Scenes from a Marriage: How We Found Our Way from Experimental Psychology to Social Neuroscience

Uta Frith[1] and Chris D Frith[2,3]

[1]Institute of Cognitive Neuroscience, University College London, London, United Kingdom; email: u.frith@ucl.ac.uk

[2]Institute of Philosophy, School of Advanced Study, University of London, London, United Kingdom; email: cfrith@ucl.ac.uk

[3]Functional Imaging Laboratory, University College London, London, United Kingdom

Annual Review of Psychology 2025

Uta Frith Chris Frith

Keywords

autism, cognition, cooperation, culture, dyslexia, schizophrenia, autobiography

Abstract

Looking back on our life and work, we reflect on the changes in our thinking due to three scientific and technological revolutions. These are information processing, computers, and brain imaging, and together they ousted behaviorism from its dominant position in experimental psychology. We champion a model of the mind that is hierarchically organized with both a robust unconscious and a harder-to-pin-down conscious mode of operation. Our studies were inspired by disorders that made us realize that cognitive processes at all levels of the information processing hierarchy impact social interactions. We locate the influence of culture at the highest level of this processing hierarchy. Here we see the interface between different minds and the importance of norms when regulating the opposing trends in our complex and even contradictory social nature.

1. INTRODUCTION

What task is both pleasant and troubling and only happens in the last decade of your life? Having to face your past and write about it! We were both pleased and surprised to be asked by the editors "to provide a personal overview of your involvement in the discipline and its development to the current state of the art, set in historical context." Having just finished a project that lasted about seven years, our book *What Makes Us Social?* (Frith & Frith 2023), we couldn't resist—not only because it provided us with a new project, but also because it allowed us to place ourselves in the context of our peers and see how we fit into this bigger picture. This was an opportunity to go on a quasi-archaeological expedition and to dig into our past. Might it be possible to find a coherent thread leading from the past to the present?

We had already had a chance to look back when we delivered Bartlett Lectures for the Experimental Psychology Society (U. Frith 2012, C. Frith 2023), but these lectures were narrowly focused on seemingly separate threads of our work. The truth is that we cannot reminisce about our working life without reminiscing about our married life. It is one and the same thing. Our history is an endorsement of the We-mode (Gallotti & C. Frith 2013). We worked in collaboration, rather than in competition, and we benefited from having different skills and different prejudices.

2. OUR ORIGIN STORY

The wonder is that we ever met at all. When we were born, our two countries were at war. This was not a propitious start for joining our lives together. But thankfully, the world had changed by 1965, and despite the significant cultural differences we found we had many things in common—love, actually. Uta was determined to adapt to English language, culture, and customs and could build on her passion for classic detective novels. All this is documented in a graphic nonfiction book produced together with our son Alex (Frith et al. 2022), with a glimpse of our younger and older selves shown in **Figure 1**.

Figure 1 The Friths's origin story. Illustration by Daniel Locke from *Two Heads* (Frith et al. 2022), with apologies to Bill Finger and Bob Kane

We had both chosen to be trained in clinical psychology, not with a view to practice, but with a view to satisfying our vague and romantic longings to understand the puzzle of the human mind. Such longings needed some solid ground in which to grow, and talking to each other was essential for finding this ground. It was this and the happenstance of the books and papers we were reading that led us to our supervisors, Hans Eysenck, Beate Hermelin, and Neil O'Connor. How lucky we were that they took us on! They led by example while being remarkably hands-off. In line with research culture at the time, they were not even coauthors on our first papers.

We were not daunted by our lack of knowledge, nor were we daunted by an obvious lack of resources.[1] On the contrary, we were attracted by simple psychological tasks that could be made up DIY style and that nevertheless would allow us to explore complex mental processes. In this, we closely followed in the footsteps of our mentors, who pointed us to Peter Medawar's depiction of science as "the art of the soluble" (Medawar 1967).

We got married in 1966, while we were both students at the Institute of Psychiatry, then part of the University of London Postgraduate Medical School (now part of King's College, London). Uta's topic was pattern detection in autism, while Chris's was motor skill learning. We lived nearby in a Victorian worker's cottage with two downstairs and two upstairs rooms and a tiny garden. Chris had inherited a baby grand piano from his grandmother that just about fitted into the living room. Our first-ever purchase together was a piano stool that was suitable for playing duets.

Our first joint paper (Frith & Frith 1972) was the result of a happy accident. We played the game Solitaire during the Christmas holidays in Germany on a board with white and blue colored glass marbles. Although we knew that they were exactly equal in number, we could not help the strong sensation that in some configurations there were more marbles of one color. We designed some simple experiments to test and confirm the idea that this illusion was a result of the Gestalt properties created by continuous lines.[2] However, we did not seriously start producing joint papers until about two decades later.

Why did it take us so long? One reason was that we were stumbling along in the dark ages of neuroscience. When we started out, the search for brain abnormalities in psychiatric conditions was still rare. Work by pathologists and anatomists was controversial and did not immediately yield any striking results. In the early 1970s, the intellectual discourse was dominated by those who, in the extreme, denied any biological causes of mental disorders and were fixated on social and psychological causes. For example, in those dark ages, autism was considered the product of "refrigerator mothers." Dyslexia was a handy excuse for middle-class parents whose children

[1] The lab displayed the slogan, "We have no money, so we will have to think (Ernest Rutherford)."

[2] This illusion was subsequently found to be experienced by monkeys, fish, and insects (Howard & Dyer 2024).

performed poorly at school. Schizophrenia was the result of contradictory communication within the family and society, the so-called double bind.

These were strong sociopolitical trends that informed whole research programs. They were strong enough to lead to the closing of psychiatric institutions. The idea was that providing care in the community would solve the problems of mental illness. Movies such as *One Flew over the Cuckoo's Nest* were immensely successful in supporting these antipsychiatry sentiments. Subsequently, the high costs and undesirable consequences of care in the community became all too apparent (Coid 1994). Defiantly, we were confident in our belief that psychiatric conditions had a basis in the brain, even though we knew hardly anything about the brain. Moreover, it took a revolution in our thinking to realize that the crucial link between brain and behavior was cognition, or, in other words, information processing.

3. THREE REVOLUTIONS THAT SHOOK OUR WORLD

Looking back over the decades we can identify three major developments that turned us into the experimental psychologists we eventually became. The first of these was the rise of cybernetics and information theory, which led to thinking of the brain as a control system. It emboldened us to focus on invisible cognitive processes that influence perception and action. The second revolution was the introduction and rapid spread of computers and the Internet. This massively increased our ability to conduct experiments, analyze data, and publish results. It also enabled us to develop computational models of potentially fragile cognitive processes, that is, processes that could go wrong and lead to psychiatric symptoms, such as delusions of control. The third revolution, brain imaging, was due to an unexpected technological advance in physics and engineering. Brain imaging opened a new testing ground for our theories and even allowed us to visualize the brain basis of the cognitive processes we had been dreaming about.

These were the cultural changes that left their imprint on our thinking and set the tone for our working life. They also coincided with turning points in our married life, in the form of three house moves. These moves could be seen as symbolic of a rise in fortune. We started in a Victorian cottage, progressed to an Edwardian terraced house, and ended up in a part of what is jokingly referred to as an ancient pile. Amazingly, we are still there.

We happily acknowledge that we have been continuously influenced by our mentors and colleagues, but the technological advances brought by the three revolutions have left a deep impact on what we did and how we did it. They shaped not only what we could do in the lab but also the questions that we could ask. Whether these questions have been answered is another matter. In a different world at a different time, we would surely have done very different things. Like fish swimming in water, we are unaware of the cultural medium we are embedded in. The history of science provides plenty of examples of how powerfully science is dependent on changes in technology. One example is the arrival of email, which facilitated communication between researchers working in distant labs; another is the crucial role that X-ray crystallography played in the history of DNA.

3.1. Mental Models, Cybernetics, and Information Theory

Behaviorism was still the dominant approach in the 1960s: identifying the rules that linked a stimulus with a response, without much concern for mechanisms. However, the information processing revolution was approaching, and we were keen to look inside the black box. What were the cognitive processes hidden in this box, and how did they map onto the brain?

Kenneth Craik was one of the heroes of the new cognitive approach.[3] He introduced the idea of mental models, that is, internal representations of how the world works (Craik 1943), and he explained human behavior in terms of an engineering control system rather than a stimulus and response chain. Information theory, developed by Shannon & Weaver (1949), is fundamental to the cognitive approach, and we both read and loved the small book by Attneave (1959) about the application of information theory to psychology. One of the topics covered in this book was the "cloze procedure," whereby one must fill in the final word, given *n* previous words. The more previous words there are, the greater the constraint on possibilities for the final word. We did not realize that, decades later, this technique would become the basis for large language models (LLMs) and ChatGTP. At that time, neither the computing power nor the gigantic text corpuses that proved necessary for this development were dreamt of.

The new cognitive approach was, at heart, computational. This was very attractive to Chris, who had a background in math. Here is an example of how this interest crucially supported Uta's

[3] Chris is particularly proud of his 1990 Kenneth Craik Award from St John's College, Cambridge.

work (U. Frith 1971). Uta had a long-standing obsession with letters of the alphabet, how children acquire their knowledge of letters, and why they sometimes reverse letters. She asked children to match and copy letters and letter-like symbols in both their normal (e.g., N) and reversed (И) orientation. Chris suggested using signal detection theory (Luce 1963) to interpret the results. This approach makes it possible to measure response bias, favoring normal letter orientation, independently of the ability to discriminate the two graphic forms. This enabled Uta to show that this bias increased over the first school year. To her this was a revelation, since the results were now catching something about the elusive internal representations of letters: Children learned to expect letters to be oriented in particular ways, and some took longer to learn this than others (IU. Frith 1974b).

The application of signal detection theory had been an early milestone for us and was never very far from our thoughts from then on. However, experiments on making decisions about stimuli go back to the beginnings of experimental psychology, with Gustav Fechner's psychophysical studies (Fechner 1860). Even though these experiments were concerned with the measurement of subjective experience (e.g., the loudness of a sound), they are among the most replicable of all psychological experiments (Read 2015). Detecting signals is an example of decision-making in the presence of noise. We need signal detection theory to understand this process, because reports of detection depend not only on the ability to discriminate the stimulus but also on response bias (Tanner & Swets 1954). In situations of danger, it is better to say "wolf" when no wolf is there than to miss a real wolf. The response bias, to say "wolf," makes us safer.

This approach influenced Uta's further work on letters and letter-like shapes (U. Frith 1974a). The response bias she was estimating is a hidden property of the reader and not a property of the stimulus. In the case of participants reporting an N, when an И was presented, this is a response bias based on expectation. The reader has learned that Иs are extremely unlikely (except, of course, in Russian) and has a model of the world in which Иs hardly ever occur. The same idea also led to an experiment that contrasted memory for conventionally and unconventionally edited film sequences (U. Frith & Robson 1975).

The idea that detection (and perception) depends upon expectation relates to another of our heroes, Hermann von Helmholtz. Helmholtz proposed that perception depends on making inferences [Helmholtz 1948 (1867)]: Given the bottom-up information provided by our senses, what is it that we are seeing? To make this inference we must take account of what is likely in the circumstances, using our current model of the world. The animal we can vaguely see approaching in the desert sandstorm is more likely to be a camel than a polar bear (Yon & Frith 2021).

These expectations based on our model of the world, which also exert top-down effects on our perception, are often called priors. This term is associated with our next hero, the Reverend Thomas Bayes. Bayes theorem (see Barnard & Bayes 1958) provides a computational basis for Helmholtz's perceptual inferences. It tells us how much we should change our model of the world given the new evidence coming from our senses. Helmholtz and Bayes were particularly important to Chris during his studies of schizophrenia. The key symptoms associated with schizophrenia are hallucinations and delusions, often classified as false perceptions and false beliefs, respectively. From a Bayesian standpoint, however, this distinction dissolves, since both perceptions and beliefs are based on prior expectations (Fletcher & C. Frith 2009).

There is a particularly striking false belief, termed a delusion of control. It involves patients feeling that their actions or even their thoughts are controlled by other agents. Somebody else is thinking their thoughts and making them do things they feel they had no part in doing. How can this be? Such symptoms have been described as "ununderstandable" (Jaspers 1962), that is, not explainable, but Chris found a cognitive account through the ideas of Helmholtz married with Craik's control system approach. Helmholtz noted that, when we move our eyes, the world stays still, even though the image is jumping about on our retina. He suggested that a signal arising in the motor system (corollary discharge; Crapse & Sommer 2008) must be being sent to the perceptual system to indicate that sensory changes are about to occur. These changes can be discounted since they are self-produced.

This mechanism allows us to distinguish between signal changes caused by our own actions (e.g., moving our eyes) and signal changes caused by external forces. This is why we can discount the effects of our own eye movements on perception and why we cannot tickle ourselves. Chris suggested that, in patients with delusions of control, the distinction between self and other has been lost. This idea was tested by an experiment that confirmed that schizophrenic patients were not able to feel the difference between tickling themselves and being tickled by others (Blakemore et al. 2000).

Today, mechanisms of predictive processing derived from the ideas of Bayes and Helmholtz dominate cognitive neuroscience and can be applied to perception, decision-making, learning, and brain function (see, for example, Hohwy 2013). In 2007 Chris summarized many of these ideas in his book *Making Up the Mind* (C. Frith 2007). The key insight in this book, hinted at in the title, is that perception is a fantasy that coincides with reality. This catchphrase captures the idea that we have no direct contact with the world. What we perceive is a mental model that is continually updated via the crude information provided by our senses.

The computational approach to psychology (e.g., signal detection theory, predictive coding) continued to guide our thinking throughout our working life. We think of it as the basis of social cognitive neuroscience. This is evident in the graphically presented models that we have both designed again and again. This revolution constitutes the theoretical framework of our research. Indeed, it is the basis of the work on social neuroscience that we ended up doing. An example, which we will revisit later, is shown in **Figure 2**.

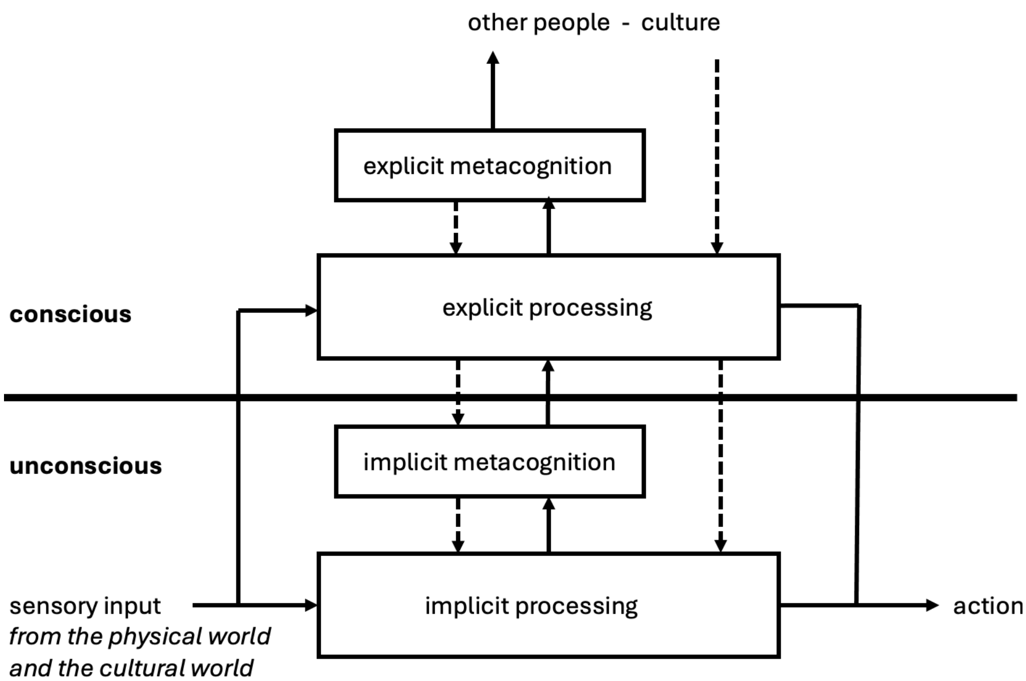

Figure 2 A hierarchy of implicit and explicit cognitive processes. Interactions among the levels of the hierarchy involve top-down prior expectations (*dashed lines*) and bottom-up evidence (*solid lines*). Figure adapted from Frith & Frith (2023).

3.2. Computers and the Internet

The appearance of computers was as dramatic in our personal life as it was in public life. Significantly, and symbolically, when we first met, Chris had an instruction manual for building a computer made from matchboxes: binary information to touch with your hands. Computers to use at home seemed far away; but when they came, they quickly replaced the equipment we had for experiments, which essentially consisted of paper and pencil, a stopwatch, a modified gramophone turntable, and a mechanical calculator.

As everyone knows, in the 1960s and 1970s London was swinging. We visited the memorable Cybernetic Serendipity exhibition at the Institute of Contemporary Arts and regularly attended free BBC concerts given by aspiring singers, including David Bowie. The world was changing as foreign travel, color TV, and computers became part of daily life. By 1972 we had moved into a bigger house with a romantic garden and a separate annex for a live-in nanny for our two young sons. This made it possible for both of us to achieve a full-on double life of home and work.

In 1966 a LINK-8 computer was first acquired by Hans Eysenck's department at the Institute of Psychiatry. Chris was immediately attracted to trying out its possible uses to make more complex paradigms possible. Rather than chasing a disk around a gramophone turntable, his volunteers could now use a joystick to follow a target moving in a complex pattern. Instead of recording whether they were on or off target, he could estimate whether the frequencies hidden in the target's movements reappeared in the participants' movements (C. Frith & Lang 1979). Yet this large machine, which occupied its own air-conditioned room, had only 8 kilobytes of memory. It was fed with punched tape.

A little later we started using a mainframe computer fed with punched cards. We remember carrying boxes of these cards to a central computing office to provide us with statistical analyses of our data. Only a few years earlier, the department had employed two awesome women solely for the purpose of analyzing data (e.g., principal component analysis) with the aid of mechanical calculators. Computers changed incredibly quickly to become smaller and more powerful. At home we had a ZX Spectrum, a BBC microcomputer, an Acorn Archimedes, and finally a Mac.

Eventually, Chris obtained his own computer, a PDP-11, for his lab at the Medical Research Council (MRC) Clinical Research Centre, where he had joined the group led by Tim Crow in 1975. This was a remarkable team of researchers from a wide range of disciplines, whose work was informed by daily encounters with psychotic patients. Chris used his computer to examine the ability of people to detect errors on a motor task in the absence of any visual feedback, since such corrections require the use of corollary discharge signals. Patients with delusions of control failed to make such corrections (C. Frith & Done 1989).

At home and at work, we could use desktop computers to manufacture test stimuli and to record responses to tests. We could also perform statistical analyses faster than ever. No wonder everyone produced more papers and at faster speed than ever before. The sinister aspect of this relentless increase and pressure on research deliverables was not immediately apparent. Only much later did we realize that we had lived through a golden era where we had time to develop and reflect on our research. Well into her retirement, Uta took up the case for slow science U. (Frith 2020), urging a move toward evaluation of output by quality rather than by quantity.

Once computers became sufficiently powerful and ubiquitous, the Internet arrived. On a global scale, the impact of the Internet and the World Wide Web was arguably even more dramatic than that of print. For us the Internet meant that we could read almost any journal on our screen immediately. Previously, we went to the library to photocopy papers, and we wrote postcards to request papers directly from the authors. Now, we could submit papers to be published on dedicated websites and we could review papers online without ever printing them out on paper. In practical terms these changes are so vast that it is quite difficult to tell younger colleagues what it was like before computers and mobile phones. Unfortunately, the ease of publishing has not been entirely beneficial. It creates the problem of too many articles of not very high quality.

3.3. Brain Imaging

The arrival of computers changed the research questions we could explore and the analyses we could apply. Above all, it was powerful computers that made brain imaging possible, and this new technology was vital for our attempt to contribute to the new idea of a social cognitive neuroscience.

In 1986 the MRC decided to close the Centre where Tim Crow had his psychiatry research group, and the members of the group were dispersed. Chris chose to join a small group of pioneering scientists who all shared one cigarette smoke–filled office at the Hammersmith Hospital. Their aim was to study the brain using positron emission tomography (PET). This was the beginning of the third revolution that we witnessed in our working life.

In 1994 the small imaging group, led by Richard Frackowiak, moved to a specially adapted building in Bloomsbury, where at first PET and functional magnetic resonance imaging (fMRI), and later magnetoencephalography (MEG), brain imaging would be used to study the relations between mind and brain. In addition to a banquet at the Dorchester Hotel, the opening was celebrated with a conference in which the latest research was presented. Chris was especially proud to present our joint study of the neural basis of theory of mind or mentalizing: the ability to make inferences about people's invisible mental states.

Briefly, using PET, we measured brain activity while our volunteers listened to stories that could only be understood in terms of what the protagonists believed (Fletcher et al. 1995). We contrasted these stories with random sentences and stories that entailed the inference of a physical cause and effect. We identified a small number of brain regions that were engaged by this task (**Figure 3**).[4] We displayed the brain images on T-shirts specially printed for the occasion. The meeting and the banquet were a great success. Much fun was had by all, even if somebody, champagne glass in hand, was heard to mutter "I'll give it five years."

However, the special glory of that occasion was not revealed until many years later. Someone who had been present at the 1994 meeting, and was now an eminent neuroscientist, said to Chris about the mentalizing study, "We all thought you were mad." So, how did we get to this point, and what happened afterward?

---

[4] The activation pattern shown in this first study with only six participants proved eminently reproducible (Schurz et al. 2021). It delineated for the first time the mentalizing system of the brain and thus gave credence to the concept of mentalizing, which had faced strong skepticism from behaviorists.

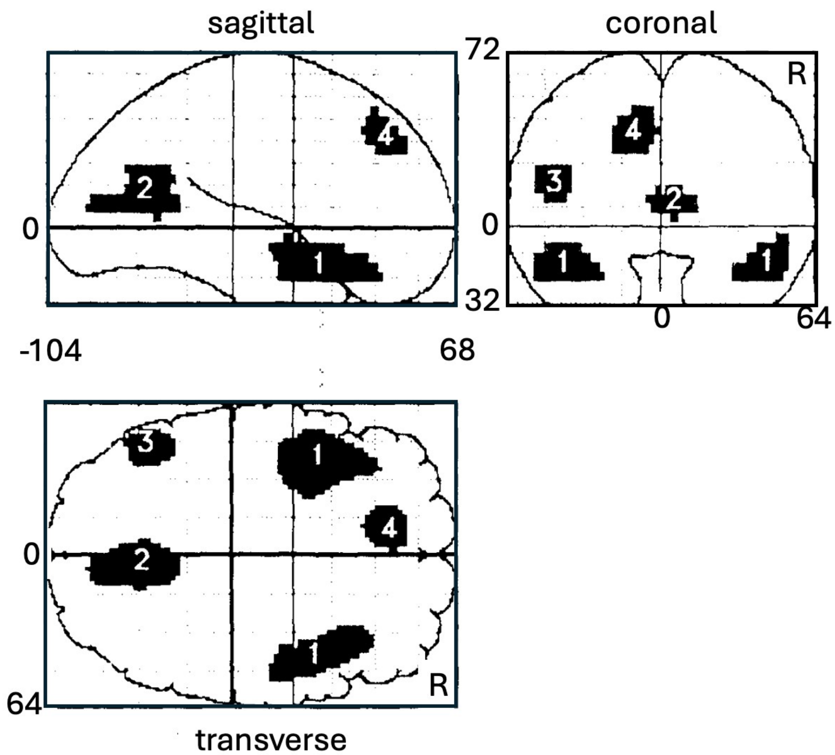

Figure 3 Activity associated with mentalizing from six subjects scanned with positron emission tomography (PET). Areas include (1) temporal poles, (2) posterior cingulate cortex, (3) superior temporal cortex, and (4) medial frontal cortex. Figure adapted from Fletcher et al. (1995).

When we started out as experimental psychologists, we were not even thinking about the brain. For us, at that time, information processing did not immediately translate into neural networks. The start of our awareness of the brain sprang from the achievements of cognitive neuropsychology. We are forever grateful to our colleagues who had started to reveal the basis of cognitive processes in patients who had suffered brain injury. This is where we learned that lesions in different parts of the brain had very specific effects (McCarthy & Warrington 1990). Some particularly striking cases showed remarkably specific impairments in object identification. For example, a patient might be able to identify inanimate objects (e.g., a torch) but not living things (e.g., a parrot) (Warrington & Shallice 1984). This unexpected distinction based on studies of patients with lesions was subsequently confirmed using brain imaging in healthy volunteers (Martin et al. 1996).

Uta was excited by the observation that children with intellectual impairment, doubtless due to abnormal brain development, could still perform well on some tasks, for instance, reading with perfect accuracy. Conversely, she tested dyslexic children, who struggled with their reading but performed well on IQ tests. The specificity of deficits, regardless of general intellectual ability, remained a strong theme in her work (U. Frith & Happé 1998). Were these specific impairments associated with specific brain abnormalities?

Chris, being equally affected by cognitive neuropsychology, had the insight that patients with schizophrenia could be studied like other neuropsychological patients, with a view to finding impaired cognitive functions that could explain their behavioral signs and subjective symptoms (Shallice et al. 1991). Brain imaging would allow us to locate such cognitive functions in undamaged brains (e.g., Posner et al. 1988).

Have brain imaging studies fulfilled their promise? Certainly, they have revolutionized cognitive science and transformed it into cognitive neuroscience. However, there have been some disappointments due to unrealistic expectations about our ability to penetrate an individual's thought processes; and, unfortunately, the data we obtain are so noisy that we need to generalize from group averages. Much to our regret, the dominance of brain imaging studies of groups has meant that single case studies are no longer used to create or eliminate cognitive theories.

A possibly unforeseen problem for brain imaging is the increasing ease of performing complex statistical analyses. Ease of analysis and faster turnaround of work are not always to the good—as seen in the now ubiquitous replication crisis. This is especially problematic with the very large and complex data sets produced by brain imaging. This is an early example of big data and all the problems associated with it. We get distracted by the large numbers as well as by attractive and seemingly intuitive brain images and think less and less about what our volunteers and their brains are actually doing. To counter this, our eminent colleague, Dick Passingham, would demand, "Show me the raw data."

4. ELECTIVE AFFINITIES

We borrowed from J.W. von Goethe's remarkable novel the title for a paper where we compared our ideas about autism and schizophrenia (Frith & Frith 1991). Was there more of an affinity between these conditions than was supposed at the time? Why was it, we wondered, that the term autism was now used to describe a specific clinical condition when the word had originally been coined by Bleuler to describe a kind of social impairment characteristic of schizophrenia?
There were other reasons, too, that drew us to the phrase "elective affinities." First, we saw

ourselves reflected in Goethe's chemistry metaphor for human relationships: "Such natures as, when they come in contact, at once lay hold of each other, and mutually affect one another, we speak of as having an affinity one for the other" (von Goethe 1872, p. 39). Second, the title is deliberately ambiguous, combining both choice and fate, and this resonated with our experience with patients at London's Maudsley Hospital. We were captivated by clinical phenomena associated with failures of social interaction that transcended diagnoses. This realization made us implicit collaborators on problems raised by studying individuals with diverse diagnoses. We were confident that at a deep level affinities could be found among them.

We readily admit that we did not presume to study disorders to find their causes and their cures. This might be the realm of molecular biologists, geneticists, and neurologists. Instead, we were fixated on finding out more about what makes us human. We can be accused of exploiting the opportunity provided by the puzzling symptoms of abnormal conditions to open, very tentatively, a window into the human mind. It follows that we would not study just one condition, or one puzzling symptom, but several. It was not strange at all that we switched freely between experiments with individuals of all ages and abilities, dyslexic children and adults, autistic children and adults, and schizophrenic patients.

4.1. Schizophrenia

Shortly after completing his PhD, Chris was asked to write the chapter on abnormalities of perception for the second edition of Hans Eysenck's *Handbook of Abnormal Psychology*. The main perceptual abnormalities he discussed in this chapter were the hallucinations associated with schizophrenia, and Chris became fascinated by the idea of explaining these false perceptions in cognitive terms. He was able to put this plan into action when he was asked to join Tim Crow's MRC unit dedicated to exploring the biological basis of schizophrenia. This was an ideal setting for him, allowing collaboration with psychiatrist Eve Johnstone and comparative psychologist Ros Ridley. Chris's lab was in the middle an acute psychiatric ward. Every day, he was interacting with psychotic patients who were often keen to come and take part in experiments.

The work of Crow's MRC unit had a major role in revealing the relevance of the brain to a better understanding of schizophrenia. The team showed that many people with a diagnosis of schizophrenia had enlarged brain ventricles (Johnstone et al. 1976) and that blocking the neurotransmitter dopamine reduced the severity of subjectively experienced hallucinations and delusions while having no effect on behavioral signs, such as poverty of speech (Johnstone et al. 1978). This latter result was a revelation to Chris, since it exposed a tight link between subjective experience (hallucinations and delusions) and brain function. This revelation fueled his subsequent pursuit of the neural correlates of consciousness (C. Frith 2021, C. Frith et al. 1999). It was in this context that Chris developed his cognitive account of symptoms such as delusions of control.

How did Chris manage to write his book, *The Cognitive Neuropsychology of Schizophrenia* (C. Frith 1992)? He took advantage of the quiet time during breakfast after the children left for school and before Uta appeared. In this book, he suggested that the cognitive abnormalities that underlie the signs and symptoms associated with schizophrenia arise from impairments in a system that constructs and monitors representations of mental events in consciousness. He speculated that this monitoring system involved interactions between prefrontal cortex and other parts of the brain, especially the temporal cortex.

4.2. Developmental Disorders

We fully adopted the idea of cognitive processes that are hidden causes of behavior. Since they are not directly observable, we must discover these processes by inference, always a dangerous procedure. Our angle was to get at these hidden processes by spotting them when they fail. We acknowledge our debt to John Morton, the director of the MRC Cognitive Development Unit that Uta joined in 1982, who strongly guided our thinking along these lines. Those brought up in the behaviorist tradition of psychology, and this includes us, often find it hard to remember that the same behavior can have very different causes and does not give a straight answer to what underlies it. To escape this ambiguity, we find simple graphic representations helpful. A generic scheme intended to help thinking about developmental disorders is illustrated in **Figure 4**.

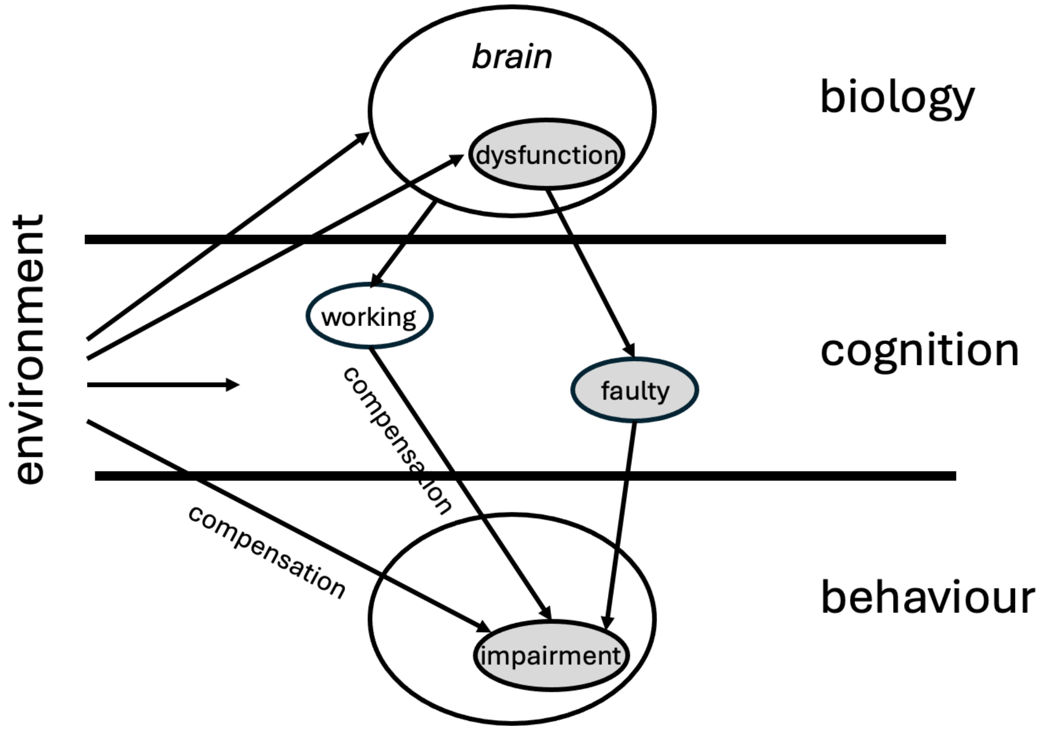

Figure 4 A simple three-level causal model of developmental disorders linking brain and behavior via cognition. A dysfunction in one brain region will lead to faulty cognitive processes, leaving other processes intact. Such a fault need not necessarily lead to behavioral impairment. There may be compensation from other cognitive process linked to intact brain regions and from the environment—for example, via special education or technical support. Figure adapted from Morton & Frith (2001).

A lightbulb moment occurred when we realized that we could use brain imaging to get evidence on the invisible layer between brain and behavior, the cognitive level. Here we could hope to find evidence for specific cognitive mechanisms that might be faulty in some clinical conditions. Indeed, we were fortunate to find neural correlates for some of these evasive cognitive processes. A series of brain imaging experiments targeting separately dyslexia and autism probably marks the peak of our research. This peak, seen in sudden increases in annual publications, was only possible because we could rely on some wonderful students, colleagues, and collaborators.

We already hinted that Uta had been attracted to reading and spelling ever since she started being interested in psychology. This attraction received a boost by her discovery that Chris, who is a superfast reader, often asked her how to spell words when doing crossword puzzles. She was more than intrigued by this contrast, since she herself had no trouble remembering spellings even when her knowledge of English was still shaky. Differences between reading and spelling soon became her favorite project. She really wanted to know why words that Chris must encounter almost daily (e.g., the word "Southern" in the sign for Southern Railways) did not stick in his memory and could not be recalled when writing.

After first toying with the idea of visual memory problems, Uta turned away from this notion. Instead, she embraced the idea of a far more interesting and deeper problem with the representation of words in the mental lexicon (U. Frith 1985, 1999). She attributes this change in research direction largely to her attendance at a Summer Institute of the Society for Research in Child Development in Delaware in the summer of 1974. The phonological theory of dyslexia was still new and the clear winner in the exciting presentations and discussions. Some of these ideas are presented in the book *Cognitive Processes in Spelling* (U. Frith 1980) that Uta edited, and to which a sizeable number of the attendants of that Institute contributed.

The phonological theory made dyslexia far more than a problem with reading and spelling. It predicted that dyslexia would affect the spoken and not just the written word. This idea was pursued by Uta's first PhD student and later collaborator, Maggie Snowling (1998). As soon as it became possible to do so, we were keen to use brain imaging to find out whether, indeed, speech- and language-related areas were activated atypically in dyslexics. Chris was vital in leading the investigation, one of the first of its kind.[6] Our friend and colleague Eraldo Paulesu was key to the interpretation of, and connection to, the few existing anatomical studies (Paulesu et al. 1996).

This was in the early days of brain imaging, before the use of MRI, and Chris and the team used the elaborate and risky PET procedure to scan just five adult dyslexics and five matched controls while performing simple phonological tasks. The results suggested a disconnection between more frontal and more posterior parts of the language network of the left hemisphere in the dyslexic volunteers. But this was just the beginning. We had bigger plans, being highly aware of the variety of languages and writing systems. For example, the writing system in Italian is far less complex, and sound–letter mappings are easier to learn than in English. Cultural differences

---

[6] Such a study needed an enormous amount of preparation, not least to identify adult dyslexics, since previous work had only been with children. Fortunately, Maggie Snowling could trace volunteers from case studies that she had carried out in the past.

in the ease of reading, as assessed by reading speed and accuracy, were indeed reflected in brain activity (Paulesu et al. 2000). However, when we scanned dyslexics from England and Italy performing phonological tasks, we found the same atypical neurophysiological patterns (Paulesu et al. 2001). Thus, dyslexia revealed to us both cultural diversity and biological unity. Dyslexics in the two countries resemble each other in terms of biology, but in Italy dyslexics are largely invisible because Italian has a more transparent orthography than English, which makes reading and spelling much easier to acquire.

Uta had another favorite project that took on ever more importance in her work. Her first meeting with an autistic child, while training in clinical psychology, led to a lifelong fascination. She was extremely fortunate to be tutored by pioneers in the field, such as Lorna Wing, Michael Rutter, Beate Hermelin, and Neil O'Connor, all then working at London's Maudsley Hospital and conducting rigorous research into possible cognitive deficits. Even before she completed her PhD, and much to her delight, Uta was offered a post by Neil O'Connor as research scientist in his newly founded MRC Developmental Psychology Unit. This Unit (1968–1982) was located at University College London. Its successor was John Morton's MRC Cognitive Development Unit (1982–1996).

During all this time Uta was able to direct her own research, and with her students and colleagues she went on to develop theories to explain the core symptoms of autism, the extraordinary difficulties in social communication on the one hand and the presence of restricted repetitive interests and savant skills on the other. In her book *Autism: Explaining the Enigma* (U. Frith 1989) she discussed various ideas to interpret and explain the core symptoms. The book appeared just at the time when public interest in autism had sharply increased. One mark of this interest was the popular film *Rainman*, which first portrayed an autistic adult with a certain degree of accuracy.

One idea to explain the nonsocial features of autism was weak central coherence or detail focus. Another idea, to explain the social features, was a deficit in mentalizing, that is, a lack of theory of mind. Uta and her colleagues conceived of detail focus as a variation in processing style ranging from prioritizing detail to prioritizing gist, with most autistic individuals showing a detail-focused bias (Happé & U.Frith 2006). Experimental evidence, for example, in tests of embedded figures, showed that many autistic people had superior ability to isolate detail (Shah & U. Frith 1983). The mentalizing hypothesis was far more ambitious, as it laid claim to a new cognitive mechanism.

5. IN PURSUIT OF THE INTENTIONAL STANCE

Given the upheavals of the technological revolutions we experienced, our work still could have proceeded uneventfully and perhaps boringly, except that from time to time we stumbled across some extraordinary and novel ideas that captured our imagination. At the time they seemed to be serendipitous, coming out of the blue. With hindsight, they were of their time and surfaced simultaneously in different labs and different countries.

One such idea was theory of mind, a term that appeared in the title of a paper by Woodruff and Premack (Premack & Woodruff 1978). They boldly asked if chimpanzees understood mental states such as intentions, knowledge, and beliefs. This paper raised a lot of interest, especially in philosophers of mind such as Daniel Dennett, who was interested in the "intentional stance" (Dennett 1987).

This concept also chimed in with experiments carried out by our brilliant young Scottish colleague, Alan Leslie. He analyzed the looking behavior of young infants and was able to pick up subtly different responses to inanimate and animate events. He found that infants responded differently when an object was moved by a human hand as opposed to being moved by another inanimate object (Leslie 1984). We were fascinated by his ideas about the origin of pretend play, which pointed to a separation of cognitive processes underlying first-order representations (I see X in the physical world), and second-order meta-representations (I believe I see X in the mental world) (Leslie 1987). Significantly, lack of pretend play was one of the distinguishing features of autistic children first identified by Wing et al. (1977).

Perhaps it was the groundbreaking paradigm invented by developmental psychologists Heinz Wimmer and Josef Perner that had the biggest impact on us. Their "Maxi and the chocolate" test showed that 4- to 6-year-old children were able to use a theory of mind, while 3-year-olds were not (Wimmer & Perner 1983). Uta, with Alan Leslie and Simon Baron-Cohen, her PhD student at the time, modified Wimmer & Perner's Maxi test into the Sally-Ann test so as to be able to use it with autistic children (Baron-Cohen et al. 1985).

It goes like this: Sally puts her marble into her basket and then leaves the scene. While Sally is out, Ann takes the marble out of the basket and puts it into her box. Then Sally comes back and wants to play with her marble. Where will Sally look for her marble? This critical question was answered correctly by most children from about the age of 4 years (i.e., she will look in her basket where she must believe her marble is, as she was not there when it was moved). In other words, they take account of Sally's now false belief to predict her behavior. Autistic children, by contrast, typically gave answers based on the actual location of the marble and showed a marked delay in passing this simple test (Happé 1995).

This was only the beginning. A range of different theory of mind tests (e.g., Baron-Cohen et al. 1986, Perner et al. 1989, Sodian & U. Frith 1992) demonstrated that in autistic children a lack of understanding of the social world could coexist with excellent understanding of the physical world. From then on, the concept of mentalizing profoundly influenced our thinking about social communication impairments in both autism and schizophrenia.

It took time to get to grips with the idea that mentalizing is a complex process and that there is probably more than one form of mentalizing. False belief tasks like the Sally-Ann test require explicit awareness of the intentional stance. There is also an implicit form, a spontaneous and nonverbal tracking of another agent's intentions and beliefs. This implicit form of mentalizing is probably present from a very early age (e.g., Onishi & Baillargeon 2005) and possibly present also in some animal species, such as birds (Clayton et al. 2007). It has proved difficult to devise tasks that reliably distinguish between implicit and explicit mentalizing, but we suspect that explicit mentalizing involves an additional layer of information processing placed on top of an implicit mentalizing system.

One paradigm that we developed to overcome the limitations of explicit verbal tests were short videos of animated triangles, as introduced by Heider & Simmel (1944). They were created by Francesca Happé and Uta, first sketching out their ideas during a long train journey. The final videos present short sequences of three types of interactions between a red and a blue triangle: random, goal-directed, and intentional. Overall, spontaneous verbal interpretations tended to fit these three types of scenarios. For example, the interpretation "bouncing" might be elicited by a random interaction, "fighting" by a goal-directed interaction, and "tricking" by an intentional interaction (Abell et al. 2000).[7] Only the intentional scenes required mentalizing, and indeed,

[7] These stimuli, known as Frith-Happé animations, have been used to generate data from large

mentalizing language was used preferentially when describing these scenarios. Our PhD student Fulvia Castelli showed the animations to volunteers in the PET scanner with the simple instruction to just watch the videos (Castelli et al. 2000). She found a remarkably similar pattern of brain activations when contrasting intentional and random sequences as the one found earlier with written stories (see **Figure 3**). Furthermore, the results revealed a disconnection between the main hubs of the mentalizing system in the autistic volunteers (Castelli et al. 2002).

The results of our early PET studies, which relied on very few participants, were fortunately replicated and confirmed by fMRI studies in other laboratories (e.g., Schurz et al. 2021). **Figure 3**, which shows the results of our first brain imaging study, already points out the three hubs that define this system. More recently we suggested a hierarchical organization for this system (Frith & Frith 2020). It includes, first, a controller, located in medial frontal regions (medial prefrontal cortex, anterior cingulate cortex); second, a connector, located in temporal regions (posterior superior temporal sulcus, temporoparietal junction); and third, a navigator, located in more posterior regions of the cortex (posterior cingulate cortex, precuneus). We suggest that the controller hub initiates or inhibits the process of mentalizing, the connector links prior expectations and incoming sensory evidence, and the third hub acts as a navigator in a social space defined by people's status and approachability.

How best to explain the intentional stance in our information processing framework? In our 2023 book we proposed that the intentional stance is the most advanced way for agents to explain and predict the behavior of other agents (Frith & Frith 2023). A simpler way is to explain behavior in terms of goals and preferences of others, things that can be observed directly. This contrasts with intentions, which can only be inferred based on context, and the implicit assumption that the other agent is reciprocally interacting with other agents. Even simpler are the explanation and prediction of behavior in terms of physical laws, affecting equally living creatures and nonliving things (**Figure 5**).

---

populations for the human connectome project (e.g., Hillebrandt et al. 2014).

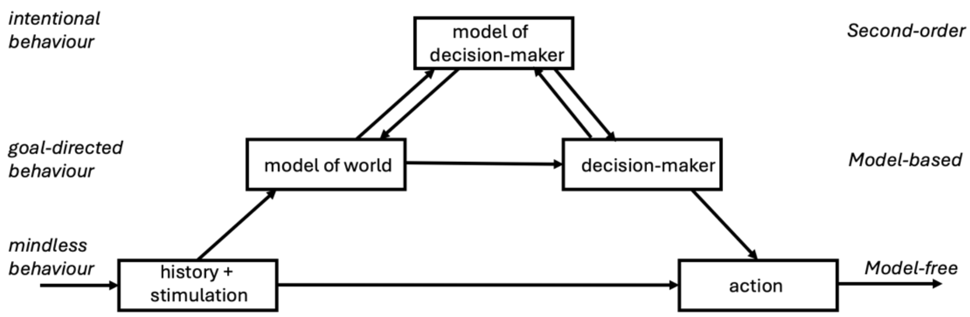

Figure 5 Predicting the behavior of agents. Figure adapted from Frith & Frith (2023).

6. HOMING IN ON SOCIAL COGNITIVE NEUROSCIENCE

We realized that we had long started to be intrigued by social cognitive processes, perhaps most obviously with our elective affinities paper (Frith & Frith 1991). Our study of the failures of social interaction and communication in autism and schizophrenia prepared us well for the question of what makes us humans the social creatures that we are. A proud moment in the realization that this was indeed the focus of our joint interests was our paper in 1999 titled "Interacting minds—a biological basis" (Frith & Frith 1999). It was a kind of manifesto that appeared at the right time when interest in social cognitive neuroscience was leading to a sudden steep increase in publications. We dared to spell out the need for a blueprint of social cognition, not social behavior. We stood by our belief that cognition is the necessary intermediate step to link brain and behavior. It is here that we need to situate abilities such as mentalizing. At the same time, we were aware that there would be many other social cognitive processes besides mentalizing waiting to be investigated (Frith & Frith 2010).

We were not alone in changing our orientation away from the individual and toward interactions between people (e.g., Schilbach et al. 2013). At the start of the twenty-first century, cognitive psychology turned into social neuroscience, gradually at first, but then fast gaining momentum. This led us to write a review of the field for the *Annual Review of Psychology* (Frith & Frith 2012).

As we reached our compulsory retirement age of 65, Uta in 2006 and Chris in 2007, we felt ready for new adventures. Chris accepted the invitation to become Nils Bohr Professor at the University of Aarhus in Denmark. Uta accepted a visiting professorship, all set in train by our colleague and friend Andreas Roepstorff. With him we set up a small group of "Interacting Minds," a group that can stand as a model of interdisciplinarity and collaboration. It included people with backgrounds in neurology, philosophy, semiotics, anthropology, music, religious studies, and biology. In group discussions we formulated the idea that the top in top-down control is not in the individual brain but at the interface with other brains (Roepstorff & C. Frith 2004).

We would move to Aarhus for up to six months each year to adapt as much as possible to a different culture and style. It was a style we liked very much, cool and uncluttered, full of new possibilities. The agreement was for five years in the first instance, but it stretched to about ten years. We still spend at least some weeks every year in Aarhus, continuing discussions with the friends we made there.

One of the projects carried out at the Interacting Minds Centre has become a milestone in the field. It showed a robust and surprising finding: Joint decisions of a pair of observers are superior to the decision of the best observer (Bahrami et al. 2010). Numerous papers since have confirmed this finding and have added to our knowledge of how to make better decisions in groups (Bang & Frith 2017). One tip is to have more diverse groups. This tied in with Uta's championing of women in science and her subsequent chairing of the Royal Society's Diversity Committee. She was able to formulate simple guidelines that were intended to prompt selection panels to overcome unconscious bias, all freely available on the Royal Society's website.

Reaching old age, we experienced many changes, none more evident than in observing our students forging their own careers. These changes produced worries and anxieties due to the increasing difficulties with funding, but this never outweighed our pride and joy. Has old age brought any wisdom? We are not convinced.

6.1. Social Cognition—What Next?

When writing our book, *What Makes Us Social?*, we had many arguments about whether selfishness was a default in human social interaction that could only be modified by the presence of social norms and by our desire to have a good reputation (Frith & Frith 2023). This was usually the position defended by Uta. Selfishness could be extended to include one's ingroup in the We-mode, and this might explain the loyalty to ingroups coupled with the disdain of outgroups. On the other hand, Chris defended the position of altruism as the default. This would explain the findings from experiments that showed that, with cognitive load, participants were likely to act spontaneously in an altruistic and fair way (Rand et al. 2012). Spontaneous empathy can result in heroic acts, another example where altruism triumphs (Brethel-Haurwitz et al. 2018). To overcome the default of altruism, and to work for our own advantage, we must use higher-order control, inhibiting otherwise unselfish behavior and justifying selfish behavior.

We have not yet found a solution to this tantalizing problem. It evokes some age-old arguments: Are we born with a moral instinct? Are moral principles universal? This may be so, but in addition, it seems we must learn quite specific rules of behavior prevalent in the society we live in. However, once these specific good behaviors have been learned they become automatic and, thus, a default (C. Frith 2023). Morality and norm-driven behavior are the topic of a growing field in philosophy and anthropology as well as experimental psychology.

6.2. Social Cognition and Culture

Where do these specific norms and moral principles come from? They come from our culture. Every culture has norms, but the precise nature of these norms will differ from one culture to another. We learn these norms from being embedded in our culture. We talked earlier of how cognition provides the bridge between brain and behavior. More recently, we have come to believe that cognition also provides the bridge between culture and behavior (Frith & Frith 2022).

Figure 2 illustrates how cultural priors might exert top-down control on our ideas, through our interactions with other people. These processes operate in the world of ideas, the top level of our hierarchy of control. Being part of a culture means sharing models in the world of ideas. We share models of how to behave and how to think (Heyes et al. 2020). With time, however, these ways of thinking and behaving become automatic and are elicited, bottom-up, by the many cultural signals that surround us.

We become aligned with our culture at many different levels. This alignment is driven by cognitive mechanisms such as predictive processing (Sørensen 2023). If our model does not align with others, this will generate a prediction error, indicating that we need to change our model. By continually minimizing these prediction errors, our models become more similar to each other (Friston & C. Frith 2015). This alignment within our cultural sphere makes communication, and joint action more generally, much easier—although, by the same process, interactions with other cultures becomes more difficult. Failure to recognize lack of alignment with another culture can lead to mutual suspicion and problems with implementing scientific best practice.

There is also always the danger that cultures can fragment. The increasing polarization we are observing today is an example of such fragmentation. Our world models are no longer aligned, and to preserve our models (our beliefs and values), we must decide which prediction errors are important and which should be ignored. Our evidence may be your fake news. We hope that exploration of the cognitive bases of these processes will help us better understand and perhaps mitigate the problems of polarization.

6.3. Social Cognition and Artificial Intelligence

Recently our culture has been invaded by a new entity, artificial intelligence (AI) in the form of LLMs. LLMs have turned out to be very useful, and this has had a rapid and dramatic impact on us all. LLMs, such as ChatGPT, can engage with us and answer questions with coherent and plausible responses in a language we can readily understand. It is probably because of this ability to interact with us with such fluency that many people (~67%) are willing to attribute some possibility of subjective experience to LLMs (Colombatto & Fleming 2024a). But what do the experts, the philosophers, and the psychologists think (LeDoux et al. 2023)? Are LLMs conscious? Do they understand what they are saying?

We believe that these are not the most important questions. What we need to explore is how to optimize human–AI interactions. Will AIs be treated as mindless assistants that are useful in cooperative situations but can be ruthlessly exploited in competitive situations? One study suggests this might be a probable scenario. People expect AIs to be as cooperative as humans but are happy to exploit them in situations where they would hesitate to exploit humans. For example, when merging lanes on a highway, people tend to refuse to give way to a self-driving vehicle compared to a vehicle with human drivers (Karpus et al. 2021). The We-mode, with its emphasis on the outcome for the group, is one cognitive mechanism for reducing such exploitation. When AIs and humans work together, can they adopt the We-mode? This would enhance cooperation and greatly increase the knowledge and action space available.

Joint actions and decisions are much enhanced by taking into account the confidence of the individuals involved (Bahrami et al. 2010). This confidence can be expressed explicitly in words (e.g., Fusaroli et al. 2012) but can also be inferred from behaviors such as response speed (Patel et al. 2012). As yet, AIs do not typically indicate their confidence, either explicitly or implicitly. At the same time, people tend to attribute high confidence to answers and solutions provided by AIs, much higher than is justified by their behavior (Colombatto & Fleming 2024b). We believe that to optimize human–AI interactions, AIs will need to signal the level of their confidence. This may require metacognition and self-reflection (Fleming 2021). But will AIs be developed that have a "self" upon which they can reflect? Ability to work with AIs successfully is a major task that has been started in pioneering studies (Harris 2024).

6.4. We Need More Social Cognition

We realize it is rather vain to claim that one's pet research field has practical significance for everyday life. But surely, we have little hope to improve our lives unless we understand ourselves better. It is our hope that some of the already existing work in social neuroscience gives pointers toward such application.

Morality, free will, responsibility, fairness, ingroup loyalty, and empathy are all dependent on unconscious brain processes that dynamically respond to cultural control. This is equally true for the dark side of our social nature, as manifest in selfishness, deception, aggression, and revenge. The importance of the law and norms to restrain these impulses and extend their control to all groups and nations cannot be exaggerated. We can only hope that cognitive neuroscience will give us some tools to better understand the dark side of our social nature, which is so apparent when we manipulate other people's beliefs for personal gain.

Self-knowledge is an essential first step toward building a better-functioning society. Some small steps may be occurring in the recognition of unconscious bias against outgroups and of the power of diversity. At the very least, we can monitor each other for bias. There seems to be a chance of increasing awareness of misinformation in a world that is more and more connected via the Internet. It would not seem impossible to develop ideas that might help to build resilience against fake news and conspiracy theories, perhaps an equivalent to vaccination (Lewandowsky & van der Linden 2021). We feel optimistic when reading about experiments where small groups discuss divisive topics and where moderate voices expressed with confidence can counteract polarization (Navajas et al. 2019).

Social cognitive neuroscience is still in its infancy. We need more developmental studies, and we need more interactive group studies, with more than just two agents. We do want to say a word of warning, too. There is so much to do, and yet we would like to urge young researchers to strive for quality and not quantity. Sadly, it is hard to resist the current publish-or-perish research culture. To some extent, this culture is an unintended side effect of the marvel of computers and the Internet. Looking back, we can appreciate how lucky we were that our own productivity was not judged by quantity, so that we felt we had space to think. The urge to publish and to rack up citations can only be tempered by the urge to publish only what will truly contribute to progress in science. How would you know what this is? There is probably no better measure of quality than the critical judgement of your expert peers (Dreber et al. 2015). You yourself may have a fine sense of what is your best work. Then we can wait for the test of time.

We would also urge young researchers not to be carried away with the popularity of big data and the ease with which such huge data sets can be analyzed. This seems to us to offer a deceptive certainty of getting the facts. Since we believe in predictive processing, we believe that priors—that is, theories, models, and testable hypotheses—are essential to finding the facts. The promise that something might emerge is as illusory as the idea that one can find an object in a messy room without having some idea of what it looks like, how big it is, and so on. If we have some idea of the desired object, we need less big data.

We also believe that instead of questionnaires to be completed by unknown users, we need new paradigms, perhaps more realistic economic games. Thinking up different paradigms was what we liked best in the pursuit of the still elusive processes involved in mentalizing. Different paradigms can be pitted against each other and can lead to convergent truths. But of course, we need enthusiasm for the work and a willingness to take risk. It is too late for us to learn from our own mistakes, but not too late for you, gentle reader.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We have greatly benefitted from helpful comments on this article by Shinobu Kitayama, Steven Fleming, Francesca Happé, Maggie Snowling, and Ros Ridley.

LITERATURE CITED

Abell F, Happé F, Frith U. 2000. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cogn. Dev.* 15(1):1–16

Attneave F. 1959. *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*. Henry Holt

Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD. 2010. Optimally interacting minds. *Science* 329(5995):1081–85

Bang D, Frith CD. 2017. Making better decisions in groups. *R. Soc. Open Sci.* 4:170193

Barnard GA, Bayes T. 1958. Studies in the history of probability and statistics: IX. Thomas Bayes's *Essay Towards Solving a Problem in the Doctrine of Chances*. *Biometrika* 45(3–4):296–315

Baron-Cohen S, Leslie AM, Frith U. 1985. Does the autistic child have a "theory of mind"? *Cognition* 21(1):37–46

Baron-Cohen S, Leslie AM, Frith U. 1986. Mechanical, behavioural and intentional understanding of picture stories in autistic children. *Br. J. Dev. Psychol.* 4(2):113–25

Blakemore SJ, Smith J, Steel R, Johnstone EC, Frith CD. 2000. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol. Med.* 30(5):1131–39

Brethel-Haurwitz KM, Cardinale EM, Vekaria KM, Robertson EL, Walitt B, et al. 2018. Extraordinary altruists exhibit enhanced self-other overlap in neural responses to distress. *Psychol. Sci.* 29(10):1631–41

Castelli F, Frith C, Happe F, Frith U. 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125(Pt. 8):1839–49

Castelli F, Happe F, Frith U, Frith C. 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12(3):314–25

Clayton NS, Dally JM, Emery NJ. 2007. Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philos. Trans. R. Soc. B* 362(1480):507–22

Coid J. 1994. Failure in community care: psychiatry's dilemma. *BMJ* 308(6932):805

Colombatto C, Fleming SM. 2024a. Folk psychological attributions of consciousness to large language models. *Neurosci. Conscious.* 2024(1):niae013

Colombatto C, Fleming SM. 2024b. Illusions of confidence in artificial systems. PsyArXiv. https://doi.org/10.31234/osf.io/mjx2v_v1

Craik KJW. 1943. *The Nature of Explanation*. Cambridge University Press

Crapse TB, Sommer MA. 2008. Corollary discharge across the animal kingdom. *Nat. Rev. Neurosci.* 9:587–600

Dennett DC. 1987. *The Intentional Stance*. MIT Press

Dreber A, Pfeiffer T, Almenberg J, Isaksson S, Wilson B, et al. 2015. Using prediction markets to estimate the reproducibility of scientific research. *PNAS* 112(50):15343–47

Eysenck, H. J. (Ed.). 1973. *Handbook of Abnormal Psychology* (2nd ed.). Sir Isaac Pitman and Sons Ltd.

Fechner GT. 1860. *Elemente der Psychophysik.* Breitkopf und Härtel

Fleming SM. 2021. *Know Thyself: The Science of Self-Awareness*. Basic Books

Fletcher PC, Frith CD. 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10(1):48–58

Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, et al. 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition* 57(2):109–28

Friston K, Frith C. 2015. A duet for one. *Conscious. Cogn.* 36:390–405

Frith CD. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Lawrence Erlbaum Associates

Frith CD. 2007. *Making Up the Mind: How the Brain Creates Our Mental World*. Blackwell

Frith CD. 2021. The neural basis of consciousness. *Psychol. Med.* 51(4):550–62

Frith CD. 2023. 49th Bartlett Lecture: consciousness, (meta)cognition, and culture. *Q. J. Exp. Psychol.* 76(8):1711–23

Frith CD, Done DJ. 1989. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol. Med.* 19(2):359–63

Frith CD, Frith U. 1972. The solitaire illusion: an illusion of numerosity. *Percept. Psychophys.* 11:409–10

Frith CD, Frith U. 1991. Elective affinities in schizophrenia and childhood autism. In *Social Psychiatry: Theory, Methodology, and Practice*, ed. P Bebbington, pp. 65–88. Transactions

Publishers
Frith CD, Frith U. 1999. Interacting minds—a biological basis. *Science* 286(5445):1692–95
Frith CD, Frith U. 2012. Mechanisms of social cognition. *Annu. Rev. Psychol.* 63:287–313
Frith CD, Frith U. 2020. Mapping mentalising in the brain. In *The Neural Bases of Mentalizing*, ed. K Ochsner, M Gilead. Springer Press
Frith CD, Frith U. 2022. The mystery of the brain-culture interface. *Trends Cogn. Sci.* 26(12):1023–25
Frith CD, Lang RJ. 1979. Learning and reminiscence as a function of target predictability in a two-dimensional tracking task. *Q. J. Exp. Psychol.* 31(1):103–9
Frith CD, Perry R, Lumer E. 1999. The neural correlates of conscious experience: an experimental framework. *Trends Cogn. Sci.* 3(3):105–14
Frith U. 1971. Why do children reverse letters? *Br. J. Psychol.* 62(4):459–68
Frith U. 1974a. A curious effect with reversed letters explained by a theory of schema. *Percept. Psychophys.* 16(1):113–16
Frith U. 1974b. Internal schemata for letters in good and bad readers. *Br. J. Psychol.* 65(2):233–41
Frith U. 1985. Beneath the surface of developmental dyslexia. In *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*, ed. KE Patterson, JC Marshall, M Coltheart, pp. 301–30. Routledge
Frith U. 1989. *Autism: Explaining the Enigma*. Blackwell
Frith U. 1999. Paradoxes in the definition of dyslexia. *Dyslexia* 5(4):192–214
Frith U. 2012. The 38th Bartlett Lecture: why we need cognitive explanations of autism. *Q. J. Exp. Psychol.* 65(11):2073–92
Frith U. 2020. Fast lane to slow science. *Trends Cogn. Sci.* 24(1):1–2
Frith U, ed. 1980. *Cognitive Processes in Spelling*. Academic
Frith U, Frith C. 2010. The social brain: allowing humans to boldly go where no other species has been. *Philos. Trans. R. Soc. B* 365(1537):165–76
Frith U, Frith CD. 2023. *What Makes Us Social?* MIT Press
Frith U, Frith CD, Frith A, Locke D. 2022. *Two Heads: Where Two Neuroscientists Explore How Our Brains Work with Other Brains*. Bloomsbury
Frith U, Happé F. 1998. Why specific developmental disorders are not specific: on-line and

developmental effects in autism and dyslexia. *Dev. Sci.* 1(2):267–72
Frith U, Robson JE. 1975. Perceiving the language of films. *Perception* 4(1):97–103
Fusaroli R, Bahrami B, Olsen K, Roepstorff A, Rees G, et al. 2012. Coming to terms: quantifying the benefits of linguistic coordination. *Psychol. Sci.* 23(8):931–39
Gallotti M, Frith CD. 2013. Social cognition in the we-mode. *Trends Cogn. Sci.* 17(4):160–65
Goethe JW von. 1872. *Elective Affinities.* Henry Holt & Co.
Happé F, Frith U. 2006. The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36(1):5–25
Happé FGE. 1995. The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Dev*. 66(3):843–55
Harris LT. 2024. The neuroscience of human and artificial intelligence presence. *Annu. Rev. Psychol.* 75:433–66
Heider F, Simmel M. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* 57:243–49
Helmholtz H von. 1948 (1867). Concerning the perceptions in general. In *Readings in the History of Psychology*, ed. W Dennis, pp. 214–30. Appleton-Century-Crofts
Heyes C, Bang D, Shea N, Frith CD, Fleming SM. 2020. Knowing ourselves together: the cultural origins of metacognition. *Trends Cogn. Sci.* 24:349–62
Hillebrandt H, Friston KJ, Blakemore S-J. 2014. Effective connectivity during animacy perception—dynamic causal modelling of Human Connectome Project data. *Sci. Rep.* 4(1):6240
Hohwy J. 2013. *The Predictive Mind*. Oxford University Press
Howard SR, Dyer AG. 2024. Quantity misperception by hymenopteran insects observing the solitaire illusion. *iScience* 27(2):108697
Jaspers K. 1962. *General Psychopathology*. Manchester University Press
Johnstone EC, Crow TJ, Frith CD, Carney MW, Price JS. 1978. Mechanism of the antipsychotic effect in the treatment of acute schizophrenia. *Lancet* 1(8069):848–51
Johnstone EC, Crow TJ, Frith CD, Husband J, Kreel L. 1976. Cerebral ventricular size and cognitive impairment in chronic schizophrenia. *Lancet* 2(7992):924–26
Karpus J, Krüger A, Verba JT, Bahrami B, Deroy O. 2021. Algorithm exploitation: Humans are keen to exploit benevolent AI. *iScience* 24(6):102679

LeDoux J, Birch J, Andrews K, Clayton NS, Daw ND, et al. 2023. Consciousness beyond the human case. *Curr. Biol.* 33(16):R832–40

Leslie AM. 1984. Infant perception of a manual pick-up event. *Br. J. Dev. Psychol.* 2(1):19–32

Leslie AM. 1987. Pretense and representation: the origins of "theory of mind." *Psychol. Rev.* 94(4):412–26

Lewandowsky S, van der Linden S. 2021. Countering misinformation and fake news through inoculation and prebunking. *Eur. Rev. Soc. Psychol.* 32(2):348–84

Luce RD. 1963. Detection and recognition. In *Handbook of Mathematical Psychology*, Vol. 1, ed. RD Luce, RR Bush, E Galanter. Wiley

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. 1996. Neural correlates of category-specific knowledge. *Nature* 379(6566):649–52

McCarthy RA, Warrington EK. 1990. *Cognitive Neuropsychology*. Academic

Medawar PB. 1967. *The Art of the Soluble*. Methuen

Morton J, Frith U. 2001. Why we need cognition: cause and developmental disorder. In *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, ed. E Dupoux, pp. 263–78. MIT Press

Navajas J, Heduan FÁ, Garrido JM, Gonzalez PA, Garbulsky G, et al. 2019. Reaching consensus in polarized moral debates. *Curr. Biol.* 29(23):4124–29

Onishi KH, Baillargeon R. 2005. Do 15-month-old infants understand false beliefs? *Science* 308(5719):255–58

Patel D, Fleming SM, Kilner JM. 2012. Inferring subjective states through the observation of actions. *Proc. R. Soc. B* 279(1748):4853–60

Paulesu E, Démonet JF, Fazio F, McCrory E, Chanoine V, et al. 2001. Dyslexia: cultural diversity and biological unity. *Science* 291(5511):2165–67

Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, et al. 1996. Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* 119(1):143–57

Paulesu E, McCrory E, Fazio F, Menoncello L, Brunswick N, et al. 2000. A cultural effect on brain function. *Nat. Neurosci.* 3(1):91–96

Perner J, Frith U, Leslie AM, Leekam SR. 1989. Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Dev.* 60(3):689–700

Posner MI, Petersen SE, Fox PT, Raichle ME. 1988. Localization of cognitive operations in the human-brain. *Science* 240(4859):1627–31

Premack D, Woodruff G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 4:515–26

Rand DG, Greene JD, Nowak MA. 2012. Spontaneous giving and calculated greed. *Nature* 489(7416):427–30

Read JCA. 2015. The place of human psychophysics in modern neuroscience. *Neuroscience* 296:116–29

Roepstorff A, Frith C. 2004. What's at the top in the top-down control of action? Script-sharing and "top-top" control of action in cognitive experiments. *Psychol. Res*. 68(2–3):189–98

Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, et al. 2013. Toward a second-person neuroscience. *Behav. Brain Sci.* 36(4):393–414

Schurz M, Radua J, Tholen MG, Maliske L, Margulies DS, et al. 2021. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* 147(3):293–327

Shah A, Frith U. 1983. An islet of ability in autistic children: a research note. *J. Child Psychol. Psychiatry* 24(4):613–20

Shallice T, Burgess PW, Frith CD. 1991. Can the neuropsychological case-study approach be applied to schizophrenia? *Psychol. Med.* 21:661–73

Shannon CE, Weaver W. 1949. *The Mathematical Theory of Communication*. University of Illinois Press

Snowling M. 1998. Dyslexia as a phonological deficit: evidence and implications. *Child Psychol. Psychiatry Rev.* 3(1):4–11

Sodian B, Frith U. 1992. Deception and sabotage in autistic, retarded and normal children. *J Child Psychol. Psychiatry* 33(3):591–605

Sørensen J. 2023. *Why Cultures Persist: Toward a Cultural Immunology*. Aarhus University Press

Tanner WP Jr., Swets JA. 1954. A decision-making theory of visual detection. *Psychol. Rev.* 61(6):401–9

Warrington EK, Shallice T. 1984. Category specific semantic impairments. *Brain* 107(3):829–53

Wimmer H, Perner J. 1983. Beliefs about beliefs—representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1):103–28

Wing L, Gould J, Yeates SR, Brierley LM. 1977. Symbolic play in severely mentally retarded and in autistic children. *J. Child Psychol. Psychiatry* 18(2):167–78

Yon D, Frith CD. 2021. Precision and the Bayesian brain. *Curr. Biol.* 31(17):R1026–32