examples comprise a participant's tendency to persevering which would have to be established in an independent task. Alternatively, use of computers and familiarity with an online environment might be considered.

Anonymized data are available from https://github.com/algorithmicnaturelab/HumanBehavioralComplexity

Author Contributions

Conceptualization: NG HZ PB FST JPD.

Data curation: NG HZ FST.

Formal analysis: NG HZ PB FST.

Funding acquisition: HZ PB.

Investigation: NG HZ PB FST JPD.

Methodology: NG HZ PB FST JPD.

Project administration: NG HZ PB.

Resources: HZ PB FST NG.

Software: HZ FST NG.

Supervision: NG HZ PB.

Validation: NG HZ PB FST.

Visualization: NG HZ FST.

Writing – original draft: NG PB HZ FST.

Writing – review & editing: NG HZ PB.

References

1. Spelke EE, Kinzler KD (2007) Core knowledge. *Dev Sci* 10:89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x PMID: 17181705

2. Izard V, Sann C, Spelke EE, Streri A (2009) Newborn infants perceive abstract numbers. *Proc of Nat Acad Sci* 106:10382–10385. https://doi.org/10.1073/pnas.0812142106

3. Rugani R, Vallortigara G, Priftis K, Regolin L (2015) Number-space mapping in the newborn chick resembles humans' mental number line. *Science* 347:534–536. https://doi.org/10.1126/science.aaa1379 PMID: 25635096

4. Halberda J, Ly R, Wilmer JB, Naiman DQ, Germine L (2012) Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc of Nat Acad Sci* 109:11116–11120. https://doi.org/10.1073/pnas.1200196109

5. Dehaene S (1997) *The number sense: How the mind creates mathematics* (Oxford, Oxford University Press).

6. Fontanari L, Gonzalez M, Vallortigara G, Girotto V (2014) Probabilistic cognition in two indigenous Mayan groups. *Proc of Nat Acad Sci* 111:17075–17080. https://doi.org/10.1073/pnas.1410583111

7. Téglás E, Vul E, Girotto V, Gonzalez M, Tenenbaum JB, Bonatti LL (2011) Pure reasoning in 12-month-old infants as probabilistic inference. *Science* 332:1054–1059. https://doi.org/10.1126/science.1196404 PMID: 21617069

8. Xu F, Garcia V (2008) Intuitive statistics by 8-month-old infants. *Proc of Nat Acad Sci* 105:5012–5015. https://doi.org/10.1073/pnas.0704450105

9. Williams JJ, Griffiths TL (2013) Why are people bad at detecting randomness? A statistical argument. *J Exp Psychol: Learn Mem Cogn* 39:1473–1490.

10. Gauvrit N, Soler-Toscano F, Zenil H, Delahaye J-P (2014) Algorithmic complexity for short binary strings applied to psychology: A primer. *Behav Res Meth* 46:732–744. https://doi.org/10.3758/s13428-013-0416-0

11. Ma L, Xu F (2012) Preverbal Infants Infer Intentional Agents From the Perception of Regularity. *Dev Psychol* 49:1330–1337. https://doi.org/10.1037/a0029620 PMID: 22889398

12. Jahanshahi M, Saleem T, Ho AK, Dirnberger G, Fuller R (2006) Random number generation as an index of controlled processing. *Neuropsychology* 20:391–399. https://doi.org/10.1037/0894-4105.20.4.391 PMID: 16846257

13. Mittenecker E (1958) Die Analyse zufälliger Reaktionsfolgen [the analysis of "random" action sequences]. *Zeit für Expe und Ang Psychol* 5:45–60.

14. Loetscher T, Bockisch CJ, Brugger P (2008) Looking for the answer: The mind's eye in number space. *Neuroscience* 151:725–729. https://doi.org/10.1016/j.neuroscience.2007.07.068 PMID: 18155853

15. Heuer H, Janczyk M, Kunde W (2010) Random noun generation in younger and older adults. *Q J Exp Psychol* 63:465–478. https://doi.org/10.1080/17470210902974138

16. Wiegersma S,Van Den Brink WAJM (1982) Repetition of tones in vocal music. *Percept Mot Skills* 55:167–170. https://doi.org/10.2466/pms.1982.55.1.167

17. Hahn U, Warren PA (2009) Perceptions of randomness: why three heads are better than four. *Psychol Rev* 116:454–461. https://doi.org/10.1037/a0017522 PMID: 19348550

18. Schulter G, Mittenecker E, Papousek I (2010) A computer program for testing and analyzing random generation behavior in normal and clinical samples: The Mittenecker Pointing Test. *Behav Res Meth* 42:333–341. https://doi.org/10.3758/BRM.42.1.333

19. Maes JH, Eling PA, Reelick MF, Kessels RP (2011) Assessing executive functioning: On the validity, reliability, and sensitivity of a click/point random number generation task in healthy adults and patients with cognitive decline. *J of Clinic and Expe Neuropsychol* 33:366–378. https://doi.org/10.1080/13803395.2010.524149

20. Towse JN (2007) On random generation and the central executive of working memory. *British J Psychol* 89:77–101. https://doi.org/10.1111/j.2044-8295.1998.tb02674.x

21. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager T (2000) The unity and diversity of executive functions and their contributions to frontal lobe tasks: A latent variable analysis. *Cog Psy* 41:49–100. https://doi.org/10.1006/cogp.1999.0734

22. Van der Linden MV, Beerten A, Pesenti M (1998) Age-related differences in random generation. *Brain and Cog* 38:1–16. https://doi.org/10.1006/brcg.1997.0969

23. Nickerson R (2002) The production and perception of randomness. *Psychol Rev* 109:330–357. https://doi.org/10.1037/0033-295X.109.2.330 PMID: 11990321

24. Baddeley A (1998) Random generation and the executive control of working memory. *Q J Exp Psychol A* 51:819–852. https://doi.org/10.1080/713755788 PMID: 9854442

25. Peters M, Giesbrecht T, Jelicic M, Merckelbach H (2007) The random number generation task: Psychometric properties and normative data of an executive function task in a mixed sample. *J of the Internat Neuropsychol Soc* 13:626–634. https://doi.org/10.1017/S1355617707070786

26. Sagaspe P, Charles A, Taillard J, Bioulac B, Philip P (2003) Inhibition et mémoire de travail: effet d'une privation aiguë de sommeil sur une tâche de génération aléatoire [Inhibition and working memory: Effect of acute sleep deprivation on an item generation task]. *Canadian J of Exp Psychol* 57:265–273. https://doi.org/10.1037/h0087430

27. Brown RG, Soliveri A, Jahanshahi M (1998) Executive processes in Parkinsons disease–random number generation and response suppression. *Neuropsychologia* 36:1355–1362. https://doi.org/10.1016/S0028-3932(98)00015-3 PMID: 9863689

28. Brugger P, Monsch AU, Salmon DP, Butters N (1996) Random number generation in dementia of the Alzheimer type: A test of frontal executive functions. *Neuropsychologia* 34:97–103. https://doi.org/10.1016/0028-3932(95)00066-6 PMID: 8852872

29. Chan K, Hui C, Chiu C, Chan S, Lam M (2011) Random number generation deficit in early schizophrenia, *Percept Mot Skills* 112:91–103. https://doi.org/10.2466/02.15.19.22.PMS.112.1.91-103 PMID: 21466082

30. Rinehart NJ, Bradshaw JL, Moss SA, Brereton AV, Tonge BJ (2006) Pseudo-random number generation in children with high-functioning autism and AspergerÕs disorder. *Autism* 10:70–85. https://doi.org/10.1177/1362361306062011 PMID: 16522711

31. Proios H, Asaridou SS, Brugger P (2008) Random number generation in patients with aphasia: A test of executive functions. *Acta Neuropsychologica* 6:157–168.

32. Brockmole JR, Logie RH (2013) Age-related change in visual working memory: a study of 55,753 participants aged 8–75. *Frontiers Psychol* 4. https://doi.org/10.3389/fpsyg.2013.00012

33. Craik FI, Bialystok E (2006) Cognition through the lifespan: mechanisms of change. *Trends in Cog Sci* 10:131–138. https://doi.org/10.1016/j.tics.2006.01.007

34. Goldberger AL, Peng CK, Lipsitz LA, Vaillancourt DE, Newell KM (2002) What is physiologic complexity and how does it change with aging and disease? Authors' reply. *Neurobiol of Aging* 23(1):27–29. https://doi.org/10.1016/S0197-4580(01)00266-4

35. Manor B, Lipsitz LA (2013) Physiologic complexity and aging: Implications for physical function and rehabilitation. *Prog in Neuro-Psychopharmaco and Biol Psych* 45:287–293. https://doi.org/10.1016/j.pnpbp.2012.08.020

36. Lipsitz LA, Goldberger AL (1992) Loss of'complexity'and aging: potential applications of fractals and chaos theory to senescence. *JAMA* 267(13):1806–1809. https://doi.org/10.1001/jama.1992.03480130122036 PMID: 1482430

37. Chater N, Vitányi P (2003) Simplicity: a unifying principle in cognitive science?. *Trends in Cog Sci* 7(1):19–22. https://doi.org/10.1016/S1364-6613(02)00005-0

38. Tamariz M, Kirby S (2014) Culture: copying, compression, and conventionality. *Cogn. Sci* 39(1):171–183. https://doi.org/10.1111/cogs.12144 PMID: 25039798

39. Delahaye J-P, Zenil H (2012) Numerical evaluation of algorithmic complexity for short strings: A glance into the innermost structure of randomness. *Applied Maths and Comput* 219:63–77. https://doi.org/10.1016/j.amc.2011.10.006

40. Soler-Toscano F, Zenil H, Delahaye J-P, Gauvrit N (2014) Calculating Kolmogorov Complexity from the Output Frequency Distributions of Small Turing Machines, *PLoS ONE* 9(5): e96223. https://doi.org/10.1371/journal.pone.0096223 PMID: 24809449

41. Zenil H, Soler-Toscano F, Delahaye J-P, Gauvrit N (2015) Two-Dimensional Kolmogorov Complexity and Validation of the Coding Theorem Method by Compressibility. *PeerJ Comput Sci* 1:e23. https://doi.org/10.7717/peerj-cs.23

42. Gauvrit N, Singmann H, Soler-Toscano F, Zenil H (2016) Algorithmic complexity for psychology: A user-friendly implementation of the coding theorem method. *Behav Res Meth* 48:314–329. https://doi.org/10.3758/s13428-015-0574-3

43. Gauvrit N, Soler-Toscano F, Zenil H (2014) Natural Scene Statistics Mediate the Perception of Image Complexity. *Visual Cognition* 22:1084–1091. https://doi.org/10.1080/13506285.2014.950365

44. Zenil H, Soler-Toscano F, Dingle K, Louis A (2014) Correlation of Automorphism Group Size and Topological Properties with Program-size Complexity Evaluations of Graphs and Complex Networks. *Physica A: Statistical Mechanics and its Applications* 404:341–358. https://doi.org/10.1016/j.physa.2014.02.060

45. Storm L, Tressoldi PE, Di Risio L (2012) Meta-analysis of ESP studies, 1987–2010: Assessing the success of the forced-choice design in parapsychology. *Journal of Parapsychology* 76.2:243–273.

46. McGraw KO, Tew MD, Williams JE (2000) The integrity of Web-delivered experiments: can you trust the data? *Psychol Sci* 11:502–506. https://doi.org/10.1111/1467-9280.00296 PMID: 11202497

47. Crump MJC, McDonnell JV, Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* 8(3):e57410. https://doi.org/10.1371/journal.pone.0057410 PMID: 23516406

48. Chetverikov A, Upravitelev P (2016) Online versus offline: the Web as a medium for response time data collection. *Behav Res Meth* 48:1086–1099. https://doi.org/10.3758/s13428-015-0632-x

49. Germine LT, Duchaine B, Nakayama K (2011) Where cognitive development and aging meet: face learning ability peaks after age 30. *Cognition* 118:201–210. https://doi.org/10.1016/j.cognition.2010.11.002 PMID: 21130422

50. Garaizar P, Reips U-D (2015) Visual DMDX: a web-based authoring tool for DMDX, a Windows display program with millisecond accuracy. *Behav Res Meth* 47:620–631. https://doi.org/10.3758/s13428-014-0493-8

51. Brugger P (1997) Variables that influence the generation of random sequences: An update. *Percept Mot Skills* 84:627–661. https://doi.org/10.2466/pms.1997.84.2.627 PMID: 9106859

52. Brugger P, Regard M, Landis T, Cook N, Krebs D, Niederberger J (1993) 'Meaningful' patterns in visual noise: effects of lateral stimulation and the observer's belief in ESP. *Psychopathol* 26:261–265. https://doi.org/10.1159/000284831

53. Brugger P, Landis T, Regard M (1990) A 'sheep-goat effect' in repetition avoidance: Extra-sensory perception as an effect of subjective probability?. *British J Psychol* 81:455–468. https://doi.org/10.1111/j.2044-8295.1990.tb02372.x

54. Wiseman R, Watt C (2006) Belief in psychic ability and the misattribution hypothesis: A qualitative review. *British J Psychol* 97:323–338. https://doi.org/10.1348/000712605X72523

55. Der G, Deary IJ (2006) Age and sex differences in reaction time in adulthood: results from the United Kingdom Health and Lifestyle Survey. *Psychol and Ageing* 21:62–73. https://doi.org/10.1037/0882-7974.21.1.62

56. Musch J, Ehrenberg K (2002) Probability misjudgment, cognitive ability, and belief in the paranormal. *British J Psychol* 93:169–178. https://doi.org/10.1348/000712602162517

57. Geisseler O, Pflugshaupt T, Buchmann A, Bezzola L, Reuter K, Schuknecht B, Weller D, Linnebank M, Brugger P (2016) Random number generation deficits in patients with multiple sclerosis: characteristics and neural correlates. *Cortex* 82:237–243. https://doi.org/10.1016/j.cortex.2016.05.007 PMID: 27403852