Human social complexity

Roozen, Mila C

DOI:
10.33612/diss.907477754

Human social complexity

Evolutionary and methodological considerations

Mila Christina Roozen

Human social complexity

Evolutionary and methodological considerations

Mila Christina Roozen

The research reported in this thesis was conducted at the Groningen Institute for Evolutionary Life Sciences of the University of Groningen, Groningen, The Netherlands.

This research was supported by the PRISM project (www.prism-project.eu).

Title: Human social complexity: Evolutionary and methodological considerations
Author: Mila Christina Roozen
Cover, layout and printing: Optima Grafische Communicatie (www.ogc.nl)

university of groningen

Human social complexity: Evolutionary and methodological considerations

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. J.M.A. Scherpen
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Tuesday 20 February 2024 at 11:00 hours

by

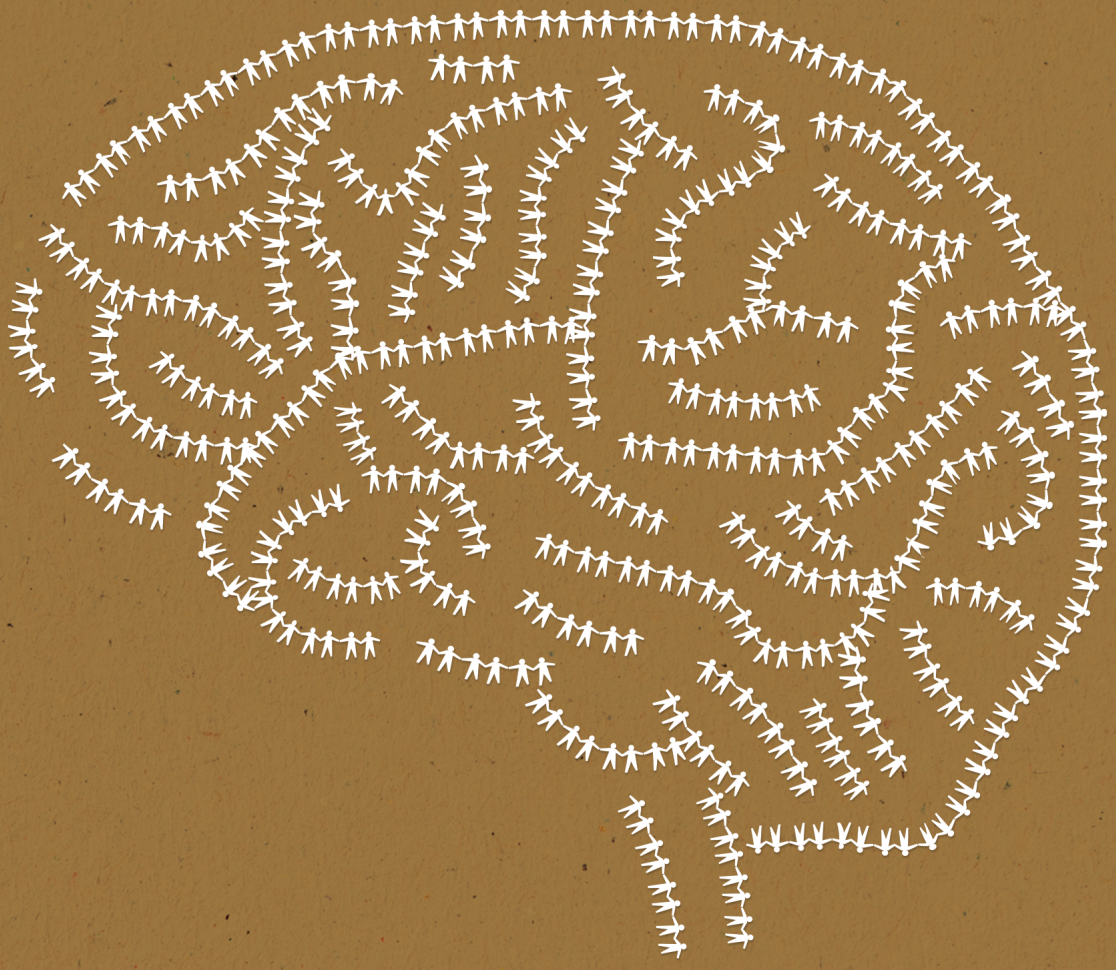
Mila Christina Roozen

born on 19 August 1994
in Amsterdam

Supervisors
Prof. M.J.H. Kas
Prof. J.A.S. Vorstman

Assessment committee
Prof. J.C. Billeter
Prof. H.L. Dugdale
Prof. A.F. Marquand

TABLE OF CONTENTS



Chapter 1

General introduction: Evolutionary dynamics of human social brain functioning

Mila C. Roozen[a], Jan Komdeur[a], Bregje Wertheim[a] and Martien J.H. Kas[a]

[a]*Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands*

In preparation for submission

ABSTRACT

Throughout the evolution of primate and early hominid species, more complex social environments have been associated with increases in the size of the neocortex. However, despite the continuing increase in the social complexity of the human environment, evidence indicates that in recent millennia the human brain has not only stopped growing, but may have started to decrease in size. Several hypotheses have been suggested to explain this apparent decrease, including the collective intelligence hypothesis and the self-domestication hypothesis. Methodological advances now allow for the detection of signals of selection in human genetic data, meaning selection pressure affecting human social complexity and structural characteristics of the human brain can be evaluated. Pleiotropy, a mechanism underlying several hypotheses explaining the evolution of brain size, can similarly be examined using genetic data. Such analyses may aid in our understanding of how the modern environment may affect the human brain in the long term.

INTRODUCTION

Researchers have wondered for decades how evolutionary processes have culminated in the large brains seen in primates and specifically humans. Based on earlier hypotheses and research, Barton and Dunbar (1997) described the 'social brain hypothesis' (SBH) (Dunbar, 1998) which put forward the hypothesis that 'primates evolved large brains to cope with their unusually complex social lives' (Dunbar, 2009). Functioning in these complex social systems might have led to increased survival odds either due to the increased protection from being hunted, or through increased access to food sources. At the time of its formulation, support for the social brain hypothesis stemmed from studies across primate species finding that group size is correlated specifically to the size of the neocortex (relative to other measures of brain size, such as total brain volume or medulla size) (Dunbar, 1998). Group size is considered a measure of the complexity of the social environment ('social complexity') (Bergman & Beehner, 2015; Dunbar, 2009; Kappeler et al., 2019a). The neocortex is mostly involved in higher-order processing functions such as cognition and learning. The large increase in relative size of the neocortex in primates is a strong indication of the importance of these functions, as brains are metabolically expensive and therefore the benefits in terms of reproductive success need to outweigh the cost of these increased metabolic requirements (Isler & van Schaik, 2006).

One of the central ideas of the social brain hypothesis is the hypothesis that social complexity has been under selection pressure, strong enough and for a long enough duration to allow primates, including humans, to develop a uniquely large neocortex. However, the environment in which humans live has changed drastically from the environment where the closest living ancestor of humans and other primates lived, including the social characteristics of this environment. Has this environmental change affected the evolution of human brain, and if so, in what manner?

In this review, we intend to highlight an aspect of human brain evolution that is relatively underexposed in scientific research. We will describe the results from research into this topic in the past decades, consistently finding strong (on an evolutionary scale) decreases in brain size over the past millennia, and summarize two prominent hypotheses offering potential explanations for these findings. In the remainder of this review, we will focus on methods which can assist in a) further solidifying the evidence that selection is acting on brain size in a manner which will reduce brain size over time and b) discovering potential processes involved in these evolutionary processes. As a final point, we will describe how pleiotropic processes are central in hypotheses attempting to explain brain evolution and how they can potentially be involved in complicated evolutionary processes which require more specialized methods to detect.

RECENT EVOLUTIONARY DYNAMICS OF HUMAN BRAIN SIZE

Recently, DeSilva et al. (2021) published their analyses on hominid skull size using skull samples ranging from 10 million years ago up to very recent samples (Figure 1). They found that a slow increase from 10 million years ago changed to a sharp increase around 2.1 million years ago, slowing somewhat around 1.5 million years ago and then changing to a very strong decrease around 3,000 years ago. This decrease was many times greater than the increase was during the periods measured in this and previous studies. However, DeSilva et al. (2021) caution that their results should be interpreted conservatively, as the results from studies such as theirs are dependent on the characteristics of the sample used.

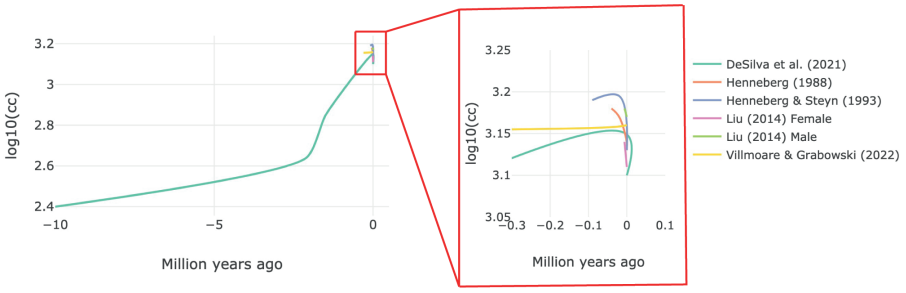


Figure 1. Examples of different findings regarding the change in brain size across the past ten million and 500000 years. Cc = Cranial capacity. Data and figures used for the creation of this figure are: DeSilva et al. (2021)[5] Figure 1 and Table 1; Henneberg (1988)[7] Table 3; Henneberg & Steyn (1993)[8] Figure 1 (data consists only of male subjects); Liu et al. (2014)[9] Table 2 and Table 3; and Villmoare & Grabowski (2022)[6] Figure 2.

One paper (Villmoare & Grabowski, 2022) showed that a reanalysis of the data from the paper by DeSilva and colleagues (2021) did not show a recent decrease in brain size. The authors bring up several valid points. First, they question the hypothesis set forward by DeSilva et al. (2021) that agriculture may have precipitated the decrease in brain size, as the results from the study by DeSilva et al. (2021) are based from samples across the world and the change to mostly agricultural food production occurred at different times across different regions. Therefore, the estimate of 3000 to 2000 years for the changepoint to the decrease in brain size does not overlap with the transition to agriculture for many samples in their data. Secondly, DeSilva et al. (2021) do not mention any p-values indicating the significance of their results, nor do they provide information why they did not or include other measures which could be used to determine to what extent their findings could have occurred by chance. The final criticism leveled is that the data is highly heteroscedastic as a result of the large difference in sample size between recent and older samples, with many more recent samples present. The solution offered by the authors is to reduce the number of samples significantly and averaging

the samples to create 100-year means. Both these actions result in a significant loss of information, although they do reduce the heteroscedasticity. Based on their reanalysis, the authors conclude no significant change occurred in brain size across the past 300,000 years or the past 30,000 years.

However, several other papers from the past decades have also found decreases in human brain size (based on different measures of cranial volume) in more homogeneous samples. For example, Henneberg (1988) found a significant decrease in both male and female cranial capacity using samples from Europe, North Africa and western Asia since the Mesolithic, which occurred between 15,000 and 5,000 years ago. In a later study, Henneberg and Steyn (1993) find a similar decrease in cranial capacity for men and women using sub-Saharan samples starting around 17,000 years ago. In a more recent analysis, Liu et al. (2014) found a similar decrease in cranial capacity between a Neolithic sample (between 6200 and 5500 years old) from the Beiqian site in China and a sample of modern-day Chinese humans.

Although the estimates for the duration of the decrease in human cranial capacity (and by association, brain size) vary significantly, many studies do appear to point at a decrease during recent evolution. However, even if the study criticizing the results from DeSilva et al. (2021) arrived at a more accurate conclusion regarding the (lack of) change in brain size across the past 300,000 and 30,000 years, the change from a significant increase in brain size in ancestral human evolution to stagnation during recent evolution is still an important change, which does not appear to be reflected in a lack of change in social environment. The SBH implies a necessity of a large neocortex size in order to deal with complex social environments. As a result of the increase in size seen in many human societies, high numbers of social relations and interactions is a common occurrence. If humans require large neocortices to possess the cognitive abilities to cope with complex social environments, why would brain size have stopped increasing or even be decreasing now, despite the social environment appearing as complex as or even more complex than ever in human evolutionary history?

Explaining potential decreases in brain size

Several researchers have attempted to offer explanations for the apparent decrease in human brain size occurring across recent millennia. Concomitant decreases in body mass or stature did not occur in a way which would explain the reduction in brain volume through allometric scaling (Hawks, 2011). Findings regarding decreases in brain size do not seem to be a local phenomenon, as evidence for decrease has been found in several different studies across several regions, including Europe, (Sub-Saharan) Africa, Asia and Australia (Brown, 1987; Henneberg, 1998; Henneberg & Steyn, 1993; Henneberg, 2004; Liu et al., 2014).

The authors from the study on ancient skulls from the Beiqian site in China and modern human skulls from Chinese individuals (Liu et al., 2014) found that although cranial capacity and several aspects of brain size had decreased throughout the Holocene, both the absolute frontal breadth and the relative frontal breadth compared to the total breadth of the brain had increased across the same period. The authors argued that the increase in size in the frontal lobe may be related to the importance of language (verbal and written) in the human environment, as the frontal lobe contains several regions with important functions for the understanding and use of language.

Associated with the possibility that language is causing the relative increase in frontal breadth in the samples in the study above, some researchers have argued that the cause of the decrease in cranial volume is that the cognitive benefits of a large brain no longer outweigh the energetic costs due to the development of a 'collective intelligence' in humans (DeSilva, 2021). The 'collective intelligence' in this hypothesis can be described as a societal source of information which can be distributed amongst its members, reducing the cognitive requirements which would otherwise be necessary to maintain this information on an individual basis, and thereby allowing specialization. Practical implications of the collective intelligence hypothesis include the ideas that i) the decrease in cranial capacity results in reduced energetic requirements, ii) that a set of individuals with smaller cranial capacity can result in similar or improved reproductive fitness compared to single individuals with increased cranial capacity, and iii) that the reductions in cranial capacity are correlated with decreases in individual cognitive performance. Some evidence for the first and second points has been gathered through studies in animals (e.g. Kortschal et al., 2013) and simulation studies (e.g. Reséndiz-Behumea et al., 2021). It might be possible to examine the third point by examining differences between humans in cranial capacity and cognitive performance, and particularly between different societies of humans, which allows for varying levels of behavioral specialization between individuals.

Another suggested hypothesis for the potential decrease in brain size is the self-domestication hypothesis (Bednarik, 2014). This hypothesis is based on observations in domesticated animals, who display a range of combined trait changes from wild animals, known as the domestication syndrome, which includes low aggression, flattened faces, smaller crania, reductions in tooth size and the retainment of youthful behavior after childhood (also called playfulness or neotenization), as well as other characteristics. Proponents of this hypothesis suggest that humans have protected themselves from natural selection by regulating access to resources and reproduction, thereby creating a situation where humans were selected not based on their adaptedness to the natural environment but on social traits and characteristics selected by humans, such as arbitrarily defined attractiveness. This then led to genetic changes which, through the presence of genes pleiotropic for the features selected for by humans and other features such as brain size, resulted in the domestication syndrome as described above. Testable

hypotheses of the self-domestication hypothesis include the existence of genes pleiotropic for the different characteristics associated with domestication syndrome, the presence of similar networks of pleiotropic genes in domesticated species on which the hypothesis is build, such as the farm fox, and the presence of signals of directional selection on these pleiotropic genes in both humans and other species for which domestication syndrome is considered present.

MEASUREMENT OF SELECTION USING INTRASPECIES GENETIC VARIATION

Recent papers have demonstrated how methodological developments now allow for the evaluation of selection pressure, both long-term and recent, on the evolution of human traits using large samples of human data including genetic information. Although between-species comparisons are absolutely vital for studies of long-term selection, including differences and similarities between species, improvements in the ability to sequence the genomes of individuals as well as advances in our knowledge about the genome have led to the creation of statistical methodologies which can be used to detect signals of selection using the genetic data from large samples of individuals from a single species. Thanks to these scientific accomplishments, it is now possible to use these databases to uncover new insights about human evolution, including the recent evolution of the human brain.

Using results from genome-wide association studies to measure recent selection

One method of measuring contemporary reproductive success is to use Mendelian Randomization (MR). MR uses genetic loci as instrumental variables to determine the presence of causal relations between two traits. As genetic variants are randomly assigned and not affected by biasing factors, the genetic variation underlying an exposure can be seen as random assignment to a condition of the exposure. By examining the relation between this genetic variation and the outcome, causal inferences can be made regarding this relation.

One aspect of fitness is reproductive fitness, the number of mates and the offspring produced by an individual. MR can be used to relate exposures (such as brain size or social complexity) to the number of sexual partners and/or the number of children, allowing for an examination of the fitness effects of the exposures (Song et al., 2021).

One characteristic of loci experiencing recent selection is that they tend to carry fewer singleton mutations (Field et al., 2016). This characteristic can be exploited to examine regions under selection using Singleton Density Score (SDS) analysis. SDS analysis can be used to infer recent selection up to about 2000-3000 years ago. By examining SDS of loci associated with

social complexity or brain size, it is possible to assess the effects of recent selection on these traits.

A notable example of a study which used a large genetic database to study human evolution using (among others) MR and SDS analyses is the study by Song et al (2021), in which the authors examined selection pressure for 870 human traits across four different time scales (since human speciation, pan-Neolithic period, 2,000-3,000 years and modern), using summary statistics from genome-wide association study (GWAS) studies carried out in previous studies. An issue with this study is that the GWAS results used for the analyses were found in previous studies, which did not appear to anticipate that these results may be used for studies about selection pressure in the future. As a result, most likely all (or most) of these studies removed SNPs which deviated significantly from Hardy-Weinberg equilibrium (HWE), as this is considered a necessary test of genotyping quality. However, deviation from the Hardy-Weinberg equilibrium may also occur in case of selection. Hence, the removal of SNPs that deviate from HWE may preclude the detection of signatures of selection. We have included a supplementary analysis at the end of this chapter to demonstrate the effect of recent selection on HWE.

Another overview of methods of measuring selection using human genome-wide data is provided in the review by Guo et al (2018). They mention that selection could affect allele frequencies at focal loci as well as at linked loci and it may affect linkage disequilibrium at loci under selection. In loci under selection, minor allele frequencies and effect sizes could become correlated, as alleles of larger effects are expected to be more affected by selection. The authors also mention several methods to leverage between-population genotypic differences, for example by examining the relation between trait-related loci and the principal components resulting from the principal components analysis (PCA) which are usually carried out in GWAS datasets to reduce the effects of population stratification. If trait-associated loci explain a disproportionally large amount of variance in the principal components, they are more likely to be under natural selection, as genotypic differences between populations solely varying due to genetic drift should be random.

DEFINING SOCIAL COMPLEXITY AS A CHARACTERISTIC OF THE INDIVIDUAL

The methods described above use variation in genetics and traits between individuals to gather new knowledge about the evolution of (human) traits. In order to use such methods to determine whether recent selection on social complexity is related to the decrease in brain size observed in humans, a useful step would be to create an individualized measure of social

complexity. Social complexity is typically discussed as a characteristic of societies or species, not of individuals (Dunbar, 1998; Bergman & Beehner, 2015; Kappeler, 2019; Kappeler et al., 2019) and the research supporting the social brain hypothesis demonstrates that social complexity is a useful measure to distinguish these groups. Kappeler (2019) provided a framework to homogenize social complexity measures across studies. While his goal was still to create a measure specific to social groups, the framework may be useful in creating a measure to distinguish different levels of social complexity across individuals.

Kappeler (2019) divided the social system into four components: Social organization, social structure, the mating system and the care system. Social organization encompasses group size, group composition and kinship pattern, which can be measured in the individual by measuring the living situation for each individual: does an individual live alone or with others, and if living together, how large is the group? Social structure on the other hand contains the content, quality and patterning of social interactions engaged in by individuals in a population. The quantity and quality of the relations in which an individual engages can be measured using self-report (e.g. questionnaires or interviews) or by objective methods, such as digital phenotyping. Another characteristic of social structure is communicative complexity, and this is similarly straightforward in its translation to individual subjects as the individuals' communicative skills. The mating system measurements can be adjusted for individuals by examining whether an individual reproduces and with how many individuals. The final component, the care system, can be integrated in individual measurement by examining whether individuals engage in alloparenting, whether this occurs only for family or also for strangers, and whether an individual has children which are alloparented by family and/or strangers. The construction of a measure of social complexity by combining characteristics of social complexity as described here and applying it in a large sample while measuring genomic data will allow for the analysis of genomic signatures of evolution in social complexity similar to the one described for different traits by Song et al. (2021) mentioned in section 3.1 of this review.

PLEIOTROPY IN THE EVOLUTION OF THE HUMAN BRAIN

Most of the hypotheses explaining either the ancient increase in primate and human brain size or the recent decrease rely on some version of assumed pleiotropic effects (two or more traits sharing part of their genetic background). The social brain hypothesis assumes pleiotropic effects between social complexity and brain size, the self-domestication hypothesis relies on a large number of pleiotropic traits, such as aggression, brain size, tooth size and craniofacial features, and collective intelligence as an explanation for the recent reduction in brain size relies on pleiotropy between brain size and intelligence.

An important characteristic of pleiotropy to take into account in evolutionary research is how genetic correlations between traits affect selection and how selection can in turn affect these genetic correlations (Svensson et al., 2021). Pleiotropy can facilitate adaptation (the change in traits in response to selection) or hamper it, depending on the direction of multivariate selection and the direction of the correlation between the traits. In the direction of the correlation, covariance is highest, allowing fast adaptation through standing variation. However, in the opposite direction of the correlation, the covariance is low, which allows for only minor amounts of adaptation on standing variation (Figure 2). Mutations (and potentially epistatic effects) can extend the trait distributions beyond what is available at a given time, but advantageous mutations are expected to be rare (Eyre-Walker & Kneightley, 2007).

Besides the pleiotropy assumed by the various hypotheses related to the evolution of the human brain, another group of traits associated with both social behavior and the brain is neuropsychiatric disorders (Porcelli et al., 2019). Many neuropsychiatric disorders are characterized by extreme variation in social behavior in humans. The overlap between neuropsychiatric disorders and social dysfunction has a genetic component (Bralten et al., 2021; Andreu-Bernabeu et al., 2022), indicating that the social dysfunction at least in part determined by the same biological pathways which affect neuropsychiatric functioning. A genetic relation between social behavior (or more specifically aggression), neuropsychiatric disorders and the brain is also part of the self-domestication hypothesis (Bednarik et al., 2022).

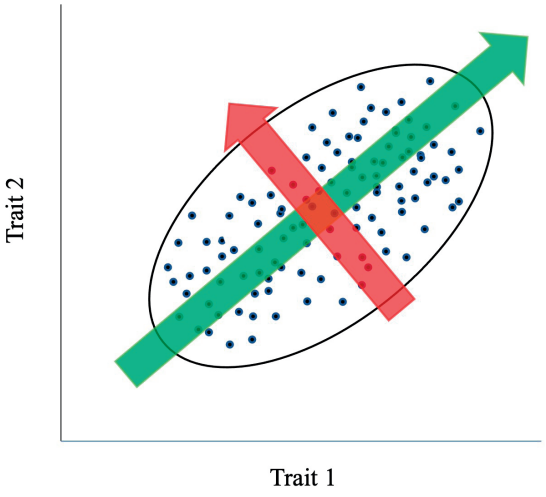


Figure 2. Adaptive response to multivariate selection on correlated traits. The green arrow represents multivariate selection in the direction of the genetic correlation, to which the population can adapt quickly as a result of high levels of genetic covariation. The red arrow indicates a direction of selection in the opposite direction of genetic covariance, which as a result has lower standing variation to serve as a basis for adaptation.

Previous studies have found that various neuropsychiatric disorders are associated with reduced fitness (reproduction and survival) (Bundy et al., 2011; Reininghaus et al., 2015; Mullins et al., 2017; Smith Dawalt et al., 2019). However, despite this evidence of selection, there is no evidence of neuropsychiatric disorders being selected out of the population, a paradox which has been described extensively in other papers (Keller & Miller, 2005; Sella & Barton, 2019) and is outside the scope of this discussion.

This may be an indication that complex selective processes are involved in the evolution of neuropsychiatric disorders which can change how selection affects genetic and phenotypic variation. Due to the pleiotropic relation between neuropsychiatric disorders, social behavior and the brain, these complex processes should be kept in mind when examining recent evolution of social complexity and brain size.

Pleiotropy can be examined using several methods which use either individual-level trait and genomic data or use only summary statistics of GWAS studies (Zhang et al., 2021). Fitness surfaces across multivariate trait such as discussed in Svensson et al. (2021) can be examined to get an idea on whether and how pleiotropy may have affected adaptation.

CONCLUSION

The question of how the human brain has evolved to its current state has resulted in many scientific studies and many hypotheses. Despite this, the more recent and contemporary evolution of the human brain has gained relatively little attention, despite evidence that significant changes are occurring across the globe. With this review, we hope to create more interest in this subject and provide information about methodological resources that can be used to answer these questions.

One reason that it is important to direct more scientific attention to contemporary evolution of the human brain is that the long-term effects of a brain size decrease such as suggested in the evidence provided in the first part of this paper are unpredictable and may have a significant impact over time. For example, the collective intelligence hypothesis implies that due to the lower necessity of individual intelligence, humans are becoming less intelligent individually. It is not known how such reduced individual intelligence would impact human societies over time.

Some proponents of the self-domestication hypothesis suggest that the results of the brain size decrease observed may be severe (Bednarik et al., 2022). They argue that the decrease in brain size is associated with increases in the prevalence of certain neuropsychiatric disorders such

as depressive disorders and schizophrenia. They argue that although the changes which have occurred in the past millennia probably cannot be undone, we can change the future trajectory of the human brain by changing the human environment.

As of yet, it is not possible to be certain what has led to the decreases in brain size and how this could potentially affect human functioning in the future. Due to the low levels of attention attracted by this subject, the level of evidence for both the collective intelligence and self-domestication hypotheses are still low, and the observed decrease in brain size may yet be explained in a very different manner. In the future, both the methods mentioned in this paper as well as other methods (such as experimental evolutionary research) can hopefully answer these questions and provide methods to address or subvert any potential detrimental effects of the recent and contemporary evolutionary trajectory of the human brain.

SUPPLEMENTARY ANALYSIS

To assess the impact of removing SNPs that deviate from HWE, we conducted a preliminary analysis on the UK Biobank genetic data. The UK Biobank contains biomedical data from over 500000 adults from the UK, including genome-wide sequencing data for over 500000 of these subjects. We used PLINK 2 (version v2.00a3.1LM) for the analyses, which were carried out on the UK Biobank Research Analysis Platform.

We corrected for population stratification by including only a homogenous set of individuals defined by Bycroft et al. (2018). We accounted for family structure by removing all individuals with 2nd degree familial relations in the dataset. For quality control, all individuals with discordant self-reported and genotypic sex were removed, SNPs with minor allele frequencies below 0.005 or missing rates over 0.05 were removed and individuals with missing variant rates of 0.1 or higher were removed from the sample. We analyzed deviation from HWE for each SNP per chromosome in order to determine the proportion of SNPs not conforming to HWE. In order to examine how selection may have affected deviations from HWE, we examined to what extent SNPs deviating from HWE were in linkage with other SNPs deviating from HWE. HWE-violating SNPs in significant linkage with other HWE-violating SNPs are especially likely to be under selection, as selection should have correlated effects on genotype frequencies in linked SNPs, while genotyping error should occur randomly.

When we assessed how many variants would be removed based on a HWE p-value threshold of 1-e06, this analysis shows that across chromosomes anywhere from 6.67 to 9.76 percent of SNPs would be removed. Moreover, anywhere between 14.66 and 30.61 percent of SNPs violating HWE were in moderate linkage ($r^2 > 0.2$) with other SNPs violating HWE, while

between 1.71 and 11.11 percent were in high linkage ($r^2 > 0.8$) with other SNPs violating HWE (the complete data can be viewed in Table 1). Particularly interesting is chromosome 6": "Particularly interesting is chromosome 6, which contains the major histocompatibility complex which is known to be evolving rapidly due to strong natural selection (Liston et al., 2021). Plotting the SNPs deviating significantly from HWE on chromosome 6 while highlighting SNPs under LD shows a clear highly colored peak at the location of the MHC, between 29 and 33 million base pairs (Figure 3B). This is a clear indication that many SNPs are removed due to violating HWE which are not genotyped erroneously but under selection. The MHC is also known to be involved in brain development (Elmer & McAllister, 2012; McAllister, 2014) and function (Nelson et al., 2013; Cebrián et al., 2014), as well as several neuropsychiatric disorders (Elmer & McAllister, 2012; Needleman & McAllister, 2012; Nelson et al., 2013; Cebrian et al., 2014; McAllister, 2014) which exacerbates the issue of not including these SNPs in GWAS studies examining neuropsychiatric disorders or neurobiological characteristics. No other chromosomes showed patterns as remarkable as chromosome 6, although some evidence of regions under selection could still be found looking at plots of other chromosomes (Figure 3A).

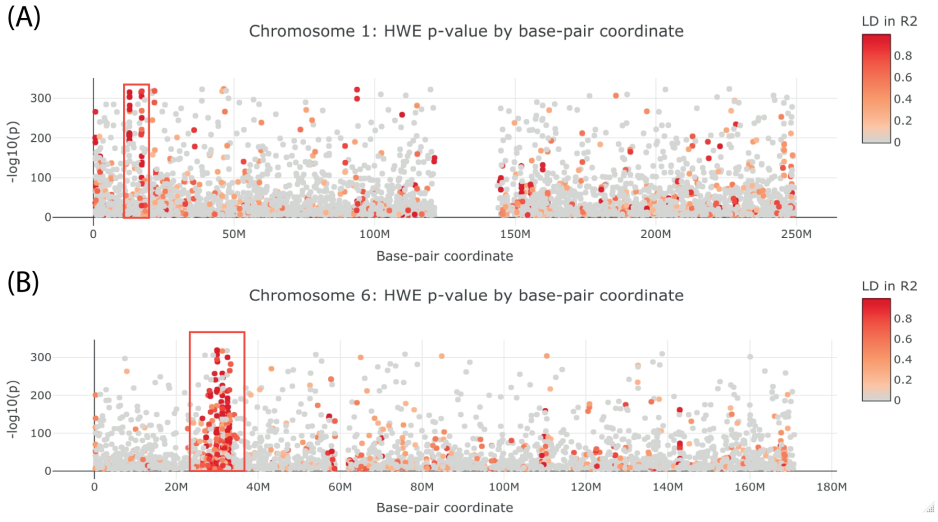


Figure 3. Evidence of selection among SNPs violating Hardy-Weinberg equilibrium (HWE). SNPs are colored by their maximum linkage with other SNPs violating HWE. SNPs violating HWE in linkage disequilibrium with other SNPs violating HWE are likely to be affected by selection and not genotyping error, as genotyping error should occur randomly while selection should have correlated effects on linked loci. Red boxes indicate regions with high numbers of SNPs violating HWE while in high linkage disequilibrium, hinting that these loci may be subject to selection pressure. HWE = Hardy-Weinberg equilibrium, LD = linkage disequilibrium.

Table 1. Numbers and percentages of SNPs violating Hardy-Weinberg equilibrium in linkage disequilibrium calculated for each chromosome.

Chromosome	# SNPs passing QC	# SNPs violating HWE	% SNPs violating HWE	# SNPs violating HWE and in high LD (R2 > 0.8)	% SNPs violating HWE in high LD (R2 > 0.8)	# SNPs violating HWE and in moderate LD (R2 > 0.2)	% SNPs violating HWE in moderate LD (R2 > 0.2)
1	50374	4140	8.22%	179	4.32%	803	19.40%
2	50785	3619	7. 13%	95	2.63%	653	18.04%
3	42685	2859	6.70%	92	3.22%	458	16.02%
4	40228	2927	7.28%	64	2.29%	493	16.84%
5	38491	2835	7.37%	68	2.40%	508	17.92%
6	44065	3267	7.41%	363	11.11%	1000	30.61%
7	34969	2707	7.74%	109	4.03%	529	19.54%
8	32973	2371	7.19%	59	2.49%	466	19.65%
9	27947	2060	7.37%	72	3.50%	348	16.98%
10	31613	2273	7.19%	84	3.70%	436	19.18%
11	31395	2306	7.35%	86	3.73%	451	19.56%
12	30062	2120	7.05%	65	3.07%	360	16.98%
13	21735	1501	6.91%	25	1.67%	228	15.19%
14	20501	1480	7.22%	44	2.97%	270	18.24%
15	19981	1436	7.19%	73	5.08%	293	20.40%
16	22484	1673	7.44%	68	4.06%	339	20.26%
17	20578	1500	7.29%	37	2.47%	262	17.47%
18	18930	1286	6.79%	22	1.71%	196	15.24%
19	17292	1675	9.76%	109	6.51%	377	22.51%
20	16566	1105	6.67%	19	1.72%	162	14.66%
21	9528	689	7.23%	23	3.34%	102	14.80%
22	10330	859	8.32%	26	3.03%	158	18.39%

REFERENCES

Andreu-Bernabeu, Á., Díaz-Caneja, C.M., Costas, J., De Hoyos, L., Stella, C., Gurriarán, X., Alloza, C., … González-Peñas, J. (2022). Polygenic contributions to the relationship of loneliness and social isolation with schizophrenia. *Nature Communications, 13*, 51. doi:10.1038/s41467-021-27598-6

Barton, R. A. & Dunbar, R. I. M. (1997). Evolution of the social brain. In A. Whiten and R. W. Byrne (Eds.), *Machiavellian Intelligence II: Extensions and evaluations.* (pp. 240-263). Doi:10.1017/CBO9780511525636.010

Bednarik R. G. (2014). Doing with less: hominin brain atrophy. *Homo : internationale Zeitschrift fur die vergleichende Forschung am Menschen, 65*(6), 433–449. doi:10.1016/j.jchb.2014.06.001

Bednarik, R. G., Saniotis, A., & Henneberg, M. (2022). Auto-domestication hypothesis and the rise in mental disorders in modern humans. *Medical Hypotheses, 164*:110874. doi:10.1016/j.mehy.2022.110874

Bergman, T. J., Beehner, J. C. (2015). Measuring social complexity. *Animal Behaviour, 103*, 203-209. doi:10.1016/j.anbehav.2015.02.018

Bralten, J., Mota, N. R., Klemann, C. J. H. M., De Witte, W., Laing, E., Collier, D. A., … Poelmans, G. (2021). Genetic underpinnings of sociability in the general population. *Neuropsychopharmacology 2021 46:9*, *46*(9), 1627–1634. doi:10.1038/s41386-021-01044-z

Brown, P. (1987). Pleistocene homogeneity and Holocene size reduction: the Australian human skeletal evidence *Archeology in Oceania, 22*(2), 41-67. doi:10.1002/j.18344453.1987.tb00166.x

Bundy, H., Stahl, D., & MacCabe, J. H. (2011). A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatrica Scandinavica*, *123*(2), 98–106. doi:10.1111/J.1600-0447.2010.01623.X

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature, 562*(7726), 203–209. doi: 10.1038/s41586-018-0579-z

Cebrián, C., Loike, J. D., & Sulzer, D. (2014). Neuronal MHC-I expression and its implications in synaptic function, axonal regeneration and Parkinson's and other brain diseases. *Frontiers in neuroanatomy, 8*, 114. doi:10.3389/fnana.2014.00114

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*(5),178-190. doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M. (2009). The social brain hypothesis and its implications for social evolution. *Annals of Human Biology, 36*(5), 562-572. Doi: 10.1080/03014460902960289

DeSilva, J. M., Traniello, J. F. A., Claxton, A. G., & Fannin, L. D. (2021). When and Why Did Human Brains Decrease in Size? A New Change-Point Analysis and Insights From Brain Evolution in Ants. *Frontiers in Ecology and Evolution, 9*, 712. doi:10.3389/fevo.2021.742639/bibtex

Elmer, B. M., & McAllister, A. K. (2012). Major histocompatibility complex class I proteins in brain development and plasticity. *Trends in neurosciences, 35*(11), 660–670. doi:10.1016/j.tins.2012.08.001

Eyre-Walker, A. & Kneightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics, 8,* 610-618. doi:10.1038/nrg2146

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M. I., & Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science (New York, N.Y.), 354*(6313), 760–764. doi:10.1126/science.aag0776

Guo, J., Yang, J., & Visscher, P. M. (2018). Leveraging GWAS for complex traits to detect signatures of natural selection in humans. *Current Opinion in Genetics & Development, 53*, 9-14. doi:10.1016/j.gde.2018.05.012

Hawks, J. (2011). Selection for smaller brains in Holocene human evolution. *arXiv [preprint].* Available at: Available online at: https://arxiv.org/abs/1102.5604#:~{}:text=Background%3A%20Human%20populations%20during%20the,linear%20measurements%20of%20the%20cranium (accessed 2022).

Henneberg, M. (1988). Decrease of human skull size in the Holocene. *Human Biology, 60*(3), 395-405.

Henneberg, M., & Steyn, M. (1993). Trends in cranial capacity and cranial index in Subsaharan Africa during the Holocene. *American journal of human biology: the official journal of the Human Biology Council, 5*(4), 473–479. doi:10.1002/ajhb.1310050411

Henneberg, M. (2004). The rate of human morphological microevolution and taxonomic diversity of hominids. *Studies in Historical Anthropology, 4*, 49-59.

Isler, K., & van Schaik C. (2006). Metabolic cost of brain size evolution. *Biology letters, 2*(4), 557-560. doi: 10.1098/rsbl.2006

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., … Kolm, N. (2013). Artifical selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology, 23*(2), 168-171. doi:10.1016/j.cub.2012.11.058

Kappeler, P. M. (2019). A framework for studying social complexity. *Behavioral ecology and Sociobiology, 73*(13). Doi:10.1007/s00265-018-2601-8

Kappeler, P. M., Clutton-Brock, T. Shultz, S., & Lukas, D. (2019). Social complexity: Patterns, processes, and evolution. *Behavioral Ecology and Sociobiology, 73*:5. doi:10.1007/s00265-018-2613-4

Keller, M. C., & Miller, G. (2006). Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *Behavioral and Brain Sciences*, *29*(4), 385–404. doi:10.1017/S0140525X06009095

Liston, A., Humblet-Baron, S., Duffy, D., & Goris, A. (2021). Human immune diversity: from evolution to modernity. *Nature immunology, 22*(12), 1479–1489. doi:10.1038/s41590-021-01058-1

Liu, C. Tang, Y., Ge, H., Wang, F., Sun, H., Meng, H., … Liu., S. (2014). Increasing breadth of the frontal lobe but decreasing height of the human brain between two Chinese samples from a Neolithic site and from living humans. *American Journal of Physical Anthropology, 154*, 94-103. Doi:10.1002/ajpa.22476

McAllister A. K. (2014). Major histocompatibility complex I in brain development and schizophrenia. *Biological psychiatry, 75*(4), 262–268. doi:10.1016/j.biopsych.2013.10.003

Mullins, N., Ingason, A., Porter, H., Euesden, J., Gillett, A., Ólafsson, S., … Stefansson, K. (2017). Reproductive fitness and genetic risk of psychiatric disorders in the general Population. *Nature Communications, 8*. doi:10.1038/NCOMMS15833

Nelson, P. A., Sage, J. R., Wood, S. C., Davenport, C. M., Anagnostaras, S. G., & Boulanger, L. M. (2013). MHC class I immune proteins are critical for hippocampus-dependent memory and gate NMDAR-dependent hippocampal long-term depression. *Learning & memory, 20*(9), 505–517. doi:10.1101/lm.031351.113

Needleman, L. A., & McAllister, A. K. (2012). The major histocompatibility complex and autism spectrum disorder. *Developmental neurobiology, 72*(10), 1288–1301. doi:10.1002/dneu.22046

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., … Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience & Biobehavioral Reviews*, *97*, 10–33. doi:10.1016/J.NEUBIOREV.2018.09.012

Reininghaus, U., Dutta, R., Dazzan, P., Doody, G. A., Fearon, P., Lappin, J., … Jones, P. B. (2015). Mortality in Schizophrenia and Other Psychoses: A 10-Year Follow-up of the ÆSOP First-Episode Cohort. *Schizophrenia Bulletin*, *41*(3), 664. doi:10.1093/SCHBUL/SBU138

Reséndiz-Benhumea, G. M., Sangati, E., Sangati, F., Keshmiri S, Froese, T. (2021) Shrunken Social Brains? A Minimal Model of the Role of Social Interaction in Neural Complexity. *Frontiers Neurorobotics*, *15*:634085. doi:10.3389/fnbot.2021.634085

Sella, G., & Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual Review of Genomics and Human Genetics, 20*, 461-493. doi:10.1146/annurev-genom-083115-022316

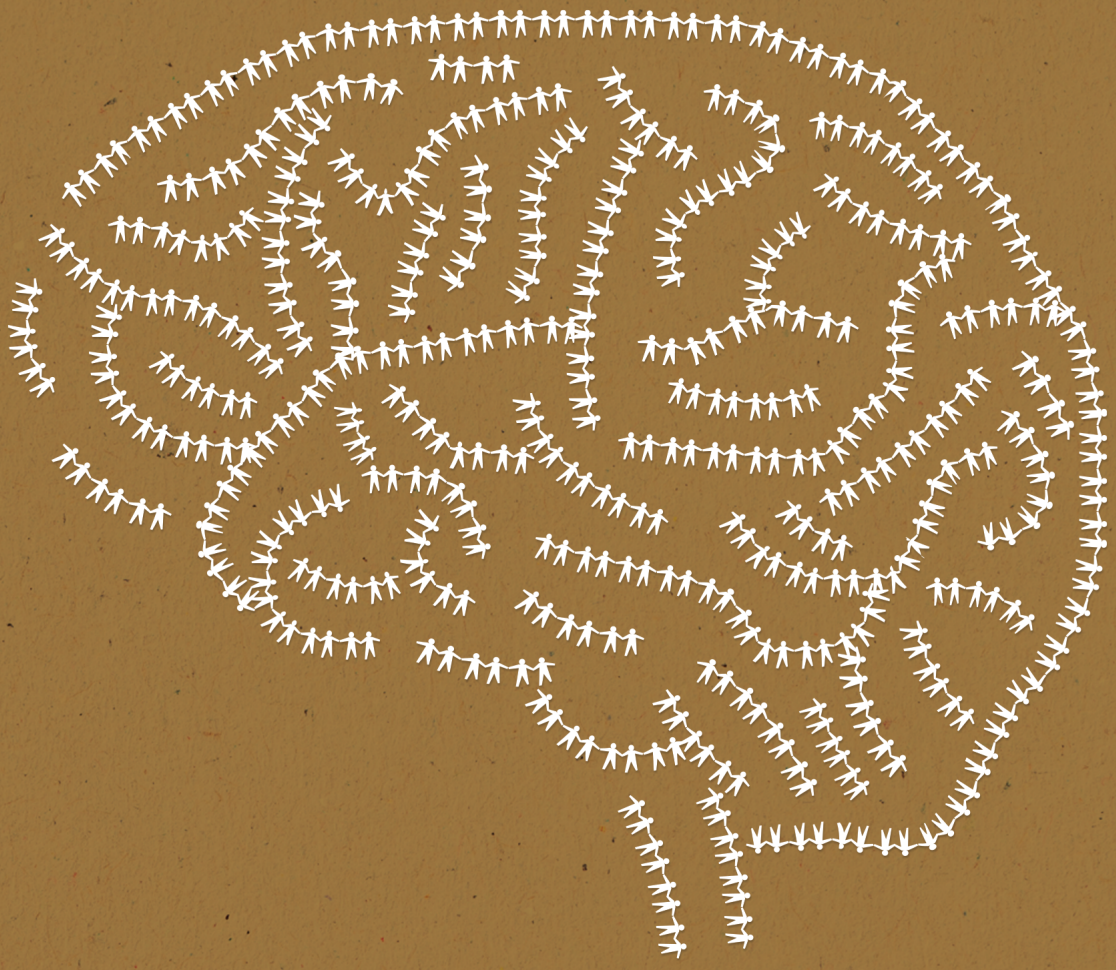
Smith DaWalt, L., Hong, J., Greenberg, J. S., & Mailick, M. R. (2019). Mortality in individuals with autism spectrum disorder: Predictors over a 20-year period. *Autism : The International Journal of Research and Practice*, *23*(7), 1732–1739. doi:10.1177/1362361319827412

Song, W., Shi, Y., Wang, W., Pan, W., Qian, W., Yu, S., … Lin, G. N. (2021). A selection pressure landscape for 870 human polygenic traits. *Nature Human Behaviour , 5*(12), 1731–1743. Doi:10.1038/s41562-021-01231-4

Svensson, E. I., Arnold, S. J., Biürger, R., Csilléry, K., Draghi, J., Henshaw, J. M., … Runemark, A. (2021). Correlational seledction in the age of genomics. *Nature Ecology & Evolution, 5*, 562-573. doi:10.1038/s41559-021-01413-3

Villmoare, B., & Grabowski, M. (2022). Did the transition to complex societies in the Holocene drive a reduction in brain size? A reassessment of the DeSilva et al. (2021) hypothesis. *Frontiers in Ecology and Evolution, 10*. doi:10.3389/fevo.2022.963568

Zhang, Y., Cheng, Y., Jiang, W., Ye, Y., Lu, Q., & Zhao, H. (2021). Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Briefings in Bioinformatics, 22*(5). doi:10.1093/bib/bbaa442



Chapter 2

Evolutionary origins of human social behavior: Assessing conservation of human sociability genes in *Caenorhabditis elegans*

Mila C. Roozen[1] & Martien J. H. Kas[1]

[1]*Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands*

In preparation for submission

ABSTRACT

Social behavior is a common though variable trait across animal species. How much of the variation in social behavior is due to biological common mechanisms across animal species is unknown. In this study we examined to what extent human genetic variation in sociability is affected by pathways shared with *Caenorhabditis elegans* and whether any conserved sociability genes show enhanced levels of essential functions and interactivity. We found inconsistent evidence of increased conservation with more thorough analyses resulting in no evidence of increased conservation of human sociability genes. Conserved genes were highly interactive compared to nonconserved and random genes, while only a limited number of genetic interactions were found to be conserved. No evidence was found for enrichment of social phenotypes in *C. elegans* orthologs of human sociability genes while evidence for associations with essential functions were limited. The gene *ACVR2A* appears to play a role in social behavior in both humans and *C. elegans*, making it an interesting gene for further study.

INTRODUCTION

Social or collective behavior is a trait observed across many species, from collective behaviors in single-celled slime moulds (Reid & Latty, 2016) to the complex social behaviors observed in primates and (by extension) humans (Roberts & Roberts, 2016). Some aspects of social behavior serve functions with obvious benefits to reproductive fitness, such as finding and courting potential mates and caring for children in order to ensure survival of offspring. However, several species also engage in social behaviors with less obvious effects on reproductive fitness or that may even seem detrimental. For example, the nematode *Caenorhabditis elegans* engages in a form of social feeding where the individual nematodes clump together around a food source (Thomas, 1998). As another example, bonobos, the closest living relatives to humans, share food with strangers even when there is no apparent benefit to themselves (Tan & Hare, 2013).

Although social behavior is evident in many species, the quantity of social contact engaged in by individuals from different species and the complexity of the social structures created by those individuals vary to a strong degree. Some species are mostly solitary and engage in social behavior mostly when it directly increases fitness, such as for feeding, mating or parenting (for example pumas (Elbroch et al., 2017) and koalas (Ellis et al., 2009)) while some other species spend most or all of their lives in social groups, such as chimpanzees and bonobos (Gruber & Clay, 2016) as well as several species of eusocial insects (Robinson et al., 1997).

The consistent variability in social behavior between species indicates that genetic variation in social behavior exists and has been under natural selection across speciation. In primates, which display strong variations in the extent to which they engage in social behaviors with kin and strange animals, interspecies variation in social behavior has been linked to variation in brain size, leading to the hypothesis that the large brain sizes observed in some primates (such as humans) are the result of long-term selection on social complexity (the ability to cognitively handle high numbers of complex social relationships) (Dunbar, 2009).

On the other hand, the fact that social behavior is common across species indicates some shared genetic basis across species. Evidence for such a shared basis of social behavior between humans and other species has been provided by experimental studies using genetic manipulations in model animals. Genes associated with extreme variations in social behavior in humans can be altered in model animals in order to examine whether these genes affect social behavior in these animals in a similar manner. For example, a gene known to be involved in the extreme sociability of individuals with Williams Syndrome (a deletion on a section of chromosome 7), *GTF2I*, similarly increases social behavior towards strangers in mice when gene function is disrupted (Sakurai et al., 2010) and structural gene variants in this gene are associated with the stereotypical hypersociability seen in domesticated dogs (vonHoldt et al., 2017). Several

successful mouse models (such as *Shank3* and *Foxp2*) have been created based on human mutations associated with autism spectrum disorders, which are characterized in part by aberrant social behavior (Crawley, 2022).

Although cross-species studies has elucidated important aspect of the evolution of social behavior across species, much is still unknown about the evolution of the molecular substrates underlying such behavior. While many species display some form of social behavior at the very least in the pursuit of procreation, this does not necessarily mean that the genetic background underlying such behavior is similar between species, especially those far removed from each other. However, three recent studies have provided some evidence that might point to the conservation of molecular mechanisms underlying social behavior across far removed species. In three separates studies, Kasap et al. (2018), Franklin and Dwyer (2020) and Sall et al. (2021) found evidence that putative risk genes for schizophrenia, bipolar disorder and major depressive disorder were highly conserved between *C. elegans* and humans, indicating that the molecular basis for these disorders may have been present even in the last common ancestor of humans and *C. elegans*. In these studies, they found that these genes were associated with essential functions for reproductive fitness (survival and reproduction) and had heightened levels of interactivity, which may both explain why these genes are so highly conserved despite variation in these genes being associated with deleterious effects in humans.

These findings could indicate that similar processes have affected genes related to social behavior, as these disorders are commonly associated with social dysfunction (Lahera et al., 2015; Tatay-Manteiga et al., 2018; Grant et al., 2001; Saris et al., 2022; Kupferberg et al., 2016) and genetic studies hint at shared molecular substrates (Álvaro Andreu-Bernabeu et al., 2022; Bralten et al., 2021).

In this study, we will examine to what extent genes associated with human variability in sociability are conserved between humans and *C. elegans*. We expect conservation to be heightened for sociability genes compared to the total human protein-coding genome, similar to the findings for neuropsychiatric disorders. We will then examine whether any conserved genes constitute a shared genetic basis for sociability between humans and *C. elegans.* Finally, we will examine whether conserved genes are enriched for lethal and sterile phenotypes in *C. elegans* and whether conserved genes show enhanced interactivity compared to non-conserved sociability genes to examine potential causes of the long-term conservation.

METHOD

Data collection

Seventy-six genes related to sociability were extracted from the sociability GWAS performed by Bralten et al. (2021). Seven of these were excluded from further analysis, either because they were not included in the current Ensembl (Cunningham et al., 2022) database (3 genes) or because they did not code for any proteins, indicating they could not be analyzed using the methods planned for this study. This resulted in a total of 69 sociability genes which were included in the following analyses.

Ortholog detection

A stepwise approach was taken to determine whether genes were conserved. The first step was to search for established orthologues in the Ensembl database (Cunningham et al., 2022). The second step was to search for established orthologs in the WormBase database (WormBase version WS285, http://www.wormbase.org; Davis et al., 2022). Finally, if no ortholog was found in the first steps, BLASTP was used (in WormBase) to determine whether the largest transcript of each gene had a clear functional counterpart in *C. elegans*. The total overview of the genes with their respective orthologues and the method through they were discovered, see Supplementary Table 1.

In order to determine whether BLASTP results of similar proteins could indeed be considered functional counterparts, BLASTP hits were evaluated against criteria based on those by Kasap et al. (2018) and based on the discussion by Pearson (2013). Transcripts were considered functional counterparts if 1) they did not differ over 100 amino acids in length; 2) the E-value of the hit was below 10^{-4}; 3) the identity was at least 20% for a segment of at least 50 amino acids in length and 4) for at least three species with established orthologs of the human gene, the same *C. elegans* gene was also a hit using BLASTP.

Statistical analyses

All analyses were carried out in R using RStudio (R version 4.2.1, RStudio version 2022.07.2). The number of orthologs were compared with data from previous efforts to study the level of conservation found across the human genome. In a large effort to determine orthologs between human and *C. elegans* genes, Kim et al. (2018) found that approximately 52.6% (10678 out of 20310 genes) of the human protein-coding genome had recognized orthologs in *C. elegans* at that time. We used Fisher's exact test to examine whether the differences between the proportion of human genes with *C. elegans* orthologs according to Ortholist II (Kim et al., 2018) and the proportion of human protein-coding sociability genes with a *C. elegans* ortholog. The genes from the sociability set were removed from the total human gene set to create independent gene sets.

However, as the determination of homologs in the study by Kim et al. (2018) was not directly comparable to the methods used in this study, we also compared the sociability set to ten random sets of human genes created using the Molbiotools Random Sequence Generator (https://molbiotools.com/randomsequencegenerator.php) with the same number of genes as the sociability set. The random human gene sets created for the analysis of the level conservation and the interactivity can be found in Supplementary Table 2. The comparisons were planned to be carried out using chi-squared tests unless any expected counts in the contingency table are below 10, in which case Fisher's exact test will be utilized. In addition, the z-score of the sociability set was compared to the distribution of all sets (random and sociability) to determine whether the sociability set is an outlier.

The function of conserved genes was examined in *C. elegans* using the WormBase Worm Phenotype Ontology database (Schindelman et al., 2011) to determine whether a) conserved sociability genes constitute a shared genetic basis for sociability between humans and *C. elegans* and b) whether conserved sociability genes may have been conserved as a result of functions essential to fitness. A type of social behavior which has been previously studied in *C. elegans* is social feeding. Social feeding can be measured by the aggregation of the nematodes or by 'bordering', a behavior typical of social feeding animals whereby the animals feeds from the edges of a bacterial lawn (the food source). The WormBase Phenotype Ontology dataset includes four phenotypes related to social feeding: social feeding, solitary feeding, bordering and non-bordering. We used WormMine (http://intermine.wormbase.org/tools/wormmine/begin.do) to query the *C. elegans* genes in the WormBase dataset for these phenotypes and we examined the presence of social phenotypes in *C. elegans* orthologs of human sociability genes to determine whether these genes constitute a shared genetic basis of sociability.

In order to determine whether human sociability genes were conserved between humans and *C. elegans* as a result of essential functions, we repeated the above analysis but instead queried WormBase for the five phenotypes related to fitness: 'lethal', 'embryonic lethal', 'larval lethal', 'sterile' and 'sterile progeny'. For each phenotype, a subset was retrieved from the data for *C. elegans* orthologs of human sociability genes which were associated with the phenotype. We then compared the proportion of each of these phenotypes in the orthologues of the human sociability genes with the proportion of these phenotypes in the total set of 19985 *C. elegans* protein-coding genes available in the WormBase dataset (Davis et al., 2022) to determine whether the conserved genes are more likely to have any of these essential functions in *C. elegans*. The analyses were performed using chi-squared tests of independence if all expected cell counts were above 10 or Fisher's exact test if this was not the case.

As a final step we wanted to examine whether certain sociability genes might be conserved between *C. elegans* and humans as a result of having particularly high numbers of interactions

with other genes. Interactions were examined using GeneMania (Wade-Farley et al., 2010). The 'max resultant gene' setting, which determines how many genes from outside the gene sets could be used to create interactions was set to 0, as well as the 'max resultant attributes'. We created two human gene sets, one with the conserved sociability genes and one with the non-conserved sociability genes, in order to examine whether sociability genes in general or conserved sociability genes specifically showed increased levels of interactions. We compared the gene sets each to the ten random gene sets created for the comparison regarding the level of conservation, where the sets were split two sets to conform to the sizes of the conserved and non-conserved sets. In order to perform a second check whether any potential increased interactivity between sociability genes is a result of the genes being involved in the same phenotype, we performed the analysis a second time, changing the 'max resultant gene' setting to 20. In this manner, highly interactive genes from the random sets would be detected even if these interactions were found outside of the random genes with which they are assigned. Statistical tests will be performed using ANOVA models. If assumptions of the ANOVA model are violated, generalized linear models with appropriate model specifications will be used.

The Molbiotools Random Sequence generator does not include an option to create random sets for *C. elegans*. In order to be able to compare the interactions found in human sociability genes with those found in the *C. elegans* orthologs, we downloaded the full set of *C. elegans* protein-coding genes from WormBase and sampled 10 random sets, each of the same size as the set of *C. elegans* sociability orthologs. The random gene sets can be found in Supplementary Table 3. In order to create an approximately fair comparison, we grouped the orthologs per human gene for which they were homologous, because *C. elegans* genes which were homologous to the same human gene were highly interactive amongst themselves. We then counted the interaction with other gene groups. In total, this led to 49 gene groups (one for each human gene minus one, as two human genes shared the same *C. elegans* ortholog). Statistical comparisons are carried out similar to those for human genes. Finally, we examined whether any interactions between genes were conserved between humans and *C. elegans.*

RESULTS

Conservation

Thirty-five out of 69 human protein-coding sociability genes (51%) had a registered *C. elegans* ortholog in the Ensembl database. The search for orthologs of human sociability genes in *C. elegans* using WormBase led to the addition of 13 out of the remaining 35 genes, leading to a total of 48 out of 69 sociability genes (70%) with a known ortholog. The search for functional counterparts using BLASTP led to the discovery of 2 orthologs in *C. elegans*, resulting in a total of 50 out of 69 (72%) conserved sociability genes in *C. elegans*. Two human sociability genes

were orthologs of the same *C. elegans* gene. As a result of one-to-many and many-to-many orthologs, the list of 50 conserved human sociability genes resulted in 70 *C. elegans* sociability orthologs, of which 4 were pseudogenes, which were not included in subsequent analyses, while the remaining 66 genes were protein coding genes. Based on the comparison with the Kim et al. (2018) data, the proportion of *C. elegans* orthologs appears significantly increased in the set of human sociability genes compared to the total human protein-coding gene set (excluding the sociability genes) ($X^2 = 10.20$, $p = 0.001$). However, the comparison to the ten random sets showed zero pairwise differences between the sociability set and the random set (see Table 1 for the full results). The sociability set had a z-score of 0.49, indicating that there is no evidence that the sociability score deviates from the norm in regards to level of conservation.

Table 1. Pairwise chi-squared tests comparing gene conservation in random sets to conservation in the sociability set.

Set	# Conserved (%)	X^2	p-value
1	44 (63.77%)	0.83	0.361
2	51 (73.91%)	0	1
3	53 (76.81%)	0.15	0.696
4	47 (68.12%)	0.14	0.710
5	41 (59.42%)	2.07	0.151
6	49 (71.01%)	0	1
7	50 (72.46%)	0	1
8	51 (73.91%)	0	1
9	44 (63.77%)	0.83	0.361
10	50 (72.46%)	0	1
Sociability	50 (72.46%)	-	-

None of the examined social phenotypes showed a significant increase in the *C. elegans* orthologs of human sociability genes. In fact, only one of the *C. elegans* orthologs of the human sociability genes, *daf-1* (which is an ortholog of the human gene *ACV2RA*), was associated with any of the four social behavior phenotypes examined. However, the WormBase catalogue showed a dearth of the social behavior phenotypes in general, with no results based on RNAi studies for solitary feeding and no results in any type of study for the non-bordering phenotype. Therefore, this result may simply point to a lack of research into social phenotypes in *C. elegans.*

The results of the analysis regarding the presence of essential phenotypes among *C. elegans* orthologs of human sociability genes are covered in Table 2. Several *C. elegans* orthologs of human sociability genes were associated with one of the essential phenotypes. All essential phenotypes had expected cell counts below ten, therefore all analysis were carried out using Fisher's exact test. Using the full set of 66 protein-coding orthologs, none of the phenotypes showed enrichment in orthologs of human sociability genes.

One human sociability gene which may have affected the results of the analysis about essential phenotypes is *OR5B17*. WormBase indicates that this gene is an orthologue of a group of 13 genes from the *srsx* family, 9 of which are protein-coding. The *srsx* gene family is one of the families of highly divergent chemoreceptor gene families in *C. elegans* (Robertson & Thomas, 2006). We therefore also carried out the analysis excluding the *srsx* genes which did not have any information available, removing all *srsx* genes from the data. This does remove one gene with lethal and sterile phenotypes from the data (*srsx-39*), but it also removes 8 genes with no essential phenotypes. Although most of the results do not change based on this adjustment, the enrichment of lethal genes in *C. elegans* orthologs of human sociability genes became significant (Odds ratio = 0.42, 95% CI = [0.22 - 0.91], p = 0.018).

Table 2. Enrichment of essential phenotypes in *C. elegans* orthologs of human sociability genes, including and excluding the *srsx* gene family. OR = Odds Ratio; CI = Confidence Interval.

Includes *srsx* gene family	Phenotype	Proportion in total protein-coding genome	Proportion in sociability orthologs	OR (95% CI)	p-value
Yes	Lethal	0.09	0.17	0.51 [0.26 – 1.08]	.507
	Embryonic lethal	0.15	0.21	0.66 [0.36 – 1.29]	.168
	Larval lethal	0.04	0.05	0.90 [0.29 – 4.49	0753
	Sterile	0.12	0.18	0.62 [0.33 – 1.28]	.131
	Sterile progeny	0.04	0.02	2.42 [0.42 – 97.11]	.733
No	Lethal	0.09	0.19	0.42 [0.22 – 0.91]	**.018***
	Embryonic lethal	0.15	0.25	0.54 [0.29-1.08]	.061
	Larval lethal	0.04	0.05	0.77 [0.25-3.86]	.509
	Sterile	0.12	0.21	0.51 [0.27 – 1.08]	.063
	Sterile progeny	0.04	0.02	2.08 [0.36 – 83.87]	.724

The random human gene sets created for the analysis of the level of interaction can be found in Supplementary Table 2. The 50 conserved sociability genes were compared with ten sets of 50 random human genes each and the 19 non-conserved genes were similarly compared with ten sets of 19 random genes as the total possible number of interactions is constrained by the total number of genes, therefore comparing sets of 19 to sets of 50 genes would lead to biased results.

Levene's test for equality of variances showed that for both the conserved (F = 12.12, p < .001) and non-conserved (F = 2.98, p = .002) groups of sets, the sets varied significantly in their variances. As a result, generalized linear models using Poisson distributions were considered. However, significant overdispersion (difference between mean and variance) was observed for the conserved set and its comparison sets (dispersion = 1.76, p < .001), which violates the assumptions of the Poisson model. This was not the case for the non-conserved set and its

comparison sets (dispersion = 1.09, p = .151). Instead, a quasipoisson model was used which estimates the variance instead of assuming equal mean and variance. This method will be used for the non-conserved set and associated random sets as well.

The conserved set had significantly more interactions between genes compared to each random set separately (Figure 1A through D and Figure 2). The non-conserved set had the fewest total number of interactions and significantly fewer interactions compared to 7 of the random sets. In *C. elegans,* orthologs of human social complexity were significantly more interactive compared to random gene sets (Figure 1E & Figure 1F). Allowing for the inclusion of maximum 20 genes outside the gene set did not change the results in a significant way, the conserved set was still significantly more interactive compared to all associated random sets while the non-conserved set was now significantly less interactive compared to 5 of the associated random sets (Supplementary Figure 1). Very few of the interactions found in the human data were also found in the *C. elegans* data. Only 9 out of 133 human sociability gene interactions were also found in *C. elegans* orthologs of these genes, whereas for *C. elegans* this was 9 out of 44. The interactions found across species are displayed in Table 3.

Table 3. Interactions discovered both between human sociability genes and between their orthologs in *C. elegans*. Square brackets indicate that this human gene has several *C. elegans* orthologs. [1]*jmjd-1.1* and *jmjd-1.2* are both orthologs of *PHF2*.

Human gene 1	Human gene 2	C. elegans gene 1	C. elegans gene 2
ACVR2A	*GBE1*	*daf-1*	*T04A8.7*
APLNR	*DRD2*	*npr-33*	*ser-5*
ARNTL	*SIM1*	*aha-1*	*hif-1*
ARNTL	*SIPA1L1*	*aha-1*	*F53A10.2*
FBXL17	*UBE3B*	*fbxl-1*	*oxi-1*
GRIK2	*NRXN1*	*glr-1*	*nrx-1*
ISL-1	*LPXN*	*lim-7*	*pxl-1*
PARVA	*SMNTL*	*pat-6*	*T15B12.1*
PHF2	*SSRP1*	*jmjd-1.1 & jmjd-1.2*[1]	*athp-1*

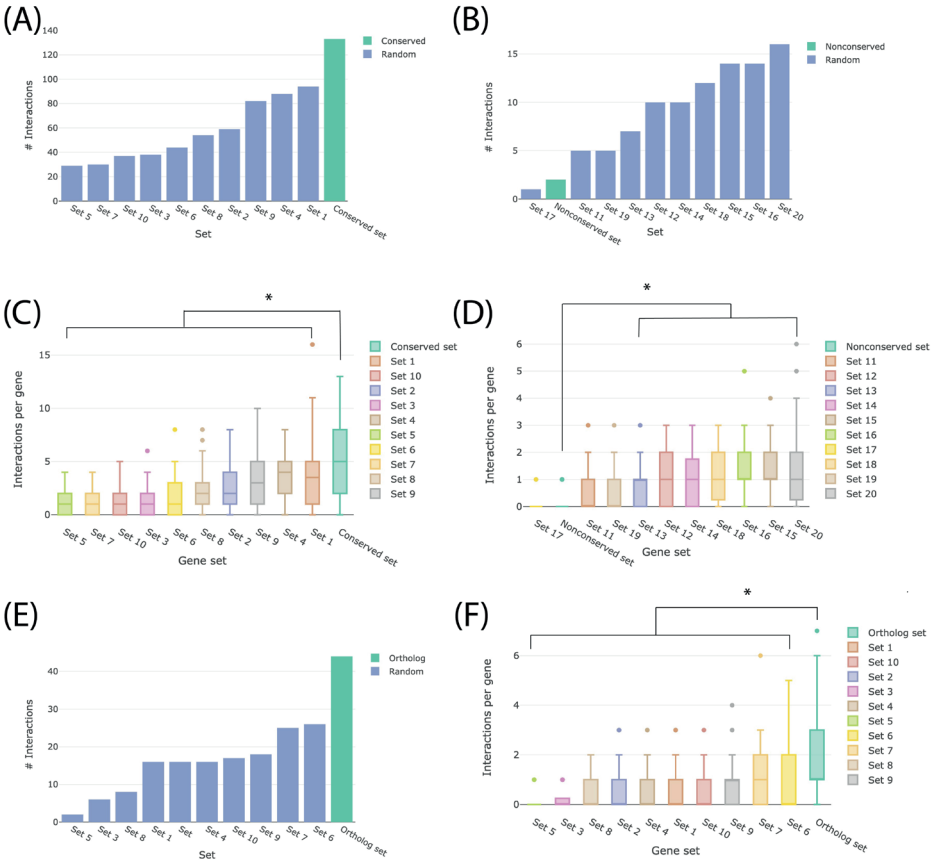


Figure 1. Interaction analyses showed increased interactivity between human sociability genes conserved between humans and *C. elegans*. A) Total interactions between conserved sociability genes and random sets of 50 human genes; B) Total interactions between non-conserved sociability genes and random sets of 19 human genes; C) Interactions per gene for conserved sociability genes and random sets of 50 human genes; D) Interactions per gene for non-conserved sociability genes and random sets of 19 human genes; E) Total interactions between *C. elegans* orthologs of human sociability genes and random sets of 49 *C. elegans* genes; F) Interactions per gene for *C. elegans* orthologs of human sociability genes and random sets of 49 *C. elegans* genes. * = $p < .05$

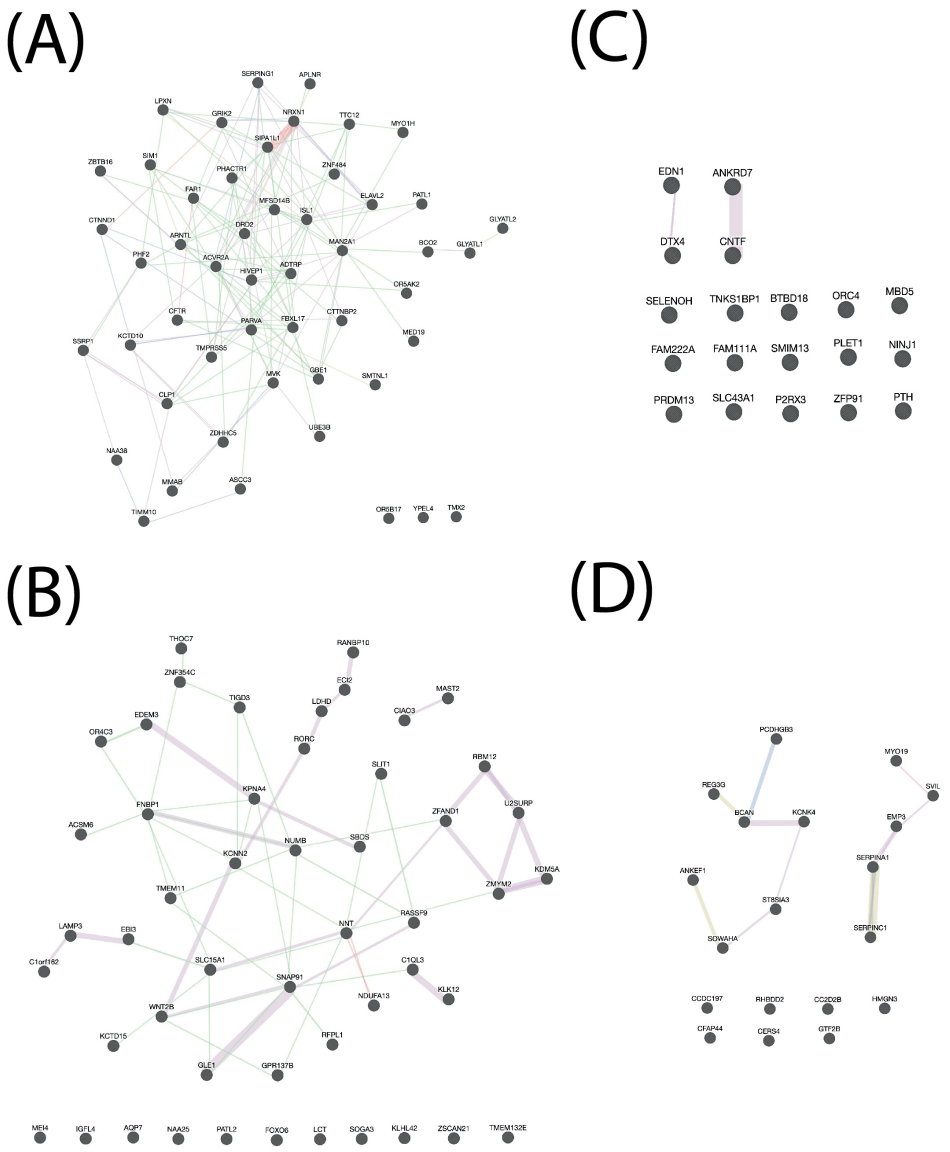


Figure 2. Network images from sociability gene sets (A for conserved genes and C for non-conserved genes) and random gene sets of the same size (B and D) as the conserved (A) and non-conserved (C) sociability gene sets. Random sets 8 (B) and 12 (D) were used for visual comparison as their total number of interactions was the mode across all gene sets of their respective sizes.

DISCUSSION

In this study, we examined to what extent genes associated with variation in human sociability are conserved between humans and *C. elegans*. Using two different analyses, we found evidence both in favor of and in contradiction with the expectation that human sociability genes would show heightened conservation between humans and *C. elegans*. As the analysis using data from Kim et al. (2018) relies on a comparison between differently constructed sets of homologs, we consider the comparison with the random sets more reliable and therefore regard these findings as evidence against heightened conservation of human sociability genes in *C. elegans*. These findings also highlight the importance in choice of methods for such studies, as suboptimal methods can lead to inaccurate results.

We found only one gene associated with social behavior in both humans and *C. elegans,* specifically *ACVR2A*. Unlike the studies on neuropsychiatric disorders, we found no evidence of enrichment for lethal or sterile phenotypes in the *C. elegans* orthologs of the human sociability genes.

Similar to the findings regarding neuropsychiatric disorders in previous studies, we found that conserved genes showed increased interactivity, while non-conserved genes actually showed reduced interactivity compared to random sets. Although most of the interactions between human sociability genes were not present in their homologs, we did find several conserved interactions, potentially indicating ancient molecular pathways which may form part of the basis of human sociability.

Limitations

During our examination of social behavior phenotypes associated with the *C. elegans* orthologs of human sociability genes it came to light that very few *C. elegans* genes have known associations with any social behavior phenotype. Although several studies have examined some effect of genetic variation on social feeding behavior in *C. elegans* according to the WormBase database, very few looked at more than a limited number of genes. Therefore, it is as of yet unclear whether the limited number of genes known to be related to social behavior in the WormBase database is the result of a dearth of research or actually an indication that the *C. elegans* social behavior is regulated by a limited number of genes. More research into social behavior in *C. elegans* is required to examine the genetic basis of social behavior in *C. elegans* and to determine the presence and extent of a common genetic basis for social behavior among humans and *C. elegans*.

Also, while we did attempt to create fair comparisons by creating random sets of protein coding genes, this method still has limitations which may have affected the results. First, ten sets of

69 might be too low a number to get an accurate view of the level of conservation considering the size of the human protein coding genome. However, manual scoring of all human genes with *C. elegans* homologs based on our operationalization would have been untenable. This could potentially be addressed in the future by comparing genetic conservation data based on common operationalizations used as in a database of homologs of the total human protein coding genome.

Discussion

In this study, we found no or limited evidence of heightened conservation of human sociability genes in *C. elegans*. Also, we did not find evidence that the conservation of sociability genes in *C. elegans* may have been the result of lethal or sterile phenotypes as was found in the examinations of conservations of human genes for neuropsychiatric disorders. One potential explanation for the unexpected results might be the methodological variation between this study and the studies previously performed regarding the conservation of genes associated with schizophrenia (Kasap et al., 2018), bipolar disorder (Franklin & Dwyer, 2020) and major depressive disorder (Sall et al., 2021). While these studies searched for homologs using methods similar to those used in this study, their comparisons were similar to the one we carried out where the conservation in the sociability set was compared to that found in a previous study, as they used various earlier studies on gene conservation between humans and *C. elegans* to examine to what extent the conservation in their gene set differed from the total. However, as these studies used different methods to determine homologs, these comparisons may be biased and result in an increased likelihood to find significant results. The second method used in this study, utilizing random gene sets for the comparison may constitute a fairer comparison. A similar explanation could explain the difference between our study and the studies mentioned above in the difference in essential phenotypes occurrence between gene sets and the total genome. Considering we used the same database to gather data for both the sociability set and the total protein-coding genome, this comparison may be considered more equal than comparing the data from WormBase to data from other studies. It may be the case that the enrichment for essential phenotypes found in the previous studies regarding the neuropsychiatric disorders might be reduced or disappear when they would be compared to the WormBase data. On the other hand, it could be the case that this difference underlies true variation in the nature of genes orthologous to human sociability genes and those orthologous to human genes associated with neuropsychiatric disorders. It may be that genes associated with variation in the risk of neuropsychiatric disorders have larger effects on reproductive fitness compared to genes associated with reduced sociability, resulting in stronger constraints on the conservation of the former and therefore more likelihood that variation serves some essential function.

On the other hand, we did find evidence that human sociability genes conserved between humans and *C. elegans* were more interactive both within the set as well as when taking into

account interactions outside the set, compared to random human protein-coding genes and compared to non-conserved human sociability genes, indicating that the increased interactivity was not a result of the fact that the genes were related to a single phenotype but specifically to the conserved nature of the genes. This is to be expected; highly interactive genes are more likely to be involved in important biological processes and therefore also more likely to be conserved (Brown & Jurisica, 2007). However, our examination of the overlap in genetic interactions between humans and *C. elegans* indicates that although the individual genes are highly conserved, the interactions themselves are rarely conserved between *C. elegans* and humans. It appears that these interactive networks do not represent biological functions important for fitness in both humans as well as *C. elegans.*

One human gene which appears of particular interest based on our study is the activin receptor type 2-A (*ACVR2A*). This gene, which is conserved between humans and *C. elegans*, is the only gene which was found to have functions related to social behavior in both humans as well as in *C. elegans*. It was also found to be highly interactive (in fact, the most interactive gene among the conserved sociability genes), interacting with 16 out of 49 (33%) of other conserved sociability genes. Previous animal studies have demonstrated an important function of *ACVR2A* in regulating fertility and sexual behavior (Matzuk et al., 1995; Wreford et al., 2001; Ma et al., 2005). In *C. elegans*, the two orthologues of human *ACVR2A*, *daf-1* and *sma-6*, have previously been implicated in egg laying behavior (Larsen et al., 1995), brood size (Maduzzia et al., 2005) and sperm recruitment (McKnight et al., 2014), indicating that these orthologues affect reproductive fitness in *C. elegans* as well as in other model animals. In humans, *ACVR2A* is also known to be associated with human reproduction, for example through regulation of follicular development and oocyte maturation (Wang et al., 2022). These may indicate that the gene is conserved between humans and *C. elegans* as a result of very basic functions in reproduction, explaining its relevance to fitness. *ACVR2A* may constitute a common basis for social behavior across species and could therefore be an interesting target for studies of social behavior using animal models (although the relation between *ACVR2A* and reproductive health may hamper the use of the orthologues of this gene as a good model). It may also be interesting to examine how gene-gene interactions affect the function of the *ACVR2A* gene, both at a pathway level and at a behavioral level, to examine how intricate networks of genes can together affect complex phenotypes such as social behavior. Finally, future studies could examine biological differences between conserved and non-conserved human sociability genes in order to examine how different biological pathways affected social behavior evolved across time and how this relates to the evolution of other individual or environmental characteristics.

Acknowledgements

We thank Professor Donard Dwyer for his fast and helpful responses to our questions regarding previous studies carried out by his lab. We thank Professor Jean-Christophe Billeter for providing us with feedback to earlier versions of the manuscript.

Data availability statement

All data used for this article are available in the supplementary materials. Code used to generate the results is available upon request.

REFERENCES

Andreu-Bernabeu, Á., Díaz-Caneja, C. M., Costas, J., De Hoyos, L., Stella, C., Gurriarán, X., Alloza, C., Fañanás, L., Bobes, J., González-Pinto, A., Crespo-Facorro, B., Martorell, L., Vilella, E., Muntané, G., Nacher, J., Molto, M. D., Aguilar, E. J., Parellada, M., Arango, C., & González-Peñas, J. (2022). Polygenic contribution to the relationship of loneliness and social isolation with schizophrenia. *Nature communications*, *13*(1), 51. https://doi.org/10.1038/s41467-021-27598-6

Bralten, J., Mota, N. R., Klemann, C. J. H. M., De Witte, W., Laing, E., Collier, D. A., de Kluiver, H., Bauduin, S. E. E. C., Arango, C., Ayuso-Mateos, J. L., Fabbri, C., Kas, M. J., van der Wee, N., Penninx, B. W. J. H., Serretti, A., Franke, B., & Poelmans, G. (2021). Genetic underpinnings of sociability in the general population. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *46*(9), 1627–1634. https://doi.org/10.1038/s41386-021-01044-z

Brown, K. R., & Jurisica, I. (2007). Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome biology*, *8*(5), R95. https://doi.org/10.1186/gb-2007-8-5-r95

Chang, S. W., Brent, L. J., Adams, G. K., Klein, J. T., Pearson, J. M., Watson, K. K., & Platt, M. L. (2013). Neuroethology of primate social behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl 2*(Suppl 2), 10387–10394. https://doi.org/10.1073/pnas.1301213110

Crawley J. N. (2012). Translational animal models of autism and neurodevelopmental disorders. *Dialogues in clinical neuroscience*, *14*(3), 293–305. https://doi.org/10.31887/DCNS.2012.14.3/jcrawley

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., … Flicek, P. (2022). Ensembl 2022. *Nucleic acids research*, *50*(D1), D988–D995. https://doi.org/10.1093/nar/gkab1049

Davis, P., Zarowiecki, M., Arnaboldi, V., Becerra, A., Cain, S., Chan, J., Chen, W. J., Cho, J., da Veiga Beltrame, E., Diamantakis, S., Gao, S., Grigoriadis, D., Grove, C. A., Harris, T. W., Kishore, R., Le, T., Lee, R. Y. N., Luypaert, M., Müller, H. M., Nakamura, C., … Sternberg, P. W. (2022). WormBase in 2022-data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics*, *220*(4), iyac003. https://doi.org/10.1093/genetics/iyac003

Dunbar R. I. (2009). The social brain hypothesis and its implications for social evolution. *Annals of human biology*, *36*(5), 562–572. https://doi.org/10.1080/03014460902960289

Ellis W.A.H, Melzer A., Bercovitch F.B. (2009). Spatiotemporal dynamics of habitat use by koalas: The checkerboard model. *Behavioral Ecology and Sociobioly, 63*, 1181-1188. https://doi.org/10.1007/s00265-009-0761-2

Elbroch, L. M., Levy, M., Lubell, M., Quigley, H., & Caragiulo, A. (2017). Adaptive social strategies in a solitary carnivore. *Science advances*, *3*(10), e1701218. https://doi.org/10.1126/sciadv.1701218

Franklin, C., & Dwyer, D. S. (2021). Candidate risk genes for bipolar disorder are highly conserved during evolution and highly interconnected. *Bipolar disorders*, *23*(4), 400–408. https://doi.org/10.1111/bdi.12996

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., & Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature*, *408*(6810), 325–330. https://doi.org/10.1038/35042517

Grant, C., Addington, J., Addington, D., & Konnert, C. (2001). Social functioning in first- and multiepisode schizophrenia. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *46*(8), 746–749. https://doi.org/10.1177/070674370104600808

Gruber, T., & Clay, Z. (2016). A Comparison Between Bonobos and Chimpanzees: A Review and Update. *Evolutionary anthropology*, *25*(5), 239–252. https://doi.org/10.1002/evan.21501

Hedrick P. W. (2007). Balancing selection. *Current biology : CB*, *17*(7), R230–R231. https://doi.org/10.1016/j.cub.2007.01.012

vonHoldt, B. M., Shuldiner, E., Koch, I. J., Kartzinel, R. Y., Hogan, A., Brubaker, L., Wanser, S., Stahler, D., Wynne, C. D. L., Ostrander, E. A., Sinsheimer, J. S., & Udell, M. A. R. (2017). Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. *Science advances*, *3*(7), e1700398. https://doi.org/10.1126/sciadv.1700398

Kasap, M., Rajani, V., Rajani, J., & Dwyer, D. S. (2018). Surprising conservation of schizophrenia risk genes in lower organisms reflects their essential function and the evolution of genetic liability. *Schizophrenia research*, *202*, 120–128. https://doi.org/10.1016/j.schres.2018.07.017

Kim, W., Underwood, R. S., Greenwald, I., & Shaye, D. D. (2018). OrthoList 2: A New Comparative Genomic Analysis of Human and *Caenorhabditis elegans* Genes. *Genetics*, *210*(2), 445–461. https://doi.org/10.1534/genetics.118.301307

Kozel, B. A., Barak, B., Kim, C. A., Mervis, C. B., Osborne, L. R., Porter, M., & Pober, B. R. (2021). Williams syndrome. *Nature reviews. Disease primers*, *7*(1), 42. https://doi.org/10.1038/s41572-021-00276-z

Kupferberg, A., Bicks, L., & Hasler, G. (2016). Social functioning in major depressive disorder. *Neuroscience and biobehavioral reviews*, *69*, 313–332. https://doi.org/10.1016/j.neubiorev.2016.07.002

Kemphues K. (2005). Essential genes. *WormBook : the online review of C. elegans biology*, 1–7. https://doi.org/10.1895/wormbook.1.57.1

Lahera, G., Herrera, S., Reinares, M., Benito, A., Rullas, M., González-Cases, J., & Vieta, E. (2015). Hostile attributions in bipolar disorder and schizophrenia contribute to poor social functioning. *Acta psychiatrica Scandinavica*, *131*(6), 472–482. https://doi.org/10.1111/acps.12399

Lanooij, S. D., Eisel, U. L. M., Drinkenburg, W. H. I. M., van der Zee, E. A., & Kas, M. J. H. (2023). Influencing cognitive performance via social interactions: a novel therapeutic approach for brain disorders based on neuroanatomical mapping?. *Molecular psychiatry*, *28*(1), 28–33. https://doi.org/10.1038/s41380-022-01698-1

Larsen, P. L., Albert, P. S., & Riddle, D. L. (1995). Genes that regulate both development and longevity in Caenorhabditis elegans. *Genetics*, *139*(4), 1567–1583. https://doi.org/10.1093/genetics/139.4.1567

Ma, X., Reyna, A., Mani, S. K., Matzuk, M. M., & Kumar, T. R. (2005). Impaired male sexual behavior in activin receptor type II knockout mice. *Biology of reproduction*, *73*(6), 1182–1190. https://doi.org/10.1095/biolreprod.105.043794

Maduzia, L. L., Roberts, A. F., Wang, H., Lin, X., Chin, L. J., Zimmerman, C. M., Cohen, S., Feng, X. H., & Padgett, R. W. (2005). C. elegans serine-threonine kinase KIN-29 modulates TGFbeta signaling and regulates body size formation. *BMC developmental biology*, *5*, 8. https://doi.org/10.1186/1471-213X-5-8

Matzuk, M. M., Kumar, T. R., & Bradley, A. (1995). Different phenotypes for mice deficient in either activins or activin receptor type II. *Nature*, *374*(6520), 356–360. https://doi.org/10.1038/374356a0

McKnight, K., Hoang, H. D., Prasain, J. K., Brown, N., Vibbert, J., Hollister, K. A., Moore, R., Ragains, J. R., Reese, J., & Miller, M. A. (2014). Neurosensory perception of environmental cues modulates sperm motility critical for fertilization. *Science (New York, N.Y.)*, *344*(6185), 754–757. https://doi.org/10.1126/science.1250598

Pearson W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, *Chapter 3*, 3.1.1–3.1.8. https://doi.org/10.1002/0471250953.bi0301s42

Porcelli, S., Van Der Wee, N., van der Werff, S., Aghajani, M., Glennon, J. C., van Heukelum, S., Mogavero, F., Lobo, A., Olivera, F. J., Lobo, E., Posadas, M., Dukart, J., Kozak, R., Arce, E., Ikram, A., Vorstman, J., Bilderbeck, A., Saris, I., Kas, M. J., & Serretti, A. (2019). Social brain, social dysfunction and social withdrawal. *Neuroscience and biobehavioral reviews*, *97*, 10–33. https://doi.org/10.1016/j.neubiorev.2018.09.012

Reid, C. R., & Latty, T. (2016). Collective behaviour and swarm intelligence in slime moulds. *FEMS microbiology reviews*, *40*(6), 798–806. https://doi.org/10.1093/femsre/fuw033

Roberts, S. G., & Roberts, A. I. (2016). Social Brain Hypothesis: Vocal and Gesture Networks of Wild Chimpanzees. *Frontiers in psychology*, *7*, 1756. https://doi.org/10.3389/fpsyg.2016.01756

Robinson, G. E., Fahrbach, S. E., & Winston, M. L. (1997). Insect societies and the molecular biology of social behavior. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *19*(12), 1099–1108. https://doi.org/10.1002/bies.950191209

Sakurai, T., Dorr, N. P., Takahashi, N., McInnes, L. A., Elder, G. A., & Buxbaum, J. D. (2011). Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. *Autism research : official journal of the International Society for Autism Research*, *4*(1), 28–39. https://doi.org/10.1002/aur.169

Sall, S., Thompson, W., Santos, A., & Dwyer, D. S. (2021). Analysis of Major Depression Risk Genes Reveals Evolutionary Conservation, Shared Phenotypes, and Extensive Genetic Interactions. *Frontiers in psychiatry*, *12*, 698029. https://doi.org/10.3389/fpsyt.2021.698029

Saris, I. M. J., Aghajani, M., Jongs, N., Reus, L. M., van der Wee, N. J. A., Bilderbeck, A. C., Winter van Rossum, I., Arango, C., de la Torre-Luque, A., Malik, A., Raslescu, A., Dawson, G. R., Ayuso-Mateos, J. L., Kas, M. J., & Penninx, B. W. J. H. (2022). Cross-disorder and disorder-specific deficits in social functioning among schizophrenia and alzheimer's disease patients. *PloS one*, *17*(4), e0263769. https://doi.org/10.1371/journal.pone.0263769

Schindelman, G., Fernandes, J. S., Bastiani, C. A., Yook, K., & Sternberg, P. W. (2011). Worm Phenotype Ontology: integrating phenotype data within and beyond the C. elegans community. *BMC bioinformatics*, *12*, 32. https://doi.org/10.1186/1471-2105-12-32

Tan, J., & Hare, B. (2013). Bonobos share with strangers. *PloS one*, *8*(1), e51922. https://doi.org/10.1371/journal.pone.0051922

Tatay-Manteiga, A., Correa-Ghisays, P., Cauli, O., Kapczinski, F. P., Tabarés-Seisdedos, R., & Balanzá-Martínez, V. (2018). Staging, Neurocognition and Social Functioning in Bipolar Disorder. *Frontiers in psychiatry*, *9*, 709. https://doi.org/10.3389/fpsyt.2018.00709

Thomas J. H. (1998). Social life and the single nucleotide: foraging behavior in C. elegans. *Cell*, *94*(5), 549–550. https://doi.org/10.1016/s0092-8674(00)81595-0

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., & Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome research*, *19*(2), 327–335. https://doi.org/10.1101/gr.073585.107

Wang, S., Fang, L., Cong, L., Chung, J. P. W., Li, T. C., & Chan, D. Y. L. (2022). Myostatin: a multifunctional role in human female reproduction and fertility - a short review. *Reproductive biology and endocrinology : RB&E*, *20*(1), 96. https://doi.org/10.1186/s12958-022-00969-4

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., & Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, *38*(Web Server issue), W214–W220. https://doi.org/10.1093/nar/gkq537

Wreford, N. G., Rajendra Kumar, T., Matzuk, M. M., & de Kretser, D. M. (2001). Analysis of the testicular phenotype of the follicle-stimulating hormone beta-subunit knockout and the activin type II receptor knockout mice by stereological analysis. *Endocrinology*, *142*(7), 2916–2920. https://doi.org/10.1210/endo.142.7.8230

SUPPLEMENTARY MATERIALS

Supplementary Table 1: Human protein coding sociability genes and *C. elegans orthologs*

Gene	Ensembl ortholog	Wormbase ortholog (non-BLAST)	Wormbase BLASTP counterpart	Ortholog used for subsequent analyses	Ortholog present
ANKRD7	X	X	X	X	0
BTBD18	X	X	X	X	0
CNTF	X	X	X	X	0
DTX4	X	X	X	X	0
EDN1	X	X	X	X	0
FAM111A	X	X	X	X	0
FAM222A	X	X	X	X	0
MBD5	X	X	X	X	0
NINJ1	X	X	X	X	0
ORC4	X	X	X	X	0
P2RX3	X	X	X	X	0
PLET1	X	X	X	X	0
PRDM13	X	X	X	X	0
PTH	X	X	X	X	0
SELENOH	X	X	X	X	0
SLC43A1	X	X	X	X	0
SMIM13	X	X	X	X	0
TNKS1BP1	X	X	X	X	0
ZFP91	X	X	X	X	0
ARNTL	*aha-1*	X	Not performed	*aha-1*	1
TTC12	*unc-45* & *sti-1*	X	Not performed	*unc-45* & *sti-1*	1
ACVR2A	*daf-1* & *sma-6*	*daf-4*	Not performed	*daf-1* & *sma-6*	1
ADTRP	*C37E2.10*	*C37E2.2*	Not performed	*C37E2.10*	1
APLNR	X	*npr-15*, *npr-31* & *npr-33*	Not performed	*npr-15*, *npr-31* & *npr-33*	1
ASCC3	*Y54E2A.4*	*Y54E2A.4*	Not performed	*Y54E2A.4*	1
BCO2	X	*bcmo-1* & *bcmo-2*	Not performed	*bcmo-1* & *bcmo-2*	1
CFTR	*cft-1*	*cft-1*	Not performed	*cft-1*	1
CLP1	*clpf-1*	*clpf-1*	Not performed	*clpf-1*	1
CTNND1	*jac-1*	*jac-1*	Not performed	*jac-1*	1
CTTNBP2	X	*C49H3.6*	Not performed	*C49H3.6*	1
DRD2	*ser-5*	*dop-3*	Not performed	*ser-5*	1
ELAVL2	X	*exc-7*	Not performed	*exc-7*	1
FAR1	*fard-1*	*fard-1*	Not performed	*fard-1*	1
FBXL17	*fbxl-1*	X	Not performed	*fbxl-1*	1
GBE1	*T04A8.7*	*T04A8.7*	Not performed	*T04A8.7*	1

Supplementary Table 1: Human protein coding sociability genes and *C. elegans orthologs* (*continued*)

Gene	Ensembl ortholog	Wormbase ortholog (non-BLAST)	Wormbase BLASTP counterpart	Ortholog used for subsequent analyses	Ortholog present
GLYATL1	*T10B10.4*	*T10B10.4*	Not performed	*T10B10.4*	1
GLYATL2	*T10B10.4*	*T10B10.4*	Not performed	*T10B10.4*	1
GRIK2	*glr-1*	*glr3* & *glr-5*	Not performed	*glr-1*	1
HIATL1 (aka MFSD14B)	*T25D3.4*	*T25D3.4*	Not performed	*T25D3.4*	1
HIVEP1	X	*sma-9*	Not performed	*sma-9*	1
ISL1	*lim-7*	*lim-7*	Not performed	*lim-7*	1
KCTD10	*D2045.8*	X	Not performed	*D2045.8*	1
LPXN	*pxl-1*	*pxl-1*	Not performed	*pxl-1*	1
MAN2A1	*aman-2*	*aman-2*	Not performed	*aman-2*	1
MED19	*mdt-19*	*mdt-19*	Not performed	*mdt-19*	1
MMAB	*mmab-1*	*mmab-1*	Not performed	*mmab-1*	1
MVK	*mvk-1*	*mvk-1*	Not performed	*mvk-1*	1
MYO1H	X	X	*hum-5*	*hum-5*	1
NAA38	*natc-3*	*natc-3*	Not performed	*natc-3*	1
NRXN1	X	*nrx-1*	Not performed	*nrx-1*	1
OR5AK2	X	X	*ckr-2*	*ckr-2*	1
OR5B17	X	*srsx-27* through *srsx-39*	Not performed	*[srsx-27* through *srsx-39]*	1
PARVA	*pat-6*	*pat-6*	Not performed	*pat-6*	1
PATL1	*patr-1*	*patr-1*	Not performed	*patr-1*	1
PHACTR1	*F26H9.2*	*F26H9.2*	Not performed	*F26H9.2*	1
PHF2	*jmjd-1.1*, *jmjd-1.2* & *jhdm-1*	*jmjd-1.2*	Not performed	*jmjd-1.1*, *jmjd-1.2* & *jhdm-1*	1
SERPING1	X	*srp-2*, *srp-6* & *srp-7*	Not performed	*srp-2*, *srp-6* & *srp-7*	1
SIM1	*hif-1*	*hlh-34*	Not performed	*hif-1*	1
SIPA1L1	*F53A10.2*	*sipa-1*	Not performed	*F53A10.2*	1
SMTNL1	X	*T15B12.1*	Not performed	*T15B12.1*	1
SSRP1	*athp-1*	*hmg-3* & *hmg-4*	Not performed	*athp-1*	1
TIMM10	*tin-10*	*tin-10*	Not performed	*tin-10*	1
TMPRSS5	X	*try-1*	Not performed	*try-1*	1
TMX2	*C35D10.10*	*C35D10.10*	Not performed	*C35D10.10*	1
UBE3B	*oxi-1*	*oxi-1*	Not performed	*oxi-1*	1
YPEL4	*M04B2.4*	*F37A8.5*	Not performed	*M04B2.4*	1
ZBTB16	X	*zfp-2*	Not performed	*zfp-2*	1
ZDHHC5	X	*dhhc-8*	Not performed	*dhhc-8*	1
ZNF484	X	*znf-782*	Not performed	*znf-782*	1

Supplementary Table 2: Random sets for comparison of conservation and interactivity

Interaction set comparison	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10
Conserved	*ZNF521*	*SOX5*	*PCBP2*	*UROD*	*OS9*	*SRR*	*PCID2*	*NDUFA13*	*ASB13*	*TMPRSS3*
	ERVK3-1	*SETDB2*	*CSGALNACT1*	*CHRNA9*	*QPCTL*	*C17orf96*	*PIGU*	*NARFL*	*CCDC88B*	*GSAP*
	CRY2	*OR13C4*	*CSRNP2*	*TRIM50*	*PRF1*	*WDR33*	*GSX2*	*IGFL4*	*HLA-DRA*	*TM9SF1*
	VPS37D	*TMEM100*	*MTMR12*	*MC2R*	*CSNK1G1*	*CELSR3*	*IYD*	*ZMYM2*	*ING1*	*RDH13*
	SI	*PRR20C*	*CHD4*	*VAMP5*	*OR13D1*	*TUBB1*	*CMTM2*	*NAA25*	*TINCR*	*NAP1L5*
	MINOS1	*EIF4B*	*HOXB4*	*SOBP*	*SRSF3*	*FAM162A*	*DEFB107B*	*U2SURP*	*C8B*	*UGT1A4*
	IL10RB	*BCKDHA*	*DCAF4*	*CTRB2*	*XRN1*	*ZNF48*	*GNAT1*	*FOXO6*	*EXTL1*	*AP5B1*
	SLC35C2	*LPL*	*MTHFD2*	*GABRA1*	*SLFN14*	*ELP5*	*RAB12*	*C1orf162*	*TACC1*	*WIPF2*
	SLC39A7	*THAP8*	*MICAL1*	*CECR6*	*CA1*	*CARM1*	*FIBP*	*LDHD*	*VTN*	*ZNF610*
	NDUFAF2	*DSG4*	*CDC42EP3*	*WDR74*	*TSGA13*	*TBKBP1*	*TBL1XR1*	*GLE1*	*ABCF3*	*STARD9*
	CBS	*CRB1*	*CBLN4*	*RGS11*	*CETP*	*AEBP2*	*RPP25*	*ZFAND1*	*MT3*	*GALNTL5*
	MRPS18B	*DNAJC4*	*HACE1*	*KDELC2*	*ILVBL*	*REL*	*GCSH*	*KCNN2*	*TAC4*	*CANT1*
	C1orf100	*PATJ*	*B3GALT6*	*ABCC10*	*GSC2*	*RNASE10*	*NPIPA1*	*RANBP10*	*C17orf64*	*CPNE6*
	HIST1H2BN	*C12orf56*	*MBOAT4*	*TSSK4*	*LRRN4*	*MAPK12*	*WDR34*	*KLK12*	*CDHR1*	*KRTAP5-3*
	PSMB8	*OR4C12*	*OA3*	*NDUFS1*	*S100A7*	*GKN2*	*OCLN*	*GPR137B*	*SRD5A3*	*SFTPC*
	RIMS1	*RNF213*	*AK7*	*GDF11*	*ZBTB25*	*RPS6KB1*	*DLEU7*	*KLHL42*	*LRRN4CL*	*UBE2I*
	TNKS1BP1	*WFDC11*	*TEX36*	*MED18*	*HIST1H1D*	*DNASE2*	*PIGN*	*SLC15A1*	*ITIH1*	*PFDN4*
	LSM6	*AMPD3*	*TPCN2*	*SYNJ2BP*	*SSR2*	*XPO4*	*MAPKAPK5*	*EDEM3*	*VEGFC*	*CAMKMT*
	CGGBP1	*ZNF598*	*YWHAG*	*PXMP4*	*SOX15*	*ZNF775*	*PIP5K1A*	*KDM5A*	*SLC5A9*	*MED20*
	CAPN8	*DYRK2*	*ADGRG7*	*PRKG2*	*PTBP1*	*USP17L23*	*CCDC141*	*WNT2B*	*NCF2*	*DNAJC15*
	VPS4B	*B3GNT5*	*LDAH*	*LCN10*	*NFRKB*	*GTF2IRD2B*	*TMEM8C*	*NNT*	*PPP1R14D*	*SAMHD1*
	RARRES1	*FAM71C*	*DCDC2B*	*OTOF*	*VMAC*	*C19orf52*	*BRDT*	*PATL2*	*VARS*	*TTC26*
	DNAH7	*ST18*	*CDK5RAP3*	*PRKAR1B*	*MAEL*	*TTC30B*	*GADD45GIP1*	*RASSF9*	*LOH19CR1*	*USP20*

Supplementary Table 2: Random sets for comparison of conservation and interactivity (*continued*)

Interaction set comparison	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10
	LRRC58	*SHANK1*	*P2RX7*	*CBLB*	*ABI1*	*USP7*	*MIER3*	*AQP7*	*LRRC4*	*EBF2*
	C7orf76	*OR2H1*	*PTGDR*	*TBC1D10A*	*SGCD*	*MAP4K4*	*TNFSF10*	*RORC*	*TRPC3*	*GET4*
	PCDH7	*RAB4A*	*CDC42SE1*	*PLA2G2C*	*TEX19*	*TNF*	*GNG7*	*ACSM6*	*NCBP2-AS2*	*LACTB*
	HN1	*ZNF830*	*HIST1H2AM*	*CASP8AP2*	*KIR2DL4*	*RAD51AP2*	*MUC4*	*ZSCAN21*	*NAT14*	*SETSIP*
	MAPK4	*TCP11L1*	*MRPL51*	*DCSTAMP*	*TMEM190*	*USP19*	*SLC37A4*	*ECI2*	*CHST15*	*PLXNC1*
	TESC	*SRSF2*	*SPEN*	*IFNA17*	*BUD31*	*AP5B1*	*OST4*	*TMEM132E*	*PIGBOS1*	*ITGB3BP*
	IFNGR2	*EIF1B*	*HIC1*	*MMP27*	*CD300LD*	*LYN*	*SORBS3*	*RBM12*	*GRIN3B*	*CNTLN*
	MGAT4B	*C20orf203*	*ZNF708*	*ACVR2B*	*ATP9B*	*LRP6*	*CBWD1*	*KCTD15*	*ARHGEF7*	*AK1*
	TLDC2	*KRTAP4-8*	*YPEL3*	*OR51I1*	*HAUS4*	*HIRIP3*	*UBA3*	*LCT*	*STIM1*	*ENTPD5*
	RIMBP2	*OR8D1*	*EFCAB5*	*PCDHA2*	*ADAMTSL5*	*PTBP1*	*CD38*	*ZNF354C*	*PRRT4*	*VNN2*
	TRIM51	*TLE4*	*SSMEM1*	*ANKRD37*	*TKFC*	*PLRG1*	*ISG20*	*LAMP3*	*MEI4*	*AHR*
	KANSL1L	*PTPN5*	*SPRTN*	*RPL34*	*LYNX1*	*TIFAB*	*PLIN4*	*NUMB*	*FNBP1*	*TYRO3*
	ZNF317	*MSH6*	*ISLR*	*OSBPL10*	*SLC25A10*	*OR51A7*	*REV1*	*KPNA4*	*RFPL1*	*MIER3*
	PPARG	*GMFG*	*PRLH*	*TC2N*	*SCARB2*	*HIST2H2AA4*	*CRYGD*	*TMEM11*	*SNAP91*	*CORO7-PAM16*
	HSPH1	*GTF2A1L*	*ANKRD63*	*TEX29*	*SLCO2B1*	*PJA2*	*HES4*	*SLIT1*	*EBI3*	*RAB40B*
	CFAP157	*SMOX*	*HSD17B7*	*SLAIN1*	*PLET1*	*SOX1*	*PANK1*	*C1QL3*	*SBDS*	*POU4F1*
	DOCK2	*AHSP*	*ZNF837*	*GZMB*	*CNTNAP3B*	*P3H2*	*BTN3A3*	*SOGA3*	*ATP2B4*	*CNIH1*
	DUSP10	*PNPLA3*	*NPRL2*	*FITM2*	*SIGLEC8*	*EXT2*	*BTG1*	*MAST2*	*NR0B2*	*AEN*
	EXOSC7	*WDR27*	*FFAR4*	*SLC35A5*	*LGALS14*	*RNF4*	*TCTN3*	*THOC7*	*JAK3*	*SPIC*
	FADD	*NUTM1*	*HPS6*	*POLR3D*	*RNF41*	*OR9I1*	*LCLAT1*	*TIGD3*	*ZNF593*	*GRB2*
	LYSMD2	*NECTIN1*	*TCEB3*	*FAM45A*	*NCOR1*	*SNAP23*	*POLR3C*	*OR4C3*	*ADAMTS15*	*ALDH9A1*
	PDXP	*WDR12*	*PLSCR4*	*ZNF836*	*RIPPLY3*	*ALKBH8*	*RGS17*	*MEI4*	*GBAS*	*IGLL5*
	NR1I2	*GFAP*	*WDFY4*	*BATF3*	*EQTN*	*SYNRG*	*RP1*	*FNBP1*	*KLK11*	*AKR1D1*
	DCTN1	*ROBO1*	*PLXNA1*	*EVA1A*	*FUT5*	*BBS9*	*OR9G4*	*RFPL1*	*ZNF695*	*BHLHE22*

Supplementary Table 2: Random sets for comparison of conservation and interactivity (*continued*)

Interaction set comparison	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10
	PPP1R42	*ITGAM*	*PTPRN*	*EEF1A2*	*LGALS7B*	*SLC1A5*	*KMT5B*	*SNAP91*	*LRRFIP2*	*FOXD4L1*
	DPH5	*APOL4*	*SLITRK6*	*CAB39*	*STMN4*	*ZNF699*	*BET1L*	*EBI3*	*WNT10A*	*SLC2A8*
	NUAK1	*SMIM11A*	*CTSK*	*COLGALT2*	*ANXA4*	*CCDC126*	*TMX4*	*SBDS*	*SH2D5*	*MESDC2*
Nonconserved	*GTF2IRD2B*	*RHBDD2*	*CELF3*	*NUDT4*	*NOXA1*	*CD69*	*NCMAP*	*ADGRA3*	*C12orf49*	*IL17RE*
	VCAM1	*SVIL*	*LRCH3*	*STRIP1*	*PDLIM4*	*OR13C5*	*KCNJ1*	*DISP1*	*FKBP6*	*DNAJC12*
	HOXD3	*KCNK4*	*MS4A14*	*C19orf81*	*MBD2*	*TMEM230*	*PRUNE*	*MEIKIN*	*DAPL1*	*SWT1*
	C6orf48	*CFAP44*	*CDK4*	*CLDN9*	*CD1D*	*MTR*	*ZMPSTE24*	*GATA3*	*CST11*	*SAAL1*
	SAFB	*EMP3*	*CNTNAP1*	*INO80E*	*TMEM14B*	*ENY2*	*ABCB9*	*RHD*	*TMPRSS4-AS1*	*CILP2*
	DPP8	*MYO19*	*COX11*	*SPATA31A5*	*ZFAND2A*	*ZBTB20*	*HTR2A*	*CD86*	*SBDS*	*MYH15*
	BRAT1	*SERPINC1*	*DNM1*	*ARPC2*	*TMEM86A*	*TEC*	*UBTD2*	*ZNF3*	*RALA*	*BCL11A*
	NAGLU	*ST8SIA3*	*BHMG1*	*PDK4*	*FAM3D*	*MTF1*	*SLC2A6*	*CCDC59*	*SOAT1*	*TPD52L2*
	OSBPL6	*CERS4*	*ESPL1*	*EEF1G*	*MRPL30*	*RCBTB1*	*FAM172A*	*POLR1A*	*FOS*	*UGCG*
	IQCC	*SERPINA1*	*RTN3*	*PPP6C*	*GPR161*	*EFCAB6*	*DSC1*	*WBSCR22*	*COG6*	*XPO1*
	PFDN5	*C10orf131*	*CDK2*	*KRTAP10-12*	*CCDC169-SOHLH2*	*TJP1*	*PRR29*	*ATP7B*	*TSPAN2*	*PTPDC1*
	C2orf68	*ANKEF1*	*CD1A*	*FAU*	*PTCRA*	*GAA*	*TMEM52*	*HMG20B*	*STOML3*	*AGAP1*
	SYT14	*HMGN3*	*CLEC18C*	*ZNF184*	*ADAR*	*PRTFDC1*	*KDELR2*	*LAG3*	*ADAM32*	*C11orf58*
	CIART	*REG3G*	*C4orf3*	*FKBP11*	*RASGRF1*	*SCUBE2*	*CFAP54*	*DEFB135*	*SCAMP3*	*KANSL1*
	ORM1	*LINC00521*	*ADAMTS6*	*HMOX2*	*TIMMDC1*	*WWOX*	*EMX2*	*SLC16A8*	*ULK3*	*LRRC19*
	FAM50B	*BCAN*	*RMND1*	*SVOP*	*RND2*	*CERKL*	*FMN1*	*PNPLA2*	*C1RL*	*GTF3C4*
	MGAT4D	*GTF2B*	*IDH3A*	*PSRC1*	*HOXC4*	*PSMA1*	*CREB3L3*	*CC2D2A*	*AP2B1*	*HIBADH*
	CAMLG	*SOWAHA*	*ANKMY1*	*CYP26B1*	*RYR3*	*THOP1*	*KLF13*	*KAT2B*	*TAS2R41*	*IGF2*
	PDZK1IP1	*PCDHGB3*	*IWS1*	*C16orf45*	*DDX10*	*C21orf62*	*ZSWIM8*	*DOCK6*	*UPF1*	*AGAP6*

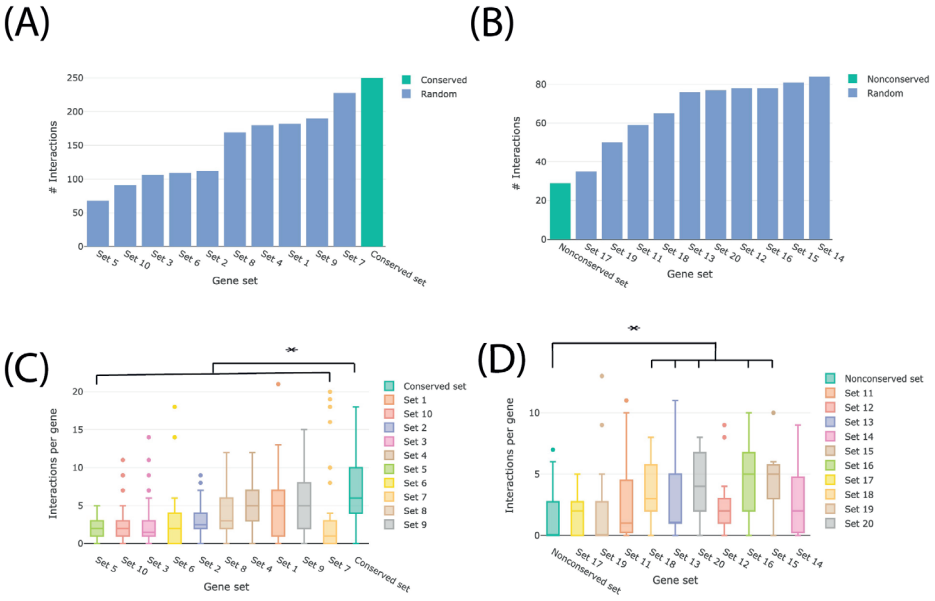
Supplementary Table 3: Random *C. elegans* gene sets for interactivity analyses

Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	Ortholog set
H39E20.1	*lron-3*	*ctf-4*	*srbc-30*	*ugt-20*	*F53E4.2*	*T23F11.2*	*C34C6.3*	*Y48A5A.1*	*adss-1*	*aha-1*
W05E10.2	*phy-4*	*F07G6.10*	*kin-34*	*Y46G5A.14*	*srw-77*	*mam-3*	*F20D6.2*	*Y71F9AL.12*	*C09B9.1*	*unc-45* & *sti-1*
haf-9	*C50B8.4*	*dmsr-8*	*srh-97*	*ksr-1*	*uaf-2*	*ucr-2.2*	*Y39A1A.27*	*srxa-19*	*Y71H2B.5*	*daf-1* & *sma-6*
epg-4	*srw-89*	*zig-12*	*Y39A3A.2*	*F07G6.10*	*Y59A8B.24*	*ZK973.9*	*W02B3.4*	*gcc-1*	*slc-17.1*	*C37E2.10*
W08E12.2	*F43D2.3*	*Y82E9BL.19*	*lgc-31*	*C17H12.5*	*pho-5*	*F20D6.10*	*K06B9.6*	*K11D12.12*	*C03B1.9*	*npr-15*, *npr-31* & *npr-33*
Y67D8B.2	*asp-19*	*F13H8.3*	*K07D4.5*	*pudl-1*	*K07H8.2*	*K10H10.6*	*clec-57*	*nhr-126*	*R07B7.8*	*Y54E2A.4*
Y105C5A.1	*lido-4*	*K05C4.7*	*F11E6.10*	*gex-2*	*prp-21*	*C08E8.3*	*B0303.11*	*tsp-19*	*Y54F10BM.6*	*bcmo-1* & *bcmo-2*
srx-31	*C45E1.5*	*rpoa-1*	*F48G7.7*	*fbxb-60*	*C40H1.9*	*T27A1.3*	*pkg-2*	*zyg-1*	*srw-135*	*cft-1*
C31C9.2	*T03F6.6*	*F47B10.9*	*C49A9.10*	*C49C3.11*	*E02H9.9*	*ceeh-2*	*Y57E12B.4*	*madd-3*	*spp-14*	*clpf-1*
hcp-3	*K07C11.7*	*R07C12.3*	*srh-291*	*Y45G12C.3*	*R05G9.5*	*C14C6.5*	*fbxb-41*	*nhr-209*	*F56H9.9*	*jac-1*
srw-41	*rpn-13*	*srbc-51*	*irld-29*	*F40H3.2*	*set-30*	*nhr-79*	*F56A8.9*	*F42A10.6*	*grd-2*	*C49H3.6*
C49A1.10	*F19B10.5*	*srh-52*	*eri-7*	*srx-117*	*acl-1*	*srz-62*	*aly-1*	*Y37F4.8*	*cyp-35A4*	*ser-5*
K09E9.3	*hex-4*	*F44E7.4*	*D1007.4*	*Y73B6BL.289*	*mlt-9*	*K12D12.4*	*F53H4.3*	*M03C11.9*	*bris-1*	*exc-7*
F18A11.2	*K09C8.8*	*T06A4.3*	*C24A3.2*	*C53A5.17*	*Y48G8AR.8*	*pbrm-1*	*gst-43*	*sqrd-1*	*K02A6.2*	*fard-1*
clec-102	*R151.8*	*F46H5.4*	*D1005.2*	*T26C5.5*	*tag-164*	*snx-1*	*oatr-1*	*poml-1*	*C25G4.7*	*fbxl-1*
F15D3.9	*lact-8*	*oac-26*	*F56C3.8*	*C55B6.4*	*F46A9.1*	*F27B10.1*	*pam-1*	*srh-239*	*pcca-1*	*T04A8.7*
B0024.15	*sre-21*	*ZC84.1*	*gpb-1*	*vha-1*	*nhr-108*	*ugt-53*	*ZK899.6*	*plpp-1.3*	*C03C10.7*	*T10B10.4*
homt-1	*Y97E10AR.4*	*clh-6*	*mkk-4*	*W02D9.2*	*egl-1*	*C01B10.11*	*M02D8.2*	*msp-38*	*sra-9*	*glr-1*
pap-1	*eor-2*	*T02G6.11*	*sma-4*	*ucr-2.3*	*nfyb-1*	*pho-1*	*C06E4.2*	*Y47C4A.1*	*F08F1.4*	*T25D3.4*
Y54F10AM.5	*fbxc-29*	*sre-48*	*tin-44*	*flp-28*	*cyp-33B1*	*vha-20*	*ceh-19*	*K07A12.8*	*R03D7.2*	*sma-9*
Y53H1A.7	*T03E6.9*	*C47F8.3*	*pgp-4*	*algn-2*	*exos-9*	*flp-13*	*ZC487.2*	*vha-2*	*tmem-131*	*lim-7*
C01B9.4	*anmt-2*	*fecl-1*	*rab-1*	*C49A9.5*	*best-12*	*Y57A10C.1*	*stdh-4*	*Y41C4A.21*	*C33D3.3*	*D2045.8*
nas-25	*uri-1*	*C01G10.17*	*F33D11.6*	*unc-2*	*calm-1*	*sre-33*	*Y38H8A.2*	*srw-33*	*cash-1*	*pxl-1*
Y69E1A.3	*czw-1*	*C16C8.7*	*Y92H12A.5*	*nhr-71*	*Y73B6BL.289*	*W01B6.5*	*mrpl-54*	*dod-22*	*B0207.5*	*aman-2*

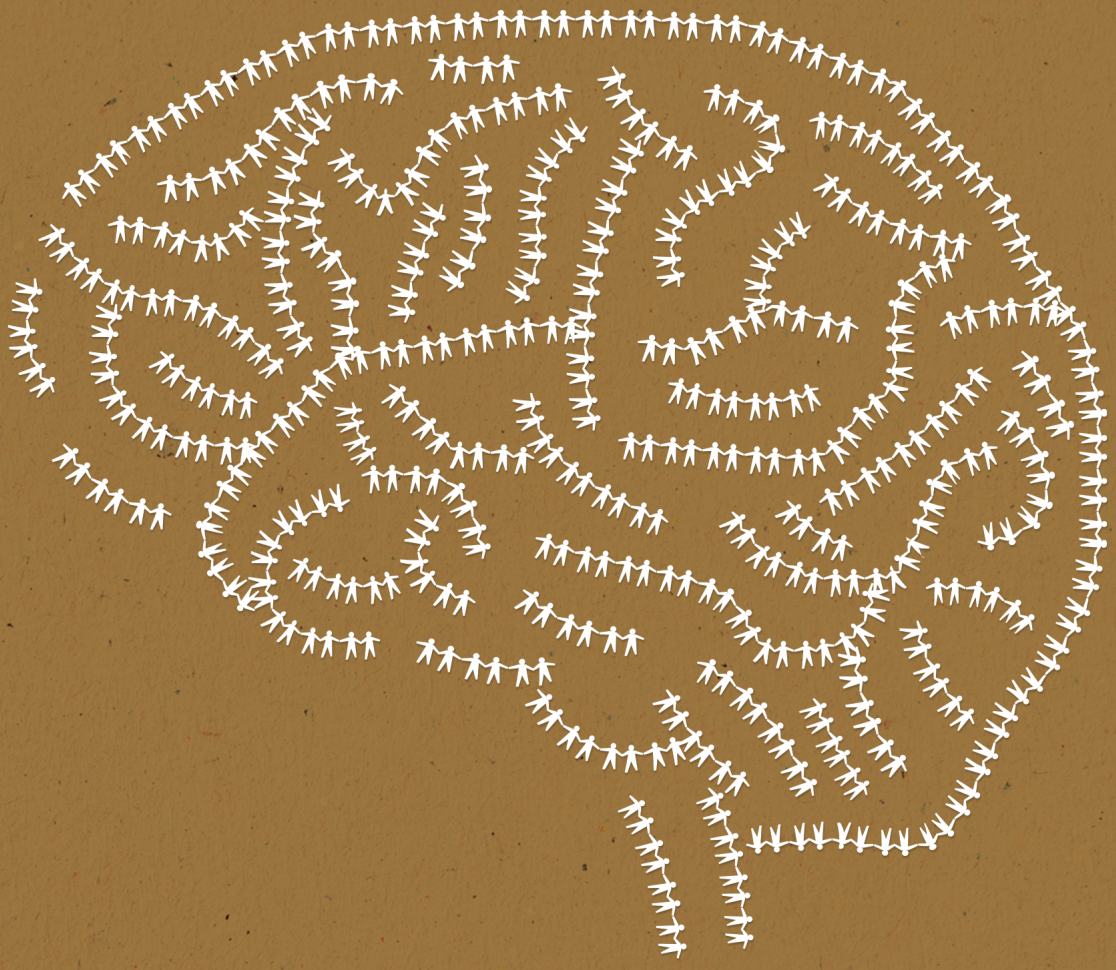
Supplementary Table 3: Random *C. elegans* gene sets for interactivity analyses (*continued*)

Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	Ortholog set
C15C8.6	*C11D2.7*	*D2030.11*	*F53C3.7*	*F44A2.3*	*unc-41*	*F44E5.1*	*T28F4.1*	*nlp-61*	*ccch-1*	*mdt-19*
F01D5.6	*cdc-25.4*	*srw-144*	*adm-2*	*R02F2.6*	*T19D2.3*	*ifa-4*	*nhr-195*	*nhr-143*	*dhhc-7*	*mmab-1*
col-109	*ZC443.7*	*T08D2.4*	*F31B9.4*	*T19A5.3*	*algn-10*	*C40C9.4*	*fbxa-118*	*lmtr-5*	*T21C9.9*	*mvk-1*
erfa-3	*C26D10.7*	*Y73B3A.20*	*srh-68*	*ttr-2*	*del-5*	*T10C6.7*	*sru-48*	*sas-6*	*Y116A8C.24*	*hum-5*
F22H10.11	*H20J04.7*	*F40A3.4*	*syp-3*	*srt-37*	*aho-3*	*best-18*	*mom-5*	*zmp-4*	*unc-9*	*natc-3*
C27A2.12	*srm-3*	*Y61B8A.6*	*phat-6*	*ZK430.7*	*syp-2*	*mct-5*	*K10G6.9*	*C33G8.2*	*F56B3.9*	*nrx-1*
mboa-7	*rbm-34*	*fbxb-25*	*T03G6.3*	*mcm-7*	*F56B3.11*	*srh-20*	*rsu-1*	*nep-22*	*K04C1.3*	*ckr-2*
F16B3.3	*T06G6.6*	*sma-1*	*dct-5*	*R08A2.2*	*F14F7.5*	*srm-1*	*chpf-2*	*R10E4.7*	*kcc-1*	*srsx-27* through *srsx-39*
nyn-1	*sdhd-1*	*C50F7.3*	*F38E9.1*	*C41G6.13*	*clec-165*	*C50F4.16*	*Y92H12A.2*	*gtsf-1*	*nep-16*	*pat-6*
syf-2	*R11A8.1*	*rcan-1*	*Y57E12AR.1*	*ZK1010.4*	*rdl-1*	*Y95D11A.3*	*srh-235*	*K10D11.2*	*Y47G6A.19*	*patr-1*
cdc-48.2	*Y40B10A.9*	*F25H5.10*	*tmc-2*	*scl-24*	*Y49E10.29*	*R52.6*	*T26C11.9*	*best-7*	*srd-45*	*F26H9.2*
pitp-1	*ttx-3*	*ZK666.15*	*F23D12.1*	*F27D4.1*	*W08F4.3*	*kars-1*	*ZC412.5*	*atfs-1*	*ZK930.6*	*jmjd-1.1*, *jmjd-1.2* & *jhdm-1*
sqst-3	*fbxa-211*	*str-52*	*thn-4*	*Y39B6A.7*	*ZK380.4*	*spe-29*	*T27A10.5*	*F29D10.1*	*rps-18*	*srp-2*, *srp-6* & *srp-7*
cps-6	*C55C3.1*	*col-86*	*impt-1*	*nhr-77*	*atp-6*	*F46B3.9*	*pmp-1*	*taf-10*	*srbc-36*	*hif-1*
zim-3	*C27B7.2*	*T28B8.4*	*Y119C1B.10*	*T09D3.3*	*T16G1.7*	*emb-1*	*C49F5.7*	*unc-25*	*arl-1*	*F53A10.2*
C47A4.5	*qars-1*	*C55A6.6*	*F13B6.3*	*unc-53*	*nhr-233*	*F14D12.1*	*C44C1.1*	*Y53F4A.2*	*T05F1.8*	*T15B12.1*
C02F5.13	*C50F4.16*	*ZK909.3*	*R08C7.4*	*B0563.18*	*his-64*	*tank-1*	*fbxb-17*	*cbs-2*	*tag-340*	*athp-1*
fezf-1	*Y54G2A.12*	*srx-46*	*tbx-8*	*ZK1290.11*	*ins-19*	*srh-228*	*F39F10.5*	*grl-3*	*F37D6.4*	*tin-10*
dmd-6	*T18D3.1*	*E04F6.15*	*D2096.13*	*K03H6.1*	*Y57A10C.1*	*F25G6.1*	*M110.8*	*F49F1.10*	*pgp-1*	*try-1*
F42A9.6	*clec-180*	*Y110A2AL.5*	*tbx-30*	*ida-1*	*C55A1.6*	*T27E7.1*	*emb-1*	*rab-21*	*nhr-104*	*C35D10.10*
nhr-285	*srsx-13*	*srj-1*	*chil-17*	*EEED8.3*	*K02A6.4*	*F37C12.3*	*nyn-1*	*F07C4.11*	*F34D10.3*	*oxi-1*
bed-3	*anmt-3*	*F43C9.2*	*Y105E8B.7*	*M03C11.1*	*T15B7.1*	*srx-31*	*R07E4.3*	*F07F6.1*	*mig-10*	*M04B2.4*
mltn-5	*Y119D3B.21*	*egas-3*	*F53F1.4*	*C36B7.3*	*C36B1.11*	*F52C6.3*	*R02D5.10*	*ZK512.8*	*nduf-2.2*	*zfp-2*
tpxl-1	*Y34D9A.7*	*mrpl-42*	*W10G6.1*	*srm-4*	*nono-1*	*nlp-71*	*mbk-1*	*F36D3.16*	*F21A3.4*	*dhhc-8*
fbxa-75	*C43H6.6*	*Y38F2AL.12*	*bed-2*	*egl-5*	*set-6*	*flp-7*	*Y51F10.15*	*best-21*	*C43H6.4*	*znf-782*

Supplementary Figure 1



Supplementry Figure 1. Interactivity between conserved sociability genes, nonconserved sociability genes and random gene sets allowing 20 resultant genes to be included in GeneMania. A) Total interactions between conserved sociability genes and random sets of 50 human genes; B) Total interactions between nonconserved sociability genes and random sets of 19 human genes; C) Interactions per gene for conserved sociability genes and random sets of 50 human genes; D) Interactions per gene for nonconserved sociability genes and random sets of 19 human genes.



Chapter 3

Creation and evaluation of measures of individualized social complexity in the UK Biobank: Associations with grey matter volume and cognition

Mila C. Roozen[1] & Martien J. H. Kas[1]

[1]*Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands*

In preparation for submission

ABSTRACT

The social brain hypothesis (SBH) is a popular hypothesis explaining the evolution of large brains in primates through selection on social complexity. Novel methods using genetic information from a single species to examine evidence of evolution have created interesting opportunities to evaluate hypotheses such as the SBH. However, these methods rely on information regarding phenotypic variation within species, while such data regarding social complexity is lacking. We used data from the UK Biobank to create measures of social complexity and examined how these relate to grey matter volume and cognition. Associations were found between social complexity and total grey matter volume as well as regional grey matter volumes in the cerebellum, pallidum and inferior lateral occipital cortex. Social cognition was broadly associated with cognition, although the number of friends and family visits showed inverse associations compared to the other measures. These findings add to the knowledge of the neurobiological and cognitive background of social complexity and provide a basis for future studies examining social complexity in the context of the SBH in humans.

INTRODUCTION

The human brain has undergone significant expansion since before the divergence between humans and other primates (Montgomery et al., 2010). The evolutionary factors driving this expansion are still up for debate, although some well-supported hypotheses have surfaced across the past decades. One example of such a hypothesis is the social brain hypothesis (SBH), synthesized from previous work and hypotheses by Barton and Dunbar (1997). The main idea of the SBH is that the large brains (and particularly the large neocortices) found in primates are a result of selection pressure for the ability to cope with their highly complex social environments (Dunbar, 2009).Complex social lives are hypothesized to be an evolutionary advantage either because it reduced risk of predation and/or increased foraging success, although evidence for the association between group size (a common operationalization of complex sociality) and predation risk are more strongly substantiated (Dunbar, 2009). Benefits of complex sociality in terms of fitness then translated to increases in brain size as the cognitive demands of increasingly complex social environments required high levels of cognitive skills necessary for maintaining large amounts of complicated social relationships (Dunbar, 2009; Shultz & Dunbar, 2007; Shultz & Dunbar, 2012).

While traditional methods of examining evolution relied on phylogenetic comparisons between species with varying levels of the trait of interest, in recent years new methods have arisen which allow for the examination of evolution using genetic data from a single species, such as singleton density scores (Field, 2016). Song et al. (2021) recently combined established and new single-species genetic methods to create estimates of how traits evolved throughout human evolution. Such methods raise the possibility of examining how traits such as social complexity have been affected by selection and other evolutionary processes from before human speciation (for an example using genetic risk for major depressive disorder see Sall et al., 2021) up to more recent evolutionary history as can be analyzed using the singleton density method by Field (2016). However, an important requirement for such studies is that first, high-quality genetic studies are performed to determine what genes contribute to the genetic background of social complexity.

In turn, such studies rely on the availability of a clear and operationalizable definition of social complexity. However, both theoretical definitions and methodological operationalizations of 'social complexity' vary. From an early moment, group size was a common measure used in studies examining the social brain hypothesis (e.g. Barton, 1996; Dunbar, 1992; Dunbar, 1998). However, several criticisms have been leveled at the idea that the association between brain size and social complexity is merely a quantitative relation between group size and brain size. Dunbar and Shultz (2007) found that while in primates a quantitative association might be an adequate description of the association between social complexity and brain size, in other

animal species the association might take a more qualitative form. Freeberg et al. (2012) argue that the size of the social grouping need not necessarily indicate complex social behavior and uses the example of a large group of ungulates which move together but do not have any real interaction with many of the individuals in the group. In the past decade, several researchers have attempted to create a definition of social complexity which lends itself for use across species in order to facilitate the ability to compare future studies of social complexity. Such attempts to define social complexity may also be useful for the creation of social complexity measures in single-species studies. Harmonizing the measures used in between-species and within-species studies may facilitate the comparison of the results from such varying studies. In the next paragraph, we will discuss several proposed definitions and how these can relate to within-species studies of social complexity.

Definitions of (individualized) social complexity

Freeberg et al. (2012) have provided a broad description of a complex social environment as "those [social systems] in which individuals frequently interact in many different contexts with many different individuals, and often repeatedly interact with many of the same individuals over time". While this definition is useful, operationalization is hampered somewhat by the vagueness of the terms "frequently" and "many". Bergman and Beehner (2015) mention that while the definition by Freeberg et al. (2012) has the advantage of encouraging objective measurement, a definition of social complexity should incorporate the association with cognition as this association is one of the most important tenets of the social brain hypothesis, while not all measures of the social environment are necessarily related to cognition. Based on this, they define social complexity as the number of differentiated relationships an individual has. One clear distinction between the definitions of Freeberg et al. (2012) and Bergman and Beehner (2015) is that while the definition by Freeberg et al. (2012) clearly focusses on the social complexity of the social system, the definition by Bergman and Beehner (2015) specifically mentions the individual, indicating that such social complexity can vary within a species.

A more thorough discussion on the measurement of social complexity specifically was provided by Kappeler (2019). Kappeler (2019) mentions the definition of Bergman and Beehner (2015) but acknowledges that the definition is not very specific about how to operationalize 'differentiated relationships' across species and that it might be difficult to examine it in similar ways across species. In order to create a more consistent and operationalizable definition for the study of social complexity across species, based on previous work in the field Kappeler (2019) created a framework for social complexity consisting of four components: social organization, social structure, mating system and care system. Social organization consists of aspects such as group size, group composition and kinship pattern, social structure describes variation in the content, quality and patterning of social relationships, the mating system regards mating patterns and reproductive skew and the care system is simply about who takes caretaking roles in regards to

offspring (parents and/or alloparents). Although the framework provides good handholds for how individualized measurements of social complexity could be designed and Kappeler (2019) mentions that intraspecific variation in social complexity is significant and understudied, the focus of the discussion is mostly on interspecific variation. However, Aureli & Schino (2019) discuss how social structure and social organization affect the individual and how this is related to variation in cognitive skills. Social structure is relevant in individuals as variation between and within relationships require the ability to differentiate between individuals and determine which behavioral strategy is ideal for each at a specific time. Social organization affects individuals through the cohesion of social relationships and through the composition of social groups, affecting the number and characteristics of conspecifics an individual can interact with. These definitions and discussions are a valuable resource for researchers attempting to create their own measures of social complexity.

Associations between social complexity, cognition and brain size

The SBH posits the idea that certain between-species differences in social complexity are related to between-species variation in brain size, at least in primates (Dunbar, 1992; Barton, 1996; Dunbar, 1998) although evidence has also been found in other species (e.g. ground squirrels, Matějů et al., 2016). One previous study has already shown that in humans, 'sociability' is associated with grey matter volumes in several brain regions expected to be involved in social behavior (Horváth et al., 2011). However, this study was quite small (25 subjects) and didn't attempt to measure social complexity specifically, although their measure of sociability did include measures which could be used to measure social complexity such as a subject's number of friends and the amount of time spent with them. This study did demonstrate that the associations hypothesized by the SBH may also be applicable to human interindividual variation.

The association between brain size and social complexity is supposedly mediated by the cognitive requirements of high social complexity, which implies both an association between certain cognitive skills and social complexity as well as between brain size and those specific cognitive skills (Dunbar, 2009; Shultz & Dunbar, 2012). An association between social complexity and cognition is imperative when studying the social brain hypothesis (Bergman & Beehner, 2015). Therefore, an important evaluative method for measures of social complexity is to what extent they relate to cognition.

The main cognitive skill supposed to be involved in social complexity is theory of mind (ToM; also referred to as mentalizing) (Dunbar, 2009; Shultz & Dunbar, 2012), the ability to recognize the intentions and emotions of other individuals. ToM is a higher-order cognitive function which depends on several lower-order cognitive skills, such as attention (Fahie & Symons, 2003; Lin et al., 2010; Mary et al., 2016), memory (Davis & Pratt, 2007; Ciaramelli et al.,

2013; Laillier et al., 2019) and processing speed (Charlton et al., 2009; Ayesa-Arriola et al., 2016; Laillier et al., 2019).

In the current study we will attempt to create an individualized measure of social complexity using data available in the UK Biobank as an example for potential future studies examining the genetic background of social complexity in the context of human evolutin. In order to examine whether the associations between brain size, social complexity and cognition are retained and may affect evolution in modern humans, we will examine whether such associations exist between our measure of social complexity, various measures of grey matter size across the brain and cognitive measures available in the UK Biobank.

METHOD

Subjects

The data was provided by the UK Biobank, a large population cohort from the United Kingdom including over 500000 adults measured across a large number of phenotypes (Bycroft et al., 2018). A complete overview of the data collection performed by the UK Biobank can be found on the UK Biobank website (https://www.ukbiobank.ac.uk/enable-your-research/about-our-data) and the UK Biobank showcase (https://biobank.ndph.ox.ac.uk/showcase/). The data collection was approved by the National Research Ethics Service Committee North West Multi-Centre Haydock and informed consent was required from all participants. Only subjects who participated in the first wave of MRI data acquisition (N = 42802) could be included in the analyses regarding the association between social complexity and (regional) grey matter volumes. In order to control for potential biasing factors, subjects were excluded if they had a BMI below 15 or above 40, if they reported having severe tinnitus, if they suffered from narcolepsy, if they had suffered a stroke, if they had been diagnosed with brain cancer, if they were blind, if they were deaf or if they were unable to walk. Subjects were also excluded if they were reported to have any disease or disorder affecting the central nervous system (for specific ICD-10 codes, see supplementary materials).

Social complexity scores

We created three measures which capture aspects of social complexity as defined by Kappeler (2019). The first measure was created by combining the number of leisure locations visited by the subject (minimum 0, maximum 6), the employment status (1 if employed, 0 otherwise) of the subject and the student status of the subject status (1 if student, 0 otherwise) into a variable we termed 'number of social contexts' (minimum 0, maximum 8). This was based on the idea that individuals who attend more locations with access to social contact likely spend more time engaging in social contact and with more variation in contacts compared to individuals with

fewer of such locations. The employment and student status of the subject were included as the work environment and study environment typically also include contact with colleagues or fellow students (although this does not have to be the case, in the case of self-employed subjects or subjects following their studies online, for example). In the rest of this paper we will term this variable 'number of social contexts'.

As a second measure of social structure, we included the amount of time spent with family and friends. This was a measure implemented as is in the UK Biobank. The categories provided to the subjects were 'No friends/family outside household' (scored as 0), 'Never or almost never' (scored as 0), 'Once every few months' (scored as 1), 'About once a month' (scored as 2), 'About once a week' (scored as 3), '2-4 times a week' (scored as 4) and 'Almost daily' (scored as 5).

Finally, we included a measure of Kappeler's (2019) 'social organization'. The key feature of social organization is the living situation of individuals of a species. This aspect of social complexity can be viewed as the aspect of social complexity typically used by studies using 'group size' as an operationalization of social complexity. In order to create a similar measure in the UK Biobank, we used the number of individuals living in the household of the subject as a measure of social organization. However, as the measure was extremely skewed with very few individuals having household sizes over 6, we created a score between 0 and 5 indicating the number of individuals sharing a living situation with the subject, with 5 indicating that the subject lived with 5 or more others.

We also created a composite measure for which each separate measure was standardized to a range of 0 to 1. The scores were then added to create the composite score, resulting in a score ranging between minimum 0 and maximum 3. Although a composite score is less likely to show how exactly different aspects of social complexity relate to brain size, it might better represent differences between individuals as separate scores might be more likely suffer from coincidentally high scores. For example, someone might live with individuals outside the direct family unit involuntarily and therefore score highly on the living situation variable, but if they do not also voluntarily spend time with friends or family and in social contexts outside the house, they would still score relatively low on the composite score.

As the UK Biobank data consists of several measurements for most individuals, we had to make a decision which measurement was used for the analyses. For the separate measures, we used the data gathered during the third instance, which was when the MRI measurements were performed. If the data were not available for that timepoint, we used the second instance data, or the first if neither the third nor second instance had the required data for the subject. The fourth instance occurred after the MRI measurement, and was only used if data was not available for the first three instances; only very few subjects had data for this instance. For the

composite measure, we created the composite score for each instance which had data for each separate measure and then chose the maximum as a subject would need the cognitive skills to cope with the most socially complex situation they choose to be in while lower scores could reflect potentially involuntary or temporary changes to the social environment of the subject. Both the separate social complexity measures as well as the composite measure were normalized and demeaned for use in the MRI analysis.

Cognition

The UK Biobank contains multiple measures of cognitive skills. A full overview of the cognitive domains measured and the tests used can be found in Table 1. The names for the cognitive domains are adapted from Fawns-Ritchie et al. (2020). Cognitive tests used by the UK Biobank are not standardized and at the time of administration had not been tested for psychometric quality. Since then, psychometric quality of the UK Biobank cognitive tests has been assessed in several studies (Lyall et al., 2016; Fawns-Ritchie et al., 2020; Ciobanu et al., 2023).Psychometric quality in terms of internal consistency, short-term test-retest reliability and concurrent validity of the UK Biobank cognitive tests have been found to be reasonable to good in general (Lyall et al., 2016; Fawns-Ritchie, 2020), although the long-term test-retest reliability of some of the tests may not be ideal (Lyall et al., 2016).

Table 1. Cognitive domains measured (adapted from Fawns-Ritchie et al., 2020) and the tests used by the UK Biobank.

Cognitive domain	*UK Biobank test name*
Visual declarative memory	Pairs matching test
Processing speed	Reaction time test
Prospective memory	Prospective memory test
	Symbol digit substitution
Verbal and numerical reasoning	Fluid intelligence test
Executive function	Trail making test
	Tower rearranging test
Verbal declarative memory	Paired associate learning test
Non-verbal reasoning	Matrix pattern completion

MRI imaging

The imaging protocol followed by the UK Biobank can be found at https://www.fmrib.ox.ac.uk/ukbiobank/protocol/. We included 66 imaging-derived phenotypes (IDPs), including total grey matter volume and regional grey matter volumes. A full list of all included IDPs is included in the Supplementary Table 1. All IDPs were demeaned before inclusion in the statistical analyses.

Statistical analyses

We used linear regression to examine the association between social complexity and MRI-derived regional and total grey matter volumes. We performed these analyses twice, once with the separate social complexity measures as predictors and once with the composite social complexity measure as predictor. Predictors were demeaned and normalized to create predictors with mean 0 and with a range of 1. The demeaned IDPs were used as outcome variables. Due to the large number of statistical tests for the IDPs, we used a Bonferroni correction based on the number of IDPs for the linear regressions examining associations between social complexity and grey matter volume. The new alpha values for the associations between the IDPs and the social complexity measures therefore became 0.05/66 = 0.0008. P-values are reported adjusted for the correction.

We examined demographic variables and BMI to examine whether they were associated with social complexity to determine whether they should be included as covariate. The variables assessed were sex, age, income, education, ethnicity and BMI. Besides demographic covariates and BMI we also controlled for head size and for the associations between social complexity and regional grey matter volumes, for total grey matter volume.

Associations between social complexity variables and cognition were assessed using Pearson correlations, with the exception of the associations between the social complexity measures and prospective memory, as prospective memory was assessed using a categorical measure. For prospective memory, an ANOVA was used. Post-hoc comparisons were made using Tukey's studentized pair-wise t-test. As the number of cognitive measures does result in several comparisons, the alpha value for the determination of significance was adjusted using a Bonferroni correction. The adjusted alpha value for the cognition-social complexity association is 0.005.

RESULTS

Social complexity measure distributions

Visual representations of the distributions of the social complexity measures can be found in Figure 1. All three separate measures had skewed distributions, especially social context score and household number scores. However, the composite social complexity scores did conform to a normal distribution.

We examined associations between the social complexity measures and demographic variables to determine which demographic variables should be included in the linear regressions assessing the associations between social complexity and grey matter volume in order to reduce the potential for biasing factors. We included analyses for age, sex, ethnicity, education, income

and additionally body mass index. The full results are displayed in Supplementary Table 2. Social contexts score was significantly associated with each of the demographic variables and BMI (p-values all <0.001). Household number score was associated with all demographic variables (p-values all <0.001) but not with BMI. Friend and family visits score was associated with all demographic variables as well as BMI (p-values range <0.001 – 0.005). Due to the associations between the demographic variables and BMI with the social complexity variables, all demographic variables plus BMI were included as covariates. Age squared was also included in order to account for nonlinear associations between age and grey matter volume.

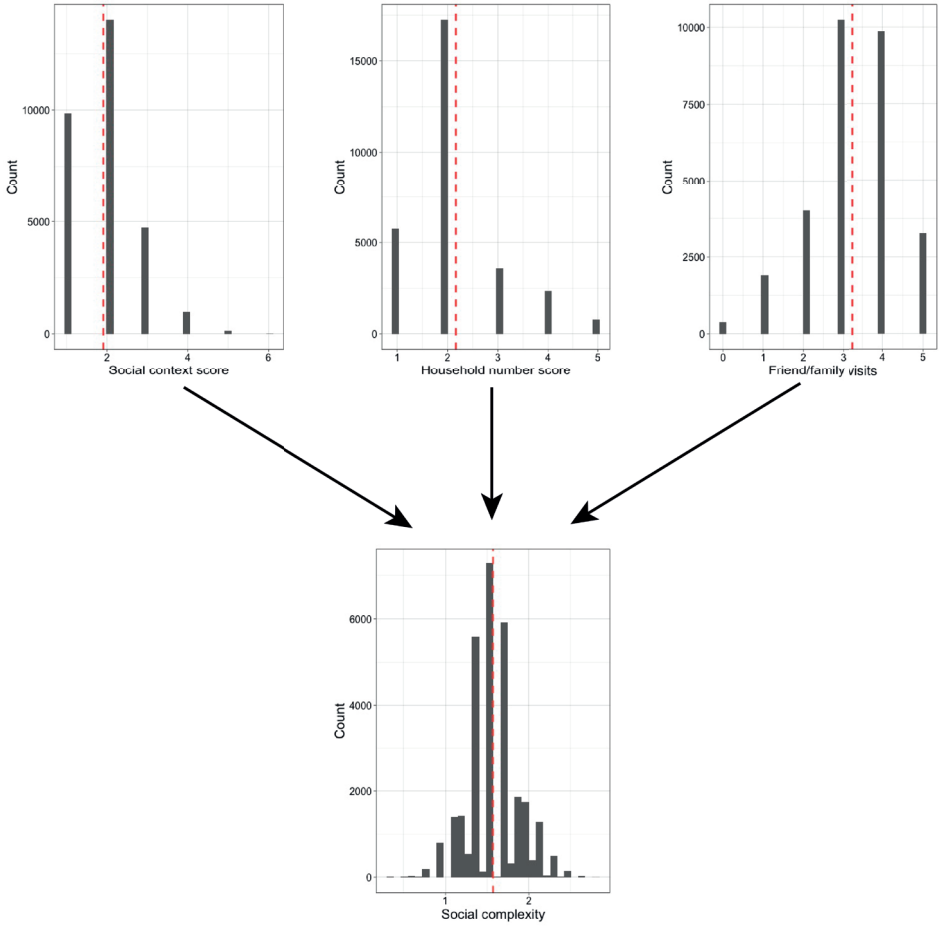


Figure 1. Distributions of separate social complexity measures and composite social complexity measure before normalization and demeaning. Red dashed lines indicate the means of the respective distributions.

Social complexity and grey matter volumes

After the Bonferroni correction, the total number of individuals with data for both grey matter volumes and social complexity was N = 29703. Total grey matter was significantly associated with the composite social complexity score (β = 0.02, F (1, 25928) = 27.68, p (adjusted) < 0.001), but not with the separate social complexity measures. The right side of the inferior lateral occipital cortex was associated with the friends and family visits social complexity variable (β = 0.02, F (1, 25922) = 13.42, p (adjusted) = 0.013), as was the right hemispheric side of the pallidum (β = -0.02, F (1, 25921) = 15.10, p (adjusted) = 0.007). The data showed a very outlying data point (see figure 2), however removing the datapoint did not change the results so we have reported the results with the datapoint included. The cerebellum lobules VIIIa (right side only, β = 0.02, F (1, 25924) = 14.45, p (adjusted) = 0.007) and VIIIb (right side: β = 0.02, F (1, 25924)= 12.04, p (adjusted) = 0.033; left side: β = 0.03, F (1, 25924) = 19.59, p (adjusted) < 0.001) as well as the cerebral crus I (right side: β = 0.03, F (1, 25924)= 26.07, p (adjusted) < 0.001; left side: β = 0.02, F (1, 25924) = 17.39, p (adjusted) = 0.002) were significantly associated with the composite social complexity measure. Scatterplots with regression lines for the significant results can be found in Figure 1. Plots showing the regression between the social complexity variables and the predicted results are included in Supplementary Figure 1. Similar plots using the predicted values (which take into account the covariates) are presented in Supplementary Figure 1. None of the other regional volumes were significantly associated with the social complexity measures after the Bonferroni correction. The full overview of the results including coefficients can be found in Supplementary Table 1.

Social complexity and cognition

The analyses regarding the associations between the various social complexity measures and cognitive skills reveal consistent associations between the social complexity measures and various cognitive abilities. The correlation coefficients (or in the case of prospective memory the F- statistic) and p-values are reported in Table 2. The number of social contexts, the size of the household and the composite social complexity score were all significantly associated with each of the cognitive measures in the same direction, indicating a positive association between cognition and social complexity. Friends and family visits was significantly associated with processing speed, verbal and numerical reasoning, executive functioning as measured by the tower rearranging task and non-verbal reasoning. However, each of these associations was in the opposite direction from the other three measures. Results from the ANOVA post-hoc tests for the prospective memory analysis can be viewed in Table 3.

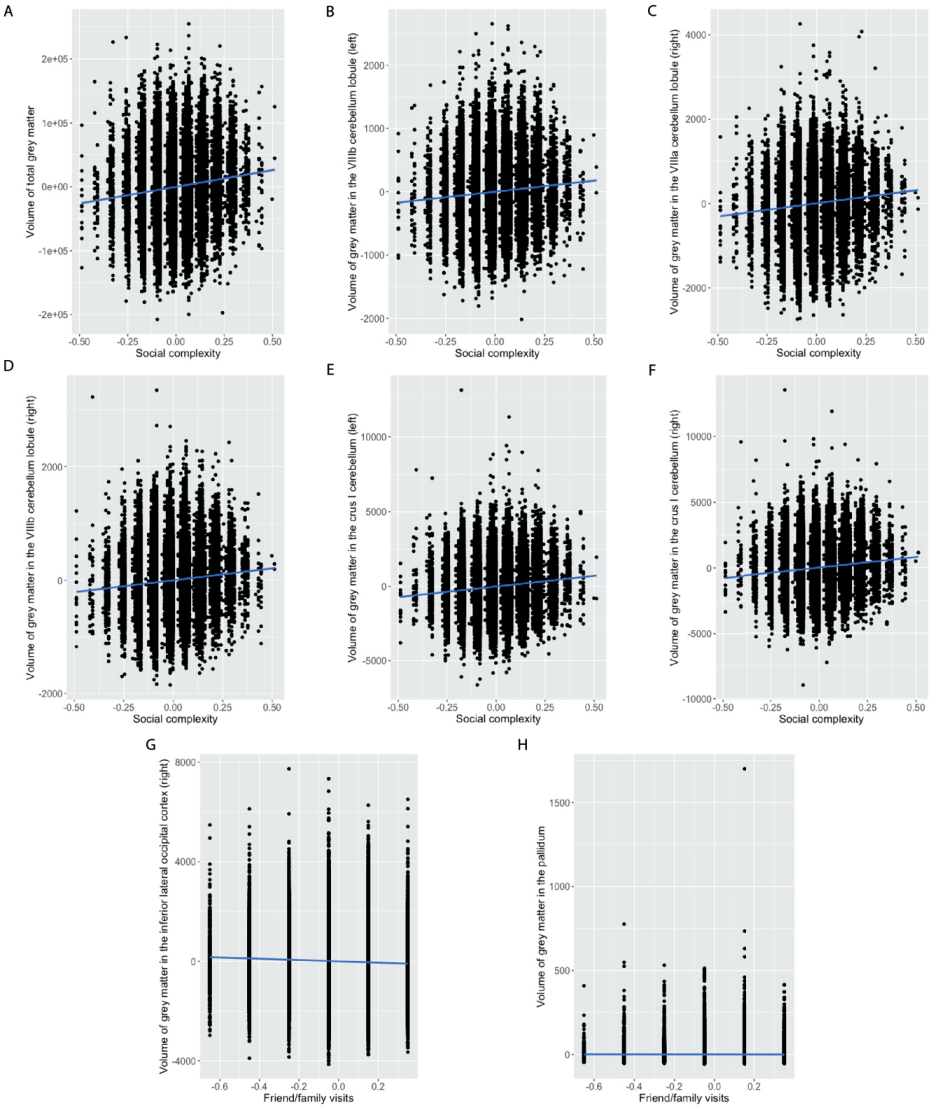


Figure 2. Associations between total and regional grey matter volumes and social complexity measure scores. Plots A through F show associations between the IDPs and the composite social complexity score, plots G and H show associations between IDPs and the number of friend and family visits. Blue lines show the regression lines.

Table 2. Results of the analyses regarding associations between social complexity measures and cognitive skills.

	Social context score	Household number	Friends/family visits	Composite social complexity
Visual declarative memory	r = -0.04, p < .001*	r = -0.05, p < .001*	r = 0.01, p = 0.158	r = -0.04, p < .001*
Processing speed	r = -0.09, p < .001*	r = -0.12, p < .001*	r = 0.05, p < .001*	r = -0.09, p < .001*
Prospective memory[1]	F (2, 27971) = 47.98, p < .001*	F (2, 27960) = 28.18, p < .001*	F (2, 27965) = 0.01, p < .992	F (2, 27954) = 45.01, p < .001*
Verbal and numerical reasoning	r = 0.08, p < .001*	r = 0.03, p < .001*	r = -0.04, p < .001*	r = 0.03, p < .001*
Executive functioning: Tower rearranging	r = 0.07, p < .001*	r = 0.08, p < .001*	r = -0.03, p < .001*	r = 0.07, p < .001*
Executive functioning: Trail making 1	r = -0.11, p < .001*	r = -0.10, p < .001*	r = 0.02, p = 0.009	r = -0.11, p < .001*
Executive functioning: Trail making 2	r = -0.04, p < .001*	r = -0.04, p < .001*	r = 0.02, p = 0.016	r = -0.03, p < .001*
Verbal declarative memory	r = 0.13, p < .001*	r = 0.07, p < .001*	r < 0.01, p = 0.804	r = 0.10, p < .001*
Non-verbal reasoning	r = 0.12, p < .001*	r = 0.09, p < .001*	r = -0.06, p < .001*	r = 0.07, p < .001*

Table 3. Results of the post-hoc analysis for the ANOVA assessing the association between social complexity and prospective memory. * = Significant at the p=0.005 level.

	Social context score	Household number	Friends/family visits	Composite social complexity
Correct on second attempt – Correct on first attempt	Δ = -0.02, p < .001*	Δ = -0.03, p < .001*	N.S.	Δ = -0.01, p < .001*
Incorrect/skipped - Correct on first attempt	Δ = -0.04, p < .001*	Δ = -0.04, p < .001*	N.S.	Δ = -0.03, p < .001*
Incorrect/skipped - Correct on second attempt	Δ = -0.02, p = 0.005*	Δ = -0.01, p = 0.623	N.S.	Δ = -0.01, p = 0.006

DISCUSSION

The goal of this study was to create measures capable of measuring individual variation in social complexity between human subjects and to examine whether such measures would be usable in studies examining recent selection on social complexity in relation to the SBH. We found that indeed our measures of social complexity captured significant variety, although the social context score and household number score were highly skewed with most individuals falling in the lower scores. However, the distribution of the composite social complexity score resembles a normal distribution with individuals spanning all potential scores.

The analyses examining the associations between social complexity and total and regional grey matter volumes resulted in several interesting findings. First, as could be expected if the concur-

rent selection on social complexity and brain size hypothesized in the social brain hypothesis indeed resulted in pleiotropy between the two variables, the composite social complexity measure was associated with the total grey matter volume. This was not the case for the separate social complexity measures, however. One possibility is that, as described by Kappeler (2019), social complexity is a complicated trait which can be divided into several subdomains, and scoring high on a single of the subdomains does not necessarily indicate high social complexity if scores on the other characteristic are low. For example, someone might not always have the ability to decide with how many people they live (in case of financial difficulty, people can for example choose to have roommates out of necessity) and in that case high household number size scores do not necessarily reflect high social complexity, while scoring relatively high on household size, social locations visited and the number of visits with friends and family do likely reflect a genuine increase in social complexity.

Besides total grey matter, associations between regional values and the composite social complexity score shows a potential involvement of the cerebellum in social complexity. Particularly, lobules in the inferior posterior lobes of the cerebellum (VIIIa and VIIIb) appear involved in social complexity. Many previous studies in humans and non-human animals have pointed to a role of the cerebellum in social cognition and social behavior (e.g. Carta et al., 2019; Heleven et al., 2019; Hoche et al., 2016; Metoki et al., 2022; Van Overwalle et al., 2019). In a meta-analysis of 350 studies, Van Overwalle et al. (2013) found that the cerebellum is implicated in social cognition with high consistency across studies. Reductions in grey matter in lobules VIIIa and VIIIb have been found to be associated with symptoms of autism spectrum disorder (Stoodley, 2014), which is characterized by social and sensory dysfunction. The cerebellar crus I has been found to be transdiagnostically associated with social cognition (Brady et al., 2020) and is grey matter abnormalities in this region are also associated with autism spectrum disorder (D'Mello et al., 2015). Our analyses provide further evidence of the role of the cerebellum in social functions. While papers treating the SBH typically focus on the evolution of the neocortex, the association between social complexity and grey matter volumes in the cerebellum (particularly the inferior posterior cerebellum) found in this study may make it more interesting for future studies to also examine whether co-evolution between social complexity and the cerebellum could have taken place.

The number of visits with friends and family also showed associations with regional grey matter volumes, specifically in the right inferior lateral occipital cortex and the right side of the pallidum. The lateral occipital cortex has been found to be involved in the face perception network (Nagy et al, 2012). The inferior occipital gyrus, which is a part of the inferior lateral occipital cortex, has also been found to play a role in face-evoked potentials (Jacques et al., 2019), suggesting a role between face perception and the social structure aspect of social complexity as described by Kappeler (2019). The inferior lateral occipital cortex also appears

to be involved in determining spatial relations between social stimuli (Abassi & Papeo, 2019). These findings indicate that perception may play an important role in interindividual variation in social complexity in humans. Social perceptual abilities are known to play an important role in ToM (Meinhardt-Injac et al., 2018), which is supposed to be one of the social cognitive abilities which developed throughout evolution according to the SBH. This may explain why our findings point to a role of the inferior lateral occipital cortex in human variation in social complexity.

Previous findings regarding the involvement of the pallidum in social functions are less clear, although some evidence has been presented. Lim and Young (2004) found that in monogamous prairie voles, the pallidum is part of a circuit involved in pair bonding. Pair bonding has been hypothesized to be the basis for non-reproductive relationships (Dunbar, 2009), which might explain why we find an involvement of the pallidum in social complexity, specifically in the frequency of visits with friends and family. Secondly, the pallidum plays an important role in the default mode network (DMN) (Klaassen et al., 2021), which is known to be involved in social cognition (Fareri et al., 2020; Mars et al., 2012; Wen et al., 2020). However, while we found a positive association between grey matter volume in the pallidum and the number of friend/family visits, other studies have found opposite associations between pallidum volume and social behavior, even finding associations between pallidum volume and social anhedonia (Wang et al., 2014). However, the results from the study regarding anhedonia found an effect specifically for the pallidum in the left hemisphere, while the association between social complexity and pallidum volume in this study was specific to the right hemisphere. This indicates that perhaps lateralization is relevant for the role of the pallidum in social cognition and behavior. As it stands, the underlying mechanisms of the association between the pallidum volume and social complexity found in this study are unclear and will require further research.

The lateralization mentioned in the previous paragraph is an interesting finding of our analyses in general, as for all but two of the associations regarding regional grey matter volumes, significant effects were only found for the right hemisphere. Previous studies have provided evidence of a lateralization of social functions in the right ride of the brain in humans (Hewetson et al., 2021; Rajimehr et al. (2022) as well as in other animals (Giljov et al., 2018; Salva et al., 2012). Our findings add to the literature documenting a right hemispheric dominance in social functions. It may be interesting in the future to examine whether the extent of lateralization of social functions itself might have evolved as part of selection on social complexity, as the extent of lateralization might be related to cognitive functioning (Gotts et al., 2013).

Unexpectedly, the associations between social complexity and regional grey matter volumes found in this study were mostly situated in the more 'ancient' parts of the brain, with the exception of the inferior lateral occipital cortex. This is unlike the majority of the findings

supporting the SBH, which typically point at the disproportionally expanded neocortex as the main aspect of brain evolution related to selection for high social complexity. Due to the novel nature of our measures, it is unclear whether this is a result of our measures not capturing the variation in social complexity aimed at by the cross-species studies, whether this is associated by a genuine change in how interindividual variation in social complexity is related to brain size or whether perhaps another explanation can account for our findings. However, it might be interesting for future studies to examine the (recent) evolution of the size of the basal ganglia and cerebellum in relation to selection on social complexity.

As could be expected based on the SBH, social complexity was consistently associated with cognition across the different cognitive tests. In fact, the consistency of the associations exceeded expectations, as the measures were not directly associated with social cognition and although general cognitive skills are likely important for social cognition, the indirect association with social complexity could have led to low or nondetectable associations. This was not the case however. The consistent associations between cognition and social complexity found in our study provide an indication that modern human variation in social complexity is indeed associated with cognitive skill. However, associations between social complexity and cognition were small, indicating that the variation in cognitive skills between humans may only determine variation in social complexity to a minor extent. If this is the case, selection on social complexity might result in only very small or no evolutionary change in cognitive skills. However, the associations could also have been small due to suboptimal measurement of social complexity (see below). Another reason why correlations might be smaller than expected is the indirect association between general cognitive skills and social complexity, as these should be mediated through social cognitive skills. Future studies could examine whether using direct measurements of social cognition result in higher correlations with social complexity. While the existence of a neurobiological overlap between social complexity and cognition in humans is outside of the scope of this study, a previous study has found that in rodents, cognition and sociality show considerable overlap in the neurobiological substrates underlying these domains (Lanooij et al., 2023). This could be an interesting avenue for further research regarding social complexity.

Surprisingly, the number of visits with friends and family showed an inverse relation with cognition compared to both the number of social contexts and household size (as well as the composite measure), with reduced cognitive abilities in those with higher frequency of visits with friends and family. Although the reason for this is unclear, as the frequency of visits with friends and family appears a relatively straightforward translation of traditional measures of social complexity such as group size or 'number of differentiated relationships', we can speculate that perhaps individuals with reduced cognitive abilities receive more attention from friends and family as a form of care. Another option is that perhaps individuals with reduced cognitive

functioning are less accurate in their reporting of their social activities, as has been reported in individuals with Alzheimer's disease and schizophrenia (Jongs et al., 2022). These findings do however call into question the quality of this self-report measure of the frequency friends and family visits as a measure of social complexity, particularly in the context of the SBH. It also casts doubt on the association between the pallidum and the inferior lateral occipital cortex and social complexity, as grey matter volumes in these regions were solely associated with this particular measure of social complexity.

An important limitation for this study was the novelty of the measures and the lack of options to validate them against existing measures of social complexity. This is unfortunately hard to overcome, as few studies at this point in time have examined social complexity specifically in terms of interindividual variation in humans. However, future studies could perform such validations by comparing the measures of social complexity used in this study to measures of other aspects of social behavior, such as the Social Functioning Scale (Birchwood et al., 1990). Similarly, the reliability of these novel measures has yet to be established.

Secondly, the analyses regarding cognition and social complexity could have been biased by the self-reported nature of the variables underlying the social complexity measures. Previous work on self-reported social complexity has shown that neuropsychiatric disorders affecting cognition such as schizophrenia and Alzheimer's disease could also cause inaccurate self-report of social behavior (Jongs et al., 2022). It is unknown whether this extends to normal variation in cognition, but if this is the case, it might have created artificial associations between cognition and social complexity.

Future studies can address these limitations and further the exploration of (recent) human evolution in the context of the SBH. Potential issues with the measurement can be addressed by carrying out studies examining the psychometric qualities of interindividual human social complexity as well as determine the ideal method to measure social complexity (e.g., objective vs self-report). One type of methodology which is promising for the measurement of social behavior in an objective manner is passive digital phenotyping. Algorithms using data from smartphones or (wearable) sensors can be used to detect behavioral variation between individuals, for example using GPS (Jagesar, 2021; Muurling et al., 2022). Such methods may also perform better in subject who might not be able to accurately report their behavior as a result of cognitive issues (Jongs et al., 2022). As mentioned, to determine whether interhuman variation in social complexity resembles interspecies variation in social complexity between primates and the association with social cognition, studies should be carried out which specifically examine this association, for examples by studying theory of mind. The associations between regional brain volumes and social complexity found in this study should be further examined to determine how the pallidum, cerebellum and inferior lateral occipital cortex volumes are

relevant to social complexity. Finally, examining the genetic background of social complexity is necessary to be able to perform future studies examining (recent) selection on social complexity using methods such as those utilized in Song et al. (2021).

In this study we performed a first examination of how the evolutionary associations between social complexity, cognition and brain size are applicable to individual human variation. We found highly consistent associations between cognition and social complexity and some associations between brain size and social complexity. This is a promising first step in creating measures of social complexity which can be used in future studies to examine recent human evolution related to social complexity.

REFERENCES

Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. The Journal of neuroscience : he official journal of the Society for Neuroscience, 40(4), 852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

Aureli, F., Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. Behav Ecol Sociobiol 73, 6. https://doi.org/10.1007/s00265-018-2604-5

Ayesa-Arriola, R., Setién-Suero, E., Neergaard, K. D., Ferro, A., Fatjó-Vilas, M., Ríos-Lago, M., Otero, S., Rodríguez-Sánchez, J. M., & Crespo-Facorro, B. (2016). Evidence for Trait Related Theory of Mind Impairment in First Episode Psychosis Patients and Its Relationship with Processing Speed: A 3 Year Follow-up Study. Frontiers in psychology, 7, 592. https://doi.org/10.3389/fpsyg.2016.00592

Barton R. A. (1996). Neocortex size and behavioural ecology in primates. Proceedings. Biological sciences, 263(1367), 173–177. https://doi.org/10.1098/rspb.1996.0028

Barton, R. A., & Dunbar, R. I. M. (1997). Evolution of the social brain. In A. Whiten & R. W. Byrne (Eds.), Machiavellian intelligence II: Extensions and evaluations (pp. 240–263). Cambridge University Press. https://doi.org/10.1017/CBO9780511525636.010

Bergman, T.J. & Beehner, J.C. (2015). Measuring social complexity. Animal behaviour, 103, 203-209. https://doi.org/10.1016/j.anbehav.2015.02.018

Birchwood, M., Smith, J., Cochrane, R., Wetton, S., & Copestake, S. (1990). The Social Functioning Scale. The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. The British journal of psychiatry: the journal of mental science, 157, 853–859. https://doi.org/10.1192/bjp.157.6.853

Brady, R. O., Jr, Beermann, A., Nye, M., Eack, S. M., Mesholam-Gately, R., Keshavan, M. S., & Lewandowski, K. E. (2020). Cerebellar-Cortical Connectivity Is Linked to Social Cognition Trans-Diagnostically. Frontiers in psychiatry, 11, 573002. https://doi.org/10.3389/fpsyt.2020.573002

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Carta, I., Chen, C. H., Schott, A. L., Dorizan, S., & Khodakhah, K. (2019). Cerebellar modulation of the reward circuitry and social behavior. Science (New York, N.Y.), 363(6424), eaav0581. https://doi.org/10.1126/science.aav0581

Charlton, R. A., Barrick, T. R., Markus, H. S., & Morris, R. G. (2009). Theory of mind associations with other cognitive functions and brain imaging in normal aging. Psychology and aging, 24(2), 338–348. https://doi.org/10.1037/a0015225

Ciaramelli, E., Bernardi, F., & Moscovitch, M. (2013). Individualized Theory of Mind (iToM): When Memory Modulates Empathy. Frontiers in psychology, 4, 4. https://doi.org/10.3389/fpsyg.2013.00004

Ciobanu, L. G., Stankov, L., Ahmed, M., Heathcote, A., Clark, S. R., & Aidman, E. (2023). Multifactorial structure of cognitive assessment tests in the UK Biobank: A combined exploratory factor and structural equation modeling analyses. Frontiers in psychology, 14, 1054707. https://doi.org/10.3389/fpsyg.2023.1054707

Davis, H.L. & Pratt, C. (1995) The development of children's theory of mind: The working memory explanation. Australian Journal of Psychology, 47, 1, 25-31. https://doi.org/10.1080/00049539508258765

D'Mello, A. M., Crocetti, D., Mostofsky, S. H., & Stoodley, C. J. (2015). Cerebellar gray matter and lobular volumes correlate with core autism symptoms. NeuroImage. Clinical, 7, 631–639. https://doi.org/10.1016/j.nicl.2015.02.007

Dunbar, R.I.M. (1992). Neocortex size as a constraint on group size in primates. Journal of Human Evolution, 22(6), 469-493. https://doi.org/10.1016/0047-2484(92)90081-J

Dunbar, R.I.M. (1998). The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews, 6(5), 178-190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R.I.M. (2009) The social brain hypothesis and its implications for social evolution. Annals of Human Biology, 36:5, 562-572.https://doi.org/10.1080/03014460902960289

Dunbar, R. I., & Shultz, S. (2007). Understanding primate brain evolution. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 362(1480), 649–658. https://doi-org.proxy-ub.rug.nl/10.1098/rstb.2006.2001

Fahie, C.M. & Douglas, K.S. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. Journal of Applied Developmental Psychology, 24, 1, 51-73. https://doi.org/10.1016/S0193-3973(03)00024-8

Fareri, D. S., Smith, D. V., & Delgado, M. R. (2020). The influence of relationship closeness on default-mode network connectivity during social interactions. Social cognitive and affective neuroscience, 15(3), 261–271. https://doi.org/10.1093/scan/nsaa031

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the UK Biobank cognitive tests. PloS one, 15(4), e0231627. https://doi.org/10.1371/journal.pone.0231627

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M. I., & Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. Science (New York, N.Y.), 354(6313), 760–764. https://doi.org/10.1126/science.aag0776

Freeberg, T. M., Dunbar, R. I., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 367(1597), 1785–1801.https://doi.org/10.1098/rstb.2011.0213

Giljov, A., Karenina, K., & Malashichev, Y. (2018). Facing each other: mammal mothers and infants prefer the position favouring right hemisphere processing. Biology letters, 14(1), 20170707. https://doi.org/10.1098/rsbl.2017.0707

Gotts, S. J., Jo, H. J., Wallace, G. L., Saad, Z. S., Cox, R. W., & Martin, A. (2013). Two distinct forms of functional lateralization in the human brain. Proceedings of the National Academy of Sciences of the United States of America, 110(36), E3435–E3444. https://doi.org/10.1073/pnas.1302581110

Heleven, E., van Dun, K., & Van Overwalle, F. (2019). The posterior Cerebellum is involved in constructing Social Action Sequences: An fMRI Study. Scientific reports, 9(1), 11110. https://doi.org/10.1038/s41598-019-46962-7

Hewetson, R., Cornwell, P., & Shum, D. H. K. (2021). Relationship and Social Network Change in People With Impaired Social Cognition Post Right Hemisphere Stroke. American journal of speech-language pathology, 30(2S), 962–973.https://doi.org/10.1044/2020_AJSLP-20-00047

Hoche, F., Guell, X., Sherman, J. C., Vangel, M. G., & Schmahmann, J. D. (2016). Cerebellar Contribution to Social Cognition. Cerebellum (London, England), 15(6), 732–743. https://doi.org/10.1007/s12311-015-0746-9

Horváth, K., Martos, J., Mihalik, B., & Bódizs, R. (2011). Is the social brain theory applicable to human individual differences? Relationship between sociability personality dimension and brain size. Evolutionary psychology: an international journal of evolutionary approaches to psychology and behavior, 9(2), 244–256.

Jacques, C., Jonas, J., Maillard, L., Colnat-Coulbois, S., Koessler, L., & Rossion, B. (2019). The inferior occipital gyrus is a major cortical source of the face-evoked N170: Evidence from simultaneous scalp and intracerebral human recordings. Human brain mapping, 40(5), 1403–1418. https://doi.org/10.1002/hbm.24455

Jagesar, R. R., Roozen, M. C., van der Heijden, I., Ikani, N., Tyborowska, A., Penninx, B. W. J. H., Ruhe, H. G., Sommer, I. E. C., Kas, M. J., & Vorstman, J. A. S. (2021). Digital phenotyping and the COVID-19 pandemic: Capturing behavioral change in patients with psychiatric disorders. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, *42*, 115–120. https://doi.org/10.1016/j.euroneuro.2020.11.012

Jongs, N., Penninx, B., Arango, C., Ayuso-Mateos, J. L., van der Wee, N., Rossum, I. W., Saris,I. M. J., van Echteld, A., Koops, S., Bilderbeck, A. C., Raslescu, A., Dawson, G. R., Sommer, B., Marston, H., Vorstman, J. A., Eijkemans, M. J., & Kas, M. J. (2022). Effect of disease related biases on the subjective assessment of social functioning in Alzheimer's disease and schizophrenia patients. Journal of psychiatric research, 145, 302–308. https://doi.org/10.1016/j.jpsychires.2020.11.013

Kappeler, P.M. (2019). A framework for studying social complexity. Behav Ecol Sociobiol 73, 13. https://doi.org/10.1007/s00265-018-2601-8

Klaassen, A. L., Heiniger, A., Vaca Sánchez, P., Harvey, M. A., & Rainer, G. (2021). Ventral pallidum regulates the default mode network, controlling transitions between internally and externally guided behavior. Proceedings of the National Academy of Sciences of the United States of America, 118(36), e2103642118. https://doi.org/10.1073/pnas.2103642118

Laillier, R., Viard, A., Caillaud, M., Duclos, H., Bejanin, A., de La Sayette, V., Eustache, F., Desgranges, B., & Laisney, M. (2019). Neurocognitive determinants of theory of mind across the adult lifespan. Brain and cognition, 136, 103588. https://doi.org/10.1016/j.bandc.2019.103588

Lanooij, S. D., Eisel, U. L. M., Drinkenburg, W. H. I. M., van der Zee, E. A., & Kas, M. J. H. (2023). Influencing cognitive performance via social interactions: a novel therapeutic approach for brain disorders based on neuroanatomical mapping? Molecular psychiatry, 28(1), 28–33. https://doi.org/10.1038/s41380-022-01698-1

Lim, M. M., & Young, L. J. (2004). Vasopressin-dependent neural circuits underlying pair bond formation in the monogamous prairie vole. Neuroscience, 125(1), 35–45. https://doi.org/10.1016/j.neuroscience.2003.12.008

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. Journal of Experimental Social Psychology, 26, 3. 551-556. https://doi.org/10.1016/j.jesp.2009.12.019

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., Anderson, J., Fawns-Ritchie, C., McIntosh, A. M., Deary, J., & Pell, J. P. (2016). Correction: Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. PloS one, 11(5), e0156366. https://doi.org/10.1371/journal.pone.0156366

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain". Frontiers in human neuroscience, 6, 189. https://doi.org/10.3389/fnhum.2012.00189

Mary, A., Slama, H., Mousty, P., Massat, I., Capiau, T., Drabs, V., & Peigneux, P. (2016). Executive and attentional contributions to Theory of Mind deficit in attention deficit/hyperactivity disorder (ADHD). Child neuropsychology: a journal on normal and abnormal development in childhood and adolescence, 22(3), 345–365. https://doi.org/10.1080/09297049.2015.1012491

Meinhardt-Injac, B., Daum, M. M., Meinhardt, G., & Persike, M. (2018). The Two-Systems Account of Theory of Mind: Testing the Links to Social- Perceptual and Cognitive Abilities. Frontiers in human neuroscience, 12, 25. https://doi.org/10.3389/fnhum.2018.00025

Metoki, A., Wang, Y., & Olson, I. R. (2022). The Social Cerebellum: A Large-Scale Investigation of Functional and Structural Specificity and Connectivity. Cerebral cortex (New York, N.Y.: 1991), 32(5), 987–1003. https://doi.org/10.1093/cercor/bhab260

Muurling, M., Reus, L. M., de Boer, C., Wessels, S. C., Jagesar, R. R., Vorstman, J. A. S., Kas, M. J. H., & Visser, P. J. (2022). Assessment of Social Behavior Using a Passive Monitoring App in Cognitively Normal and Cognitively Impaired Older Adults: Observational Study. *JMIR aging*, *5*(2), e33856. https://doi.org/10.2196/33856

Nagy, K., Greenlee, M. W., & Kovács, G. (2012). The lateral occipital cortex in the face perception network: an effective connectivity study. Frontiers in psychology, 3, 141. https://doi.org/10.3389/fpsyg.2012.00141

Rajimehr, R., Firoozi, A., Rafipoor, H., Abbasi, N., & Duncan, J. (2022). Complementary hemispheric lateralization of language and social processing in the human brain. Cell reports, 41(6), 111617. https://doi.org/10.1016/j.celrep.2022.111617

Sall, S., Thompson, W., Santos, A., & Dwyer, D. S. (2021). Analysis of Major Depression Risk Genes Reveals Evolutionary Conservation, Shared Phenotypes, and Extensive Genetic Interactions. Frontiers in psychiatry, 12, 698029. https://doi.org/10.3389/fpsyt.2021.698029

Salva, O.R., Regolin, L., Mascalzoni, E., & Vallortigara, G. (2012). Cerebral and behavioural asymmetries in animal social recognition. Comparative Cognition & Behavior Reviews, 7, 110-138. https://doi.org/10.3819/ccbr.2012.70006

Shultz, S., & Dunbar, R. I. (2007). The evolution of the social brain: anthropoid primates contrast with other vertebrates. Proceedings. Biological sciences, 274(1624), 2429–2436. https://doi.org/10.1098/rspb.2007.0693

Shultz, S., & Dunbar, R. I. M. (2012). The social brain hypothesis: An evolutionary perspective on the neurobiology of social behaviour. In Richmond, S.D., Rees, G, & Edwards, S. J. L. (Eds.), I Know What You're Thinking: Brain imaging and mental privacy (pp.12-28). Oxford Academic. https://doi.org/10.1093/acprof:oso/9780199596492.003.0002

Song, W., Shi, Y., Wang, W., Pan, W., Qian, W., Yu, S., Zhao, M., & Lin, G. N. (2021). A selection pressure landscape for 870 human polygenic traits. Nature human behaviour, 5(12), 1731–1743. https://doi.org/10.1038/s41562-021-01231-4

Stoodley C. J. (2014). Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia. Frontiers in systems neuroscience, 8, 92. https://doi.org/10.3389/fnsys.2014.00092

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. NeuroImage, 86, 554–572. https://doi.org/10.1016/j.neuroimage.2013.09.033

Van Overwalle, F., De Coninck, S., Heleven, E., Perrotta, G., Taib, N. O. B., Manto, M., & Mariën, P. (2019). The role of the cerebellum in reconstructing social action sequences: a pilot study. Social cognitive and affective neuroscience, 14(5), 549–558. https://doi.org/10.1093/scan/nsz032

Wang, Y., Deng, Y., Fung, G., Liu, W. H., Wei, X. H., Jiang, X. Q., Lui, S. S., Cheung, E. F., & Chan, R. C. (2014). Distinct structural neural patterns of trait physical and social anhedonia: evidence from cortical thickness, subcortical volumes and inter-regional correlations. Psychiatry research, 224(3), 184–191. https://doi.org/10.1016/j.pscychresns.2014.09.005

Wen, T., Mitchell, D. J., & Duncan, J. (2020). The Functional Convergence and Heterogeneity of Social, Episodic, and Self-Referential Thought in the Default Mode Network. Cerebral cortex (New York, N.Y.: 1991), 30(11), 5915–5929. https://doi.org/10.1093/cercor/bhaa166

Supplementary Table 1

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes.

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Total grey matter	0.010	5.890	0.015	0.010	4.930	0.026	0.010	10.330	0.001	0.020	27.680	0.000*
Amygdala (l)	-0.010	1.280	0.026	-0.002	0.150	0.694	-0.010	1.930	0.164	-0.004	0.057	0.414
Amygdala (r)	-0.004	0.790	0.373	-0.010	1.930	0.165	-0.010	4.010	0.045	-0.004	0.600	0.440
Angular gyrus (l)	0.010	4.180	0.041	0.010	0.610	0.436	0.004	0.570	0.449	0.010	3.410	0.064
Angular gyrus (r)	0.010	0.830	0.362	-0.010	1.240	0.266	0.002	0.140	0.711	0.010	1.040	0.308
Brain stem	0.000	0.040	0.849	0.010	2.540	0.111	0.010	1.700	0.192	0.020	7.980	0.005
Caudate (l)	0.001	0.010	0.907	0.000	0.000	0.983	-0.010	1.290	0.256	-0.010	0.820	0.367
Caudate (r)	-0.003	0.210	0.646	0.003	0.180	0.674	-0.010	3.110	0.078	-0.010	2.260	0.133
Central opercular cortex (l)	0.003	0.360	0.548	-0.010	3.360	0.067	0.010	2.470	0.116	0.002	0.210	0.647
Central opercular cortex (r)	-0.004	0.650	0.419	-0.010	4.090	0.043	0.003	0.310	0.576	-0.010	1.860	0.173
Anterior cingulate gyrus (l)	0.000	0.000	0.986	-0.010	1.000	0.318	0.002	0.110	0.738	0.002	0.100	0.748
Anterior cingulate gyrus (r)	-0.010	1.680	0.196	-0.010	2.430	0.119	-0.003	0.160	0.689	-0.004	0.530	0.468
Posterior cingulate gyrus (l)	0.001	0.040	0.852	-0.001	0.060	0.802	0.004	0.830	0.364	0.004	0.570	0.449
Posterior cingulate gyrus (r)	0.004	0.660	0.417	0.003	0.220	0.640	0.010	1.680	0.195	0.010	2.820	0.093
Crus I cerebellum (l)	0.010	1.280	0.258	0.010	1.410	0.235	0.010	2.110	0.146	0.020	17.390	0.000*
Crus I cerebellum (r)	0.010	1.000	0.318	0.010	3.320	0.068	0.010	5.250	0.022	0.030	26.070	0.000
Crus II cerebellum (l)	0.020	8.670	0.003	0.010	1.150	0.283	0.010	1.380	0.240	0.020	9.280	0.002
Crus II cerebellum (r)	0.010	4.260	0.039	0.003	0.240	0.621	0.010	1.700	0.193	0.020	6.480	0.011
Cuneal cortex (l)	0.000	0.004	0.950	-0.010	1.240	0.265	0.001	0.060	0.811	-0.003	0.340	0.561
Cuneal cortex (r)	0.001	0.040	0.843	-0.001	0.005	0.942	0.010	1.130	0.287	0.004	0.480	0.486
Frontal medial cortex (l)	0.010	4.360	0.037	0.010	1.170	0.279	0.010	0.940	0.331	0.010	3.640	0.057
Frontal medial cortex (r)	0.010	0.820	0.364	0.003	0.280	0.595	-0.004	0.370	0.545	0.004	0.390	0.535
Frontal operculum cortex (l)	0.003	0.350	0.552	-0.010	0.770	0.382	-0.004	0.670	0.413	0.010	1.000	0.317
Frontal operculum cortex (r)	0.010	3.420	0.064	0.005	0.490	0.482	0.004	0.550	0.457	0.010	3.660	0.056

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes. (*continued*)

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Frontal orbital cortex (l)	-0.003	0.390	0.531	-0.003	0.370	0.542	0.010	4.200	0.041	0.010	1.240	0.266
Frontal orbital cortex (r)	0.004	0.580	0.448	-0.003	0.360	0.551	0.010	1.350	0.245	0.010	2.240	0.134
Frontal pole (l)	0.010	2.320	0.128	0.001	0.070	0.789	0.003	0.570	0.449	0.010	3.160	0.075
Frontal pole (r)	0.010	2.480	0.115	0.000	0.001	0.977	0.002	0.120	0.729	0.010	4.000	0.045
Heschl's gyrus (l)	0.000	0.000	0.989	0.004	0.450	0.503	0.010	1.400	0.236	0.010	2.560	0.110
Heschl's gyrus (r)	0.020	7.450	0.006	-0.004	0.350	0.553	0.002	0.220	0.637	0.010	3.810	0.051
Hippocampus (l)	-0.003	0.410	0.524	0.003	0.350	0.556	-0.004	0.540	0.463	-0.010	0.910	0.339
Hippocampus (r)	0.003	0.300	0.583	0.003	0.300	0.581	-0.003	0.340	0.559	-0.003	-0.400	0.528
I-IV cerebellum (l)	-0.010	1.950	0.163	0.010	1.180	0.280	0.004	0.590	0.443	0.010	5.590	0.018
I-IV cerebellum (r)	-0.010	1.870	0.172	0.003	0.160	0.686	0.010	1.830	0.176	0.010	3.070	0.080
IX cerebellum (l)	0.000	0.003	0.954	0.010	2.460	0.117	0.003	0.290	0.589	0.020	7.640	0.006
IX cerebellum (r)	0.001	0.040	0.840	0.010	3.300	0.069	0.000	0.004	0.952	0.020	7.210	0.007
Inferior frontal gyrus. po (l)	0.010	2.450	0.118	0.005	0.500	0.481	-0.004	0.410	0.523	0.010	1.800	0.180
Inferior frontal gyrus. po (r)	0.004	0.420	0.510	0.010	1.110	0.291	0.002	0.100	0.757	0.010	0.980	0.321
Inferior frontal gyrus. pt (l)	0.000	0.001	0.978	0.003	0.180	0.671	0.003	0.250	0.618	0.010	0.710	0.399
Inferior frontal gyrus. pt (r)	0.010	0.640	0.422	0.000	0.001	0.972	0.004	0.430	0.511	0.005	0.529	0.467
Anterior inferior temporal gyrus (l)	0.010	0.790	0.375	0.000	0.000	0.999	0.002	0.080	0.776	-0.004	0.374	0.541
Anterior inferior temporal gyrus (r)	0.002	0.150	0.698	-0.010	3.290	0.070	-0.010	4.130	0.042	-0.010	3.610	0.057
Posterior inferior temporal gyrus (l)	-0.010	1.550	0.213	0.004	0.500	0.479	0.002	0.110	0.742	-0.002	0.150	0.698
Posterior inferior temporal gyrus (r)	-0.003	0.280	0.597	-0.010	1.780	0.182	-0.010	1.730	0.189	-0.010	6.210	0.013
Temporooccipital inferior temporal gyrus (l)	-0.010	1.240	0.266	0.002	0.130	0.713	0.010	3.830	0.050	0.001	0.040	0.836
Temporooccipital inferior temporal gyrus (r)	-0.003	0.290	0.590	-0.002	0.120	0.729	0.010	1.770	0.184	0.004	0.620	0.432
Insular cortex (l)	0.003	0.340	0.558	-0.002	0.120	0.730	0.002	0.130	0.715	0.010	3.510	0.061

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes. (*continued*)

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Insular cortex (r)	0.000	0.000	0.999	0.003	0.290	0.588	0.003	0.510	0.474	0.004	0.770	0.382
Intracalcarine cortex (l)	-0.001	0.040	0.843	0.010	1.160	0.281	0.003	0.200	0.653	0.010	1.270	0.259
Intracalcarine cortex (r)	0.001	0.040	0.847	0.010	3.080	0.079	0.010	2.320	0.127	0.010	3.050	0.080
Juxtapositional lobule cortex (l)	0.001	0.050	0.828	-0.001	0.120	0.724	-0.004	0.520	0.471	0.004	0.360	0.550
Juxtapositional lobule cortex (r)	-0.010	0.750	0.386	-0.002	0.060	0.799	0.002	0.150	0.713	0.002	0.140	0.704
Inferior lateral occipital cortex (l)	-0.001	0.020	0.902	-0.001	0.060	0.808	0.020	10.020	0.002	0.010	2.550	0.111
Inferior lateral occipital cortex (r)	-0.004	0.720	0.397	0.004	0.380	0.538	0.020	13.420	0.000*	0.010	5.370	0.020
Superior lateral occipital cortex (l)	-0.005	0.880	0.348	0.010	1.690	0.194	0.010	2.250	0.133	0.010	3.360	0.067
Superior lateral occipital cortex (r)	-0.010	1.880	0.180	0.020	10.290	0.001	0.008	3.000	0.083	0.010	3.020	0.082
Lingual gyrus (l)	0.010	2.930	0.087	0.001	0.010	0.909	-0.005	1.020	0.312	0.003	0.370	0.543
Lingual gyrus (r)	0.010	1.790	0.181	-0.010	1.440	0.231	0.003	0.400	0.525	0.002	0.220	0.650
Middle frontal gyrus (l)	0.002	0.200	0.657	0.010	1.880	0.171	-0.002	0.090	0.762	0.004	0.560	0.456
Middle frontal gyrus (r)	0.004	0.570	0.449	0.002	0.160	0.694	-0.004	0.560	0.456	-0.002	0.100	0.755
Anterior middle temporal gyrus (l)	0.001	0.010	0.929	0.001	0.010	0.943	0.010	3.390	0.066	0.003	0.300	0.581
Anterior middle temporal gyrus (r)	-0.001	0.010	0.939	-0.010	1.310	0.252	0.001	0.070	0.794	0.000	0.000	0.998
Posterior middle temporal gyrus (l)	0.010	3.950	0.047	-0.003	0.220	0.641	0.010	4.150	0.042	0.010	2.590	0.107
Posterior middle temporal gyrus (r)	0.010	0.920	0.336	0.010	1.160	0.280	0.010	2.020	0.155	0.010	3.280	0.070
Temporooccipital middle temporal gyrus (l)	0.003	0.290	0.589	-0.010	0.660	0.416	0.020	7.420	0.006	0.010	2.350	0.125
Temporooccipital middle temporal gyrus (r)	0.010	2.310	0.128	0.010	0.720	0.397	0.010	1.470	0.226	0.010	3.510	0.061
Occipital fusiform gyrus (l)	0.004	0.530	0.468	0.010	1.260	0.261	0.010	1.620	0.203	0.010	0.900	0.342
Occipital fusiform gyrus (r)	0.010	1.040	0.308	0.005	0.640	0.424	0.010	2.700	0.100	0.010	2.630	0.105
Occipital pole (l)	-0.010	1.480	0.224	0.010	0.970	0.325	0.010	6.430	0.011	0.010	5.450	0.020
Occipital pole (r)	-0.010	2.230	0.136	-0.002	0.070	0.795	0.010	8.450	0.004	0.010	2.200	0.138

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes. (*continued*)

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Pallidum (l)	0.010	1.420	0.233	0.010	0.910	0.341	0.020	9.140	0.003	0.020	6.830	0.009
Pallidum (r)	0.010	2.160	0.142	0.010	2.340	0.126	0.020	15.100	0.000*	0.020	7.270	0.007
Paracingulate gyrus (l)	0.010	5.13	0.024	0.010	2.430	0.119	0.004	0.710	0.398	0.002	0.240	0.627
Paracingulate gyrus (r)	0.010	8.110	0.004	0.002	0.200	0.655	0.004	0.700	0.404	0.010	3.440	0.064
Anterior parahippocampal gyrus (l)	0.002	0.090	0.767	0.004	0.460	0.500	0.000	0.001	0.976	0.010	1.100	0.295
Anterior parahippocampal gyrus (r)	0.010	1.850	0.173	0.001	0.040	0.843	0.005	0.810	0.368	0.010	1.270	0.259
Posterior parahippocampal gyrus (l)	-0.001	0.050	0.823	0.001	0.050	0.824	0.020	7.970	0.005	0.010	0.920	0.338
Posterior parahippocampal gyrus (r)	0.010	2.850	0.092	-0.003	0.180	0.671	0.010	5.160	0.023	0.003	0.290	0.591
Parietal operculum cortex (l)	-0.002	0.090	0.761	0.001	0.050	0.819	-0.002	0.120	0.733	0.005	0.830	0.361
Parietal operculum cortex (r)	0.004	0.490	0.485	0.000	0.001	0.977	0.010	0.990	0.320	0.010	3.200	0.074
Planum polare (l)	0.010	2.120	0.145	0.000	0.001	0.974	-0.002	0.080	0.778	0.004	0.045	0.502
Planum polare (r)	-0.010	1.350	0.246	-0.010	0.980	0.323	-0.002	0.410	0.521	0.002	0.100	0.756
Planum temporale (l)	-0.004	0.610	0.435	0.002	0.130	0.722	-0.004	0.510	0.475	0.002	0.160	0.687
Planum temporale (r)	0.005	0.740	0.390	-0.01	4.430	0.035	0.000	0.001	0.976	0.010	0.845	0.358
Postcentral gyrus (l)	0.010	2.060	0.151	0.005	0.670	0.414	0.010	8.030	0.005	0.010	1.020	0.312
Postcentral gyrus (r)	-0.001	0.010	0.904	-0.001	0.050	0.817	0.010	2.100	0.147	-0.010	1.860	0.170
Precentral gyrus (l)	0.010	1.370	0.242	-0.001	0.020	0.899	0.002	0.230	0.633	0.005	0.780	0.376
Precentral gyrus (r)	0.003	0.380	0.536	-0.001	0.030	0.866	0.004	0.470	0.494	0.010	1.630	0.202
Precuneous cortex (l)	0.001	0.010	0.921	0.004	0.500	0.480	0.010	2.020	0.156	0.010	3.620	0.057
Precuneous cortex (r)	-0.002	0.090	0.763	0.01	0.850	0.356	0.010	2.400	1.122	0.010	3.890	0.049
Putamen (l)	-0.002	0.100	0.750	0.005	0.470	0.492	-0.020	8.540	0.003	-0.010	1.710	0.191
Putamen (r)	-0.001	0.010	0.910	0.0003	0.002	0.963	-0.010	5.420	0.020	-0.003	0.180	0.667
Subcallosal cortex (l)	0.001	0.050	0.821	0.004	0.500	0.481	0.003	0.480	0.487	0.010	8.150	0.004

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes. (*continued*)

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Subcallosal cortex (r)	0.001	0.030	0.865	0.004	0.630	0.426	-0.002	0.190	0.665	0.010	3.890	0.048
Superior frontal gyrus (l)	0.004	0.580	0.448	-0.010	0.990	0.321	0.010	1.150	0.284	0.004	0.590	0.441
Superior frontal gyrus (r)	0.010	2.800	0.094	-0.004	0.480	0.488	0.004	0.580	0.446	0.010	0.800	0.371
Superior parietal lobule (l)	0.010	2.350	0.125	0.010	1.840	0.174	0.010	1.070	0.301	0.010	2.890	0.089
Superior parietal lobule (r)	-0.003	0.290	0.591	0.010	0.660	0.418	0.010	1.840	0.175	0.010	3.380	0.066
Anterior superior temporal gyrus (l)	0.001	0.010	0.908	0.000	0.001	0.971	-0.001	0.010	0.904	-0.001	0.010	0.919
Anterior superior temporal gyrus (r)	-0.003	0.290	0.590	0.000	0.000	0.993	-0.003	0.240	0.624	0.010	1.850	0.174
Posterior superior temporal gyrus (l)	0.010	0.820	0.364	-0.010	3.280	0.070	0.010	0.930	0.334	0.004	0.550	0.458
Posterior superior temporal gyrus (r)	0.010	3.290	0.070	-0.002	0.130	0.722	0.010	2.370	0.123	0.010	4.340	0.037
Supracalcarine cortex (l)	0.010	1.030	0.310	0.004	0.390	0.530	0.001	0.040	0.848	0.001	0.030	0.873
Supracalcarine cortex (r)	0.004	0.530	0.467	0.010	4.300	0.038	0.010	1.670	0.196	0.010	3.430	0.064
Anterior supramarginal gyrus (l)	-0.010	1.880	0.170	0.000	0.000	0.986	0.010	4.620	0.032	0.002	0.100	0.751
Anterior supramarginal gyrus (r)	0.010	1.460	0.227	0.005	0.480	0.488	0.010	1.900	0.168	0.010	1.450	0.228
Posterior supramarginal gyrus (l)	-0.001	0.010	0.919	0.003	0.260	0.613	0.004	0.364	0.546	-0.010	1.750	0.186
Posterior supramarginal gyrus (r)	0.005	0.560	0.453	-0.003	0.170	0.682	0.010	1.110	0.292	0.010	3.570	0.059
Anterior temporal fusiform cortex (l)	0.010	2.580	0.108	-0.004	0.410	0.522	-0.001	0.060	0.813	0.004	0.510	0.474
Anterior temporal fusiform cortex (r)	-0.002	0.090	0.759	0.001	0.020	0.894	0.010	1.080	0.298	0.010	2.800	0.090
Posterior temporal fusiform cortex (l)	-0.010	4.870	0.027	0.003	0.270	0.601	0.004	0.680	0.409	0.010	0.870	0.351
Posterior temporal fusiform cortex (r)	-0.010	1.760	0.184	-0.010	0.880	0.347	0.000	0.002	0.963	-0.010	1.040	0.308
Temporal occipital fusiform cortex (l)	0.010	4.080	0.042	0.020	9.390	0.002	-0.002	0.130	0.713	0.010	4.130	0.042
Temporal occipital fusiform cortex (r)	0.002	0.150	0.703	0.002	0.140	0.707	0.005	0.730	0.392	0.010	2.430	0.119
Temporal pole (l)	0.010	2.580	0.108	-0.010	1.930	0.164	0.010	6.900	0.009	0.010	4.720	0.030
Temporal pole (r)	0.010	3.510	0.061	0.010	1.020	0.312	0.003	0.400	0.527	0.010	1.070	0.301

Supplementary Table 1. Complete overview of results from the liner models examining the associations between social complexity measures and grey matter volumes. (*continued*)

Measure												
IDP	Social contexts B	Social contexts F	Social contexts p	Household number B	Household number F	Household number p	Friends/ family visits B	Friends/ family visits F	Friends/ family p	Composite social complexity B	Composite social complexity F	Composite social complexity p
Thalamus (l)	0.003	0.220	0.628	0.030	6.290	0.012	-0.004	0.460	0.500	0.010	2.060	0.151
Thalamus (r)	0.000	0.005	0.944	0.010	1.500	0.221	-0.010	2.010	0.156	0.001	0.040	0.849
V cerebellum (l)	0.002	0.110	0.740	0.010	0.920	0.337	0.003	0.230	0.632	0.010	4.970	0.026
V cerebellum (r)	0.002	0.110	0.744	-0.004	0.390	0.530	-0.001	0.030	0.871	0.003	0.350	0.554
VI cerebellum (l)	0.010	4.310	0.038	0.004	0.360	0.551	-0.004	0.560	0.455	0.010	1.980	0.160
VI cerebellum (r)	0.010	3.930	0.047	0.001	0.030	0.863	-0.001	0.060	0.803	0.010	1.500	0.221
VIIIa cerebellum (l)	0.010	5.030	0.025	0.010	1.670	0.196	-0.001	0.030	0.860	0.020	8.890	0.003
VIIIa cerebellum (r)	0.010	6.240	0.012	0.010	2.220	0.137	0.003	0.370	0.545	0.020	14.450	0.000*
VIIIb cerebellum (l)	0.005	0.600	0.437	0.010	4.430	0.035	0.001	0.040	0.839	0.020	12.040	0.0005*
VIIIb cerebellum (r)	0.010	0.800	0.370	0.020	8.510	0.004	0.003	0.220	0.637	0.030	19.590	0.000*
VIIb cerebellum (l)	0.020	7.290	0.007	-0.002	0.100	0.757	-0.010	0.940	0.333	0.010	2.520	0.113
VIIb cerebellum (r)	0.010	5.890	0.015	0.000	0.002	0.964	-0.001	0.010	0.925	0.010	5.160	0.023
Ventral striatum (l)	-0.003	0.330	0.564	0.010	1.540	0.214	-0.010	1.980	0.160	-0.002	0.140	0.709
Ventral striatum (r)	-0.002	0.110	0.742	0.002	0.070	0.785	-0.010	6.370	0.012	-0.010	1.960	0.162
X cerebellum (l)	0.010	3.230	0.072	0.010	2.720	0.100	0.010	2.500	0.114	0.010	2.280	0.131
X cerebellum (r)	0.010	1.280	0.258	0.002	0.110	0.738	0.010	3.660	0.056	0.010	2.630	0.105

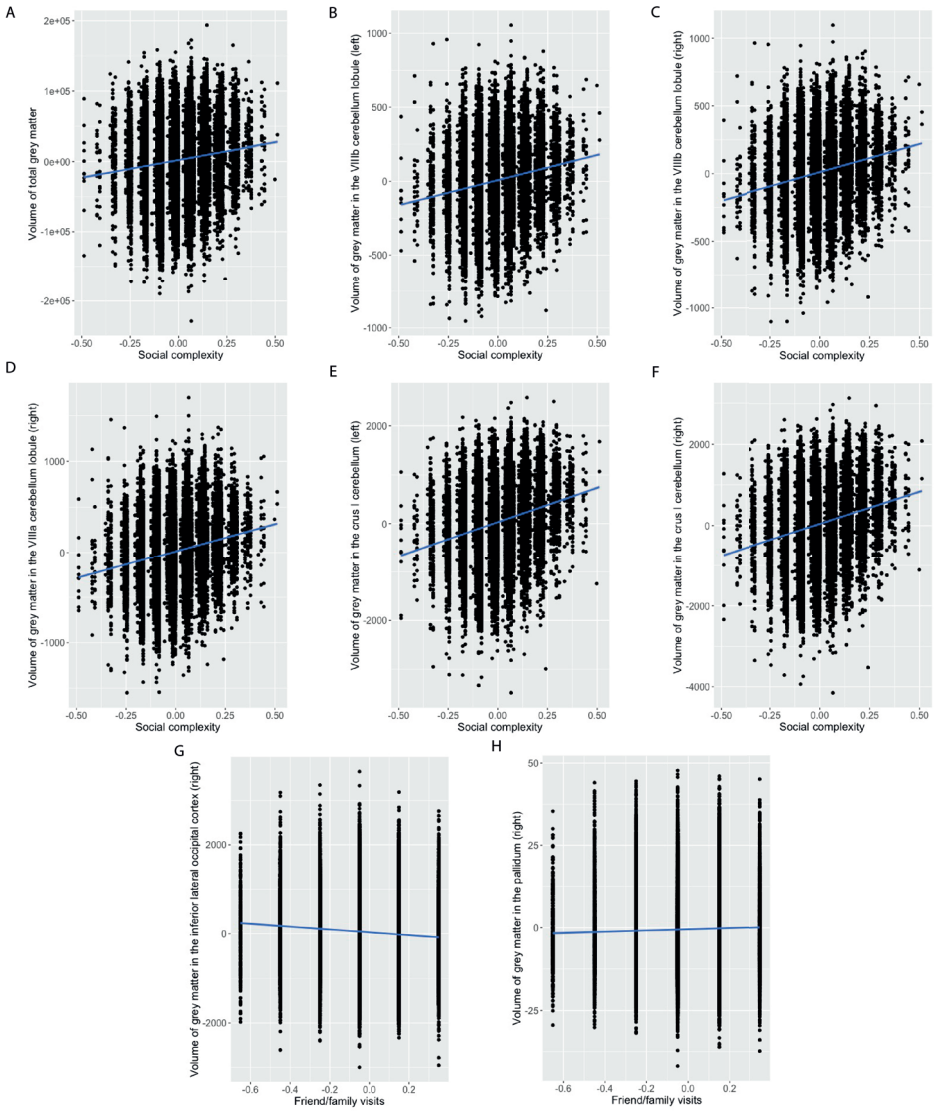
* = significant after Bonferroni correction

Supplementary Table 2. Associations between demographics/BMI with social complexity measures.

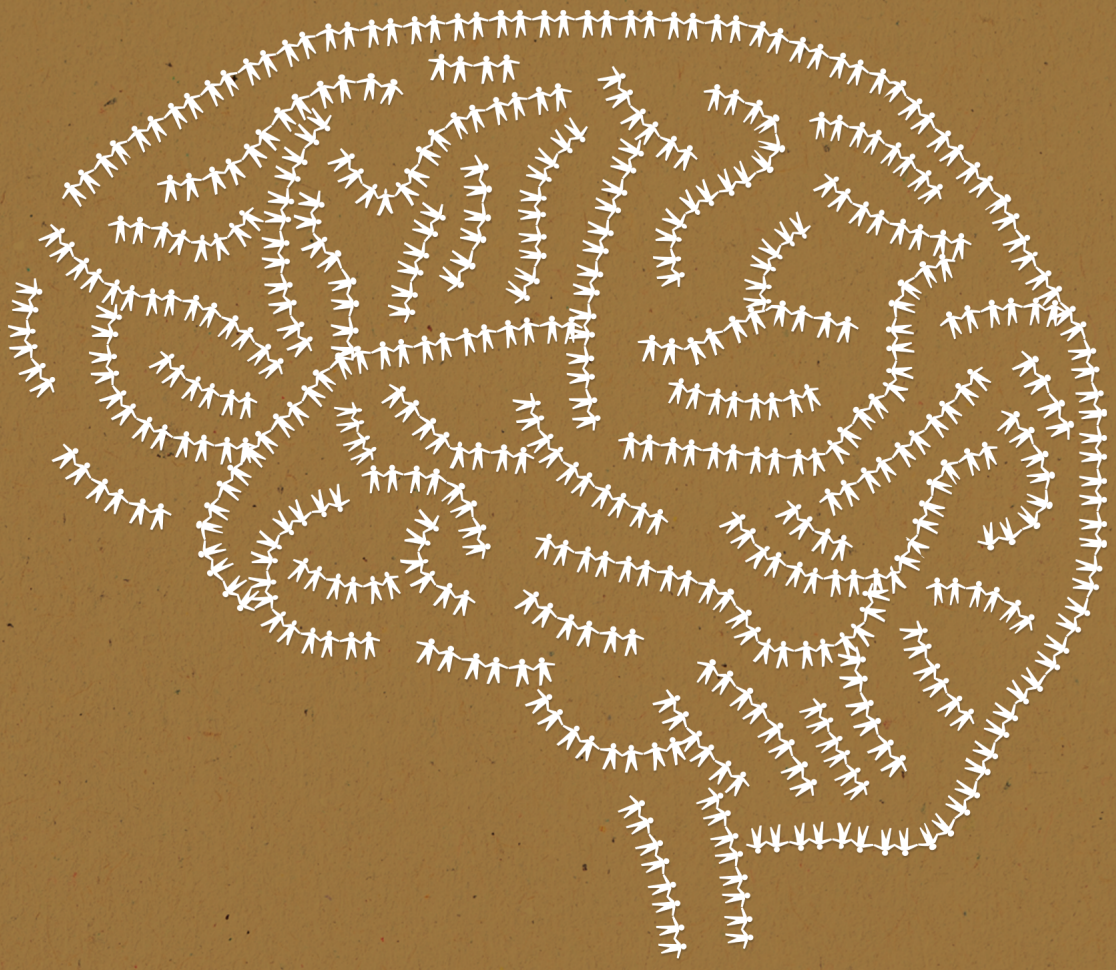
	Social contexts ES[1]	Social contexts F	Social contexts p	Household number ES[1]	Household number F	Household number p	Friends// family visits ES[1]	Friends/family visits F	Friends/family visitsy p	Composite social complexity ES[1]	Composite social complexity F	Composite social complexity p
Age	-0.30[2]	2957.000	< .001*	-0.36[2]	4464.700	< .001*	0.12[2]	425.960	< .001*	-0.29[2]	2679.000	< .001*
Sex	0.07[2]	135.220	< .001*	-0.07[2]	139.240	< .001*	0.14[2]	620.200	< .001*	0.09[2]	241.520	< .001*
Ethnicity	< 0.01[3]	5.580	< .001*	0.01[3]	22.200	< .001*	0.01[3]	12.890	< .001*	< 0.01[3]	5.830	< .001*
Education	-0.16[2]	766.370	< .001*	-0.06[2]	98.590	< .001*	0.05[2]	66.650	< .001*	-0.07[2]	143.810	< .001*
Income	-0.20[2]	1101.600	< .001*	0.41[2]	5302.100	< .001*	-0.10[2]	254.830	< .001*	0.24[2]	1697.600	< .001*
BMI	-0.02[2]	15.590	< .001*	0.01[2]	0.760	0.383	0.02[2]	7.750	0.005*	0.01[2]	4.020	0.045*

[1] ES = effect size. [2] Effect size measured in beta. [3] Effect size measured in eta squared. * Significant at alpha = 0.05.

Supplementary Figure 1



Supplementary Figure 1. Associations between predicted total and regional grey volumes retrieved from the linear models. Plots A through F show associations between the IDPs and the composite social complexity score, plots G and H show associations between IDPs and the number of friend and family visits. Blue lines show the regression lines.



Chapter 4

Assessing data collection and data quality in passive smartphone behavioural monitoring

Mila C. Roozen[1*], Raj J. Jagesar[1*], Jacob A. Vorstman[2,3#], and Martien J. Kas[1,#]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands

[2]Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, Canada

* *shared first authors*

shared last authors

ABSTRACT

Background: Literature on digital phenotyping is rapidly expanding with the development of smartphone applications for active and passive monitoring of human behaviour. Much of the interest in the field goes out to their promising potential to capture clinically relevant outcome measures. This is understandable given the excitement around the measurement potential of this novel set of tools. However, as for all novel research tools, it is crucial that all methodological aspects are carefully evaluated. Regarding smartphone-based passive monitoring, a particular challenge relates to the ability to maintain stable application performance as well as completeness and accuracy of data collected. This is important because these raw data are the basis from which any behavioural outcome measure is derived. Hence, establishing adequate practices for the collection and use of raw data is an essential first step required towards the reliability of scientific output generated downstream.

Objective: To explore aspects of passive sensing data collection and data quality by comparing smartphone based behavioural data with alternative methods for the measurement of mobility, calling behaviour, foreground app usage and Wi-Fi access point sensing.

Methods: The study is conducted using BEHAPP, a mobile passive monitoring platform. First, we outline the design and development challenges of such platforms. Second, we compare between three different methods of measuring human behaviour: BEHAPP, self-reports and independent GPS tracking from 22 participants over a period of ten consecutive days.

Results: In general levels of agreement for location data and foreground app usage points were adequate. Wi-Fi access points did not match well between BEHAPP and the questionnaire.A more thorough examination of the Wi-Fi data showed large differences in the number of access points at single locations. The mobility features showed large differences regarding the measurement of trajectories between the measurement modalities. Missing data affected all data types to some extent, and gaps in data shows that not all measurements were gathered as programmed.

Discussion: The results elucidate several important aspects of data collection and quality to optimize the reliability of scientific results using passive sensing methodologies. The agreement between methods indicates that mobile passive monitoring has the potential to measure real-time and real-world behaviour. The missingness of data from different modalities highlights the importance of data quality control measures. Passive smartphone monitoring offers both additional benefits such as real-time longitudinal and high-resolution data, expanding our opportunities to measure human behaviour for research and, potentially, clinical applications.

INTRODUCTION

Digital phenotyping is a scientific field which is rapidly amassing large amounts of scientific studies. A definition of digital phenotyping was offered by Torous et al. (2016), specifically as "moment-by-moment quantification of the individual-level human phenotype *in situ* using data from smartphones and other personal digital devices". In other words, digital phenotyping refers to the continuous ("moment-by-moment") measurement of phenotypes of individuals using smartphones or other types of personal digital devices. Passive smartphone behavioural monitoring is one specific type of digital phenotyping which uses sensors and logs from the smartphone without active input, which can be distinguished from other methods which rely on additional digital devices such as wearables or require active input from the user (for example, through digitally delivered surveys).

Digital phenotyping has several advantages over traditional methods such as surveys or clinical examinations, which explain why the field is expanding at its current pace. First, the continuous nature of data collection, not limited to fixed moments such as is the case with surveys and clinical measurements, allows for a more complete image of the behaviour of the participant over a prolonged period of time, which is less likely to be affected momentary circumstances at the moment of measurement itself (although longer lasting circumstances, such as vacations, might still affect continuous measurement). Secondly, digital phenotyping methodologies take place in real world conditions, providing a more accurate overview of the daily lives of subjects. These first two advantages of digital phenotyping can be summarized as their ability to generate real-time and real-world data. Thirdly, one particular method of digital phenotyping relies on passive monitoring, which means that no active input from the subject is required. This particular approach has additional advantages in that it is objective, reduces input biases or errors such as social desirability bias, recall bias or malingering. One example which shows possible advantages is the study by Toepoel et al. (2020), where data gathered through passive GPS measurements improved the detection of short travelling instances.

The passive aspect also reduces the burden on both subjects and researchers, as they are not required to spend time performing the measurement. In his treatise on the possibilities of and challenges for digital phenotyping, Insel (2018) describes how the mentioned benefits of digital phenotyping tools can change the approach to mental health research and practice, for example by allowing early detection of many different types of disorders by long-term monitoring of high-risk groups and by making possible the prevention of relapse for those already in care. The paper by Sugle (2016) demonstrates how smartphone-based measurement can have a positive effect on the participation rates in hard-to-reach populations and can assist in detecting rare and/or temporary occurrences.

The number of studies in the field of digital phenotyping is increasing fast. Where a Google Scholar search using the search terms 'digital phenotyping' limited to 2016, the year Torous et al. (2016) provided their definition, yields 3780 results, the same search for the year 2020 yields 10200. Several of these studies have already provided interesting results. For example, Berrouiguet et al. (2018) demonstrate in a sample of 5 patients with mood disorders that data passively gathered from smartphones using GPS can be used to detect changes in mobility patterns, which would require either large amounts of questionnaires to record or would require subjects to recall their behaviour during long periods of time. Doryab et al. (2019) use a smartphone application and a wearable sensor to measure a range of variables including communication and phone usage, activity and mobility and sleeping behaviours to predict loneliness. With only minimal burden for researchers and subjects, the data from 160 subjects measured for 16 weeks could be used to predict loneliness with 88.4% accuracy. As a final example, Barnett et al. (2018) analysed long-term measurements of 15 patients suffering from schizophrenia using an algorithm which can detect behavioural changes over time and found that for those experiencing relapse, the rate of behavioural changes increased by 71% in the two weeks leading up to the relapse. More examples can be found for example in Torous et al. (2021), who provide a very comprehensive overview of many different ways in which digital phenotyping methods can be used in mental health research.

Digital phenotyping, including passive monitoring, has opened the door for longitudinal real-world and real-time measurements of large numbers of variables and is saleable to large sized cohorts. Considering the high stakes involved in the assessment of mental health, data quality is essential. In this regard, it is somewhat surprising that research on data quality aspects of digital phenotyping methods is scarce. In fact, none of the three studies cited above mention data quality assessment, and only one specifies how missing GPS data was addressed (Barnett et al. (2018)). Recently, Bähr et al. (2020) has specifically examined data quality of GPS data collected through smartphone sensors, and demonstrate proneness of this method for missing data and invalid coordinates. Their results demonstrate that many different causes of GPS data errors exist, suggesting the importance of data quality assessment for any research relying on smartphone-based (GPS) data collection.

Furthermore, there are broader factors are at play which may negatively affect the data collection and quality. First, mobile operating systems such as Android and iOS change continuously. Adjustments to regulations regarding privacy and battery consumption limit developers in their ability to access certain data sources or require them to explicitly specify the resolution, frequency and continuity at which they need access to these data sources. Examples of limitations are restrictions regarding access to sensors such as the accelerometer and microphone (*Android Pie (9.0): Behavior Changes*, 2019) and imposed limits to the frequency at which location updates are delivered to any third party app (*Android Oreo (8.0): Behavior Changes*, 2020),

specifically when apps are running in the background of a device (see Supplementary table 1 for a full overview of Android platform changes with an impact on data collection and data accuracy). Measures that specifically target background apps are particularly difficult because, inherent to passive monitoring of behaviour, they require unobtrusive and *backgrounded* operation of mobile applications. Regarding iOS, strict restrictions are in place limiting the types of data available for passive sensing, reducing the number of behaviours which can be measured.

Second, in line with the aforementioned, smartphone manufacturers often implement strategies to conserve battery consumption leading to diminished sample frequencies or, in some cases, complete inactivation of the app. These measures even cause app failures amongst apps that are fully supported by Google such as coronavirus tracing apps (Hofmans, 2020). Third and last, user error can also cause problems with data collection and quality. Specific to sensors or sensor applications, participants may unintentionally cause the data to be inaccurate or incomplete while the data collection continues, making it difficult for the researchers to assess the extent of data loss. For example, data loss can occur because users accidentally retract data access permission for a specific sensor modality mid-study or fail to keep the device on their person.

With regard to traditional survey research, previous studies have examined the predictors of data quality and quantity and ways to address these appropriately (e.g. Shin et al (2011), Mavletova (2013), & Meterko et al. (2018)), and some of the insights derived from this field are likely informative for digital phenotyping methods. For example, Amaya et al. (2020) described a so-called "total survey error framework" to categorize error types in large scale data gathered from third parties (data gathered by other entities than the researchers analysing the data, such as data gathered by governments). Although this type of data has some differences compared to digital phenotyping data, the steps from the point of data gathering onwards are comparable. Therefore, several of the error types mentioned in this paper may also occur in digital phenotyping, suggesting the relevance of evaluating such data errors in studies relying on digital phenotyping.

The goal of this study is to examine the data quantity and quality of three methods compared to each other: a passive smartphone monitoring application called BEHAPP, a questionnaire and an independent GPS sensor. We want to examine the agreement between the methods in order to assess their advantages, disadvantages and interchangeability for the measurement of behaviour.

ABOUT BEHAPP

The data for this study was collected using BEHAPP, a smartphone based behavioural monitoring platform implemented in multiple scientific studies (Jagesar et al. 2021).

BEHAPP shares the same goal of collecting objective, real-world passive monitoring data with existing and emerging platforms such as Health by MindStrong (Insel, 2017), Beiwe (Torous et al., 2016), eB2 (Berrouiguet et al., 2018), NIIMA (Aledavood et al., 2017) and the AWARE framework (Ferreira et al., 2015). With four years of service BEHAPP is host to multiple studies with data collected (or ongoing data collection). The service is designed for scale, security and ease of use following the *software as a service* (SaaS) paradigm (Ma, 2007). Capable of hosting multiple studies, the responsibility of managing participants is delegated to study managers through the use of an administrative interface. The mobile application, which currently is Android only, guides the participant through a short onboarding process after which the app is activated and retreats to the background of the device for the set duration of the study.

Mobile app data collection

BEHAPP's primary mode of data collection is a mobile app specifically built for the Android platform. The app taps into various data sources of a participant's smartphone, selected based on their putative relevance to a participant's social behaviour in terms of smartphone behaviours and mobility. Smartphone behaviours covers activities in which participants interact with the smartphone or others through the smartphone and is captured by monitoring the frequency and duration app usage events.

Mobility is concerned with data expressive of how a participant moves around physically, such as GPS-based location data, but also Wi-Fi access point data. In our current study, we specifically examined collected raw data relevant to both smartphone behaviour and mobility, and meaningful features computed from this raw data from the following (and most commonly used) data sources mined by BEHAPP.

Location and mobility data based on GPS

Real-time location data is requested by the mobile application to record the mobility of a participant. BEHAPP implements the FusedLocationProvider which is a subsystem of Android responsible for handling the collection and dissemination of location data on Android based smartphones (*Simple, Battery-Efficient Location API for Android*, n.d.). The app requests an update from the FusedLocationProvider every 30 seconds or if the displacement of the smartphone is over 30 metres. However, the FusedLocationProvider only delivers a location point if the battery is sufficient, if a method of measuring location (GPS, Wi-Fi or cellular) is available and if the measurement does not return missing values for the latitude and longitude.

The app records geographical coordinates (latitude / longitude), speed, altitude and accuracy measures for each data point. The temporal resolution and accuracy of location data streams depend on factors such as battery savings measures and GPS satellite visibility, and thus may vary over time.

Wi-Fi Access Points

Wi-Fi access points (AP's) are fixed internet connected points that enable devices such as smartphones to connect to the internet. The data is collected by scanning for AP's in the vicinity of the device, this is done on a fixed interval of approximately ten minutes. Sensitive identifiers such as the BSSID / MAC address are obfuscated using a technique known as *one-way-hashing*, which uses a cryptographic method to encode the information provided meaning that if anyone with malicious intent gains entry to the database somehow the data does not convey meaningful and traceable information, thereby safeguarding privacy while preserving unicity. Considering the amount of Wi-Fi access points is dependent on the amount of Wi-Fi access point generating devices (such as modems) in an area, we examine it as a measurement of the social 'density' of the environment.

Foreground app usage

The use of mobile applications is monitored in real-time recording the name of the app and the duration of use in seconds. The data provides a broad perspective on device usage as well as in-depth app usage patterns on the level of app categories such as 'communication' and 'entertainments' apps and specific apps such as 'WhatsApp', 'Facebook' and 'YouTube'.

Security & privacy

BEHAPP is built with security in mind at each layer of the service. A *defense in depth* strategy combined with the principle of *least privileges* are applied to realize a redundant security structure minimizing the overall attack surface. The strategy is based on four pillars: 1) Data is to be handled in an encrypted state at all times; 2) Data may not be permanently stored in servers that can be reached over public networks; 3) Upon receiving data the data may only flow in one direction. From publicly connected servers to servers in private networks; 4) Every study has a separate database for which specific access controls can be set, this is further enforced by a segmented encryption hierarchy. On an organizational level, every researcher with sensitive data access is required to attend a privacy and security briefing outlining responsible use of their access credentials, hardware and participant data. The briefing is based on an information security policy requiring signed endorsement by every researcher.

METHODS

In order to examine how characteristics of data collection and outcomes are affected by the method of data collection, we conducted a study in which BEHAPP data were compared against data from independent, parallel sources informative on the same type of behaviours under study, i.e. activities regarding calling, texting, app usage, and patterns of location and mobility.

We used three independent, parallel sources for our study: 1) BEHAPP, 2) a GPS tracking device, 3) a questionnaire with questions regarding the participants' behaviour on the ninth day of participation. The questionnaire (see Appendix II) contains two subsets of questions.

The first subset is a diary, which prompts each participant to report the number of available Wi-Fi access points visible on their phone and to open an application and specify which app was opened. The diary probes each participant three times during day nine of the study, with a minimum time interval of three hours between probes. The second subset of the questionnaire consists of questions about other behaviours during day nine of the study, such as calling and texting behaviour.

Study design

The study consisted of two phases. In the first, data collection phase participants were included on-site and assisted with installation of the BEHAPP app configured to collect data over a period of ten days. At the same time, each participant was provided the Tractive GPS tracker along with user instructions. Participants were provided with charging equipment for the tracker and instructed to keep the GPS device (and their smartphone) with them for ten consecutive days. Additionally, participants were asked to fill in a questionnaire on the ninth day of the study, probing for specific behaviours during that day (Appendix II). Next in the second phase of the study we analysed the resulting data from the BEHAPP app, the GPS tracker and the questionnaire, with the aim to examine the degree of agreement between the three independent data sources.

Participants

Students from the University of Groningen were approached with the request to participate in this study. Other participants included employees of the Groningen Institute for Evolutionary Life Sciences (GELIFES), PhD students from several departments of the University of Groningen and acquaintances of the authors. Participants were provided with a form providing information on the study and asked for informed consent.

The BEHAPP application was installed on their own phone. In total, 22 participants were recruited for this study, of whom 54.5% were male. Device manufacturers were relatively equally distributed over participants with Samsung having a majority share.

Ethics

Data collection for this study did not require approval from the concerned institutional ethics review boards. Written informed consent was provided by all subjects.

Instrumentation

We used an independent GPS tracker from Tractive, n.d. (Tractive Classic, Pasching, Austria) as a control data source for location data. Tractive products are targeted towards pet owners to be able to examine the location and past locations of their pets. The Tractive Classic GPS tracker weighs 35 grams and measures at 51x41x15 (height x width x depth) mm (*Tractive Realtime GPS Tracking*, n.d.).

Measures

The first step in the analysis was to examine the extent of the data gathered by examining the number of days in which BEHAPP data was recorded by the participants' smartphones. After this, we started the comparison between the BEHAPP data and the data from the concurrent methods. We created specific measures for each of the data sources in comparison to BEHAPP data

1) For our questionnaire data we defined the measures 'number of measurements agreement in foreground app usage' and 'number of measurements agreement in Wi-Fi access points observed'. Regarding the first, we determined the extent of overlap between the activities self-reported by participants through the activity diaries, and the data collected by BEHAPP. To this end, we examined whether the apps recorded by participants in the diary part of the questionnaire coincided with the activated apps identified by BEHAPP at the recorded times. We used these features for different reasons. First, the average hourly sampling frequency gives an indication of the information density of the method in question. The features regarding places visited and time spent at home gives an indication of the variability of the environment of the subject. Finally, the total number of trajectories, time spent at stationary and perimeter of operation give an indication of the mobility of the subject.
2) Finally, we attempted to replicate part of the analysis of Bähr et al. (2020) by examining gaps in data over. As our location data is not gathered at continuous rates but the Wi-Fi data is, we used the Wi-Fi interval to determine gaps. We examined the presence gaps of at least 30 (3 times the interval of 10 minutes) minutes in the Wi-Fi data and the total data.

1 Unfortunately, the overlap in Wi-Fi categories was not noticed in time, meaning the data does contain the overlap. However, only two BEHAPP measurement of 5 Wi-Fi APs occurred. The participants scored it as 5-10 and the measurements have been treated as correct.

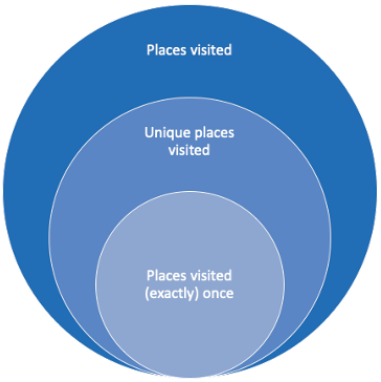


Figure 1: Visualization of the 'places visited' features. This figure shows that the features 'unique places visited' and 'places visited exactly once' are subsets of respectively 'places visited' and 'unique places visited'.

RESULTS

Of the 22 participants that were recruited for this study three participants did not fill in the questionnaire, bringing the questionnaire data total at 19 participants. Of those 19, one participant omitted the third measurement of Wi-Fi points and app usage on the questionnaire.

Number of days with BEHAPP data

19 out of 22 participants had BEHAPP data for at least 10 unique days (the maximum was 11 unique days, as the study duration of ten days always started somewhere during the first day, so 11 unique days were measured). Participants 1, 9 and 11 had nine, four and five unique days of BEHAPP data respectively. Each of these participants had data following the days with missing data. The days with missing data appeared randomly distributed and did not follow any consistent pattern across participants.

App usage

Table 1 shows the percentage of agreement between the self-reported foreground app usage and the same as recorded by BEHAPP. Average agreement on app usage was 65%. However, the spread is large with nine participants showing total (100%) agreement, whereas three participants' (participant numbers 9, 15 and 19) data showed no agreement. In one case (participant 12) there was no app data recorded at all. Average agreement excluding participant 12 was 69%.

Table 1: Agreement app-usage based on BEHAPP vs based on self-report.

	No agreement	1 measurement in agreement	2 measurements in agreement	3 measurements in agreement
Frequency	4*	2	4	9

* Participant 12 had no app data recorded by BEHAPP

Wi-Fi Access Point scans

Wi-Fi Access Point data also shows a wide range of agreement between BEHAPP and the participants' questionnaire data (Table 2). Four participants showed total agreement and seven participants showed zero agreement. Mean agreement is 42%. One participant had no BEHAPP-recorded Wi-Fi data at all (participant 9) and one of the participants (participant 19) had no BEHAPP-recorded Wi-Fi data on the day of the questionnaire. Average agreement excluding participants 9 and 19 was 47%.

Table 2: Agreement Wi-Fi access point quantity based on BEHAPP vs based on self-report.

	No agreement	1 out of 3 measurement in agreement	2 out of 3 measurements in agreement	100% agreement**
Frequency	7*	4	4	4

* Participants 9 and 19 had no Wi-Fi data recorded by BEHAPP on the day of the questionnaire

** Not all participants had three measurements in both the BEHAPP data and the questionnaire, however, they may still have full agreement in the numbers reported by BEHAPP and the questionnaire. None of the participants had 1 out of 2 measurements agreement, which is why the category is not included in the table.

Compared to the questionnaire, BEHAPP found higher values in 25 cases, lower values in 3 cases and equal values in 20 cases.

The agreement in general was quite poor. Participants 4, 10, 15, 17 and 21 in particular stood out. Despite complete and consistent data from both the questionnaire and BEHAPP, the agreement between both modalities was null. This prompted further investigation into the level of variation of Wi-Fi access points measured on one location, under the assumption that automatically reported Wi-Fi access points are likely (relatively) stable over time at a given location. However, contrary to our expectations, the number of available Wi-Fi access points recorded by BEHAPP on fixed geographical points varied largely over time, with a large variability of the spread of data between participants (Figure 2). Also contrary to our expectations, the number of access points found per scan was much higher than expected, which may also have affected the results of the comparison between the questionnaire and the application data, as BEHAPP often showed higher measurements compared to the questionnaire.

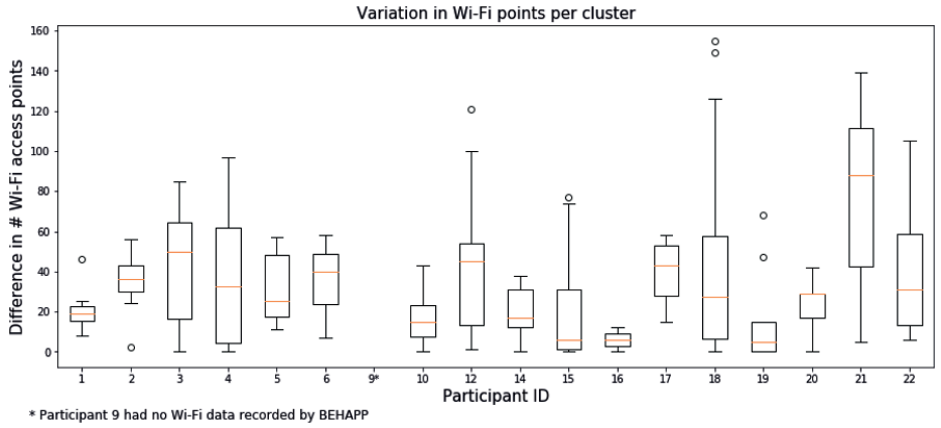


Figure 2: Variation in BEHAPP-reported number of Wi-Fi access points recorded at different times across locations. This figure shows the differences in amounts of Wi-Fi access points found across the places visited by the participants. The line within the boxes is the mean, the box contains all values within the interquartile range (the middle fifty percent) and the whiskers span between the first points within 1.5 times the IQR lower than the first quartile point up to the last point within 1.5 times the IQR higher than the third quartile point. The circles indicate outliers.

Sleep activity

The day nine questionnaire also included questions about when participants went to sleep the preceding night and when they woke up. As sleep is a period during which in theory the smartphone is not activated by the participant, we examined whether calls, apps and location trajectories were present in the BEHAPP data during the recorded sleeping period. Participants 5 and 21 were removed from this examination, as their recorded sleep times were very long (18.5 hours and 21.5 hours respectively) and for a large part during the day. In accordance with expectations, no calls or displacements were detected for any participants during their recorded sleeping period. However, several participants displayed app usage in the very first hour after the recorded sleeping time and in the very first hour prior to the recorded waking up time. In Figure 3, the yellow bars show the total amount of app usage during the recorded sleep time, with the orange bars displaying the same after removal of the first and last hour of the reported sleep.

GPS location data

Two participants had either insufficient BEHAPP (participant 9) or insufficient independent GPS tracking (participant 22) location data points and were thus excluded from the analysis of location data-based features. Of the 21 remaining participants, one person had fewer than ten unique days of independent GPS tracking location data (participant 13). In the BEHAPP data two participants had fewer than ten unique days of location data. Specifically, participants 1 and 11 had six and four unique days of data respectively. Location data measurement frequency differed between BEHAPP and the independent GPS tracker, with BEHAPP generally

recording more location data points per hour compared to the independent tracker (see Figure 4). The GPS tracker gathered data quite equally across participants while BEHAPP location sample frequency varied more between participants. When a participant remains stationary, the BEHAPP application gathers data somewhat sparsely, on average about 9.39 times per hour. When movement is noticed, location sampling is increased drastically, to on average 163.66 times per hour. The exact sample frequency may vary and is determined by the Android

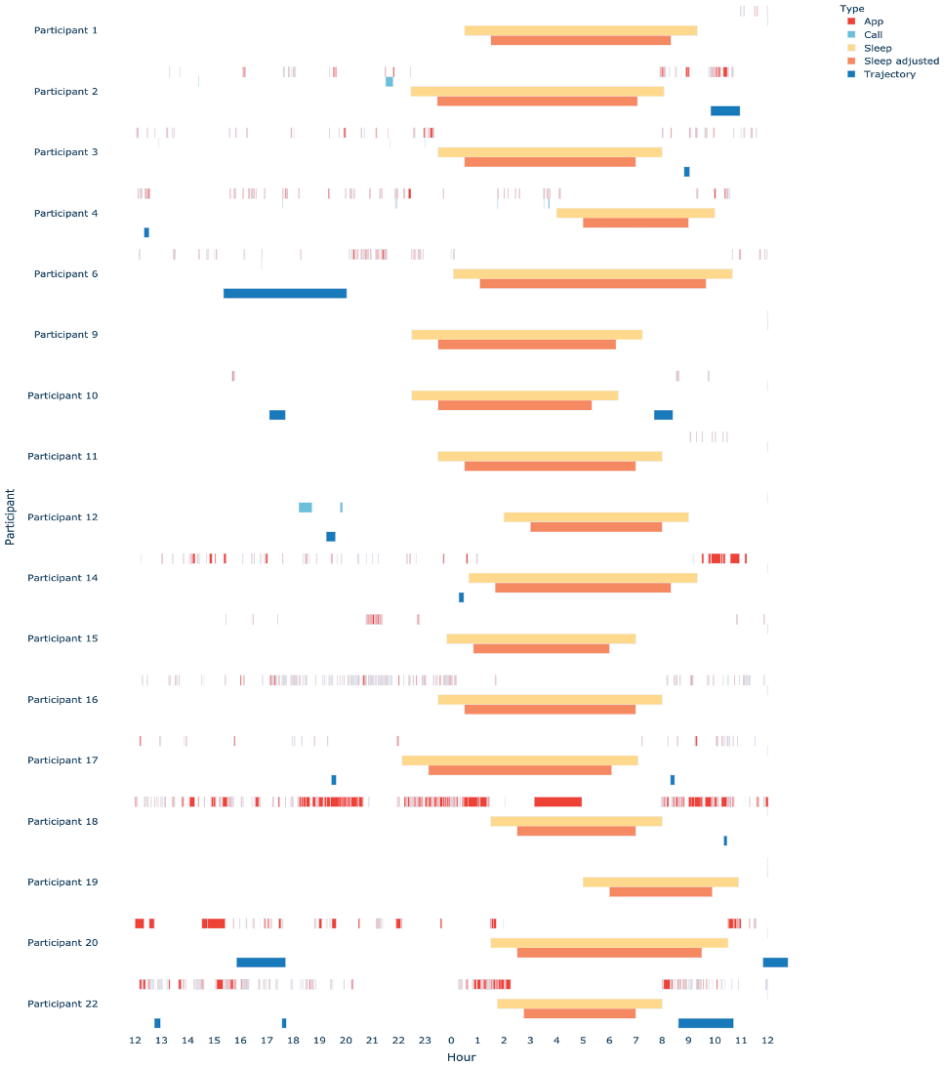


Figure 3: App (red), call (light blue), trajectory (dark blue), self-reported sleep (yellow) and adjusted self-reported sleep with the first and last hour removed (orange) plotted between 12 PM of the eighth day and 12 PM of the ninth day of the study. This graph illustrates the overlap time between self-reported sleep and activities suggesting wakefulness, and the improved fit of the data after removal of the first and last hours of reported sleep (orange).

operating system and specific manufacturer settings. In contrast, during moments when movement is minimal, Tractive based GPS trackers gather location data samples on average about 2.55 times per hour, with only small deviations from this interval. When significant movement is noticed relative to the original location, the GPS tracker increases the sampling frequency up to a maximum sample frequency of once every two minutes (mean hourly sampling frequency during trajectories = 9.07).

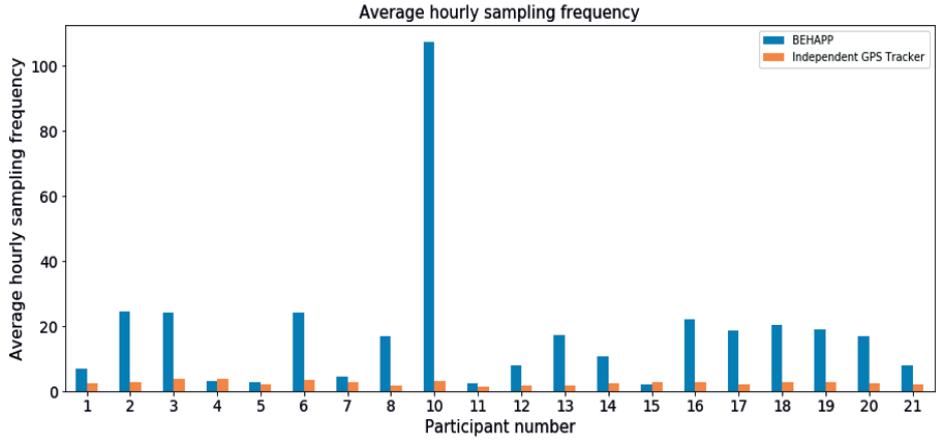


Figure 4: Average number of location data samples per hour for BEHAPP (blue) and the GPS tracker (orange). The figure shows the average of the amount of GPS samples per hour for all participants for the BEHAPP (blue) and independent GPS tracker (orange) data.

Places visited

Regarding the locations visited by participants, we looked at the total number of visited locations, the number of unique locations visited (ignoring repeated visits to the same location), the number of locations visited only once, the total time spent stationary and finally the total time spent at home. Generally, agreement between the methods was reasonable.

The absolute total number of locations visited was similar (r = 0.78) between BEHAPP and the GPS tracker (Supplementary figure 1). When plotting the BEHAPP scores and the GPS tracker scores against one another, the data shows a linear positive relation between the methods (Supplementary figure 2). In absolute numbers, the values of the unique number of locations visited did differ to an extent between BEHAPP and the GPS tracker (Supplementary figure 3). However, plotting the BEHAPP values against the GPS tracker (Supplementary figure 4) shows that the measures are strongly linearly related to one another (r = 0.78). A similar pattern of differing absolute values but a strong linear relationship was observed for the single visit locations (r = 0.73) (Supplementary figured 5 and 6).

Total time spent stationary showed only small differences (r = 0.92) between the modalities compared to differences between the participants (Supplementary figures 7 and 8). The data showed a clear outlier, specifically participant 21. In order to give a clearer picture of the data, plots excluding participant 21 can be found in Supplementary figures 9 and 10. These also show adequate agreement between measurement modalities (r = 0.56). Finally, the total amount of total time spent at home showed slight differences and a similar linear relation between the two methods (r = 0.87) (Supplementary figures 11 and 12).

Movement

Two features were examined regarding the movement of the participants, specifically the number of trajectories (instances of location change) the participants exhibited and the perimeter of operation of all gps coordinates measured for the participants. The differences between BEHAPP and the GPS tracker were very large for the total number of trajectories (r = 0.20) (Supplementary figures 13 and 14). Strikingly, for three participants the BEHAPP GPS data was insufficient for the computation of trajectories, while this was also the case for eleven of the participants using the data from the GPS tracker. These cases are visible in Supplementary figure 13 as participants with no bars for a particular method. The perimeter of operation did show good agreement between methods, both in absolute numbers as well as plotted against each other (r = 0.94) (Supplementary figures 15 and 16).

Data gaps

Table 3 contains an overview of the number of gaps over 30 minutes per participant. The table shows that although most participants have few gaps over 30 minutes, some participants have very many gaps, indicating the data collection on their smartphones was not consistent.

Table 3: Number of total data points, number of total data gaps over 30 minutes, number of Wi-Fi measurements and number of Wi-Fi data gaps over 30 minutes.

Participant	Total # measurements	Total gaps > 30 min	Wi-Fi # measurements	Wi-Fi gaps > 30 min
1	745	53	61	26
2	6770	2	1091	2
3	9018	2	1262	2
4	3817	6	780	14
5	5863	0	1240	2
6	4248	32	537	42
7	3937	1	1231	1
8	6492	0	1263	0
9	26	4	0	0
10	11798	2	1281	1
11	477	34	59	33
12	2390	1	1326	1
13	7752	0	1251	0
14	4938	0	1294	0
15	1480	52	425	7
16	8893	53	27	19
17	7463	1	1406	0
18	18665	0	1291	6
19	10920	37	524	62
20	10462	30	1306	19
21	6413	13	1061	17
22	12710	0	1328	0

DISCUSSION

In the present study, we set out to examine the characteristics and quality of behavioural data passively collected using smartphones compared to data obtained by questionnaires or dedicated GPS devices. We aimed to assess the yield of the raw data across these platforms, as well as the comparability of the behavioural endpoint features extracted from them.

Our examination of the number of Wi-Fi access points showed that the number of access points at a location is not stable over time and may therefore in its current form not yet be ready for use as a measure of social density of a given location. We speculate that the location clusters with a higher degree of variability of access points indicate places such as apartment buildings, where moving within the perimeter of the same apartment or building may affect the number of recorded access points.

The comparison analysis between the Wi-Fi data from BEHAPP and the questionnaire is difficult to interpret due to this high variability and because the numbers of access points found by BEHAPP are much higher than expected when we designed the questionnaire. Of note, Wi-Fi access point data have been used more successfully in other studies, for example for the purpose of indoor positioning (e.g. Evennou & Marx, 2006; He & Gary Chan, 2016; Hilsenbeck et al., 2014; Zhou et al., 2015).

The agreement between application usage data recorded by BEHAPP versus the self-reported usage in the questionnaire was reasonable, though variable. In the absence of a gold standard it is not possible to attribute a higher degree of reliability to either of the two sources we examined. Concordance should not be assumed and further studies are required to extensively examine the causes of the observed disagreements.

The analysis of the agreement between the questionnaire items regarding sleep and sleep determined by the absence of app use and location data of the smartphone has two implications. The high levels of application usage in the first and last hour of reported sleep implies that self-reported times of start and end of sleep time may be not very precise. Secondly, it might be possible to achieve higher precision by clearly indicating that specifically the moment the participant stops all other activities in order to go to sleep is being queried.

The large differences between BEHAPP and the GPS tracker in data collection quantity exist in part due to the way in which the two methods determine when to gather location data. BEHAPP implements the FusedLocationProvider, which led to relatively high measurement frequencies as compared to the GPS Tracker. The Tractive GPS Tracker is developed with the goal is to make the battery last as long as possible. When considering GPS data collection, researchers should carefully consider the balance between measurement frequency they need for the level of detail their research requires, and its effect on battery life, which may negatively affect study attrition.

Regarding the outcomes of the GPS data, agreement between BEHAPP and the independent GPS tracker was generally acceptable, with two exceptions. For the location visits features, the methods agreed on the relative scores of participants, but there were some differences between the BEHAPP and GPS tracker methodologies in the exact scores.

Secondly, the measurement of trajectories did vary substantially between the methods. We speculate the trajectory measures may be affected by the measurement frequency in two ways. First, the increased sampling frequency may make it possible for BEHAPP to pick up smaller changes in location. Secondly, the algorithm used to discover movement instances was not tuned to low frequency data. The algorithm's definition of 'trajectories' or movement instances

required a minimum of twenty observations per trajectory in order to remove spurious changes in location from the data. The GPS device's low measurement frequency may imply that even some longer movement instances by the participants may not have been labelled as trajectories by the algorithm. We cannot conclude that measurement frequency is the only issue with the measurement of movement instances as the large disparity in measurement frequency may have obscured other differences between the methods.

We found missing data and data gaps to be common for each category of data that we collected. In one instance this could be attributed to user error where the BEHAPP app was given insufficient permissions at the start of the participation to collect foreground app usage data. In the other instances app termination was the most likely culprit underlying gaps in our data. Passive monitoring app developers can consider possible work-arounds for these issues, for example by distributing the app directly through sideloading (a direct download onto the phone, circumventing app stores) which gives more control over app behaviour. Combined with proper app whitelisting practices chances of the app being terminated can be expected to decrease substantially.

Limitations

We recognize the following limitations in our research. Our first limitation concerns the large difference in measurement frequency between BEHAPP and the independent GPS tracker. This difference did allow us to highlight the importance of high sample frequencies, but other relevant features of GPS data acquisition may be obscured due to the large differences in measurement frequency between the two methods.

The study was limited by its reliance on only a single day of questionnaire data. This was decided so as to not overburden the participants. Even so, several participants omitted to complete the questionnaire, highlighting a weakness of this more traditional approach. Data requiring active input can be collected alongside passive monitoring studies, but carry the risk of missing data.

Finally, the questionnaire contained a minor issue in that the Wi-Fi categories overlapped. However, only two concurrent BEHAPP measurements were recorded as 5 Wi-Fi access points, the participants filled in 5-10 and these were classified as correct, not severely affecting our results.

Conclusion

We conclude that although the different measurement modalities show many similarities, they cannot simply be used interchangeably. The methods have significant differences in terms of measurement frequency and precision which should be considered when choosing which

modality to use. We also conclude that there is great potential for passive monitoring platforms such as BEHAPP to serve as an instrument for the objective, naturalistic observation of participants in a real-time and real-world setting.

Secondly, data quality examinations are paramount to all types of research, but are especially important in studies using passively gathered data as errors and missing data can easily remain unnoticed, with consequences for the results and their interpretation.

Data availability statement

The data generated during the study is under a confidentiality agreement due to the sensitive, identifiable nature of the data. Data sharing is therefore not applicable for this study.

Code availability statement

The code that support the findings of this study are available on request from the corresponding author.

Author contributions

M.C.R and R.J.J. conceived of the study and wrote the manuscript. M.C.R. gathered the data and performed the analyses for the study. R.J.J. maintained the smartphone application. J.A.V. and M.J.K. supervised and provided valuable feedback and insights for the study and manuscript.

Acknowledgements

The PRISM project (www.prism-project.eu) leading to this application has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115916. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This publication reflects only the authors' views neither IMI JU nor EFPIA nor the European Commission are liable for any use that may be made of the information contained therein.

Competing interests

The authors declare no competing interests.

REFERENCES

Aledavood, T., Triana Hoyos, A. M., Alakörkkö, T., Kaski, K., Saramäki, J., Isometsä, E., & Darst, R. K. (2017). Data Collection for Mental Health Studies Through Digital Platforms: Requirements and Design of a Prototype. *JMIR Research Protocols*, *6*(6), e110.

Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology, 8*(1), 89-119. doi:10.1093/jssam/smz056

Android Oreo (8.0): Behavior changes. (2020, January 22). Android Developer Documentation. https://developer.android.com/about/versions/oreo/android-8.0-changes

Android Pie (9.0): Behavior changes. (2019, December 27). Android Developer Documentation. https://developer.android.com/about/versions/pie/android-9.0-changes-all

Bähr, S., Haas, G.-C., Keusch, F., Kreuter, F., & Trapmann, M. (2020). Missing data and other measurement quality issues in mobile geolocation sensor data. *Social Science Computer Review,* 1-24. doi:10.1177/0894439320944118

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J.-P. (2018). Relapse prediction in schizophrenia through digital phenotyping: a pilot study. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *43*(8), 1660–1666.

Berrouiguet, S., Ramírez, D., Barrigón, M. L., Moreno-Muñoz, P., Carmona Camacho, R., Baca-García, E., & Artés-Rodríguez, A. (2018). Combining Continuous Smartphone Native Sensors Data Capture and Unsupervised Data Mining Techniques for Behavioral Changes Detection: A Case Series of the Evidence-Based Behavior (eB2) Study. *JMIR mHealth and uHealth*, *6*(12), e197.

Doryab, A., Villalba, D. K., Chikersal, P., Dutcher, J. M., Tumminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J., Creswell, J. D., & Dey, A. K. (2019). Identifying Behavioral Phenotypes of Loneliness and Social Isolation with Passive Sensing: Statistical Analysis, Data Mining and Machine Learning of Smartphone and Fitbit Data. *JMIR mHealth and uHealth*, *7*(7), e13209.

Evennou, F., & Marx, F. (2006). Advanced Integration of WiFi and Inertial Navigation Systems for Indoor Mobile Positioning. In *EURASIP Journal on Advances in Signal Processing* (Vol. 2006, Issue 1). https://doi.org/10.1155/asp/2006/86706

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile Context Instrumentation Framework. In *Frontiers in ICT* (Vol. 2). https://doi.org/10.3389/fict.2015.00006

He, S., & Gary Chan, S.-H. (2016). Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons. *IEEE Communications Surveys Tutorials*, *18*(1), 466–490.

Hilsenbeck, S., Bobkov, D., Schroth, G., Huitl, R., & Steinbach, E. (2014). Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct*. https://doi.org/10.1145/2632048.2636079

Hofmans, T. (2020, August 31). *Nederlandse corona app werkt niet goed op achtergrond door accubesparingsmodus*. Tweakers.net. https://tweakers.net/nieuws/171338/nederlandse-corona-app-werkt-niet-goed-op-achtergrond-door-accubesparingsmodus.html

Insel, T. R. (2017). Digital Phenotyping: Technology for a New Science of Behavior. *JAMA: The Journal of the American Medical Association*, *318*(13), 1215–1216.

Insel, T. R. (2018). Digital phenotyping: a global tool for psychiatry. *World Psychiatry: Official Journal of the World Psychiatric Association* , *17*(3), 276–277.

Jagesar, R. R., Roozen, M. C., van der Heijden, I., Ikani, N., Tyborowska, A., Brenda W J, Ruhe, H. G., Sommer, I. E. C., Kas, M. J., & Vorstman, J. A. S. (2021). Digital Phenotyping and the COVID-19

Pandemic: Capturing Behavioral Change in Patients with Psychiatric Disorders. In *European Neuropsychopharmacology*. https://doi.org/10.1016/j.euroneuro.2020.11.012

Jongs, N., Jagesar, R., van Haren, N. E. M., Penninx, B. W. J. H., Reus, L., Visser, P. J., van der Wee, N. J. A., Koning, I. M., Arango, C., Sommer, I. E. C., Eijkemans, M. J. C., Vorstman, J. A., & Kas, M. J. (2020). A framework for assessing neuropsychiatric phenotypes by using smartphone-based location data. *Translational Psychiatry*, *10*(1), 211.

Ma, D. (2007). The Business Model of "Software-As-A-Service." In *IEEE International Conference on Services Computing (SCC 2007)*. https://doi.org/10.1109/scc.2007.118

Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review, 31*(6), 725-743. doi:10.1177/0894439313485201

Meterko, M., Restuccia, J.D., Stolzmann, K., Mohr, D., Brennan, C., Glasgow, J., & Kaboli, P. (2018). Response rates, nonresponse bias, and data quality: Results from a national survey of senior healthcare leaders. *Public Opinion Quarterly, 79*(1), 130-144. doi:10.1093/poq/nfu052

Simple, battery-efficient location API for Android. (n.d.). Fused Location Provider API. Retrieved September 4, 2020, from https://developers.google.com/location-context/fused-location-provider

Shin, E. Johnson, T. P., & Rao, K. (2011). Survey mode effects on data quality: Comparison of web and mail modes in a U.S. National Panel survey. *Social Science Computer Review, 30*(2), 212-228. doi:10.1177/0894439311404508

Sugle, N.F. (2016). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research, 47*(3), 458-491. doi:10.1177/0049124115626176

Toepoel, V., Lugtig, P., & Schouten, B. (2020). Active and passive measurement in mobile surveys. *The Survey Statistician, 20,* 14-26.

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Mental Health*, *3*(2), e16.

Torous, J., Bucci, S., Bell, I. H., Kessing, L. V., Faurholt-Jepsen, M., Whelan, P., Carvalho, A. F., Keshavan, M., Linardon, J., & Firth, J. (20210). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry, 20*(3), 318-335. doi:10.1002/wps.20883

Tractive Realtime GPS Tracking. (n.d.). Retrieved November 20, 2020, from https://tractive.com/

Zhou, Z., Wu, C., Yang, Z., & Liu, Y. (2015). Sensorless sensing with WiFi. In *Tsinghua Science and Technology* (Vol. 20, Issue 1, pp. 1–6). https://doi.org/10.1109/tst.2015.7040509

SUPPLEMENTARY MATERIALS

Supplementary Table 1 - Android Platform Changes

Table 1. Overview of Android energy savings and privacy measures impacting passive smartphone based monitoring apps

Android Version	Measure	Impact	Mitigation
Marshmallow [12]	Doze mode	Enforced device sleep state maximization which impacts the maximum frequency of sampling methods that are run on a fixed interval (e.g. WiFi Access Point scans)	Implement task scheduling system using API's that may partially override and cut through doze mode restrictions
	App Standby	Restricts network access and scheduled jobs for background apps when the device is unplugged	None, testing showed minimal impact to data accuracy
Nougat [13]	Doze mode v2	Device sleep state maximization engages earlier and is extended with another set of restrictions	Existing strategy applies with same level of effect
Oreo [14]	Background execution limits	Services belonging to an app considered to be in the background are stopped by the operating system. This will stop data observers that need to continuously run and actively monitor for sensor and device state changes	Declare a foreground service which shows a permanent notification on screen
	Background location limits	The frequency at which location updates are reported are limited for apps that are considered to be in the background	Declare a foreground service which shows a permanent notification on screen
Pie [15]	App standby buckets	Usage pattern based device resource allocation. The system imposes severe restrictions on apps that are rarely or never used in the foreground	Declare a foreground service which shows a permanent notification on screen
	Battery saver improvements	When the device is set in battery savings mode (e.g. when the battery runs low) an additional set of measures may come into effect with regards to location updates and background resource usage	None
	No access to the camera, microphone and continuously reporting sensors such as accelerometers and gyroscopes	Apps that run in the background are limited in their access to various data sources that are common to smartphone based passive monitors	Declare a foreground service which shows a permanent notification on screen
Google Play Store 2018 [16]	Restrictions to the use of READ_CALL_LOG and READ_SMS permissions	Starting from Q4 2018, when using the Google Play Store as a distribution channel, only a very small subset of apps are allowed to use permissions that allow access to phone call and text messaging history. There are no exceptions available for the use case of smartphone based behavioral monitoring	- Distribute the app directly to participants through the practice of sideloading keeping all features intact - Comply and distribute a version of the app without READ_CALL_LOG and READ_SMS permissions and functionalities for the Play Store

Supplementary Materials - Participant information for completing the questionnaire

Questionnaire

On the final full day of the ten-day study, we ask two things of you: 1) to record at three times during the day a) to record the time, b) to record how many Wi-Fi points are available for your phone at that moment (0-5, 5-10 or over 10) and c) at each of the recorded times open a random app on your phone, then record which app you opened; and 2) to fill in a questionnaire at the end of the day with some questions about communication, location and sleeping behavior during the day. The three times the app and Wi-Fi data is recorded can be at any time during the day, as long as they are at least three hours apart. The questionnaire also functions as input for the Wi-Fi and app data.

This questionnaire marks the 9th full day of participation in the study. For any questions for which you can check the answers on your phone (such as the amount of phone calls made), you are welcome to do so.

What is your participant number?

The number is included in the e-mail containing the link to this questionnaire

What brand and type of phone do you have?

E.g. Apple Iphone 7, Samsung Galaxy S7, Lenovo P2 etc.

How many calls did you make today?[1]
How many of these calls did you dial yourself?
If you made any calls, please give the start- and endtimes for 1-3 of these calls.
If you received any calls, please give the start- and endtimes for 1-3 of these calls.
How many different people did you converse with over the phone today?
What types of travel did you use today? Please select all that apply.

- Walking
- Cycling
- Car
- Bus
- Train
- Taxi
- Boat
- Skating

1 Call data was not analyzed because of restrictions placed on retrieving call data by Google during the study duration.

What type of functions do the locations you visited serve FOR YOU? Please select all that apply.

- Home
- Work
- Family/Friends
- Recreation/sports, not primarily focused on the social aspect
- Shopping (groceries or other)

At what time did you go to bed last night?
At what time did you get up this morning?

The following questions regard the app and wifi notes you made during the day.
At what times did you collect the notes on WiFi points and app use?

- Time 1
- Time 2
- Time 3

Which apps did you use?

- Time 1
- Time 2
- Time 3

How many Wi-Fi points were available?

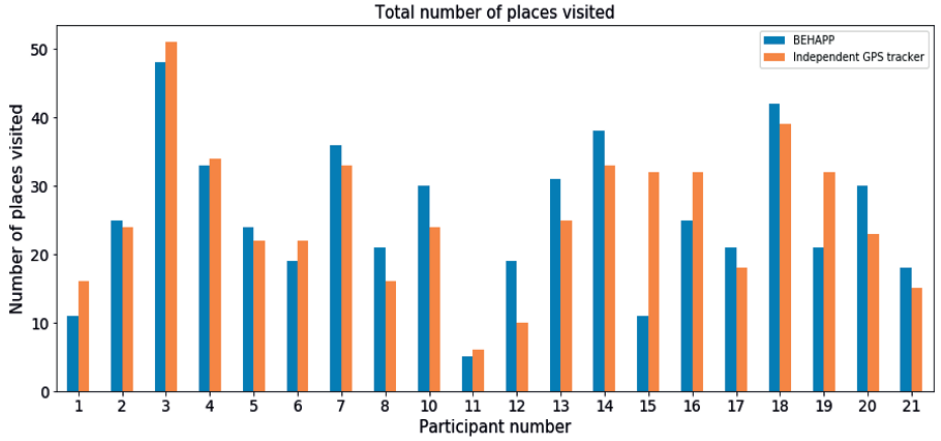
- Time 1
  - 0-5
  - 5-10
  - > 10
- Time 2
  - 0-5
  - 5-10
  - > 10
- Time 3
  - 0-5
  - 5-10
  - > 10

Do you have any additional notes?

Here you can include feedback about the Tractive device and the questionnaire or other information. Please also make a note if the Tractive device has been apart from your phone for an extended amount of time (more than 2 hours).

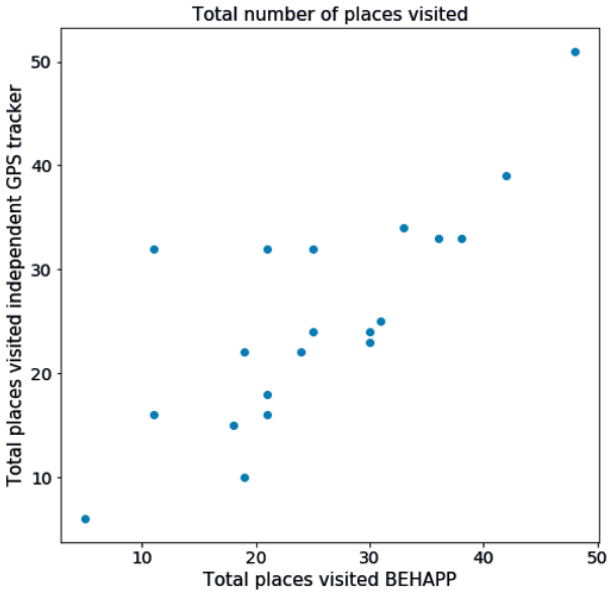
SUPPLEMENTARY FIGURES

Supplementary Figure 1



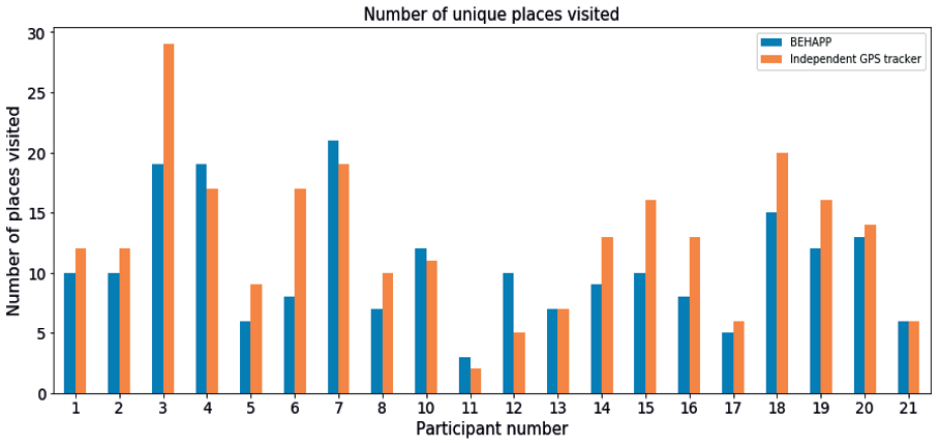
Supplementary figure 1: Barplot of total number of places visited. The figure shows the amount of places visited over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 2



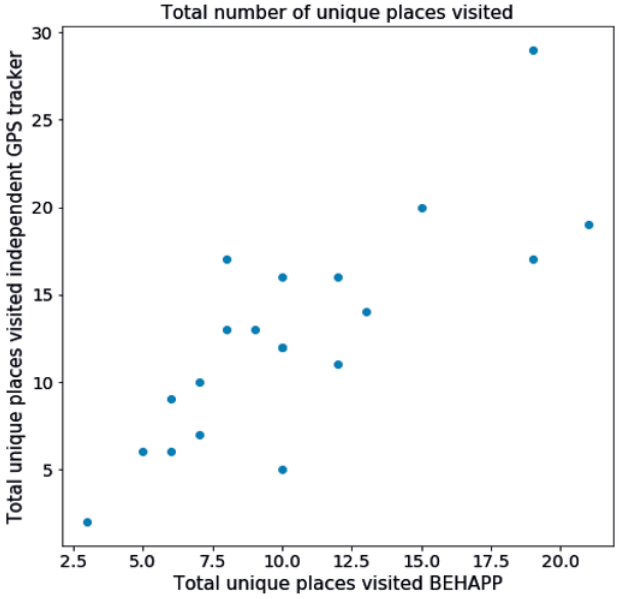
Supplementary figure 2: Scatterplot of total number of places visited. The figure shows the amount of places visited over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 3



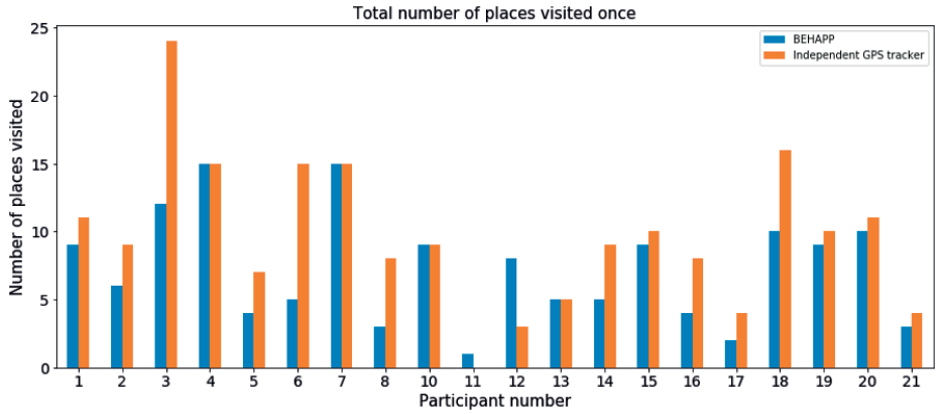
Supplementary figure 3: Barplot of the number of unique places visited. The figure shows the amount of unique places visited over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 4



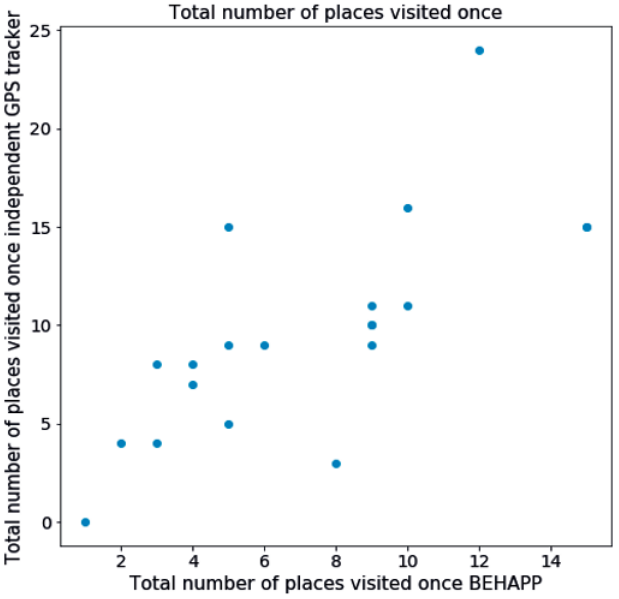
Supplementary figure 4: Scatterplot of the number of unique places visited. The figure shows the amount of unique places visited over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 5



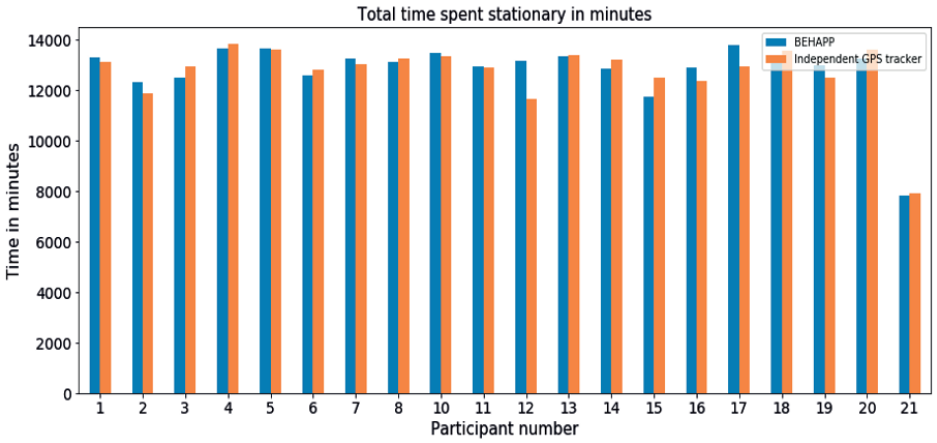
Supplementary figure 5: Barplot of number of places visited once. The figure shows the amount of places visited exactly once over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 6



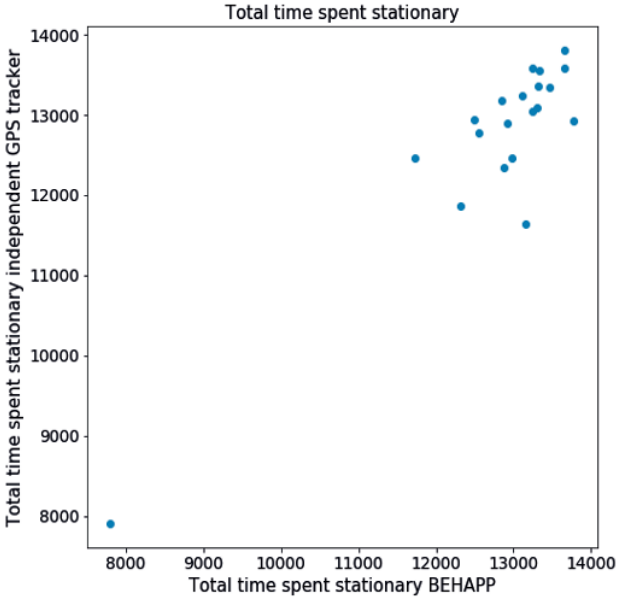
Supplementary figure 6: Barplot of number of places visited once. The figure shows the amount of places visited exactly once over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 7



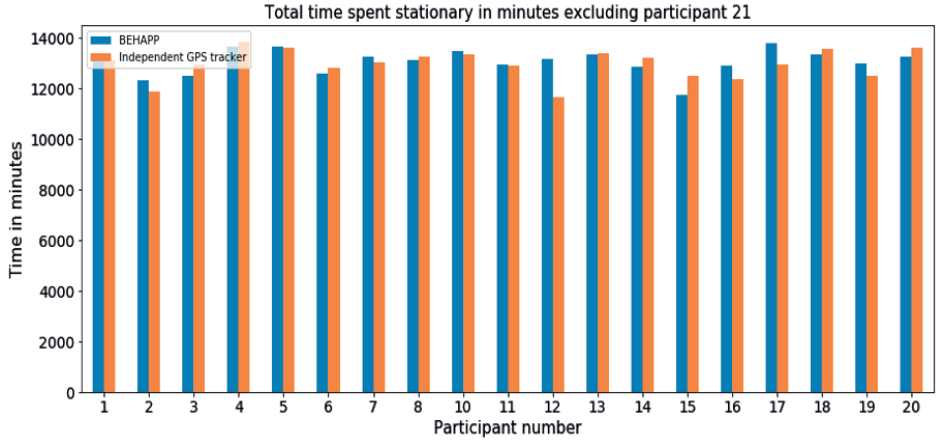
Supplementary figure 7: Barplot of time spent stationary in minutes. The figure shows the total time spent stationary in minutes over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 8



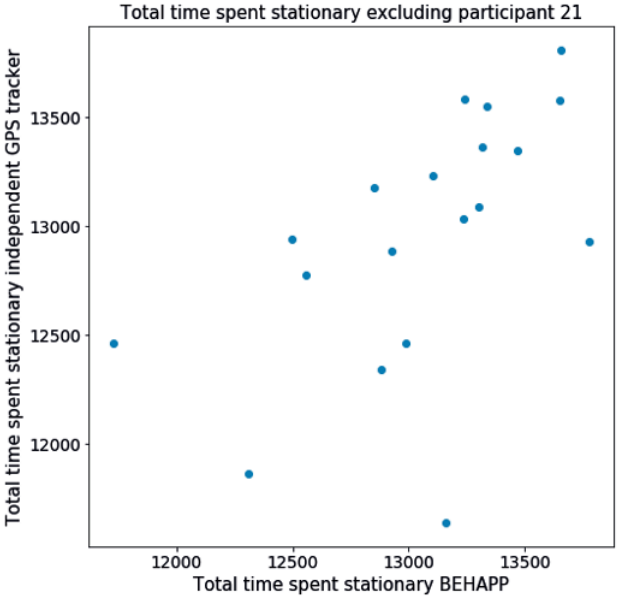
Supplementary figure 8: Scatterplot of time spent stationary in minutes. The figure shows the total time spent stationary in minutes over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 9



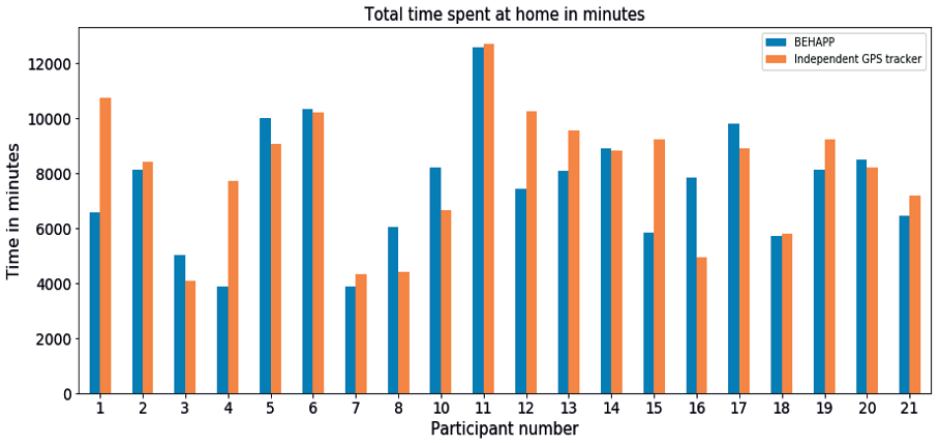
Supplementary figure 9: Barplot of time spent stationary in minutes. The figure shows the total time spent stationary in minutes over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange) excluding the outlying value of participant 21.

Supplementary Figure 10



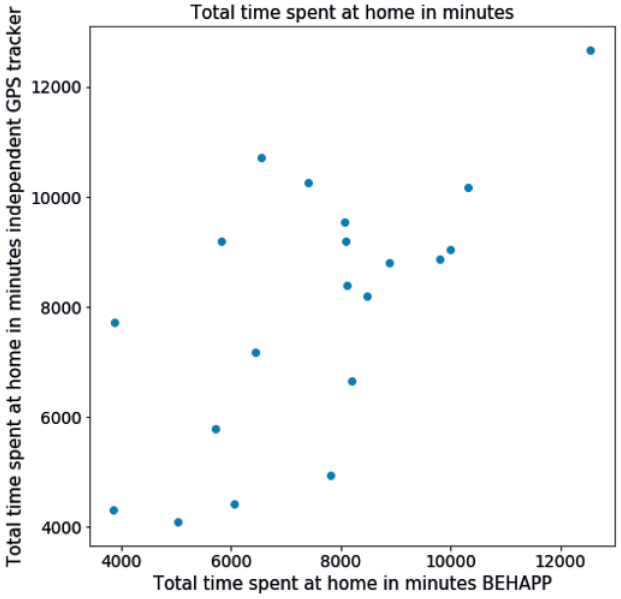
Supplementary figure 10: Scatterplot of time spent stationary in minutes. The figure shows the total time spent stationary in minutes over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis) excluding the outlying value of participant 21.

Supplementary Figure 11



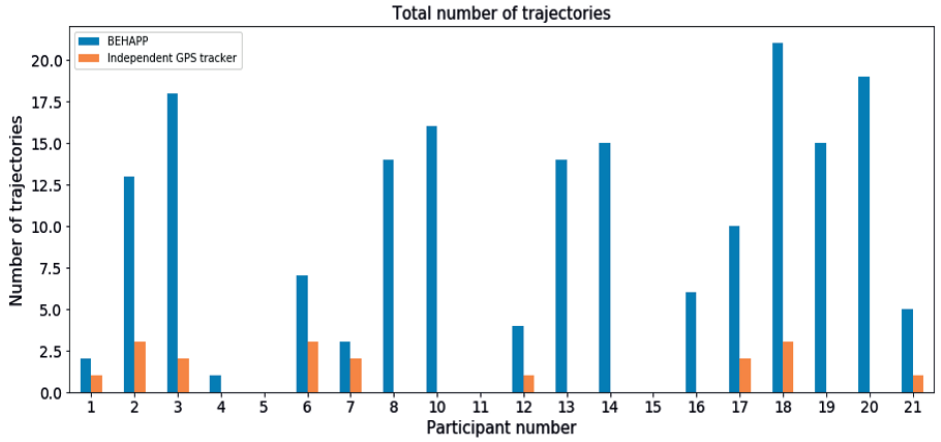
Supplementary figure 11: Barplot of the amount of time spent at home in minutes. The figure shows the total time spent at home over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 12



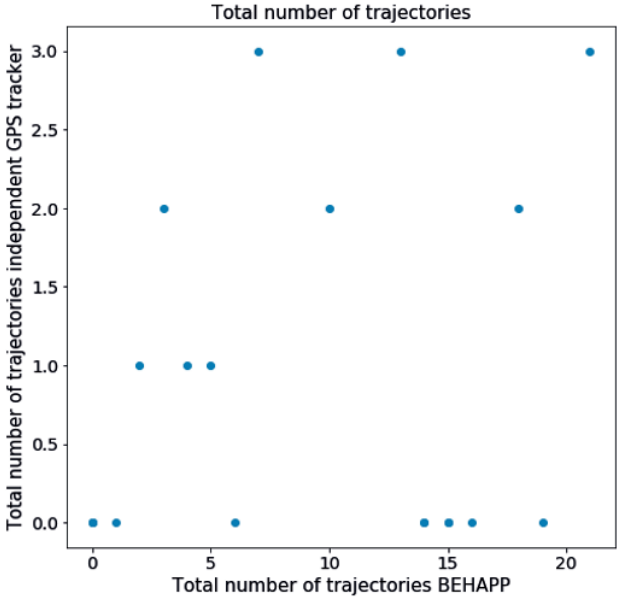
Supplementary figure 12: Scatterplot of the size of time spent at home in minutes. The figure shows the time spent at home in minutes over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 13



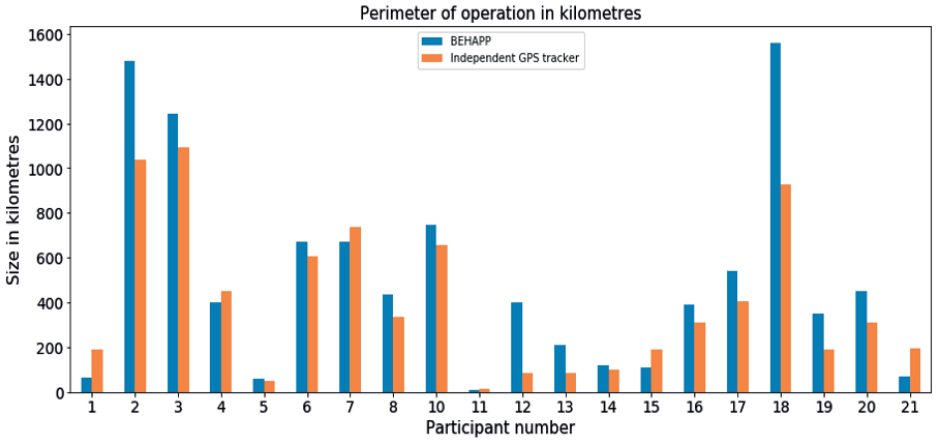
Supplementary figure 13: Barplot of the total number of trajectories. The figure shows the total number of trajectories over the study duration according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 14



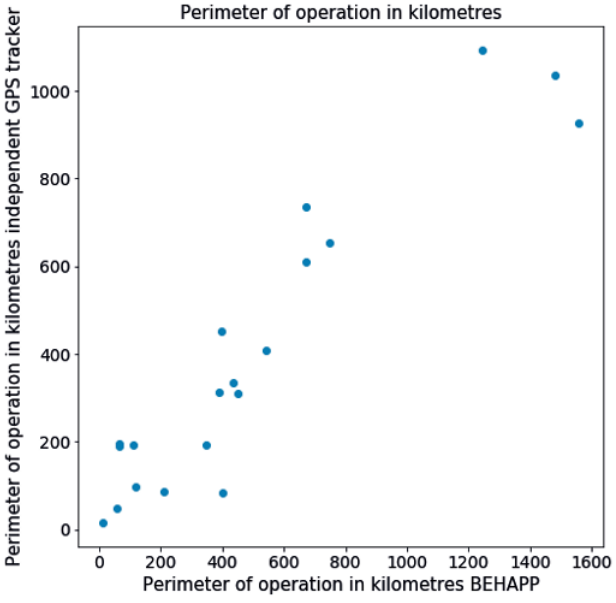
Supplementary figure 14: Scatterplot of the total number of trajectories. The figure shows the total number of trajectories over the study duration compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).

Supplementary Figure 15

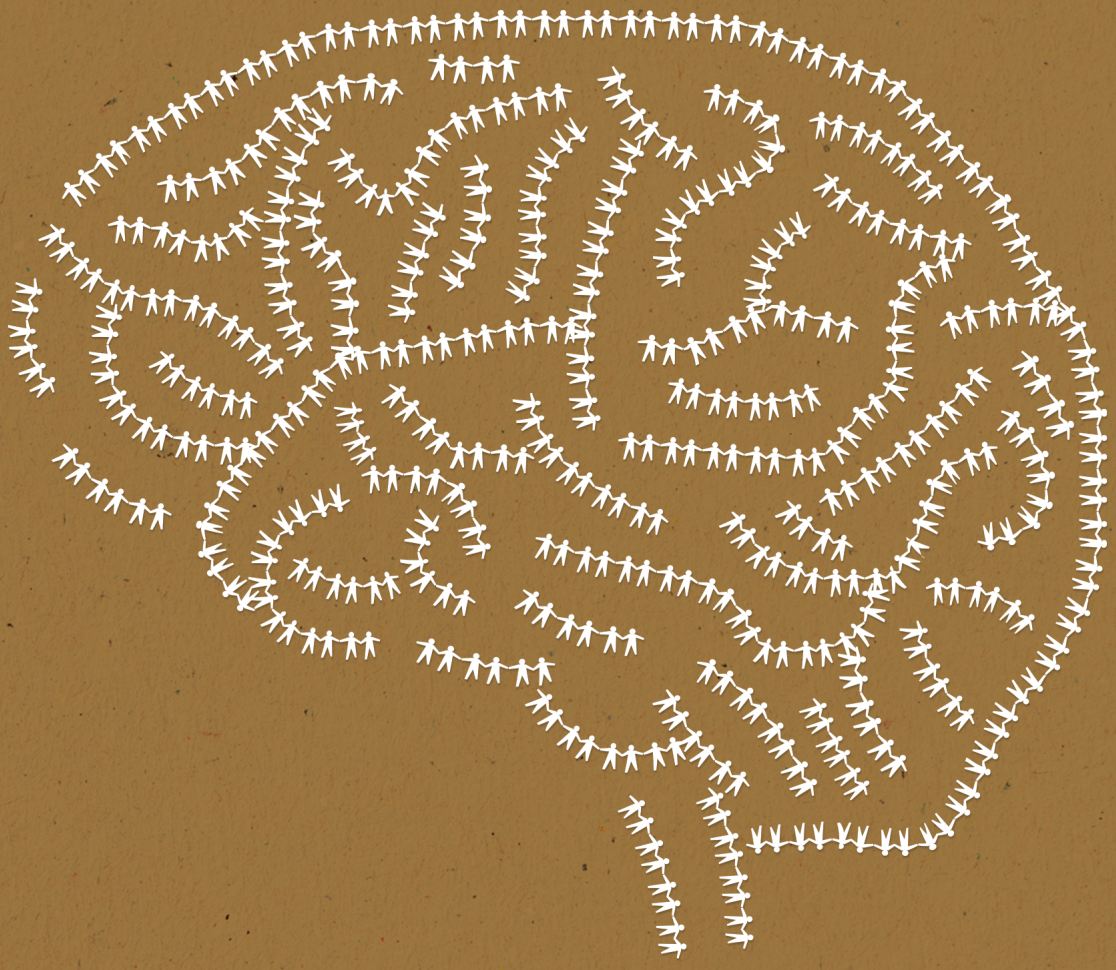


Supplementary figure 15: Barplot of the size of the perimeter of the area of operation in kilometres. The figure shows the perimeter of operation (the perimeter of the total area covered by the participant) according to BEHAPP (blue) and the independent GPS tracker (orange).

Supplementary Figure 16



Supplementary figure 16: Scatterplot of the size of the perimeter of the area of operation in kilometres. The figure shows the perimeter of operation (the perimeter of the total area covered by the participant) compared between BEHAPP (on the x axis) and the independent GPS tracker (on the y axis).



Chapter 5

General discussion

Mila C. Roozen[1]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, the Netherlands

The processes underlying the evolution of the brain across speciation, resulting in large brains in primates and particularly humans are still a hotly debated subject, despite decades of research findings and the formulation of several explanatory hypotheses. For example, while the Social Brain Hypothesis (SBH), which predicts that brain size is dependent on social complexity across species, has many proponents, research supporting alternative hypotheses continues to emerge (e.g., DeCasien, 2017), while on the other hand evidence supporting the social brain hypothesis also continues to come out (e.g., Maclaren et al., 2023). While still debated, the SBH is a popular hypothesis to explain brain evolution across species.

The SBH also has interesting implications regarding the recent evolution of the brain in humans. Dunbar (2009) mentions that explanations for the fitness benefit of high social complexity in primates and ancient hominids is that social complexity may either reduce predation risk or enhance foraging through cultural transmission of skills. Dunbar (2009) also mentions that due to the energetic expense of having a large brain, the benefits of having a large brain need to have a strong effect on fitness. However, in many modern human societies, humans face little to no risk of predation and rely in much more limited amounts on foraging. Does this indicate that the cost of having a large brain may have started to outweigh the benefits of high social complexity?

On the other hand, it may be possible that social complexity in humans has had novel benefits, but this has not been established. Across recent evolutionary history, many humans have started to congregate in large groups such as cities. Such environments may require cognitive skills capable of dealing with a large number of relationships. While the environment of many modern humans is highly social, it is perhaps questionable whether deficits in social cognition result in real deficits in reproductive fitness. Some evidence points towards reduced reproductive fitness in individuals with social dysfunction, such as those with autism spectrum disorders (Mullins et al., 2017; Power et al., 2013), however such evidence is typically found in patient samples with additional symptoms (e.g., comorbid intellectual disability) which have not been assessed separately. It is not difficult to imagine individuals with more positive social relationships having higher mating success in modern humans. Some studies have examined the association between social status and reproductive fitness in recent human history or modern humans using proxy measurements such as income or education (Hopcroft, 2019; Zhang & Santtila, 2022). However, it appears no studies have been carried out examining the association between social complexity and reproductive fitness specifically in modern humans.

While the evolution of social complexity in recent human evolutionary history has not been subject to much scientific study, recent evolutionary changes to the human brain have been examined. We describe several studies examining this evolution in **chapter 1**. Although several studies from varying samples have found evidence that brain size may have decreased across

thousands or tens of thousands of years (e.g., DeSilva et al., 2021; Liu et al., 2014), such findings are still debated, with some arguing that these results may spurious and the result of incorrect methodological choices (Villmoare & Grabowski, 2022). It is clear more research is necessary to determine whether and how human brain size changed across the past thousands of years, but the results from the studies carried out so far do inspire the question whether in the case that brain size did decrease during recent human evolutionary history, this is related to the changes in the social environment and whether such changes could affect human social cognition over time.

In order to find answers to such questions, more research into the association regarding the evolution of human social complexity and human brain size is needed. However, comparative methods cannot be used to study recent evolutionary changes within a species, and while fossil records can provide evidence of changes to brain size (DeSilva et al., 2021), such studies typically cannot be used to determine the processes underlying such changes. Fortunately, in recent years, novel analytical methods have been developed to study evolution based on the rapid development of genetic technologies. As a result of technological improvements in the ability to genotype and increased understanding of how genomes are affected by selection, methods such as singleton density scores (Field et al., 2016) can provide information on whether specific genes with known associations to phenotypes have been under the effects of selection pressure during recent millennia. Singleton density scores, for example, utilize knowledge about how the distribution of singleton mutations around genes varies dependent on the selection those genes are affected by. Then, by gathering data on which genes are involved in certain phenotypes and examining the distribution of singleton mutations around such genes, the strength of selection affecting the phenotype across the past 3000-2000 years can be studied. Several of such methods (including the singletond density scores) are described in **chapter 1,** including a description of a recent study where methods were combined to create selection pressure timelines across many phenotypes using data from the UK Biobank (Song et al., 2021). By combining genetic data across species, even more ancient evolutionary processes resulting in genetic backgrounds of phenotypes in modern species can be examined. An example of such a study can be found in **chapter 2**.

Genetic methods of studying evolution require that earlier studies have uncovered the genetic background of the phenotype being studied. However, at the previous time, no studies have been carried out examining the association between social complexity and genetic variation in humans. Related phenotypes such as sociability have been examined (Bralten et al., 2021). While social complexity and sociability are likely highly correlated (especially when overlapping measures are used, which is the case for the Bralten et al. (2021) study and the operationalization of social complexity used in **chapter 3**), variation in conceptualization and operationalization of social complexity has been discussed as a potential detrimental factor

in interspecies studies of social complexity, and could potentially lead to similar issues if this becomes typical in within-species studies of social complexity. In **chapter 3**, we have provided a proof of principle that expert frameworks of social complexity can be implemented in studies in humans. Such operationalizations can, in the future, be used to examine the genetic background underlying human variation in social complexity, which can in turn then be used in statistical methods using genetic information to study evolution.

The study of social complexity in animals typically relies on relatively objective quantitative or qualitative measures which are observed and recorded by experienced researchers. However, , measuring social complexity in individual modern humans might be a daunting task. For example, in order for a researcher to observe all 'differentiated relationships' of an individual as per the definition of Bergman and Beehner (2015), this would likely require long-term and intensive observation of individuals during their daily lives. Of course, with human subjects the option exists to simply ask these subjects about their social environment as was done in the data collection for the UK Biobank data used in **chapter 3**. However, self-report of such variables comes with its own difficulties. For example, concepts such as 'differentiated relationships' might be difficult to explain to subjects in such a way that all subjects understand it the same way. Also, biasing factors such as social desirability and recall bias may affect measurement. A potential solution to the issues surrounding the measurement of social complexity in humans might exist in the form of (passive) digital phenotyping. Smartphones, wearables or other sensors could be used to detect characteristics of subjects' social environments and interactions, both completely passively using sensors such as GPS and logs such as call logs or in combination with self-report such as through diaries or prompted recording of social interactions. In **chapter 4** we examined measurement characteristics of a smartphone app, called Behapp, which can gather data regarding social behavior passively and may in the future provide a potential low-bias measurement tool for social complexity.

Conceptual and methodological considerations in the study of the social brain hypothesis

One of the potential causes resulting in disagreement between experts regarding the validity of the social brain hypothesis is a lack of consistency in conceptualization of social complexity across studies (Kappeler, 2019; Kappeler et al., 2019). Whereas such conceptual issues have so far mostly been discussed in terms of interspecies variation, considerations regarding adequate definition and operationalization of social complexity may be equally important for future studies using intraspecific measures of social complexity such as those proposed by Aureli and Schino (2019). The importance of the choice of conceptualization of social complexity was demonstrated in **chapter 3**, where associations between social complexity measures and cognitive skills were assessed. While two measures and a composite of all three measures of social complexity used showed positive associations between cognition and social complexity, the

number of friend and family visits showed the opposite. If such measures were used separately across studies, they could lead to disagreement about associations between social complexity and cognition in modern humans. **Chapter 3** also demonstrated the importance of examining aspects of social complexity both separately and as a composite measure. On the one hand, the composite measure may have captured more variation in social complexity compared to the separate measures, potentially resulting in the finding of several associations between social complexity and grey matter volumes where separate measures found few. On the other hand, the fact that the number of visits with friends and family was the only separate measure to be associated with grey matter volumes in any region while also having opposite associations with cognition compared to the other measures may be an indication that different neurobiological processes underlie this measure. Future studies could incorporate other measures of social complexity and experiments with different compositions of composite scores in order to arrive at a more ideal method to measure social complexity.

Besides the definition and operationalization of social complexity itself, other methodological considerations also have the potential to strongly affect results in evolutionary studies. We show in **chapter 1** that the use of genome wide association study (GWAS) results from studies which have filtered single nucleotide polymorphisms based on Hardy-Weinberg equilibrium (HWE) may be biased to show low selection pressure due to the fact that HWE also correlates with selection pressure. However, filtering on HWE is a common procedure in GWAS studies, as it can also filter out poorly genotyped variants. Afterwards, imputation can be used to make sure genes are included in the analyses, however, this imputation will occur based on the assumption of HWE, resulting in inaccurate imputation for genes which deviate from HWE due to real genetic effects. Future GWAS studies carried out with the goal of performing evolutionary analyses on the results should carefully consider how to create a balance between ensuring optimal genotyping quality and including accurate data regarding the distribution of SNPs under selection pressure. One potential solution would be to carry out sensitivity analyses where GWAS is performed both with and without selection for HWE to compare the results. Variants associated with the phenotype only in the analyses without HWE selection could then be carefully examined further to determine whether HWE violations might be due to HWE violations or be a result of selection pressure. GWAS studies carried out on whole genotype sequencing (WGS) data may also benefit evolutionary analyses, as the high coverage of the WGS arrays can contribute to the reliability of genotyping results.

In **chapter 2** we found that the comparison set used for the evaluation of genetic conservation in a subset of genes of interest can substantially affect the results of such analyses. By including a comparison dataset used in previous studies (Franklin & Dwyer, 2021; Kasap et al., 2018; Sall et al., 2021) and a comparison dataset which allows for a fairer evaluation of differences, we were able to determine a more optimal way to measure the extent to which genes as-

sociated with a phenotype are enriched for genetic conservation across species. Due to the novelty of many of the methods mentioned in this thesis, few or no studies have been carried out to determine optimal parameters and methodological choices. This may result in more inconsistent results as harmonization of such choices is not supported by existing literature. The analysis of evolutionary processes using genetic data might benefit greatly from studies focusing specifically on examining how parameterization and other methodological choices affect such analyses.

Associations between brain structure and function in the context of the social brain hypothesis.

In **chapter 3** we found small but significant associations between grey matter volume in several regions of the brain as well as the total brain and measures of social complexity. While the association between total brain volume and social complexity was as expected, the findings regarding regional brain volumes did not conform to expectations based on the SBH. Specifically, research supporting the SBH appears to show a link specifically between the size of the neocortex and social complexity, which can be found in analyses of total brain volume as a result of the high correlation between total brain volume and neocortex volume (Dunbar, 2009). However, all but one of the associations found between regional brain volumes and social complexity were found in regions of the brain outside the neocortex, specifically in the basal ganglia (the pallidum) and various lobules of the cerebellum. While controlling for total grey matter volume is important in analyzing associations between regional grey matter volumes and other phenotypes, the high correlation between neocortex volume and grey matter volume may have resulted in the masking of associations between social complexity and regional volumes in the neocortex if such associations were spread out across the neocortex.

It may also be the case that associations between brain volume and social complexity is different for human variation compared to the within-species variation used to support the SBH. While in older studies the cerebellum was assumed to be associated mostly with motor behavior, recent findings have amended this to include an important role for the cerebellum in social cognition (Van Overwalle et al., 2020). Although evidence of the role of the pallidum in social behavior is limited to its role in pair bonding in animals (Lim and Young, 2004), the pallidum is known to play a role in the functioning of the default mode network (DFM) (Klaassen et al., 2021), which in turn may play a role in social behavior (Fareri et al., 2020; Mars et al., 2012; Wen et al., 2020). The cerebellum is also functionally connected to the DFM, and dysfunction of this connection is associated with various forms of neuropsychiatric disorders characterized by social dysfunction (Guo et al., 2015; Luo et al., 2018; Wang et al., 2014). Possibly, functioning of parts of the neocortex such as those involved in the DFM could be dependent on the cerebellum and basal ganglia (pallidum). It might be interesting to examine how social

complexity is associated with neural functioning and connectivity and whether such factors might be affected by selection on social complexity in recent or ancient human evolution.

Potential of (passive) digital phenotyping for social complexity research in humans

One weakness of phylogenetic studies, but also of studies such as those in **chapter 3,** is that the results are dependent on measurements performed by individuals (such as the researcher or the subject) who are per definition not objective. Researchers judging the social behavior of others may have preconceptions which could affect their results, while subjects themselves might report on their own social behavior inadequately due to reasons such as social desirability or recall bias. The latter might especially be an issue in studies including individuals with neuropsychiatric disorders, as such individuals might be affected by cognitive symptoms resulting in inaccurate reporting of social behavior (Jongs et al., 2022), potentially creating spurious associations between cognition and social complexity. Whereas objective reporting may not have been an option in the past, recent developments in digital phenotyping have created the possibility of monitoring subjects' behavior without relying on subjective accounts. Typical measures of social behavior either provide only a snapshot of a relatively short time period or require the subject to remember activities which occurred a long time ago, increasing the odds of inaccurate reporting. Smartphone applications such as those examined in **chapter 4** have the potential to not only measure behavior in a more objective manner but also longitudinally and in a real life setting. On the other hand,

However, as **chapter 4** highlights, it is important to be aware of the methodological characteristics of digital phenotyping applications, and the objective nature of the methodology should not lure researchers into believing that the measures are necessarily always accurate. Missing data, and variability of the amount of missing data between subjects, could impact measurements in such a way that it could bias interpretation. It is also important to verify whether passive measurements work as intended, as we found that for example the number of Wi-Fi access points is highly unstable and therefore this measure should not be used without extensive adaptations as a measure of characteristics of the environment. While keeping these caveats in mind, (passive) digital monitoring may be an interesting methodological consideration for the measurement of social complexity for the reasons mentioned previously. In order to provide a first test of this possibility, we performed a preliminary analysis in the Lifelines biobank from the Netherlands. This analysis is included as a supplementary analysis below.

SUPPLEMENTARY ANALYSIS

Although the measures one can retrieve from sensors such as GPS or logging devices may not directly translate to definitions of social complexity such as those by Bergman & Beehner (2015) or Kappeler (2019), it may be possible to find associations between digital phenotyping measures and measures of social complexity which potentially can be used to measure social complexity objectively and longitudinally in humans. The preliminary analysis was performed in the Lifelines biobank, which is a multigenerational cohort containing medical and behavioral data from subjects from the northern provinces of the Netherlands. For this analysis, data from adult subjects was used only. Associations between three indices of social complexity based on the measures created in the UK Biobank in **chapter 3** and digital phenotyping measures of social behavior were examined, as well as associations between social complexity measures and cognition and between digital phenotyping measures of social behavior and cognition. Digital phenotyping was performed using a new version of the Behapp application examined in **chapter 4.** This measurement resulted in 89 measures describing application usage, calling behavior, location visitations, sleeping behavior and movement behavior. The three indices of social complexity were the number of social contexts, which is a combination of the time people spent in locations which are not work, education or home; the number of social contacts; and the number of individuals in the subjects' household. Three tests assessing cognitive skills were included. The first measure was the Mini Mental State Examination, which has been developed to detect cognitive functioning issues in elderly subjects. Secondly, we included the Cogstate, which is intended to measure several cognitive skills in order to make an assessment of brain functioning and finally the Ruff Figural Fluency Test, which is intended to assess executive functioning specifically. The associations were evaluated using correlation matrices for each combination of modalities (Behapp features, social complexity measures and cognition measurements). Sample sizes varied per comparison due to the different numbers of individuals having participated in each of the modalities. The sample sizes for correlations between Behapp features and cognition varied between 146 and 168 for all tests beside the Cogstate, for which the sample sizes varied between 82 and 98. The associations between Behapp features and social complexity measures were based on sample sizes varying between 1361 and 1603. Finally, correlations between social complexity measures and cognitive measurements were based on sample sizes varying from 2882 and 5493.

The full results from the correlation analyses can be found in Supplementary Tables 1, 2 and 3. These results should be interpreted carefully, as they contain many comparisons. Due to the preliminary nature of these analyses, we did not correct for multiple comparisons, consequently it is likely that some of the results with p-values below 0.05 contain spurious results. We found that most passive digital phenotyping measures were not associated with the social complexity measures. However, some interesting correlations were revealed. The passive digital phenotyp-

ing measures contain measures of the number of locations visited during leisure time (during the evening), which at face value bears similarity with the social complexity measure regarding the number of social contexts. The number of locations visited during leisure time was indeed significantly correlated with the number of social contexts (r = 0.07, p = 0.006). Similarly, as expected, the number of social contexts was associated with the time spent travelling (r = 0.06, p = 0.010). The number of social contexts also showed some negative associations with passive digital phenotyping features relating to the time spent on the phone, such as the app addition score, which is a measurement of the interval between smartphone application usage instances (r = -0.05, p = 0.028), and the total duration social media apps were opened (r = -0.08, p = 0.002). The self-reported number of contacts was consistently associated to digital phenotyping features relating to the calling behavior of the participants, with the number of calls (r = 0.08, p = 0.002) and the number of contacts (r = 0.09, p = 0.001), indicating that the number of real-life social contacts was associated with the number of contacts over the phone, which was expected and hints at the validity of the Behapp call measurements as a measure of social behavior. The low number of significant associations between the number of individuals in the household of the subject and the digital phenotyping measures is unsurprising, as the latter are not intended to measure particular characteristics of the home location.

The MMSE scores have limited associations with digital phenotyping features, apart from the recall score and the language score. Recall showed strong positive associations with several of the app usage digital phenotyping features, such as the app addiction score (r = 0.27, p = < 0.001) and the total duration of communication apps opened (r = 0.22, p = 0.004), while showing strong associations with the features describing the time spent in movement (e.g., mean duration walking in seconds, r = 0.29, p < 0.001; mean time spent stationary r = -0.32, p = < 0.001) with more time spent in movement having positive associations with recall. The MMSE language score, which is dependent on questions targeting language production and understanding, was associated with some app and movement measures, but more consistently with digital phenotyping measures concerning calling behavior (such as the total duration of calls in seconds, r = 0.21, p = 0.006) as well as associations with several features describing movement and sleeping times. The Cogstate scores, which represent attention, psychomotor function, working memory and visual learning, showed few strong associations with the digital phenotyping score, and no consistent associations between Cogstate measures and sets of digital phenotyping measures based on the same behavior. Finally, we included two outcomes from the RFFT, which measures executive functioning. The main outcome is the total number of unique designs, which showed several strong associations with calling behavior (e.g., number of calls, r = 0.24, p = 0.002), locations visited (e.g., number of leisure time locations visited, r = 0.28, p < 0.001) and movement (e.g., mean duration walking in seconds, r = 0.29, p < 0.001). The second outcome, the number of perseverative errors, on the other hand, showed no associations with digital phenotyping features whatsoever.

The results from the preliminary analyses show some promising associations between social complexity measures and digital phenotyping measures, but they also show that some measures of social complexity are not simply measured by smartphone behavioral measures, such as the number of individuals in the household. Similarly, although some of the cognitive measures did show associations with behaviors measurable through digital phenotyping, for several measures such associations we inconsistent or absent. The latter includes cognitive skills for which associations were previously found with self-reported social complexity in **chapter 3**. It may be interesting for future studies to assess to what extent aspects of social complexity can be measured using digital phenotyping devices such as smartphones and whether such measures also capture the associations of social complexity with cognition.

Future directions

With the goal of examining how social complexity has evolved through (recent) human evolution, the research carried out in this thesis can be used as a basis for future studies to build on. More studies are required to determine how social complexity measurement can be carried out in humans. These should include psychometric studies to ensure that the tools are reliable and valid. After the development of social complexity measures for individual human variation has been furthered, steps can be taken to examine how genetic variation associates with these measures, for example using GWAS studies. As mentioned before, while carrying out the GWAS, thought has to be put into the analyses to ensure that valuable genetic data which may have been under selection is not filtered out.

To better understand whether potential evolutionary changes in brain size and social complexity are associated, pleiotropy between the phenotypes can be examined. Genetic correlations can be examined to determine the extent to which genetic variation underlying either of the phenotypes is associated with the other. In this process it might also be interesting to look at the role of (social) cognition, for example by examining whether social cognition mediates the association between (regional) brain volume(s) and social complexity. As discussed in **chapter 1**, knowledge regarding pleiotropy is relevant because it may alter how selection affects genetic variation underlying phenotypes (Svensson et al., 2021).

After the determination of genetic variation underlying social complexity has been carried out, methodologies such as those discussed in **chapter 1** can be used to examine how genes involved in human social complexity have evolved during (recent) human evolution. By combining various methods as done in the study by Song et al. (2021) it is possible to create more certainty regarding the validity of the results. By examining specifically genes which affect both social complexity and brain size, it may be possible to determine whether these factors have co-evolved over time and whether potential reductions in brain size in recent human evolutionary history are associated with changes to the social environment of modern humans.

REFERENCES

Aureli, F., Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. *Behavioral Ecology and Sociobiology, 73*, 6. https://doi.org/10.1007/s00265-018-2604-5

Bergman, T.J. & Beehner, J.C. (2015). Measuring social complexity. *Animal behaviour, 103*, 203-209. https://doi.org/10.1016/j.anbehav.2015.02.018

Bralten, J., Mota, N. R., Klemann, C. J. H. M., De Witte, W., Laing, E., Collier, D. A., … Poelmans, G. (2021). Genetic underpinnings of sociability in the general population. *Neuropsychopharmacology*, *46*(9), 1627–1634. https://doi.org/10.1038/s41386-021-01044-z

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature ecology & evolution*, *1*(5), 112. https://doi.org/10.1038/s41559-017-0112

DeSilva, J. M., Traniello, J. F. A., Claxton, A. G., & Fannin, L. D. (2021). When and Why Did Human Brains Decrease in Size? A New Change-Point Analysis and Insights From Brain Evolution in Ants. *Frontiers in Ecology and Evolution, 9*, 712. doi:10.3389/fevo.2021.742639/bibtex

Dunbar, R.I.M. (2009) The social brain hypothesis and its implications for social evolution. *Annals of Human Biology*, 36:5, 562-572.https://doi.org/10.1080/03014460902960289

Fareri, D. S., Smith, D. V., & Delgado, M. R. (2020). The influence of relationship closeness on default-mode network connectivity during social interactions. Social cognitive and affective neuroscience, 15(3), 261–271. https://doi.org/10.1093/scan/nsaa031

Field, Y., Boyle, E. A., Telis, N., Gao, Z., Gaulton, K. J., Golan, D., Yengo, L., Rocheleau, G., Froguel, P., McCarthy, M. I., & Pritchard, J. K. (2016). Detection of human adaptation during the past 2000 years. *Science (New York, N.Y.), 354*(6313), 760–764. https://doi.org/10.1126/science.aag0776

Franklin, C., & Dwyer, D. S. (2021). Candidate risk genes for bipolar disorder are highly conserved during evolution and highly interconnected. *Bipolar disorders*, *23*(4), 400–408. https://doi.org/10.1111/bdi.12996

Guo, W., Liu, F., Liu, J., Yu, M., Zhang, Z., Liu, G., Xiao, C., & Zhao, J. (2015). Increased cerebellar-default-mode-network connectivity in drug-naive major depressive disorder at rest. *Medicine*, *94*(9), e560. https://doi.org/10.1097/MD.0000000000000560

Henneberg, M., & Steyn, M. (1993). Trends in cranial capacity and cranial index in Subsaharan Africa during the Holocene. *American journal of human biology: the official journal of the Human Biology Council, 5*(4), 473–479. https://doi.org/10.1002/ajhb.1310050411

Hopcroft R. L. (2019). Sex Differences in the Association of Family and Personal Income and Wealth with Fertility in the United States. *Human nature (Hawthorne, N.Y.)*, *30*(4), 477–495. https://doi.org/10.1007/s12110-019-09354-4

Jongs, N., Penninx, B., Arango, C., Ayuso-Mateos, J. L., van der Wee, N., Rossum, I. W., Saris,I. M. J., van Echteld, A., Koops, S., Bilderbeck, A. C., Raslescu, A., Dawson, G. R., Sommer, B., Marston, H., Vorstman, J. A., Eijkemans, M. J., & Kas, M. J. (2022). Effect of disease related biases on the subjective assessment of social functioning in Alzheimer's disease and schizophrenia patients. *Journal of psychiatric research, 145*, 302–308. https://doi.org/10.1016/j.jpsychires.2020.11.013

Kappeler, P.M. (2019). A framework for studying social complexity. *Behavioral Ecology and Sociobiology, 73*, 13. https://doi.org/10.1007/s00265-018-2601-8

Kappeler, P.M., Clutton-Brock, T., Shultz, S., & Lukas, D. (2019). Social complexity: patterns, processes and evolution. *Behavioral Ecology and Sociobiology, 73,* 5. https://doi.org/10.1007/s00265-018-2613-4

Kasap, M., Rajani, V., Rajani, J., & Dwyer, D. S. (2018). Surprising conservation of schizophrenia risk genes in lower organisms reflects their essential function and the evolution of genetic liability. *Schizophrenia research*, *202*, 120–128. https://doi.org/10.1016/j.schres.2018.07.017

Liu, C. Tang, Y., Ge, H., Wang, F., Sun, H., Meng, H., … Liu., S. (2014). Increasing breadth of the frontal lobe but decreasing height of the human brain between two Chinese samples from a Neolithic site and from living humans. *American Journal of Physical Anthropology, 154*, 94-103. https://doi.org/10.1002/ajpa.22476

Luo, X., Chen, G., Jia, Y., Gong, J., Qiu, S., Zhong, S., Zhao, L., Chen, F., Lai, S., Qi, Z., Huang, L., & Wang, Y. (2018). Disrupted Cerebellar Connectivity With the Central Executive Network and the Default-Mode Network in Unmedicated Bipolar II Disorder. *Frontiers in psychiatry*, *9*, 705. https://doi.org/10.3389/fpsyt.2018.00705

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the "default mode network" and the "social brain". *Frontiers in human neuroscience, 6*, 189. https://doi.org/10.3389/fnhum.2012.00189

Maclaren, N.G., Meng, L., Collier, M., & Masuda, N. (2023). Cooperation and the social brain hypothesis in primate social networks. arXiv:2302.00075. https://doi.org/10.48550/arXiv.2302.00075

Mullins, N., Ingason, A., Porter, H., Euesden, J., Gillett, A., Ólafsson, S., Gudbjartsson, D. F., Lewis, C. M., Sigurdsson, E., Saemundsen, E., Gudmundsson, Ó. Ó., Frigge, M. L., Kong, A., Helgason, A., Walters, G. B., Gustafsson, O., Stefansson, H., & Stefansson, K. (2017). Reproductive fitness and genetic risk of psychiatric disorders in the general population. *Nature communications*, *8*, 15833. https://doi.org/10.1038/ncomms15833

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Långström, N., Landen, M., McGuffin, P., Lewis, C. M., Lichtenstein, P., & Svensson, A. C. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry*, *70*(1), 22–30. https://doi.org/10.1001/jamapsychiatry.2013.268

Sall, S., Thompson, W., Santos, A., & Dwyer, D. S. (2021). Analysis of Major Depression Risk Genes Reveals Evolutionary Conservation, Shared Phenotypes, and Extensive Genetic Interactions. *Frontiers in psychiatry*, *12*, 698029. https://doi.org/10.3389/fpsyt.2021.698029

Svensson, E. I., Arnold, S. J., Biürger, R., Csilléry, K., Draghi, J., Henshaw, J. M., … Runemark, A. (2021). Correlational seledction in the age of genomics. *Nature Ecology & Evolution, 5*, 562-573. https://doi.org/10.1038/s41559-021-01413-3

Van Overwalle, F., Manto, M., Cattaneo, Z., Clausi, S., Ferrari, C., Gabrieli, J. D. E., Guell, X., Heleven, E., Lupo, M., Ma, Q., Michelutti, M., Olivito, G., Pu, M., Rice, L. C., Schmahmann, J. D., Siciliano, L., Sokolov, A. A., Stoodley, C. J., van Dun, K., Vandervert, L., … Leggio, M. (2020). Consensus Paper: Cerebellum and Social Cognition. *Cerebellum (London, England)*, *19*(6), 833–868. https://doi.org/10.1007/s12311-020-01155-1

Villmoare, B., & Grabowski, M. (2022). Did the transition to complex societies in the Holocene drive a reduction in brain size? A reassessment of the DeSilva et al. (2021) hypothesis. *Frontiers in Ecology and Evolution, 10*. https://doi.org/10.3389/fevo.2022.963568

Wang, L., Zou, F., Shao, Y., Ye, E., Jin, X., Tan, S., Hu, D., & Yang, Z. (2014). Disruptive changes of cerebellar functional connectivity with the default mode network in schizophrenia. *Schizophrenia research*, *160*(1-3), 67–72. https://doi.org/10.1016/j.schres.2014.09.034

Wen, T., Mitchell, D. J., & Duncan, J. (2020). The Functional Convergence and Heterogeneity of Social, Episodic, and Self-Referential Thought in the Default Mode Network. *Cerebral cortex (New York, N.Y. : 1991)*, *30*(11), 5915–5929. https://doi.org/10.1093/cercor/bhaa166

Zhang & Santtila (2022). Social status predicts different mating and reproductive success for men and women in China: evidence from the 2010–2017 CGSS data. *Behavioral Ecology and Sociobiology, 76*(101). https://doi.org/10.1007/s00265-022-03209-2

SUPPLEMENTARY MATERIALS

Supplementary Table 1

Supplementary Table 1. Associations between Behapp digital phenotyping measures and social complexity measures.

Social complexity measure / Digital phenotyping measure	Social contexts *r*	Social contexts *p*	Number of contacts *r*	Number of contacts *p*	Number of individuals in household *r*	Number of individuals in household *p*
App addiction score	-0,05	0.028*	0,01	0,600	0,06	0,022*
Total duration all apps opened in seconds	-0,06	0.013*	0,01	0,807	0,04	0,111
Total duration all apps opened at night in seconds	-0,05	0.047*	-0,01	0,571	0,05	0,054
Total duration camera apps opened in seconds	0,03	0.206	0,01	0,814	0,01	0,693
Total duration clock apps opened in seconds	-0,01	0.706	0,01	0,835	0,00	0,937
Total duration communication apps opened in seconds	-0,02	0.391	-0,02	0,524	0,03	0,198
Total duration entertainment apps opened in seconds	-0,01	0.705	0,00	0,859	0,04	0,106
Total duration health/fitness apps opened in seconds	0,04	0.108	-0,02	0,422	0,02	0,501
Total duration news/magazines apps opened in seconds	0,01	0.581	-0,01	0,823	0,05	0,059
Total duration social media apps opened in seconds	-0,08	0.002**	-0,02	0,386	0,02	0,431
Total duration communication apps opened in seconds	0,05	0.059	-0,03	0,219	0,03	0,181
Mean duration entertainment apps opened in seconds	< 0,01	0.848	-0,01	0,681	0,06	0,020*
Mean duration health/fitness apps opened in seconds	0,02	0.472	0,02	0,416	0,01	0,565
Mean duration news/magazines apps opened in seconds	0,03	0.210	-0,01	0,806	0,00	0,853
Mean duration social media apps opened in seconds	-0,03	0.166	-0,01	0,718	0,04	0,080
Number of apps used	-0,04	0.146	0,04	0,079	0,01	0,572
Number of times all apps were opened	-0,08	0.002**	0,03	0,266	0,03	0,166
Number of times all apps were opened at night.	-0,05	0.058	-0,02	0,519	0,05	0,054
Number of times communication apps were opened	-0,07	0.008**	0,00	0,968	0,03	0,186
Number of times entertainment apps were opened	-0,04	0.151	-0,01	0,683	0,04	0,106
Number of times health/fitness apps were opened	0,04	0.147	-0,01	0,570	0,01	0,790
Number of times news/magazines apps were openeed	-0,01	0.699	0,00	0,895	0,05	0,056
Number of times social media apps were opened	-0,06	0.016*	-0,01	0,689	0,00	0,939
Total duration of calls in seconds	-0,01	0.719	0,04	0,107	0,04	0,158

Supplementary Table 1. Associations between Behapp digital phenotyping measures and social complexity measures. (*continued*)

Social complexity measure	Social contexts	Social contexts	Number of	Number of	Number of individuals in	Number of individuals in
Digital phenotyping measure	*r*	*p*	contacts *r*	contacts *p*	household *r*	household *p*
Total duration of incoming calls in seconds	-0,01	0.550	0,03	0,183	0,03	0,261
Total duration of outgoing calls in seconds	< 0,01	0.926	0,04	0,101	0,04	0,141
Mean number of call contact repeats	-0,08	0.004**	0,02	0,414	0,00	0,859
Number of calls	< 0,01	0.895	0,08	0,002**	0,03	0,276
Number of calls of nonzero duration	-0,01	0.734	0,08	0,001**	0,02	0,348
Number of incoming calls	-0,02	0.419	0,07	0,003**	0,02	0,536
Number of incoming calls of nonzero duration	-0,02	0.412	0,07	0,003**	0,02	0,545
Number of missed calls	< 0,01	0.915	0,02	0,533	0,04	0,160
Number of nonresponse on outgoing calls	0,02	0.540	0,06	0,010*	0,02	0,326
Number of outgoing calls	0,01	0.787	0,08	0,002**	0,03	0,270
Number of outgoing calls of nonzero duration	< 0,01	0.867	0,08	0,002**	0,03	0,282
Number of nonrepeated call contacts	0,05	0.045*	0,11	<.001**	0,02	0,460
Number of nonrepeated call contacts incoming calls	0,04	0.159	0,09	0,001**	0,01	0,596
Number of nonrepeated call contacts outgoing calls	0,07	0.015*	0,10	< 0,001**	0,00	0,924
Number of unique contacts incoming calls	0,02	0.439	0,09	< 0,001**	0,00	0,938
Number of unique contacts missed calls	0,01	0.777	0,03	0,339	0,02	0,433
Number of unique contacts outgoing calls	0,04	0.098	0,09	0,001**	0,01	0,706
Number of unique contacts total calls	0,04	0.141	0,09	< 0,001**	0,01	0,592
call_percentage_of_missed_calls	0,04	0.125	-0,04	0,155	0,04	0,093
Standardized number of call contact repeats	-0,05	0.068	-0,02	0,434	-0,02	0,502
Mean time spent stationary in hours	-0,06	0.023*	0,01	0,819	-0,02	0,432
Mean time spent stationary in hours excluding home	-0,05	0.047*	0,00	0,961	-0,01	0,750
Mean time travelled	0,00	0.953	0,02	0,368	0,00	0,941
Normalized entropy of the time spent stationary	0,02	0.495	0,02	0,381	0,02	0,524
Normalized entropy of the number of visits to a single location	0,01	0.748	0,03	0,171	0,00	0,961
Percentage of locations visited once	-0,01	0.769	-0,04	0,077	0,03	0,233
Percentage of time spent at home	-0,03	0.236	-0,02	0,505	-0,01	0,635
Standard deviation of time travelled in hours	0,00	0.877	0,01	0,567	-0,01	0,653
Number of leisure time locations visited	0,07	0.006**	0,03	0,298	-0,03	0,194
Number of locations visited at night excluding home	0,01	0.830	0,02	0,337	0,00	0,936
Number of locations visited once	0,06	0.010*	0,03	0,301	-0,01	0,555
Number of locations visited	0,03	0.203	-0,01	0,783	-0,01	0,813
Number of travel instances	0,04	0.113	0,01	0,648	0,01	0,601

Supplementary Table 1. Associations between Behapp digital phenotyping measures and social complexity measures. (*continued*)

Social complexity measure	Social contexts	Social contexts	Number of	Number of	Number of individuals in	Number of individuals in
Digital phenotyping measure	*r*	*p*	contacts *r*	contacts *p*	household *r*	household *p*
Number of unique locations visited during leisure time	0,05	0.041*	0,03	0,314	0,00	0,859
Number of unique locations visited at night	0,01	0.798	0,04	0,161	-0,02	0,519
Number of unique locations visited	0,07	0.003**	0,02	0,336	0,02	0,444
Total time spent at home in hours	-0,03	0.225	-0,02	0,484	-0,01	0,641
Total time spent outside including travel in hours	0,04	0.100	0,03	0,195	-0,01	0,567
Time spent stationary in hours	-0,01	0.565	-0,01	0,762	-0,02	0,443
Time spent stationary in hours excluding the home location	0,02	0.317	0,02	0,526	-0,02	0,480
Time spent travelling in hours	0,06	0.010*	0,06	0,014*	0,00	0,908
Mean durations running in seconds	0,06	0.009**	-0,02	0,392	-0,03	0,277
Mean duration still in seconds	0,02	0.431	0,00	0,873	0,02	0,442
Mean duration travelling in vehicle in seconds	-0,01	0.715	-0,01	0,574	-0,01	0,777
Mean duration walking in seconds	0,02	0.474	0,05	0,0390*	-0,02	0,497
Number of biking instances	0,00	0.968	0,01	0,687	0,03	0,183
Number of running instances	0,07	0.008**	-0,03	0,207	0,02	0,461
Number of still instances	0,02	0.539	0,01	0,582	0,01	0,648
Number of vehicle travel instances	0,00	0.862	0,03	0,198	0,03	0,228
Number of walking instances	0,01	0.713	0,05	0,040*	0,02	0,470
Average bed time	0,02	0.420	0,04	0,130	0,03	0,188
Standard deviation of bed time	-0,02	0.515	0,02	0,511	0,02	0,427
Time spent using phone in seconds	0,03	0.177	-0,01	0,682	-0,01	0,748
Number of times phone was used	-0,06	0.014*	0,01	0,730	0,04	0,092
Number of times phone was used at night	-0,07	0.003**	0,03	0,243	0,04	0,147
Time spent using phone at night in seconds	-0,05	0.066	-0,02	0,527	0,05	0,052
Average sleep duration	-0,05	0.048*	-0,01	0,567	0,05	0,051
Standard deviation of sleep duration	0,02	0.357	-0,02	0,382	0,01	0,721
Average waking time	0,04	0.116	-0,01	0,760	-0,01	0,587
Standard deviation of waking time	0,02	0.402	-0,02	0,394	0,02	0,470
Total duration of step taking in seconds	0,01	0.810	0,03	0,269	0,05	0,039*
Total duration of step taking at night in seconds	0,02	0.327	0,00	0,942	0,03	0,283
Number of steps taken	-0,02	0.457	0,01	0,783	-0,01	0,739
Number of steps taken at night	0,03	0.203	0,01	0,675	0,03	0,182

* = $p < 0.05$, ** = $p < 0.01$.

Supplementary Table 2. Associations between Behapp digital phenotyping measures and MMSE outcomes.

MMSE outcome												
Digital phenotyping measure	MMSE orientation score	MMSE orientation score *p*	MMSE registration score *r*	MMSE registration score *p*	MMSE attention / calculation score *r*	MMSE attention / calculation score *p*	MMSE recall score r	MMSE recall score *p*	MMSE language score *r*	MMSE language score *p*	MMSE total score *r*	MMSE total score *p*
App addiction score	0,06	0,433	-0,04	0,591	-0,01	0,873	0,27	< 0,001**	0,19	0,013*	0,16	0,038*
Total duration all apps opened in seconds	0,09	0,265	-0,02	0,807	-0,11	0,147	0,22	0,004**	0,11	0,141	0,11	0,175
Total duration all apps opened at night in seconds	-0,08	0,285	0,03	0,690	-0,04	0,597	0,06	0,426	-0,09	0,225	-0,06	0,459
Total duration camera apps opened in seconds	0,17	0,032	0,06	0,459	0,07	0,378	0,16	0,039*	0,09	0,249	0,21	0,005**
Total duration clock apps opened in seconds	0,07	0,368	-0,04	0,563	0,01	0,854	0,03	0,691	0,05	0,538	0,05	0,518
Total duration communication apps opened in seconds	0,06	0,462	0,02	0,763	-0,05	0,512	0,22	0,004**	0,21	0,006	0,16	0,043
Total duration entertainment apps opened in seconds	-0,02	0,807	0,01	0,874	0,02	0,773	0,07	0,391	0,02	0,809	0,03	0,687
Total duration health/fitness apps opened in seconds	0,09	0,250	0,04	0,646	-0,01	0,908	0,11	0,159	0,02	0,797	0,10	0,192
Total duration news/magazines apps opened in seconds	0,13	0,089	0,05	0,549	-0,02	0,769	-0,03	0,709	0,03	0,668	0,07	0,376
Total duration social media apps opened in seconds	0,08	0,299	-0,08	0,289	0,05	0,502	0,09	0,228	-0,05	0,505	0,05	0,527
Total duration communication apps opened in seconds	0,04	0,571	0,02	0,792	0,03	0,668	0,17	0,027*	0,22	0,005**	0,16	0,037*
Mean duration entertainment apps opened in seconds	0,07	0,399	-0,01	0,897	0,00	0,952	0,11	0,145	-0,04	0,575	0,06	0,449
Mean duration health/fitness apps opened in seconds	0,09	0,241	0,08	0,288	0,00	0,979	0,10	0,220	0,03	0,695	0,11	0,141
Mean duration news/magazines apps opened in seconds	-0,01	0,887	0,00	0,954	-0,06	0,439	0,00	0,988	0,01	0,946	-0,03	0,691
Mean duration social media apps opened in seconds	0,10	0,203	-0,03	0,669	0,01	0,939	0,12	0,132	0,10	0,198	0,11	0,147
Number of apps used	0,00	0,975	-0,04	0,579	0,08	0,322	0,18	0,022*	0,18	0,022*	0,12	0,114
Number of times all apps were opened	0,09	0,255	-0,07	0,393	-0,04	0,645	0,23	0,003**	0,13	0,091	0,13	0,101
Number of times all apps were opened at night.	-0,02	0,778	0,03	0,668	-0,15	0,059	0,19	0,014*	0,02	0,845	0,02	0,815
Number of times communication apps were opened	0,03	0,663	-0,06	0,444	-0,04	0,646	0,20	0,011*	0,21	0,007**	0,11	0,152

Supplementary Table 2. Associations between Behapp digital phenotyping measures and MMSE outcomes. (*continued*)

MMSE outcome												
Digital phenotyping measure	MMSE orientation score	MMSE orientation score *p*	MMSE registration score *r*	MMSE registration score *p*	MMSE attention / calculation score *r*	MMSE attention / calculation score *p*	MMSE recall score r	MMSE recall score *p*	MMSE language score *r*	MMSE language score *p*	MMSE total score *r*	MMSE total score *p*
Number of times entertainment apps were opened	-0,03	0,742	-0,03	0,718	0,04	0,628	0,10	0,199	0,06	0,459	0,04	0,563
Number of times health/fitness apps were opened	0,07	0,343	0,04	0,602	0,02	0,755	0,09	0,223	0,03	0,724	0,10	0,185
Number of times news/magazines apps were openeed	0,17	0,026*	0,02	0,770	0,03	0,735	0,04	0,642	-0,01	0,887	0,11	0,148
Number of times social media apps were opened	0,09	0,240	-0,08	0,328	0,07	0,363	0,10	0,182	-0,08	0,309	0,06	0,442
Total duration of calls in seconds	-0,13	0,089	0,00	0,973	0,08	0,277	0,04	0,577	0,21	0,006**	0,03	0,663
Total duration of incoming calls in seconds	-0,06	0,410	-0,01	0,932	0,09	0,265	0,05	0,541	0,17	0,027*	0,06	0,457
Total duration of outgoing calls in seconds	-0,16	0,043*	0,01	0,899	0,06	0,429	0,03	0,709	0,19	0,013*	0,00	0,949
Mean number of call contact repeats	0,00	0,989	0,00	0,970	0,06	0,465	-0,03	0,756	-0,06	0,466	0,00	0,980
Number of calls	-0,10	0,216	-0,06	0,433	0,11	0,158	0,14	0,069	0,16	0,038*	0,06	0,411
Number of calls of nonzero duration	-0,11	0,161	-0,08	0,277	0,11	0,167	0,14	0,066	0,17	0,028*	0,05	0,503
Number of incoming calls	-0,07	0,401	-0,11	0,138	0,12	0,108	0,11	0,150	0,14	0,065	0,05	0,504
Number of incoming calls of nonzero duration	-0,07	0,388	-0,11	0,153	0,12	0,109	0,11	0,155	0,14	0,064	0,05	0,504
Number of missed calls	0,01	0,904	0,05	0,521	0,12	0,118	0,04	0,583	0,08	0,275	0,10	0,184
Number of nonresponse on outgoing calls	-0,08	0,281	-0,02	0,758	0,01	0,878	0,13	0,087	0,10	0,217	0,03	0,737
Number of outgoing calls	-0,12	0,107	-0,04	0,574	0,07	0,382	0,15	0,054	0,16	0,043*	0,04	0,593
Number of outgoing calls of nonzero duration	-0,13	0,096	-0,05	0,548	0,08	0,301	0,14	0,065	0,16	0,034*	0,04	0,580
Number of nonrepeated call contacts	-0,10	0,251	-0,05	0,582	0,13	0,121	0,12	0,165	0,20	0,017*	0,07	0,424
Number of nonrepeated call contacts incoming calls	-0,18	0,031*	-0,13	0,112	0,22	0,009**	0,09	0,262	0,15	0,067	0,01	0,934
Number of nonrepeated call contacts outgoing calls	-0,12	0,133	-0,10	0,244	0,06	0,476	0,11	0,197	0,16	0,059	-0,01	0,949
Number of unique contacts incoming calls	-0,14	0,082	-0,13	0,129	0,21	0,012*	0,08	0,351	0,12	0,135	0,01	0,893
Number of unique contacts missed calls	0,03	0,754	0,05	0,549	0,15	0,074	-0,06	0,458	0,06	0,508	0,08	0,355
Number of unique contacts outgoing calls	-0,14	0,098	-0,07	0,434	0,07	0,390	0,12	0,157	0,14	0,101	0,00	0,974

Supplementary Table 2. Associations between Behapp digital phenotyping measures and MMSE outcomes. (*continued*)

MMSE outcome												
Digital phenotyping measure	MMSE orientation score	MMSE orientation score *p*	MMSE registration score *r*	MMSE registration score *p*	MMSE attention / calculation score *r*	MMSE attention / calculation score *p*	MMSE recall score r	MMSE recall score *p*	MMSE language score *r*	MMSE language score *p*	MMSE total score *r*	MMSE total score *p*
Number of unique contacts total calls	-0,12	0,154	-0,06	0,448	0,13	0,111	0,11	0,201	0,15	0,067	0,04	0,668
call_percentage_of_missed_calls	0,12	0,132	0,08	0,283	0,07	0,399	-0,01	0,885	0,11	0,159	0,14	0,072
Standardized number of call contact repeats	0,03	0,741	-0,01	0,897	0,00	0,966	-0,09	0,289	-0,02	0,820	-0,03	0,731
Mean time spent stationary in hours	-0,12	0,126	0,05	0,554	0,01	0,858	-0,32	< 0,001**	-0,03	0,716	-0,17	0,029*
Mean time spent stationary in hours excluding home	-0,05	0,507	0,01	0,894	0,12	0,108	-0,28	< 0,001**	0,04	0,577	-0,08	0,321
Mean time travelled	-0,05	0,514	0,00	0,998	0,03	0,713	0,05	0,502	0,03	0,673	0,01	0,882
Normalized entropy of the time spent stationary	0,03	0,744	-0,06	0,472	0,13	0,101	-0,10	0,191	-0,05	0,488	-0,02	0,799
Normalized entropy of the number of visits to a single location	-0,08	0,293	-0,03	0,717	0,11	0,154	-0,08	0,282	-0,11	0,141	-0,07	0,343
Percentage of locations visited once	0,06	0,454	0,02	0,802	0,05	0,488	-0,03	0,682	-0,02	0,776	0,04	0,570
Percentage of time spent at home	0,03	0,721	-0,03	0,681	-0,07	0,341	0,04	0,580	0,08	0,293	0,02	0,841
Standard deviation of time travelled in hours	-0,02	0,784	-0,10	0,216	0,01	0,931	0,03	0,704	0,02	0,752	-0,03	0,694
Number of leisure time locations visited	0,11	0,146	-0,05	0,512	0,06	0,432	0,18	0,020*	0,08	0,282	0,15	0,053
Number of locations visited at night excluding home	0,03	0,721	-0,11	0,142	0,04	0,589	0,01	0,895	0,04	0,623	0,00	0,975
Number of locations visited once	0,13	0,082	-0,03	0,656	0,05	0,522	0,20	0,008**	-0,05	0,553	0,14	0,074
Number of locations visited	0,19	0,013*	-0,05	0,550	-0,01	0,919	0,16	0,043	0,05	0,510	0,15	0,058
Number of travel instances	0,16	0,038*	-0,07	0,347	-0,04	0,590	0,19	0,011*	0,06	0,428	0,12	0,111
Number of unique locations visited during leisure time	-0,05	0,509	-0,06	0,416	0,12	0,128	-0,02	0,772	0,05	0,546	-0,01	0,926
Number of unique locations visited at night	0,15	0,046*	-0,09	0,223	-0,04	0,632	0,11	0,144	0,05	0,480	0,09	0,267
Number of unique locations visited	0,10	0,182	-0,06	0,404	0,04	0,582	0,25	0,001**	-0,01	0,907	0,13	0,090
Total time spent at home in hours	0,02	0,752	-0,03	0,659	-0,07	0,335	0,05	0,549	0,09	0,270	0,01	0,852

Supplementary Table 2. Associations between Behapp digital phenotyping measures and MMSE outcomes. (*continued*)

MMSE outcome												
Digital phenotyping measure	MMSE orientation score	MMSE orientation score *p*	MMSE registration score *r*	MMSE registration score *p*	MMSE attention / calculation score *r*	MMSE attention / calculation score *p*	MMSE recall score r	MMSE recall score *p*	MMSE language score *r*	MMSE language score *p*	MMSE total score *r*	MMSE total score *p*
Total time spent outside including travel in hours	0,11	0,148	-0,07	0,383	0,09	0,225	0,11	0,151	0,00	0,989	0,10	0,177
Time spent stationary in hours	0,07	0,361	-0,06	0,424	-0,01	0,942	0,09	0,239	0,08	0,275	0,07	0,382
Time spent stationary in hours excluding the home location	0,08	0,311	-0,05	0,535	0,11	0,147	0,08	0,322	0,00	0,976	0,09	0,247
Time spent travelling in hours	0,16	0,044*	-0,09	0,233	-0,02	0,805	0,16	0,042*	0,00	0,962	0,09	0,240
Mean durations running in seconds	0,13	0,102	0,01	0,853	0,00	0,996	-0,09	0,227	0,04	0,586	0,04	0,619
Mean duration still in seconds	0,10	0,210	-0,12	0,120	-0,02	0,798	0,02	0,844	0,00	0,972	0,00	0,962
Mean duration travelling in vehicle in seconds	-0,03	0,717	-0,03	0,664	0,04	0,581	0,13	0,081	0,24	0,002**	0,10	0,200
Mean duration walking in seconds	0,04	0,605	-0,18	0,017*	0,09	0,259	0,29	< .001**	0,13	0,098	0,14	0,079
Number of biking instances	0,03	0,743	0,01	0,944	0,11	0,155	0,12	0,129	0,16	0,034*	0,14	0,071
Number of running instances	0,05	0,546	-0,01	0,861	-0,20	0,009**	0,17	0,027	0,06	0,431	0,02	0,789
Number of still instances	0,09	0,245	-0,20	0,010*	-0,05	0,492	0,05	0,532	-0,03	0,717	-0,03	0,677
Number of vehicle travel instances	0,07	0,367	-0,04	0,644	0,06	0,444	0,21	0,006**	0,20	0,009**	0,18	0,023*
Number of walking instances	0,01	0,870	-0,12	0,117	-0,03	0,714	0,26	0,001**	0,14	0,066	0,09	0,262
Average bed time	0,07	0,397	0,01	0,941	0,08	0,323	0,20	0,011*	0,17	0,026*	0,18	0,019*
Standard deviation of bed time	-0,02	0,809	0,02	0,759	-0,06	0,418	0,16	0,036	-0,10	0,204	0,01	0,864
Time spent using phone in seconds	-0,01	0,849	-0,07	0,357	0,06	0,438	-0,10	0,202	0,11	0,145	-0,02	0,785
Number of times phone was used	0,08	0,294	-0,02	0,800	-0,11	0,159	0,21	0,005**	0,12	0,119	0,10	0,183
Number of times phone was used at night	0,09	0,262	-0,07	0,383	-0,04	0,640	0,23	0,003**	0,13	0,091	0,12	0,108
Time spent using phone at night in seconds	-0,02	0,769	0,03	0,672	-0,15	0,058	0,19	0,015*	0,02	0,843	0,02	0,827
Average sleep duration	-0,09	0,260	0,03	0,693	-0,04	0,571	0,06	0,448	-0,09	0,228	-0,06	0,426
Standard deviation of sleep duration	0,00	0,987	-0,04	0,588	0,04	0,615	0,01	0,931	0,22	0,005**	0,06	0,468

Supplementary Table 2. Associations between Behapp digital phenotyping measures and MMSE outcomes. (*continued*)

MMSE outcome												
Digital phenotyping measure	MMSE orientation score	MMSE orientation score *p*	MMSE registration score *r*	MMSE registration score *p*	MMSE attention / calculation score *r*	MMSE attention / calculation score *p*	MMSE recall score r	MMSE recall score *p*	MMSE language score *r*	MMSE language score *p*	MMSE total score *r*	MMSE total score *p*
Average waking time	-0,04	0,599	-0,04	0,574	0,04	0,597	-0,06	0,450	0,12	0,126	-0,02	0,837
Standard deviation of waking time	-0,02	0,803	-0,04	0,588	0,02	0,769	0,09	0,248	0,24	0,001**	0,08	0,315
Total duration of step taking in seconds	0,08	0,296	0,07	0,343	0,10	0,196	0,16	0,038*	0,13	0,083	0,19	0,011*
Total duration of step taking at night in seconds	0,10	0,199	0,07	0,375	0,11	0,139	0,15	0,059	0,10	0,177	0,20	0,011*
Number of steps taken	0,00	0,985	0,05	0,516	0,03	0,700	0,06	0,434	0,01	0,915	0,05	0,513
Number of steps taken at night	0,10	0,212	0,07	0,360	0,09	0,268	0,07	0,366	0,11	0,164	0,16	0,043*

* = $p < 0.05$, ** = $p < 0.01$.

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes.

Cogstate outcome	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Digital phenotyping measure										
App addiction score	0,11	0,296	0,09	0,378	0,04	0,684	-0,01	0,888	0,13	0,193
Total duration all apps opened in seconds	0,11	0,286	0,07	0,492	0,00	0,972	-0,08	0,453	0,09	0,362
Total duration all apps opened at night in seconds	-0,08	0,455	-0,06	0,576	-0,02	0,812	0,00	0,985	0,09	0,401
Total duration camera apps opened in seconds	0,07	0,508	-0,01	0,950	0,04	0,689	0,01	0,889	-0,01	0,895
Total duration clock apps opened in seconds	-0,04	0,699	-0,23	0,028	0,14	0,171	0,02	0,862	-0,04	0,668
Total duration communication apps opened in seconds	0,11	0,286	0,13	0,204	-0,08	0,423	0,00	0,969	0,04	0,674
Total duration entertainment apps opened in seconds	0,03	0,750	-0,10	0,316	0,25	0,012*	0,02	0,879	0,06	0,558
Total duration health/fitness apps opened in seconds	0,02	0,870	0,13	0,202	-0,04	0,674	-0,13	0,213	0,09	0,367
Total duration news/magazines apps opened in seconds	-0,17	0,103	-0,24	0,018*	0,03	0,788	-0,06	0,545	-0,03	0,775
Total duration social media apps opened in seconds	0,01	0,929	0,00	0,972	-0,18	0,070	-0,04	0,704	-0,01	0,895
Total duration communication apps opened in seconds	0,07	0,479	0,08	0,468	-0,01	0,910	-0,07	0,507	0,06	0,582
Mean duration entertainment apps opened in seconds	0,11	0,275	0,01	0,910	0,17	0,090	0,13	0,201	0,19	0,062
Mean duration health/fitness apps opened in seconds	0,08	0,443	0,06	0,570	-0,07	0,464	-0,08	0,431	0,08	0,427

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes. (*continued*)

Cogstate outcome Digital phenotyping measure	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Mean duration news/magazines apps opened in seconds	-0,07	0,468	-0,12	0,257	0,12	0,252	-0,04	0,730	0,04	0,678
Mean duration social media apps opened in seconds	-0,03	0,785	0,16	0,117	-0,04	0,708	-0,11	0,265	0,02	0,880
Number of apps used	0,17	0,089	0,10	0,316	0,04	0,723	-0,04	0,698	0,08	0,415
Number of times all apps were opened	0,12	0,229	0,06	0,588	0,03	0,782	0,00	0,977	0,17	0,087
Number of times all apps were opened at night.	-0,02	0,812	-0,02	0,862	0,07	0,477	0,05	0,606	0,14	0,178
Number of times communication apps were opened	0,15	0,139	0,10	0,333	-0,06	0,560	0,11	0,275	0,12	0,231
Number of times entertainment apps were opened	0,03	0,753	-0,15	0,136	0,26	0,008	0,01	0,892	0,00	0,962
Number of times health/fitness apps were opened	0,03	0,803	0,14	0,178	-0,01	0,945	-0,16	0,114	0,11	0,257
Number of times news/magazines apps were openeed	-0,04	0,726	-0,17	0,098	0,11	0,285	0,01	0,960	0,05	0,641
Number of times social media apps were opened	0,08	0,406	-0,06	0,593	-0,08	0,455	0,08	0,404	0,03	0,762
Total duration of calls in seconds	0,12	0,242	0,24	0,018*	-0,02	0,834	-0,02	0,851	0,14	0,180
Total duration of incoming calls in seconds	0,14	0,166	0,20	0,057	0,07	0,510	0,03	0,734	0,19	0,054
Total duration of outgoing calls in seconds	0,06	0,550	0,23	0,026*	-0,12	0,241	-0,08	0,450	0,03	0,768
Mean number of call contact repeats	0,07	0,523	0,12	0,288	0,20	0,066	0,20	0,072	0,21	0,056

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes. (*continued*)

Cogstate outcome Digital phenotyping measure	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Number of calls	0,08	0,456	0,21	0,042*	-0,07	0,462	0,07	0,465	0,16	0,113
Number of calls of nonzero duration	0,10	0,345	0,22	0,032*	-0,08	0,409	0,07	0,498	0,17	0,100
Number of incoming calls	0,15	0,131	0,21	0,041*	-0,01	0,946	0,10	0,316	0,23	0,021*
Number of incoming calls of nonzero duration	0,15	0,133	0,21	0,041*	0,00	0,962	0,10	0,310	0,23	0,021*
Number of missed calls	0,07	0,511	0,11	0,307	0,10	0,348	0,08	0,445	0,21	0,037*
Number of nonresponse on outgoing calls	-0,07	0,519	0,15	0,138	-0,16	0,113	0,09	0,380	0,05	0,596
Number of outgoing calls	0,02	0,872	0,20	0,056	-0,14	0,154	0,04	0,673	0,09	0,376
Number of outgoing calls of nonzero duration	0,04	0,716	0,20	0,054	-0,13	0,190	0,03	0,785	0,09	0,352
Number of nonrepeated call contacts	0,07	0,513	0,20	0,071	-0,22	0,048*	-0,07	0,508	0,02	0,832
Number of nonrepeated call contacts incoming calls	0,04	0,734	0,16	0,156	-0,15	0,163	0,02	0,876	0,13	0,238
Number of nonrepeated call contacts outgoing calls	0,07	0,548	0,23	0,042*	-0,26	0,016*	-0,06	0,616	0,07	0,543
Number of unique contacts incoming calls	0,04	0,741	0,16	0,147	-0,16	0,153	0,02	0,878	0,11	0,298
Number of unique contacts missed calls	-0,05	0,633	0,06	0,593	-0,05	0,674	0,04	0,751	0,11	0,315
Number of unique contacts outgoing calls	0,01	0,953	0,21	0,059	-0,24	0,029*	-0,02	0,861	0,05	0,676
Number of unique contacts total calls	0,01	0,897	0,21	0,057	-0,23	0,037*	-0,01	0,912	0,06	0,573
call_percentage_of_missed_calls	-0,09	0,374	-0,17	0,108	-0,02	0,827	0,12	0,232	-0,02	0,881

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes. (*continued*)

Cogstate outcome Digital phenotyping measure	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Standardized number of call contact repeats	0,11	0,321	0,06	0,573	0,24	0,026*	0,22	0,040*	0,19	0,088
Mean time spent stationary in hours	-0,09	0,393	-0,10	0,358	0,00	0,962	-0,03	0,734	-0,05	0,622
Mean time spent stationary in hours excluding home	0,17	0,094	0,10	0,356	0,08	0,457	0,03	0,739	0,13	0,187
Mean time travelled	0,04	0,709	0,00	0,980	-0,03	0,733	-0,08	0,414	-0,09	0,353
Normalized entropy of the time spent stationary	0,11	0,267	0,09	0,386	0,06	0,525	-0,01	0,889	0,15	0,143
Normalized entropy of the number of visits to a single location	0,02	0,850	0,06	0,546	0,04	0,672	-0,07	0,498	0,09	0,393
Percentage of locations visited once	0,14	0,155	-0,04	0,677	-0,01	0,883	0,00	0,999	0,20	0,042*
Percentage of time spent at home	-0,11	0,287	-0,20	0,054	0,10	0,325	-0,10	0,339	-0,08	0,459
Standard deviation of time travelled in hours	0,13	0,208	0,09	0,391	-0,03	0,792	0,16	0,109	0,18	0,073
Number of leisure time locations visited	0,04	0,661	-0,06	0,541	0,08	0,403	-0,09	0,400	0,08	0,446
Number of locations visited at night excluding home	0,12	0,244	-0,04	0,734	0,19	0,062	-0,07	0,502	0,13	0,188
Number of locations visited once	-0,03	0,785	-0,03	0,750	-0,06	0,539	-0,11	0,275	0,09	0,382
Number of locations visited	0,13	0,197	-0,02	0,864	0,10	0,337	-0,10	0,342	0,11	0,265
Number of travel instances	0,13	0,201	-0,04	0,698	0,08	0,449	-0,09	0,396	0,10	0,316
Number of unique locations visited during leisure time	0,15	0,148	0,07	0,516	0,01	0,907	0,02	0,860	0,17	0,086

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes. (*continued*)

Cogstate outcome Digital phenotyping measure	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Number of unique locations visited at night	0,00	0,968	-0,10	0,347	0,07	0,472	0,01	0,932	0,15	0,127
Number of unique locations visited	0,07	0,490	0,04	0,725	-0,03	0,786	-0,06	0,559	0,11	0,292
Total time spent at home in hours	-0,10	0,327	-0,20	0,055	0,10	0,309	-0,10	0,327	-0,07	0,505
Total time spent outside including travel in hours	0,09	0,373	-0,03	0,775	0,08	0,415	0,01	0,943	0,16	0,122
Time spent stationary in hours	-0,06	0,587	-0,20	0,057	0,14	0,172	-0,08	0,408	0,01	0,920
Time spent stationary in hours excluding the home location	0,07	0,505	-0,03	0,758	0,09	0,393	0,01	0,905	0,14	0,171
Time spent travelling in hours	0,12	0,225	-0,01	0,937	0,03	0,773	-0,01	0,918	0,13	0,198
Mean durations running in seconds	0,25	0,012*	0,20	0,054	-0,18	0,081	-0,03	0,755	0,22	0,032*
Mean duration still in seconds	-0,02	0,877	0,06	0,556	-0,09	0,392	0,07	0,473	-0,06	0,564
Mean duration travelling in vehicle in seconds	0,15	0,134	0,11	0,269	-0,05	0,625	0,01	0,954	0,09	0,399
Mean duration walking in seconds	0,11	0,263	0,09	0,381	-0,02	0,822	-0,02	0,808	0,13	0,211
Number of biking instances	0,07	0,517	0,10	0,345	-0,10	0,330	0,04	0,725	0,23	0,024*
Number of running instances	0,22	0,033*	0,17	0,098	-0,14	0,180	-0,07	0,471	0,09	0,359
Number of still instances	0,00	0,997	0,07	0,491	-0,09	0,371	0,10	0,326	0,03	0,764
Number of vehicle travel instances	0,13	0,214	0,14	0,177	-0,13	0,210	0,03	0,767	0,17	0,099
Number of walking instances	0,07	0,489	0,08	0,466	-0,11	0,283	-0,12	0,250	0,04	0,670
Average bed time	0,08	0,408	0,09	0,386	-0,10	0,330	0,07	0,514	0,20	0,047
Standard deviation of bed time	-0,16	0,114	-0,11	0,272	0,18	0,068	-0,02	0,870	0,00	0,990
Time spent using phone in seconds	0,11	0,299	0,12	0,249	-0,27	0,008**	-0,02	0,839	0,00	0,971

Supplementary Table 3. . Associations between Behapp digital phenotyping measures and Cogstate outcomes. (*continued*)

Cogstate outcome	Cogstate z-score Identification task *r*	Cogstate z-score Identification task *p*	Cogstate z-score Detection task *r*	Cogstate z-score Detection task *p*	Cogstate z-score One Card Learning task *r*	Cogstate z-score One Card Learning task *p*	Cogstate z-score One-back task *r*	Cogstate z-score One-back task *p*	Cogstate secondary outcome z-score One-back task *r*	Cogstate secondary outcome z-score One-back task *p*
Digital phenotyping measure										
Number of times phone was used	0,11	0,276	0,08	0,435	0,00	0,978	-0,08	0,453	0,10	0,344
Number of times phone was used at night	0,12	0,233	0,06	0,573	0,03	0,787	0,00	0,984	0,17	0,086
Time spent using phone at night in seconds	-0,03	0,807	-0,02	0,861	0,07	0,480	0,05	0,609	0,14	0,181
Average sleep duration	-0,08	0,453	-0,06	0,579	-0,03	0,784	0,00	0,991	0,08	0,419
Standard deviation of sleep duration	0,21	0,036*	0,16	0,122	-0,13	0,185	0,00	0,977	0,04	0,663
Average waking time	0,10	0,326	0,14	0,179	-0,24	0,018*	0,04	0,693	-0,01	0,917
Standard deviation of waking time	0,20	0,050	0,15	0,135	-0,08	0,439	-0,01	0,923	0,06	0,534
Total duration of step taking in seconds	0,06	0,572	0,05	0,601	-0,01	0,919	-0,05	0,651	0,20	0,049*
Total duration of step taking at night in seconds	0,03	0,742	0,09	0,375	-0,08	0,428	-0,05	0,622	0,19	0,056
Number of steps taken	-0,12	0,242	-0,02	0,837	-0,03	0,737	-0,06	0,533	0,05	0,591
Number of steps taken at night	0,09	0,377	0,12	0,247	-0,01	0,958	-0,05	0,640	0,20	0,048*

* = $p < 0.05$, ** = $p < 0.01$.

Supplementary Table 4

Supplementary Table 4. Associations between Behapp digital phenotyping measures and RFFT outcomes.

RFFT outcome / Digital phenotyping measure	RFFT perseverative errors *r*	RFFT perseverative errors *p*	RFFT unique designs *r*	RFFT unique designs *p*
App addiction score	0,03	0,689	0,22	0,005
Total duration all apps opened in seconds	-0,07	0,342	0,05	0,550
Total duration all apps opened at night in seconds	-0,07	0,340	-0,07	0,344
Total duration camera apps opened in seconds	0,02	0,841	0,13	0,082
Total duration clock apps opened in seconds	-0,09	0,268	-0,11	0,147
Total duration communication apps opened in seconds	-0,02	0,799	0,14	0,077
Total duration entertainment apps opened in seconds	-0,01	0,855	0,13	0,090
Total duration health/fitness apps opened in seconds	0,02	0,840	0,00	0,977
Total duration news/magazines apps opened in seconds	0,08	0,299	0,02	0,823
Total duration social media apps opened in seconds	-0,05	0,505	0,00	0,973
Total duration communication apps opened in seconds	-0,06	0,444	0,03	0,725
Mean duration entertainment apps opened in seconds	-0,05	0,543	0,19	0,014
Mean duration health/fitness apps opened in seconds	-0,03	0,669	0,02	0,753
Mean duration news/magazines apps opened in seconds	-0,03	0,706	-0,03	0,714
Mean duration social media apps opened in seconds	-0,12	0,128	0,05	0,518
Number of apps used	0,01	0,939	0,19	0,015*
Number of times all apps were opened	0,06	0,404	0,19	0,014*
Number of times all apps were opened at night.	-0,04	0,603	0,02	0,766
Number of times communication apps were opened	0,12	0,119	0,28	< 0,001**
Number of times entertainment apps were opened	0,00	0,999	0,17	0,023*
Number of times health/fitness apps were opened	0,00	0,969	-0,02	0,819
Number of times news/magazines apps were openeed	0,08	0,281	0,01	0,894
Number of times social media apps were opened	-0,04	0,633	0,05	0,520
Total duration of calls in seconds	0,05	0,500	0,23	0,002**
Total duration of incoming calls in seconds	-0,03	0,681	0,21	0,006**
Total duration of outgoing calls in seconds	0,11	0,151	0,20	0,011*
Mean number of call contact repeats	-0,02	0,833	0,17	0,040*
Number of calls	0,01	0,879	0,23	0,002**
Number of calls of nonzero duration	0,01	0,908	0,24	0,002**
Number of incoming calls	0,00	0,951	0,21	0,007**
Number of incoming calls of nonzero duration	0,01	0,947	0,21	0,008**
Number of missed calls	0,06	0,417	0,17	0,028*
Number of nonresponse on outgoing calls	-0,04	0,566	0,10	0,217
Number of outgoing calls	-0,01	0,937	0,20	0,009**
Number of outgoing calls of nonzero duration	0,01	0,937	0,22	0,004**
Number of nonrepeated call contacts	-0,01	0,914	0,21	0,010*

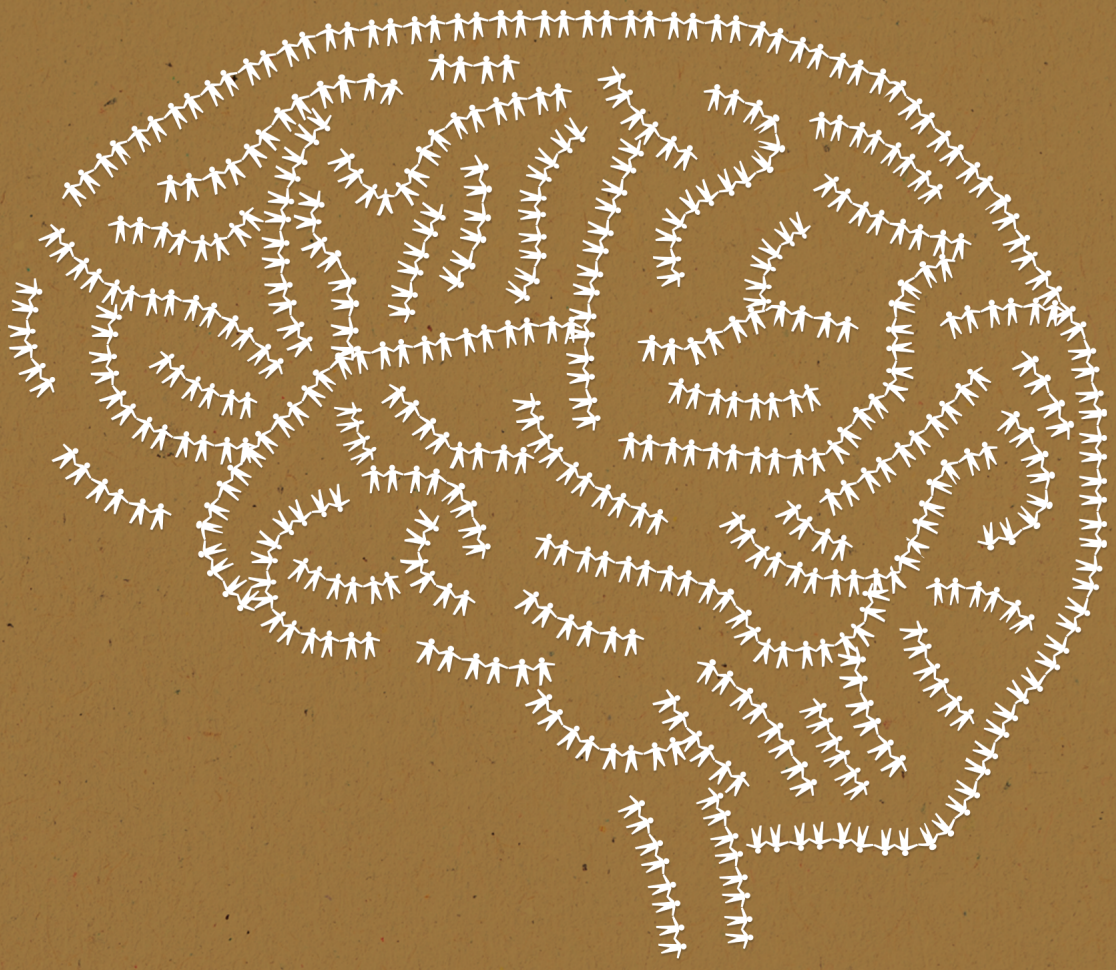
Supplementary Table 4. Associations between Behapp digital phenotyping measures and RFFT outcomes. (*continued*)

RFFT outcome Digital phenotyping measure	RFFT perseverative errors *r*	RFFT perseverative errors *p*	RFFT unique designs *r*	RFFT unique designs *p*
Number of nonrepeated call contacts incoming calls	-0,03	0,713	0,24	0,004**
Number of nonrepeated call contacts outgoing calls	-0,02	0,798	0,19	0,025*
Number of unique contacts incoming calls	-0,02	0,769	0,19	0,022*
Number of unique contacts missed calls	0,00	0,997	0,05	0,549
Number of unique contacts outgoing calls	-0,06	0,447	0,17	0,042*
Number of unique contacts total calls	-0,04	0,604	0,19	0,021*
call_percentage_of_missed_calls	0,09	0,263	0,06	0,463
Standardized number of call contact repeats	0,02	0,789	0,12	0,145
Mean time spent stationary in hours	0,12	0,112	-0,20	0,009**
Mean time spent stationary in hours excluding home	-0,05	0,560	-0,04	0,620
Mean time travelled	0,02	0,790	0,00	0,997
Normalized entropy of the time spent stationary	-0,04	0,569	0,17	0,028*
Normalized entropy of the number of visits to a single location	0,01	0,885	0,13	0,093
Percentage of locations visited once	-0,03	0,669	0,14	0,077
Percentage of time spent at home	0,05	0,534	-0,07	0,372
Standard deviation of time travelled in hours	0,07	0,342	0,17	0,029*
Number of leisure time locations visited	0,04	0,628	0,28	< 0,001**
Number of locations visited at night excluding home	-0,06	0,427	0,11	0,143
Number of locations visited once	0,03	0,671	0,27	< 0,001**
Number of locations visited	0,02	0,816	0,14	0,080
Number of travel instances	0,01	0,862	0,18	0,018*
Number of unique locations visited during leisure time	-0,06	0,431	0,15	0,058
Number of unique locations visited at night	0,03	0,658	0,18	0,019*
Number of unique locations visited	0,03	0,726	0,28	< 0,001**
Total time spent at home in hours	0,05	0,517	-0,07	0,385
Total time spent outside including travel in hours	0,07	0,361	0,29	< 0,001**
Time spent stationary in hours	0,09	0,265	0,09	0,223
Time spent stationary in hours excluding the home location	0,06	0,422	0,27	< 0,001**
Time spent travelling in hours	0,06	0,458	0,18	0,017*
Mean durations running in seconds	0,06	0,428	0,00	0,998
Mean duration still in seconds	-0,08	0,313	0,06	0,462
Mean duration travelling in vehicle in seconds	0,11	0,147	0,12	0,123
Mean duration walking in seconds	-0,11	0,151	0,29	< 0,001**
Number of biking instances	-0,02	0,811	0,21	0,006**
Number of running instances	0,14	0,068	0,12	0,131
Number of still instances	-0,07	0,366	0,06	0,417
Number of vehicle travel instances	0,03	0,743	0,20	0,010*

Supplementary Table 4. Associations between Behapp digital phenotyping measures and RFFT outcomes. (*continued*)

RFFT outcome / Digital phenotyping measure	RFFT perseverative errors *r*	RFFT perseverative errors *p*	RFFT unique designs *r*	RFFT unique designs *p*
Number of walking instances	-0,06	0,475	0,28	< 0,001**
Average bed time	0,01	0,921	0,18	0,017*
Standard deviation of bed time	-0,07	0,360	0,06	0,459
Time spent using phone in seconds	0,12	0,133	0,00	0,965
Number of times phone was used	-0,07	0,381	0,05	0,506
Number of times phone was used at night	0,07	0,389	0,19	0,015*
Time spent using phone at night in seconds	-0,04	0,605	0,02	0,775
Average sleep duration	-0,07	0,338	-0,07	0,337
Standard deviation of sleep duration	0,08	0,315	0,04	0,593
Average waking time	0,11	0,162	-0,03	0,708
Standard deviation of waking time	0,06	0,444	0,08	0,291
Total duration of step taking in seconds	-0,06	0,410	0,22	0,005**
Total duration of step taking at night in seconds	-0,02	0,818	0,23	0,003**
Number of steps taken	-0,04	0,620	0,11	0,166
Number of steps taken at night	0,01	0,929	0,18	0,019*

* = $p < 0.05$, ** = $p < 0.01$.



Appendix I

English Summary

The evolutionary processes which resulted in the large brains seen in humans and several other primates have been the subject of scientific study and debate for many decades. A hypothesis which gained traction towards the end of the twentieth century is the social brain hypothesis, which posits that these large brains were the result of selection on the ability to cognitively cope with complex social environments. Being able to navigate more complex social environments is hypothesized to have been beneficial due to the increased ability to survive predation in groups, and/or because it results in more efficient foraging strategies. The ability to navigate complex social environments is termed 'social complexity'. While evidence in favor of the social brain hypothesis continues to emerge, so does evidence in favor of other hypotheses attempting to explain the evolution of the brain in various species, resulting in continued debate surrounding the topic.

In chapter 1 technological and statistical advancements which have resulted in several methods which provide the opportunity to search for evidence of evolution across longer or shorter time spans using genetic information from single species are discussed. Such evidence could potentially result in more clarity regarding the driving forces of the evolution of the brain, both within and across species. However, several issues will have to be resolved in order to maximize the potential of these novel methods, as well as to reduce the likelihood that such studies will result in similar inconsistent findings as has been the case for comparative studies. One important step is to create definitions and operationalizations of social complexity which are applicable across species and which can be consistently used in studies in a single species. Secondly, methodological parameters and data sources have to be carefully chosen in studies attempting to study evolution through genetic data. For example, studies using results from existing GWAS to examine recent selection should carefully examine how the potential exclusion of loci violating Hardy-Weinberg equilibrium may affect their results, as while violation of this criterion may indicate poor genotyping quality, this violation can also be expected in loci experiencing the effects of selection.

In chapter 2 the evolution of genes previously found to be associated with sociability in humans across a long process of speciation was examined. Previous studies had demonstrated evidence that genes found to be associated with schizophrenia, bipolar disorder and major depressive disorder had been disproportionally conserved between humans and *Caenorhabditis elegans*, indicating that for some reason these genes had survived at an increased rate across a long evolutionary time period. These studies found that the conserved genes were highly interactive and disproportionally associated with phenotypes related to lethality or sterility, indicating that perhaps these secondary functions had resulted in the increased level of cross-species conservation. When the method used in these studies to compare conservation between the genes related to the phenotype and the total genome was replicated, we similarly found increases in conservation. However, as this method could be considered an unfair comparison due to the

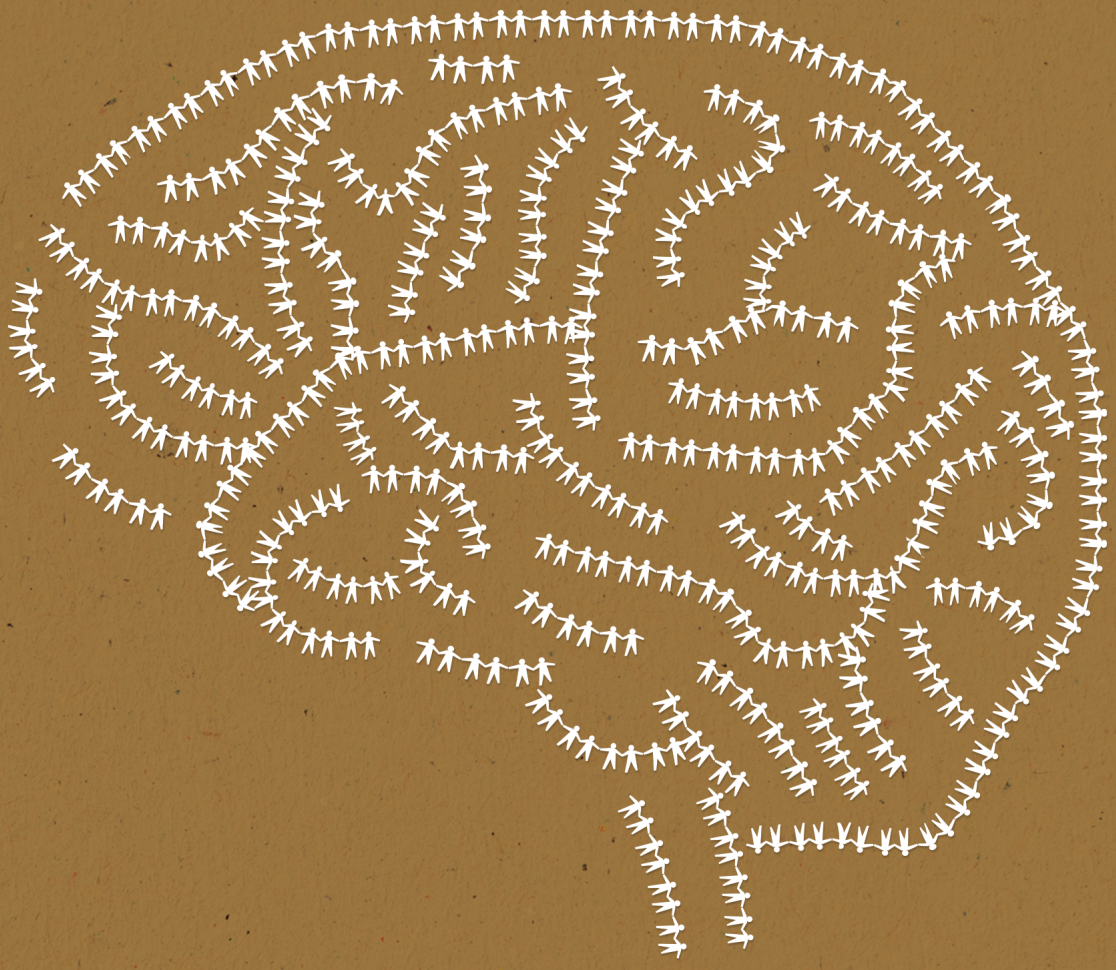
different method of determining whether genes were conserved, we performed a 'fairer' analysis which showed no evidence of increased conservation, indicating that perhaps the findings from previous studies were the result of methodological choices rather than true increases in genetic conservation, highlighting the importance of methodological considerations in studies using genetic data to examine evolution mentioned in chapter 1. The analyses did confirm that conserved genes are highly interactive, however, no indication was found that human sociability genes were related to lethal or sterile phenotypes in *C. elegans*. The dearth of knowledge regarding the association between social phenotypes and genetics in *C. elegans* hampered our analysis regarding the conservation of social function in conserved genes, however it was found that the highly interactive gene *ACVR2A* is associated with social functions in both species, making it an interesting gene for further study.

In order to address the issues with the definition and operationalization of social complexity mentioned in chapter 1, in chapter 3 data from the UK Biobank was used to create measures of social complexity in human data which aligned with recent discussions about this subject. Specifically, the number of social contexts visited by the subject, the number of individuals living in the household, the number of visits with friends and family and a composite measure where each aspect was weighted equally. The goal was to examine how modern human social complexity related to cognition and total as well as regional brain volumes, to examine how such associations compared to those suggested in the social brain hypothesis. Two out of three aspects of social complexity as well as a composite measure consisting of the three combined were consistently associated with cognition where increased cognitive function was associated with higher social complexity. However, one aspect (the number of visits with friends and family) showed nonsignificant or even opposite associations with cognitive skills, indicating that this measure does not measure the same underlying cognitive demands as the other measures of social complexity, and that the composite score is likely not a great representation of social complexity in its current form. Associations with grey matter showed an association between the composite score and total grey matter volume as well as several associations between regional grey matter volumes and either the composite measure or the number of visits with friends and family. Most of these associations were found in the cerebellum, which is interesting as evidence for the social brain hypothesis mostly points to selection for neocortex size in relation to selection for social complexity, indicating that perhaps different processes underlie the variation in social complexity in modern humans compared to the variation in social complexity selected for during the expansion of the human brain.

One potentially important difference in the measurement of social complexity between humans and other species is that while social complexity in non-human species is typically measured by researchers observing the behavior of a set of individuals from the species, in humans social behavior is typically measured subjectively, by asking individuals about their

social behavior. Such subjective measures are prone to bias and research has shown that individuals suffering from neuropsychiatric disorders affecting cognition may be more likely to incorrectly report their own social behavior, potentially affecting results in studies examining the association between the two. Therefore, the development of objective measures of social behavior, for example using smartphones, has the potential to greatly benefit studies examining human social complexity as well as comparative studies examining humans and other species. In chapter 4 we examine how data quality and quantity can affect various measurements of (social) behavior resulting from a smartphone application called Behapp. We found that while objective measurement of behavior using smartphones is promising, quality control is a very important step to make sure that methodological issues do not affect the data. In chapter 5 we performed a preliminary examination of how such objective measures relate to subjective measures of social complexity and to cognition in the Dutch biobank Lifelines. We found several weak or moderate associations between social complexity measures and smartphone-based measurements of social behavior, and various moderate to strong associations between smartphone-based measurements of social behavior and cognitive measurements. The fact that the associations between smartphone-based objective measurements of social behavior and cognition found are stronger than those found between social complexity and cognition in chapter 3 may indicate that such objective measurements are a better reflection of social behavior as affected by cognition compared to subjective measurements, but the low strength of associations between subjective social complexity and objective social behavior may also indicate that the latter is not a great measure of the former. More research in this field could determine whether and how such objective measure can be used to measure social complexity in future studies.

The research in this thesis constitutes a first step towards examining the evolution of human social complexity and the potential associations with the evolution of the human brain using genetic data. The results repeatedly showed that the importance methodological considerations in the collection, preparation and analysis of data in the study of human social complexity cannot be understated. However, if these considerations can be adequately addressed, analyses of evolution using human genetic data and human variation in social complexity could potentially breathe new life into the study of the role of social complexity in the evolution of the human brain.



Appendix II

Nederlandse samenvatting

De evolutionaire processen die hebben geleid tot de grote breinen aanwezig in mensen en verschillende andere primaten zijn al decennia onderwerp van wetenschappelijke studie. Een hypothese die sinds het eind van de twintigste eeuw populair is, is de "social brain hypothesis" (SBH), waarin wordt gesteld dat de grote breinen gezien in sommige primaten het resultaat is van selectie voor de cognitieve capaciteiten die nodig waren om op de juiste manier om te gaan met complexe sociale omgevingen. Het navigeren van complexe sociale omgevingen zou voordelig zijn geweest in verband met de verhoogde kans om aanvallen door roofdieren te overleven en/of omdat het leven in groepen het succes van voedsel verzamelen zou verbeteren. De cognitieve capaciteit om complexe sociale omgevingen succesvol te navigeren wordt "sociale complexiteit" genoemd. Hoewel regelmatig wetenschappelijk bewijs dat de SBH ondersteund verschijnt, dat is ook het geval voor hypotheses die de evolutie van het brein in verschillende soorten verklaren, wat heeft geresulteerd in doorlopend wetenschappelijk debat over het onderwerp.

In hoofdstuk 1 wordt beschreven hoe technologische en statistische vooruitgang heeft geresulteerd in verschillende nieuwe methodes die de mogelijkheid bieden om evolutie over lange en korte tijdspanne te analyseren door middel van het gebruik van genetische data van een enkele soort. Uitkomsten resulterend van zulke methodes zouden meer duidelijkheid kunnen scheppen over de drijvende krachten achter de evolutie van het (menselijk) brein. Echter, er zijn verschillende zaken die moeten worden opgelost om het potentieel van deze methodes optimaal te benutten en de kans te verminderen dat studies die deze methoden benutten resulteren in inconsistente resultaten zoals is gebeurd studies waar de evolutie van het brein werd onderzocht door middel van vergelijkingen tussen soorten. Een belangrijke stap is het creëren van definities en operationaliseringen van sociale complexiteit die bruikbaar zijn niet alleen in mensen, maar ook in andere soorten. Ten tweede, methodologische parameters en databronnen moeten zorgvuldig gekozen worden in onderzoek waar genetische data wordt gebruikt om evolutie te onderzoeken. Bijvoorbeeld, in studies die resultaten van voltooide genome-wide association analyses gebruiken om recente selectie te onderzoeken is het van belang om te evalueren hoe het uitsluiten van genetische varianten die niet voldoen aan het Hardy-Weinberg equilibrium invloed heeft op de resultaten. Schending van deze eis kan betekenen dat de genotypering van lage kwaliteit is, maar is ook verwacht in varianten die recente selectie hebben ondervonden.

In hoofdstuk twee hebben we de evolutie van genen die in een eerdere studie geassocieerd bleken te zijn met sociaal gedrag onderzocht. Eerdere studies hebben gevonden dat menselijke genen die potentieel geassocieerd zijn met schizofrenie, bipolaire stoornissen en depressie disproportioneel vaak een equivalent hadden in *Caenorhabditis elegans*, erop wijzend dat deze genen om de een of andere reden relatief vaak bewaard waren gebleven in het lange proces van soortvorming sinds de laatste gemeenschappelijke voorouder van mensen en *C. elegans*. Deze studies hadden ook ontdekt dat de geconserveerde genen vaker dan gemiddeld interacties

hadden met andere genen en vaker dan gemiddeld geassocieerd waren met dodelijkheid of vruchtbaarheid, wat mogelijk betekent dat deze genen vaak waren bewaard in verband met het belang van hun functies voor de kans op voortplanting. Toen de methoden van deze studies werden gerepliceerd voor genen geassocieerd met menselijk sociaal gedrag werd opnieuw een verhoging van de kans op conservatie van de genen in mensen en *C. elegans* gevonden. Echter, gezien deze methode kon worden gezien als 'oneerlijk' gezien de methode waarmee conservatie werd bepaald niet hetzelfde was voor de set van genen geassocieerd met sociaal gedrag versus het totale menselijke genoom, is ook een tweede methode gebruikt waarbij conservatie werd onderzocht van sets van willekeurige menselijke genen, waarbij werd gevonden dat de verhoging van conservatie die eerder gevonden werd niet gerepliceerd kon worden en mogelijk een onecht resultaat was, wat het belang aanstipt van de keuzes voor methoden in dit soort onderzoek. Er werd wel gevonden dat geconserveerde genen meer interactief waren dan gemiddeld, maar geen indicaties dat geconserveerde genen vaker geassocieerd waren met dodelijkheid of vruchtbaarheid. Aangezien er weinig genen zijn in *C. elegans* waarvan bekend is of ze gerelateerd zijn aan sociaal gedrag, was het niet mogelijk om te onderzoeken of geconserveerde menselijke genen geassocieerd aan sociaal gedrag ook deze functies hadden bewaard in *C. elegans*, maar het bleek wel dat het hoog interactieve gen *ACVR2A* sociale functies heeft in mensen en *C. elegans*, waardoor het mogelijk een interessant gen is voor toekomstig onderzoek.

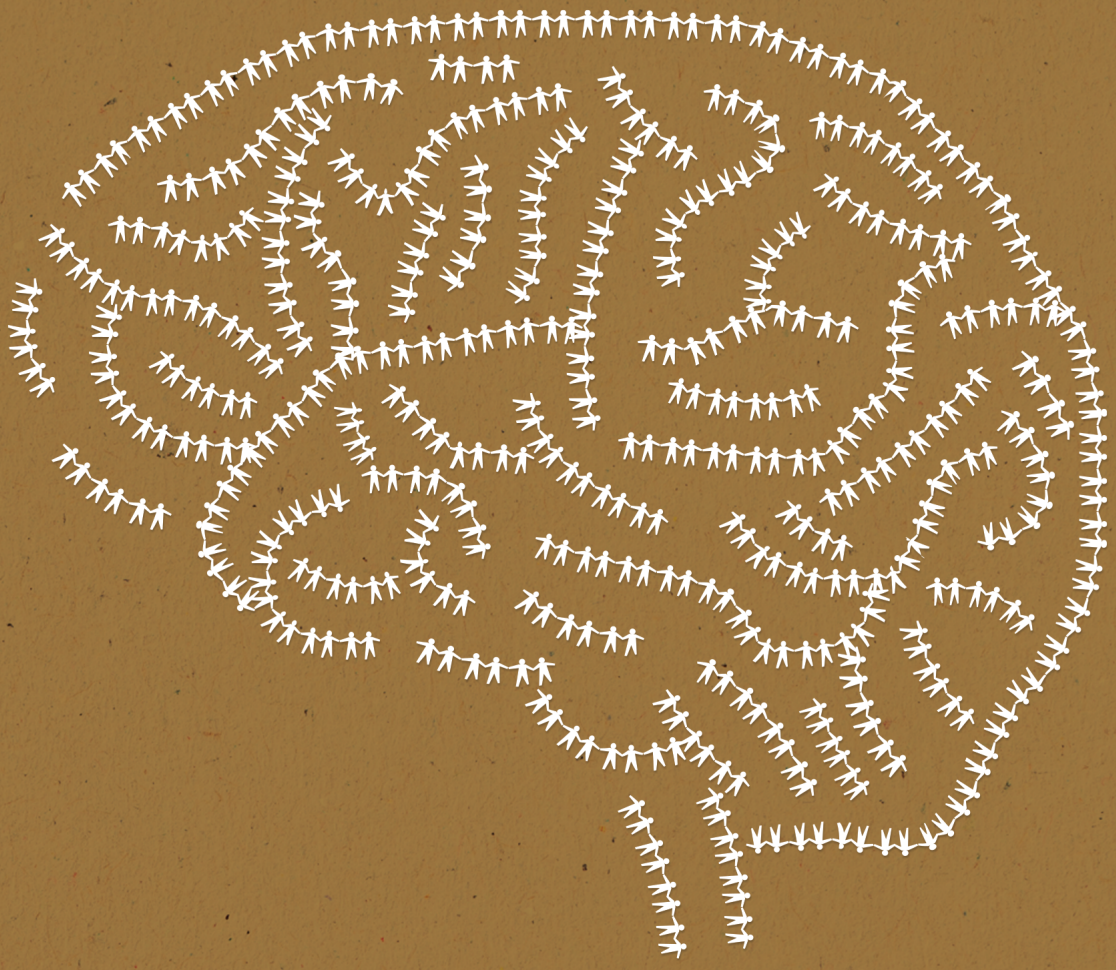
Met als doel om problemen met definitie en operationalisering van sociale complexiteit genoemd in hoofdstuk 1 te adresseren, is in hoofdstuk drie data van de UK Biobank gebruikt om maten voor sociale complexiteit te creëren uitgelijnd met recente wetenschappelijke discussies over het onderwerp. Vier maten kwamen hieruit: het aantal sociale contexten bezocht, het aantal mensen in het huishouden en de frequentie van bezoek met vrienden en familie, plus een composietmaat waarbij elke van de eerdere drie genoemde maten gelijk werd gewogen. Het doel van deze studie was om te onderzoeken hoe sociale complexiteit in moderne mensen is gerelateerd aan cognitie en totale en regionale hersengrootte, en hoe deze associaties relateren aan de associaties tussen cognities, sociale complexiteit en hersengrootte die worden verondersteld om hersenevolutie van primaten te verklaren in de social brain hypothesis. Twee van de drie losse maten en de composietmaat van sociale complexiteit waren consistent geassocieerd met cognitie waarbij verhoogde cognitieve capaciteiten gerelateerd waren aan hogere sociale complexiteit. Dit was echter niet waar voor de frequentie van bezoek met vrienden en familie, waar associaties niet significant of zelfs omgekeerd waren. Dit betekent mogelijk dat de frequentie van bezoek met vrienden en familie niet hetzelfde meet als de andere maten van sociale complexiteit, en dat een composietmaat waarbij de scores van de drie losse maten bij elkaar opgeteld worden mogelijk niet de beste manier is om verschillende aspecten van sociale complexiteit gecombineerd te meten. Analyses aangaande de associaties met de hoeveelheid grijze materie resulteerden in significante associaties voor de composietscore en enkele associaties tussen regionale grijze materie en de frequentie van bezoek met vrienden en familie.

Totale grijze materie was gerelateerd aan de composietscore. Regionale associaties waren voor het grootste gedeelte gesitueerd in het cerebellum, een interessante bevinding gezien het bewijs voor de social brain hypothesis vooral wijst op een relatie tussen grijze materie in de neocortex en sociale complexiteit. Dit betekent mogelijk dat de biologische processen die variatie in moderne menselijke sociale complexiteit bepalen niet (volledig) dezelfde zijn die de variatie in sociale complexiteit bepalen waarop geselecteerd is volgende de social brain hypothesis.

Een belangrijk verschil in de meting van sociale complexiteit tussen mensen en andere soorten is dat hoewel sociaal gedrag in andere soorten normaal gesproken wordt gemeten door middel van observatie door wetenschappers, terwijl dit in mensen vaak gebeurt door deelnemers te vragen naar hun sociale gedrag. Dit laatste is meer subjectief, terwijl het eerste meer objectief kan worden genoemd. Subjectieve meningen zijn meer vatbaar voor bias en een eerdere studie heeft gevonden dat mensen met neuropsychiatrische stoornissen geassocieerd met verslechterde cognitie mogelijk slechter zijn in het rapporteren van hun sociale gedrag dan anderen, wat mogelijk effect zou kunnen hebben op de kans om associaties te vinden in studies die kijken naar cognitie en sociaal gedrag. Een mogelijke oplossing is het creëren van objectieve maten voor sociaal gedrag, bijvoorbeeld door gebruik te maken van smartphones. In hoofdstuk 4 wordt onderzocht hoe de kwaliteit en kwantiteit van data effect kan hebben op metingen van (sociaal) gedrag resulterend van een smartphone applicatie genaamd Behapp. Hoewel het objectief meten van gedrag veelbelovend is, gaven de resultaten aan dat kwaliteitscontrole van groot belang is om te verzekeren dat methodologische problemen niet de resultaten van onderzoek dat gebruik maakt van dit soort methoden beïnvloeden. In hoofdstuk 5 worden kort de resultaten behandeld van een voorlopige analyse van data uit de Lifelines biobank. Er is gekeken naar associaties tussen zelfgerapporteerde sociale complexiteit, objectief gemeten sociaal gedrag op basis van de Behapp app en cognitieve maten. Verschillende significante associaties werden gevonden tussen sociale complexiteit en objectief gemeten sociaal gedrag volgens Behapp, hoewel de associaties van zwakke of middelmatige sterkte waren. Verschillende middelmatig sterke associaties werden gevonden tussen objectief gemeten sociaal gedrag en cognitie, sterker van effect dan de associaties tussen cognitie en sociale complexiteit gevonden in de UK Biobank. De verschillen in de grootte van de effecten zou een indicatie kunnen zijn dat zulke objectieve maten een betere weergave zijn van sociaal gedrag zoals bepaald door cognitieve vaardigheden, maar het gebrek aan sterke associaties tussen subjectief gemeten sociale complexiteit en objectief gemeten sociaal gedrag betekent mogelijk ook dat de objectieve maten niet hetzelfde concept meten als subjectief gemeten sociale complexiteit. Er is meer onderzoek nodig om te bepalen of en hoe objectieve maten gebruikt kunnen worden om menselijke sociale complexiteit te meten.

Het onderzoek uitgelijnd in deze thesis vormt een eerste stap richting verder onderzoek naar de evolutie van menselijke sociale complexiteit en de mogelijke associaties met de evolutie van het menselijk brein met behulp van genetische data. De resultaten tonen aan dat het belang

aan van methodologische overwegingen in het verzamelen, voorbereiden en analyseren van de data in studies naar menselijke sociale complexiteit niet kan worden onderschat. Echter, als deze overwegingen adequaat worden geadresseerd, zouden analyses aangaande evolutie waarbij menselijke genetische data gebruikt wordt naast menselijke variatie in sociale complexiteit nieuw leven kunnen blazen in het onderzoek naar de rol van sociale complexiteit in de evolutie van het menselijk brein.



Appendix III

Acknowledgements/Dankwoord

There are a lot of people without whom this thesis would never have been completed, and I'm very grateful to all of you for supporting me along the way.

Of course, my supervisors **Martien** en **Jacob**. **Martien**, I'm so happy that you gave me the chance to work on your amazing project, despite my academic background being quite different from the rest of your team. Your positivity and flexibility allowed me to keep moving forward, even when we encountered setbacks. **Jacob**, thank you for your supportive presence especially when I was still finding my way with the project, and it was very nice meeting you in Toronto.

Raj, you know I would never have been able to create this thesis without you, not only because we could not have Behapp without you but also because your support has been invaluable to me during the past six years. You believed in me even when I didn't myself. Thank you so much and I hope we can work together for a long time to come, but if we can't I know you will crush it regardless.

Andrea, Anna and **Renate**, thank you for being great members of the Behapp team! It is/was always nice being able to work together and you have both made great improvements to Behapp with your contributions.

Mayerli, I had never thought that when I came back to Groningen I would find one of my best friends here. Your kindness and happy disposition always make my day better. I hope to keep having lunch with you in the future because your random questions consistently make for very interesting conversations!

All the lovely people of **room 0258** over the past six years, and especially the recent crew, **Bente, Alex, Betty, Juliana, Joelle, Kong, Warner, Xin** and **Xiaodan**, it was really great coming back to the office in 2022 and being able to work with you all. You were always up for serious questions, but also for serious goofing off during lunch time or after work. ;)

I also want to thank all the people from the **Kas lab**, we didn't all have the opportunity to work together but it was always great listening to your presentations and talking about interesting research together.

Freya, I'm still amazed (and a little envious) at how you found the time and perseverance to make all those amazing knitted sweaters. I enjoyed our conversations and hope to see you again sometimes, either in the Netherlands or in Denmark!

María, I think few topics of science are cooler or more important than meta-science, and I think your dedication to research quality both in the past and in your current job is very admirable. I hope we can have one of our discussions again soon.

Lieve **Joris**, voor jou geldt misschien wel het meest dat deze thesis er niet was geweest zonder jou. Ik heb niet voor jou gekozen als paranimf omdat je mijn man bent, maar omdat niemand zoveel vragen over genetica en evolutie heeft moeten aanhoren en beantwoorden als jij, en omdat niemand zo vaak me heeft geholpen als er weer eens een tegenslag was. Ik weet zeker dat je zelf over een aantal jaar een geweldige thesis aan zal leveren. Ik hoop dat ik jou daarbij kan steunen zoals je de afgelopen jaren voor mij hebt gedaan.

Mam, ik moet toegeven dat ik niet weet of ik voor wetenschappelijk onderzoek had durven gaan als ik het jou niet eerst had zien doen. Niet omdat jouw PhD zo makkelijk leek, maar omdat jij me liet zien hoe interessant het proces van onderzoek doen is, ook al is het vaak veel werk en niet altijd leuk. Gelukkig was je er altijd voor me als het weer een keer niet leuk was, dat heeft me ontzettend geholpen. **Pap**, je hebt me altijd gepusht om het maximale uit mezelf te halen. Het doorzettingsvermogen om door te blijven gaan ook als alles tegenzat is zeker voor een deel te danken aan hoe jij me altijd aangemoedigd hebt (tijdens badminton of school). Ook heel erg bedankt voor je bereidheid om te helpen met verhuizen de afgelopen jaren, waar ik veel gebruik van heb gemaakt. Dank jullie wel voor alle goede voorbeelden en steun allebei!

Donna, ik ben zo blij dat we rond het begin van mijn PhD closer werden en dat we altijd dingen konden bespreken en samen konden gaan klimmen. Naast de gezelligheid en sportiviteit ben je ook een bron voor inspiratie voor me, en ik weet zeker dat je het nog ver gaat schoppen!

Marja en **Arie**, dank jullie wel voor het warme welkom in jullie familie. Ik vind jullie toewijding aan de wetenschap heel bewonderenswaardig en ik hoop dat ik dat ook op die manier vast kan houden in mijn verdere carrière.

Tessa, ik ben altijd ontzettend blij dat hoewel onze levens heel anders zijn verlopen na de basisschool we toch altijd in contact konden blijven. Dankjewel voor er altijd voor me zijn als ik je nodig had.

Hoewel het soms het contact met Groningen wat moeilijker maakte, ben ik toch erg blij dat ik een tijdje in Leiden heb kunnen wonen dankzij de geweldige Duivelseitjes waarmee ik daardoor in contact heb kunnen komen en blijven. **Emma** en **Jasmijn**, wij hebben duidelijk bewezen dat breien/haken, klimmen, bordspellen en soms iets te veel drinken een hele goede basis zijn voor vriendschap. Ik hoop dat we altijd af kunnen blijven spreken (ook zodat we Betrayal misschien een keer afmaken als we 70 zijn). Over te veel drinken gesproken, **Jesse** en **Nenad**,

dank jullie wel voor alle leuke avonden samen, zowel op het Duivelsei als daarbuiten. Jullie hadden altijd tijd voor me als ik afleiding nodig had. Hopelijk kunnen we nog veel spareribs eten en bordspellen spelen in de toekomst.

Twee mensen die zeker ook niet vergeten kunnen worden zijn de twee mensen die in de jaren voor mijn PhD mijn waardering voor statistiek en wetenschap hebben aangewakkerd en me hebben ondersteund bij mijn studie. **Ralph**, ik ben nog steeds ontzettend dankbaar dat je me toestond mijn thesis over statistiek te schrijven ook al was dat niet erg typisch voor de studierichting. Jij hebt me ooit aangemoedigd om de research master te gaan doen en zonder die aanmoediging denk ik niet dat ik dat ooit was gaan doen. **Mevrouw van Haeff** (ik zal u maar aanspreken zoals ik dat ook op het Veurs deed), als iemand me in de 4^{e} klas had gezegd dat wiskunde mijn favoriete vak zou worden, zou ik ze voor gek hebben verklaard, maar u heeft het toch voor elkaar gekregen. Dank u wel voor uw vertrouwen, zonder u had ik nooit het zelfvertrouwen gehad om tijdens mijn studie me te focussen op statistiekvakken.

Lieve baby, ik kan je naam hier nog niet noemen want die is nog geheim maar jij was altijd aanwezig in de maanden dat ik mijn thesis aan het afmaken was. Jij was het duwtje dat ik nodig had om extra hard te werken zodat deze thesis op tijd klaar was en ik mijn verdediging nog kon doen voordat je geboren werd. Ik kan niet wachten tot je er bent.