JNphi

A Critical Review of *Determined: Life Without Free Will* (2023)

Max Whittaker

Abstract

The current paper critically appraises Robert Sapolsky's recent popular science book, *Determined* (2023). Sapolsky's basic claim is that there is much less scope for free will than many intuitively imagine, though book's subtitle makes clear that free will is viewed as a wholly untenable concept. If there is no scientific basis for free will, Sapolsky argues, there is no basis for individual culpability, and little basis for society be organised around this principle. Modern forms incarceration, laden with notions of personal culpability, are thus unjustified. An instrumental use of quarantine, stripped of notions of guilt and shame, would represent a more rational, scientific approach towards those who behave dangerously. In response, the current review argues that the existence of free will is largely irrelevant to the question of individual responsibility. It also questions whether it is self-evident that society should be based upon a modern scientific understanding of human behaviour, while further arguing that scientific knowledge emerges from a professional culture grounded in individual accountability. Some cautionary reflections are then offered which challenge Sapolsky's belief that more humane and compassionate societies will emerge if we dispense with notions of free will and personal culpability. Lastly, it will be argued that attributions of guilt and shame may not be wholly corrosive forces, but may in fact be instrumental in the maintenance stable and humane forms of human organisation.

Key Words: free will, personal responsibility, determinism, scientific knowledge, neurobiology

DOI: 10.5281/zenodo.14546708

Introduction

The current paper is a rejoinder to Robert Sapolsky's recent popular science book, *Determined* (2023). This review does not reckon with the book's foremost contention; that humans do not, in fact, possess free will. Instead, the scope of this review is restricted to critically

Corresponding author: Dr Max Whittaker, Clinical Psychologist, Specialist Autism Team, Preston, United Kingdom
e-mail ✉ max.whittaker@lscft.nhs.uk

appraising the book's other contentions which, according to Sapolsky, follow logically and morally once we accept that free will does not exist. As Sapolsky states, with admirable clarity, his fundamental case is that "we have no free will at all." Sapolsky then argues: "here would be some of the logical implications of that being the case; that there can be no such thing as blame and that punishment as retribution is indefensible. Sure, keep dangerous people from damaging others, but do so as straightforwardly and non-judgementally as keeping a car with faulty brakes off the road." (p.5). Sapolsky also suggests there will be a net-benefit to society if our intuitive ideas about personally culpability are abandoned in favour of mechanistic models of human behaviour in which there is no scope – or, at least, a much more limited scope – for personal responsibility. In the current review it is argued that the existence of free will is largely irrelevant to the question of individual responsibility. It will also be argued that it is by no means self-evident that society should be based upon a modern scientific understanding. Rather, the modern scientific understanding upon which Sapolsky believes society should be based is predicated upon an *a priori* set of values and norms which emerge from general culture of individual responsibility. Therefore, arguing that society should abandon notions of individual responsibility because a belief in free will is irreconcilable with a modern scientific understanding of human behaviour is fundamentally paradoxical. The review will then offer some cautionary reflections on Sapolsky's view that abandoning notions of free will and personal culpability will lead to more decent, humane and compassionate forms of human organization.

Determined: core arguments

Sapolsky's core argument is that the scope of supposedly free behaviour has inevitably narrowed as our understanding of biology has increased. As he correctly highlights, societies have previously recalibrated perceptions of personal culpability in particular cases following advances in our understanding of neurobiology and its relationship to human behaviour. For example, while it is now considered self-evident that people with epilepsy are not culpable for behaviours caused by seizure activity in the brain, this was not always so. Sapolsky recounts how those with epilepsy, and later their parents, were held morally culpable for the seizures, fits or convulsions, caused by involuntary seizure activity. Sapolsky is surely correct that, within the context of neurological conditions such as epilepsy – and indeed many other neurological conditions such as the dementias – dispensing with notions of personal culpability and adopting a neurobiological and mechanistic understanding of a particular behaviour represents unambiguous progress, scientifically and morally. Sapolsky uses examples of neurological conditions such as epilepsy to illustrate his central contention; "free will" is what we call a biological system that we have not yet understood. Once we

understand the biology of a particular system which precedes a given behaviour, this understanding is ultimately exculpatory, as the biological processes which occurred prior to the behaviour are entirely mechanistic and determined by a chain of causation which fundamentally stretches back to the beginning of the universe. There is no space within this causal chain, according to Sapolsky, for free will.

Thus, just as we should no longer hold those with neurological conditions, such as epilepsy or forms of dementia, accountable for whatever behaviours are caused by the underlying, mechanistic processes associated with those particular neuropathologies, neither are we truly justified in holding *anyone* personally accountable for their behaviour. While Sapolsky acknowledges the compelling subjective sense of free will, he in steadfast in view that we must follow the implications of a modern scientific understanding of neurobiology whatever the implications may be. If that means dispensing with beliefs to which we may have strong emotional attachments, so be it! To underscore this point, Sapolsky emphasises that various sub-categories of our subjective experience, which intuitively feels as though they are synonymous with our personal agency, are illusory. He takes specific aim at a category of subjective experience to which we are likely to have a strong emotional attachment, namely "grit". Though we may experience pride in our tenacity if we have achieved a particular goal and have overcome particular obstacles, losing our sense of pride in these accomplishments is not only logically entailed by "the science", but also ultimately a price worth paying. Within the context of Sapolsky's thesis, stating that "grit" does not exist is, of course, a tautology. If free will does not exist, then by definition any sub-component of free will does not exist. However, Sapolsky's assertion that grit does not exist represents a confusion about the ontological status of subjective experiences. Grit *does exist*; it exists in the form of a subjective experience which has a particular qualitative character, which is distinct from other categories of subjective experience. Grit exists in same way physical pain exists. Both grit and pain have a first-person ontology. While the neurological activity associated with these subjective states has a third person ontology, insofar as it can be observed on neuroimaging, our subjective experiences are not directly accessible by others. However, they exist as first-person ontologies.

The key point with respect to Sapolsky's thesis, however, is that nothing is sacred; he is merely following the implications of his thesis, even if this means pulling the existential rug from underneath our most cherished accomplishments. Indeed, he locates in at least some critics of his thesis a psychological weakness, in that those who want to retain notions of personal agency are presumed to be trying to cling on to feelings of personal pride. Despite Sapolsky's objection this apparent motivation, it is not made clear why a sense of pride in one's accomplishments leads to bad outcomes. It is surely an empirical –

though fiendishly complex – question whether instrumentally validating a sense of personal achievement, in oneself and in others, incentivises the types of behaviour a decent society would wish to incentivise. It is entirely plausible that those who argue that humans have free will are arguing from a position of motivated reasoning and merely wish to evade the pain of disillusionment. However, it equally follows that those who argue against free will may be arguing from a position of motivated reasoning; perhaps they wish to less the pain associated with negative emotions such as guilt, shame or a sense of personal failure. One can thus postulate powerful psychological motivations in either direction. However, it is difficult to see how attributing a difference in perspective to a psychological weakness within others leads anywhere other than counterproductive polarisation. In some sense, of course, this is all beside the point. The key question is whether we can retain notions of individual responsibility if free will does not exist. This review will argue that notions of personal responsibility can indeed be retained, quite straightforwardly in fact. Furthermore, it will be argued that shame and stigma are not wholly corrosive forces. Instead, these attributions can be invoked for instrumental purposes and can serve as powerful disincentives for harmful behaviour.

The influence of abstract incentives on conscious, chaotic systems

The central question that Sapolsky seeks to answer is “why did that behaviour occur?” (p.3) In answering this question, Sapolsky states that we need to considerations what occurred in the immediate moments, days, weeks, centuries, and indeed millennia, prior to a particular behaviour. Sapolsky cites different factors which determine our behaviour, from the activity of neurotransmitters immediately prior to a given action, our endocrine activity in the moments, days and weeks prior to an action, the foetal environment, the socioeconomic status of our patents and indeed the behavioural traits and habits of our ancestors. As Sapolsky phrases it, “we are nothing more or less than the cumulative biological and environmental luck over which we had no control that has brought us to any moment” (p.4). It is unclear on Sapolsky’s account how much weight should be assigned to any of these factors. He does not quantify, for instance, the significance of an individual’s cortisol or testosterone levels before a given behaviour, but instead cites these hormones as prototypical examples of the mechanical factors which combine to determine our behaviour. One may be tempted to argue that while Sapolsky cites variables which are highly relevant to understanding human behaviour from the purview of neurobiology, these only form part of a very complex picture. However, on Sapolsky’s account, neurobiology *is* the whole picture. However complex and chaotic the behavioural equation undoubtedly is, our behaviour is necessarily the product of

deterministic neurobiological processes and *nothing else*. As Sapolsky correctly emphasises, though it is likely to be impossible to calculate to contribution of all the individual factors in a chaotic system, much less precisely determine how all these myriad factors interact with one another, chaos is *not* indeterminism. For Sapolsky, it is crucial that there is no extra part, no homunculus, or ghost in the machine, inhabiting the human system and thus no scope whatsoever for truly free agency. To be clear, Sapolsky is not proposing there should be no consequences to forms of antisocial or destructive behaviour. Sapolsky recognises it is morally necessary to place restrictions upon dangerous individuals. However, he argues that such restrictions should be akin to an instrumental use of quarantine. Quarantine, according to Sapolsky, is distinct from incarceration in modern penal systems, where those found guilty of harmful behaviours are given custodial sentences based on a sense that they are personally culpable for their behaviours. Attributions of persona blame, shame and stigma thus permeate the whole culture of incarceration. Sapolsky argues that, in some fundamental sense, we are no more justified in personally holding to account those who inflict harm on others than we are holding to account a malfunctioning car. A malfunctioning car needs to be taken off the road, but holding the car responsible of course makes no sense, as its failure resulted from an entirely deterministic and mechanical process. So it is too, according to Sapolsky, with humans.

One can accept Sapolsky's description of the underlying processes which precede decision making, while also noticing that absent from Sapolsky's account is any substantive consideration of how abstractions, such as norms, expectations and the spectre of punishment or social ostracism, are also part of the neurobiological picture and shape human behaviour. This is an important omission as the effects of norms and interpersonal expectations on human behaviour have profound implications for the overall thesis that individual responsibility is untenable. Indeed, Sapolsky's example of a malfunctioning car engine is quite misleading, as *mechanistic failure* is not synonymous with *moral failure*. Like all physical systems, a car engine is in a progressive state of entropy from the moment of its creation, and its eventual failure is inevitable. However, this is a very poor analogue for human behaviour. As a matter of fact, humans have the capacity to pursue many different and complex courses of action and possess cognitive faculties which enable them to respond to incentives, such as personal praise and reward, and punishments. A prototypical form of punishment in modern society is incarceration, whereby individual liberty is suspended for a period of time as specified by formal criminal codes. Punishment, however, can also come in the form of interpersonal personal admonishments, such as a stinging interpersonal rebuke, or social ostracism. One may not violate a formal law, but behave in such a way that is met with disgust or revulsion from others. The spectre of social ostracism is something

which is profoundly disturbing for a species which evolved in tight-knit communities. Whereas all car engines move inexorably towards inevitable failure, humans have the capacity to respond innovatively and pursue many different courses of action. Unlike cars, our actions are influenced by objective punishment as well as the *abstract spectre of punishment*, objective forms of social ostracism or the *anticipation of social ostracism*. Whether this is due to an intrinsic capacity for truly autonomous conduct, or merely because human brains are particularly receptive to these forms of feedback, is beside the point. The key issue is that humans, as a matter of fact, have the capacity to follow many courses of action, whereas a car engine does not. Social feedback, and the prospect of social feedback, plays no role in a car engine's inevitable progression towards mechanistic failure. Therefore, placing a sincere and earnest moral expectation on a car engine not to fail will have no functional impact upon anything, except maybe upon one's own sanity. Indeed, King Lear's beseeching of the weather, another deterministic though chaotic system, was a motif of his own descent into derangement. Human brains, by contrast, are particularly receptive to interpersonal feedback. The inputs into the neurobiological equation, therefore, include various social and interpersonal abstraction, such as the spectre of ostracism and attributions of shame. It is thus not self-evident why it is wrong to place expectations upon other conscious agents who have the capacity to pursue different courses of action, *irrespective of whether or their behaviour is free or determined*. Indeed, at various points through the book, Sapolsky himself expresses incredulity and moral indignation that certain human did not behave differently. Quite astonishingly, in a book which repeatedly affirms there is no basis for either free will or personal responsibility, Sapolsky holds to account members of the psychoanalytic school of psychiatry – who Sapolsky labels as "psychoanalytic scumbags" (p.329) – for the harm they perpetrated against the parents of schizophrenic children, by attributing this disease to pathological parenting styles. Sapolsky's disdain for the psychoanalytic school is evident as he recounts how "a sneering, pejorative term" was developed "for families, that is mothers, of schizophrenic patients, who tried to dodge responsibility by believing that it was a brain disease: dissociative organic types." (p.329). Sapolsky favourably recounts the activism from those parents in response to the stigmatisation they encountered from the psychoanalytic school of psychiatry. He sympathetically acknowledges the "emotional rage" and "bitterness" of the families who were "furious in their nice mid-western way." However, the activism of parents, driven by a deep sense of anger and injustice was presumably based on the assumption that the psychoanalytic psychiatrists who had unfairly stigmatised them were personally and professionally culpable for their conduct. Instead of providing genuine medical care to some of the most vulnerable members of society, these clinicians, in Sapolsky's view, arrogantly promulgated pseudo-scientific beliefs –

"schizophrenogenic voodoo" – that caused harm to those they were duty bound to serve. Sapolsky's anger and incredulity only makes sense, however, if he believes these professionals could have been reasonably be expected to have behaved differently. Furthermore, in using the terms "scumbag" and "voodoo", Sapolsky evidently wishes to contribute to a broader culture in which medical malpractice is shameful, personally and professionally, for the individuals responsible.

A more mundane example of every-day decision making is perhaps helpful in illustrating the ways in which social feedback and anticipated feedback shape human behaviour. In a typical workplace, there is a general expectation that employees arrive for work on time. Let us imagine that most employees at this hypothetical workplace do, in fact, generally arrive on time. Why does this occur? Presumably, employees are aware of the expectation of punctuality, and are likely to be aware of this general work-based expectation without it being explicitly articulated. Presumably they are also aware of the implications of persistent lateness, in terms of the reputational damage they may incur and the potential for this to invite scrutiny of other aspects of their performance, which may then be negatively framed by their lack of punctuality. This may ultimately lead to future opportunities for career advancement being denied and, potentially, a termination of employment. Unemployment can lead very quickly to financial instability and a loss of social standing. The everyday expectation of punctuality thus sits is within a hierarchy of human concerns. Our evolutionary heritage is clearly central to understanding human behaviour in this context; we are primates with overarching concerns about our position within hierarchical structures. However, our evolutionary heritage is only relevant only in a *general* sense, not the highly bespoke and individualised sense implied by Sapolsky. As evolved primates, humans have a variety of deep-seated fears and preoccupations. Amongst our most basic fears are fast moving objects and biological contamination. However, another of our most fundamental fears is social isolation. Crucially, human anxiety is not confined to the moment of social exclusion. Instead, we nervously anticipate the various ways in which we may be negatively evaluated and excluded by others (Baumeister & Tice, 1990). Employment is one of the ways in which humans make a distinctive contribution to communal life and attain a measure of social standing. The consequences of unemployment are thus far reaching, threatening to immediately de-stabilise our position within a socio-economic hierarchy and undermine our sense of community participation, while also undermining the possibility of attracting a mating partner. It seems self-evident, in fact, that a desire to avoid the psychological distress that accompanies the destabilising effects of unemployment is an important factor in the maintenance of punctuality as a work-based norm. It is not practically possible for any employer to consider, in its infinite complexity, our idiosyncratic social

histories and individualised evolutionary lineage, or our endocrine activity at any given moment, and weigh up all of these factors to determine how the behaviour of each individual should ultimately be appraised. For example, should employee X be granted extra leeway if their testosterone and cortisol levels differ significantly from colleague Y? And, if so, what would be the implications of inequitably treating colleagues have on general workplace norms and morale? The key question, surely, is whether a person in question possesses the *general faculties* that enable them to be receptive to an incentive structure and expectations such as punctuality. If so, other people are justified in placing a standardised expectation upon them. Crucially, if a naïve observer surveyed the vast differences between colleagues within any workplace, in terms of endocrine activity, personal histories, genetic lineage, diet and cognitive capacities, they may predict vast individual differences in punctuality, given the decision-making of any one person emerges from highly individual, chaotic processes, stretching back to the Big Bang. Yet, if most employees in a workplace generally arrive on time, this suggests that a general environment of accountability is not only part of the "what occurred in the moments before a particular behaviour" picture, but a *decisive* factor. Despite an infinite number of individual differences forming part of the behavioural equation, the *constant factor of accountability*, which is both explicit and implicit within a work-based culture, reliably predicts the behaviour of employees. Therefore, one can accept Sapolsky's view that neurobiology is the *whole* picture of human behaviour. Yet interpersonal expectations, and the abstract spectre of unemployment and its destabilising consequences such as a loss of social standing, impact upon our neurobiological systems in a way that decisively predicts behavioural outcomes.

The key point is that one need not posit the existence of free will to retain a social structure predicated on individual responsibility. Instead, we need only notice that social abstractions can decisively effect, or determine, our decisions. Further evidence for this view can be found in the classic *Obedience to Authority* studies, which are amongst the most well-known and frequently cited in social psychology. Following Milgram's initial – and unexpected – finding that a high proportion of people readily comply with instructions to administer dangerous levels of electric shocks to strangers, many subsequent studies were designed, aimed at measuring the willingness of people to obey authority figures who instruct them to perform acts which conflict with their conscience. Researchers were interested in the factors associated with either compliance or resistance. Of particular note are the experiments undertaken by Meeus & Raaijmakers (1995), who investigated whether people would administer harm to others if it entailed a threat of legal accountability. After being informed that an experimental task entailed harming another person, subjects were asked to sign a disclosure in which they acknowledged they were aware that other persons may be harmed

during the experiment and that, as the participant administering the harm, they accepted legal liability for any possible damages. Unsurprisingly, compliance rates plummeted far below those reported by Milgram. Thus, despite the presumably stark differences amongst participants in terms of their endocrine activity, and social and ancestral histories, the vast majority of people reliably defied authority when obedience involved a serious personal risk, presumably because participants were unwilling to compromise their own personal security and interests. Thus, even if conscious agency is illusory, it is still reasonable to assume that people are generally capable of avoiding behaviours that harm others. The key point is that a potential threat to one's own interests – and the prospect of *personal culpability* – appears to be an important input into the equation which determines a particular behaviour.

Scientific knowledge and the culture individual responsibility

There is also a fundamental paradox at the heart of Sapolsky's argument which is neither resolved nor even reckoned with throughout the book. To see this, it is interesting to reflect on the case cited by Sapolsky of an individual who did not take his medication despite being prone to seizures prior to getting behind the wheel of his vehicle. He subsequently experienced seizure activity while driving, resulting in a child and grandfather being killed. The reason for his non-compliance with his medication? The medication interfered with his enjoyment of liquor. At sentencing, the judge referred to him as an "abomination". However, according to Sapolsky, just as someone does not deserve to be blamed for having seizures, it is also "scientifically unjustifiable to make someone's life a living hell because they drove despite not having taken their meds, even if they did that because they did not want those meds interfering with their getting a buzz when drinking. But this is what we must do, if we are to live the consequences of what science is teaching us." (p.315). Thus, Sapolsky asserts that a modern scientific understanding of human behaviour is incompatible with the concept free. By extension, modern science does not support any notion of personal culpability. Yet, in basing his thesis upon the findings of modern science, Sapolsky relies upon a whole range of processes and values which are grounded within a culture personal responsibility. To see this, let us consider the process by which scientific knowledge is generated.

Humans face a perennial navigation problem and are constantly faced with the question of what to do next. Sapolsky, apparently, considers the answer to this question to be self-evident, insofar as he assumes it is axiomatic that our norms, customs and legal prohibitions should be based upon a modern scientific understanding. However, this starting point is not perfectly self-justifying. Instead, it presupposes a comprehensive value structure in which it is assumed that humans *ought* to adhere to a variety of

principles. For example, producing scientific knowledge not only requires individual scientists to undertake extensive training to acquire technical expertise. Scientific research also needs to be undertaken in a particular spirit. Scientists must be receptive to evidence, and value rational argument and logical consistency. This is true even in totalitarian states. Open, democratic societies, which foster intellectual liberty and spontaneous forms of collaboration, undoubtedly have an advantage over closed forms of society with respect to innovation and scientific inquiry. However, "good science" can still occur within totalitarian regimes. As the historian Robert Proctor documented, scientists and physicians in Nazi Germany were the first to promote a whole range of health reforms that are today regarded as progressive and socially responsible (Proctor, 1999). For example, it was Nazi scientists who first definitively linked cigarette smoking to lung cancer and who first substantiated the causal link between asbestos and mesothelioma. Nazi Physicians, meanwhile, were the first in the world to encourage women to undertake breast self-exams. Laws banning the x-raying of pregnant women were also first introduced in Nazi Germany after Nazi scientists discovered that radiation could harm the foetus, a discovery which occurred long before radiation was recognised as a hazard to unborn infants in the United States or United Kingdom. As discomforting as it is to acknowledge, a culture of "good science" can still be preserved in totalitarian states. Fundamentally, generating scientific knowledge is predicated upon a general culture of individual accountability. Such a culture presupposes that scientists *ought* to adhere to the principle of intellectually honesty, even if only within the narrow confines of a particular discipline. Scientists accept there is an overarching imperative to demonstrate technical competence while also adhering to certain codes of practise and conduct, and that there will be a reputational cost if such core values are contravened. The generation of knowledge is an all-too-familiar human affair and neither bad science nor bad actors are automatically marginalised. Instead, this requires other scientists to actively question, challenge and falsify dubious findings, and highlight procedural errors in the underlying methods and, in extreme cases, call attention to the deliberate manufacturing of falsehoods. Individual scientists undoubtedly wish to avoid professional embarrassment and censure, and are thus likelier monitor their own conduct to avoid this spectacle. As has already been suggested, Sapolsky's thesis is itself a testament to such a culture. Sapolsky cites several examples where scientists have acted immorally, and even seditiously, with respect to the scientific enterprise. We have already seen that Sapolsky highlighted the supposed malpractice of psychoanalysts who attributed seizure activity in children to the pathological parenting styles. He further speaks critically of the minority of scientists have advanced fraudulent findings which causally linked vaccinations to autism in children. Sapolsky expresses not merely his *disagreement* with these ideas, but

his *moral disapproval* that such ideas were disseminated at all, due to the subsequent vaccine scepticism and increased incidence of illness that followed. As Sapolsky states, it was insisted "in the face of every possible scientific refutation, that autism can be caused by vaccinations gone awry. Sapolsky goes on to refer to the people who propounded this view as "medieval witch-hunters" who were "responsible for decreased vaccination rates, a resurgence of measles and the deaths of children." (p.338). In expressing such disapproval and disdain, Sapolsky is again contributing to a general culture in which scientific malpractice is regarded as shameful and intellectually disreputable. In other words, those scientists who violate the rules of scientific inquiry should incur some kind of reputational cost. To argue, as Sapolsky does, that "the science" of behaviour does not support individual accountability, radically misses the point, as the very process by which scientific knowledge is generated requires a culture of individual accountability and the successful marginalising of those who violate the rules. Thus, any initial move in the direction of reason and scientific inquiry presupposes a value structure predicated upon personal accountability and Sapolsky is not entitled to have "the science" for free. Furthermore, as Sapolsky's own style of argumentation attests, marginalising bad actors within scientific fields is not synonymous with an instrumental use of quarantine. It is implicit in his view that those who egregiously violate the rules rightly incur both a significant reputational cost and professional stigmatisation as a result of their scientific malpractice. Presumably, these consequences constitute important tools in dis-incentivising scientific malpractice.

A desire for social distance: The dark side of reduced blame attributions?

It is also far from self-evident that abandoning notions of free will automatically leads to more decent forms of human organisation. To see this, it is instructive to consider how people with psychiatric diagnoses are perceived by lay persons. Though most people do not possess specialist knowledge of the neurobiology of schizophrenia – indeed, the neurobiology of schizophrenia remains somewhat opaque even to experts, as Sapolsky acknowledges – there is an increasing acceptance amongst the public that schizophrenia is an involuntary biological illness and that this diagnosis is highly relevant when judging a person's behaviour, and making assessments of culpability and mitigation (Mehta & Farina, 1997). For Sapolsky, dispensing with theories of schizophrenia promulgated by the school of psychoanalytic psychiatry, in favour of the theories of biological psychiatry, represents unmitigated progress. However, there may be further implications associated with adopting mechanistic, biological models of Schizophrenia which challenge Sapolsky's view that adopting mechanistic disease models leads to more humane outcomes. As

Angermeyer and Matschinger (2005) have shown, people who understand schizophrenia is a biological illness are also likelier to conclude that schizophrenic patients lack the general cognitive abilities that constitute moral competence. Naturally, when moral competence is judged to be absent, it becomes increasingly difficult to hold someone accountable for their behaviour. Interestingly, research suggests that those who view schizophrenia as an intrinsic biological illness are also likelier to favour more authoritarian restrictions to be placed on those with this diagnosis (Mehta & Farina, 1997). This is because those who are deprived of the faculties which constitute moral competence are viewed as having limited long-term association value. They are perceived to lack the general abilities to abide by societal customs, and inhibit the urge to engage in harmful behaviours (Angermeyer & Matschinger, 2005). It is certainly plausible that people are likelier to favour an instrumental use of quarantine for violent schizophrenic patients, given that their behaviour is likelier to be attributed to involuntary, mechanistic processes. However, whether people favour instrumental use of quarantine or current forms of incarceration in the correctional system is only half the question. Sapolsky does not consider how a desire for social distance from those judged to be incapable of exercising moral autonomy impacts upon preferences about the desired *duration* of this quarantine. Might we possibly desire a greater social distance from those we judge to incapable of moral decision-making and those we deem to be of low long-term association value? It is plausible that a desire for increased social distance from schizophrenics may result in people wishing for them to be detained for longer periods, even if this is within a rehabilitative context. There are thus reasons to be cautious about viewing human conduct through a mechanistic and involuntary lens. If we conclude that someone is fundamentally unable to exercise moral autonomy and agency, and that their behaviour is instead attributable to intrinsic, involuntary mechanistic processes, this may have significant implications for their perceived association value and the extent to which others desire social distance from them.

The Twentieth Century: cautionary reflections

There are additional reasons why we should be hesitant to dispense with notions of personal guilt, stigma and shame. To see this, it is worth reflecting on the peculiar and distinctive aspect of the human condition. Humans possess both *cognitive* empathy – the ability to recognize another person's mental state – and *affective* empathy – the drive to appropriately respond to another person's emotional state. The capacity for cognitive empathy, however, appears to be somewhat double-edged. Irvin Yalom reasoned that an awareness of our future death "is the primordial source of anxiety and, as such, is the primary font of psychopathology" (Yalom, 1980, p. 29). Becker (1973) likewise

proposed that human beings are distinct among species in the knowledge of their own mortality and that an awareness of one's finitude produces a constellation of fears and anxieties that are peculiar to humans. It certainly seems plausible that a deep insight into one's mortality, and an ability to articulate a foreknowledge of both our death and our progressive physical decline, is a central component of human psychopathology. The key point, however, is that humans possess this foreknowledge and this has profound implications. An intrinsic awareness of one's own vulnerability, as well as a capacity to understand the emotional states of others, entails a profound insight into the vulnerabilities of *other* people. Humans have the unique capacity to *capitalise* on their awareness of other people's vulnerability to inflict a maximum degree of harm upon others. Humans are thus able to make an *art-form* out of inflicting pain on other people that far exceeds the capacity of non-human animals which, at their worst, are predatory in a brutal manner. However, the human capacity for maximising the physical and psychological distress in other people far exceeds the predation necessary for survival. A cursory reading of twentieth century history bears this out. The ideological frenzy which characterised swathes of the twentieth century reached its apogee in the crematoria of Auschwitz, Treblinka and *Chełmno. In The Hell called Treblinka* (1994), Vassily Grossman provides one of the most vivid accounts of the appalling capacity of humans to creatively maximise suffering in others. Grossman describes how Nazi extermination camp guards, morally deranged by their adherence to a genocidal ideology, felt morally and intellectually justified in creatively maximising the psychological horror and suffering of camp prisoners. Prisoners – who were *already condemned to certain death* – were made to partake in absurd and meaningless games for the amusement of concentration guards, who, Grossman notes, were also prone to offering arrogant philosophical speeches to the prisoners in an apparent attempt to justify what Hannah Arendt described elsewhere as the *radical evil* they were perpetrating. There is little reason to think the majority of the guards described by Grossman had anything other than in-tact pre-frontal cortices. Indeed, as Grossman describes, many of guards lived upstanding lives away from the camps. They prioritised their fitness, had developed sophisticated aesthetic tastes, and cultivated general dispositions and habits of mind that reasonable people would otherwise want to emulate, and which would be associated with success in a well-functioning society. They were, it would seem, clearly in possession of the general faculties which constitute moral competence and they *could have behaved otherwise*. It is worth noting that there is not one recorded example in Nazi Germany of a soldier facing formal criminal sanctions for refraining to participate in the Nazi genocides. This fact constitutes the basis of Christopher Browning's challenge to Primo Levi's notion of a *Grey Zone*. Levi, an Italian chemist and holocaust survivor, argued that the network of human relationships inside the

concentration camps was not simple and "could not be reduced to the two blocs of victims and persecutors". Such binary thinking was, in Levi's view, "inadequate in the face of the complexity of life in the camps." (Levi, 2003, p.29). Levi argued that both concentration camp guards and those prisoner functionaries who were forced, on pain of death, to collaborate with Nazis in their industrialised slaughter, occupied a *grey zone* of human conduct, in which the dividing line between good and evil was often obscure. However, as Browning reasons, the perpetrators did not become fellow victims in the way that victims sometimes became accomplices of perpetrators. Neither did those who became accomplices occupy the same moral space as the perpetrators. Why? Because the *range of choices* available to guards and those collaborating was totally different (Browning, 2005). And therein lies the fundamental basis of our ability to make fundamental moral judgements; the intrinsic capacity of humans to pursue *multiple different courses of conduct*, irrespective of whether this is free or mechanistic. The inescapable conclusion of Browning's work is that, as barbaric the holocaust atrocities were, the atrocities were routinely committed by ordinary people. Fundamentally, all humans partake of the same essence and, so it seems, most humans generally appear capable of such behaviour. Examples of other atrocities committed during the twentieth century include the Nanking Massacre, the Armenian Genocide and the Holodomor, demonstrating that capacity to participate in atrocities in a deeply human phenomenon, not localised to any one historical period or culture. Sapolsky – quite remarkably – appears to regard shame and notions of personal culpability and guilt as *entirely* atavistic forces, based on an outdated and intuitive understanding of human behaviour. It is undoubtedly true that shame can be misattributed, and maliciously attributed, purely for the purposes of maximising the psychological distress and social isolation of other people. However, guilt and shame, and the threat of social exclusion, heavily dis-incentivise harmful behaviours, alongside formal criminal codes. Given our capacity as humans for extraordinary cruelty, the stakes are startling high for our species. Should we therefore not avail ourselves of all the available tools in order to disincentive the worst forms of human behaviour and, indeed, to encourage the best, irrespective of whether our behaviour is free or determined?

References

Angermeyer MC, Matschinger H. Causal beliefs and attitudes to people with schizophrenia. Br J Psychiatry. 2005;186(4):331–4.

Baumeister RF, Tice DM. Anxiety and social exclusion. J Soc Clin Psychol. 1990;9(2):165–95.

Baumeister RF, Leary MR. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychol Bull. 1995;117(3):497–529.

Becker E. The denial of death. 1973.

Beevor A, Vinogradova L. A Writer at War: Vasily Grossman with the Red Army, 1941–1945. J Mil Ethics. 2007;6(1):81–2.

Browning CR. Ordinary men: Reserve Police Battalion 101 and the final solution in Poland. New York: HarperPerennial; 1998.

Levi P. The drowned and the saved. In: Edinburgh University Press eBooks. 2003. p. 29–35.

Meeus WHJ, Raaijmakers QAW. Obedience in modern society: The Utrecht studies. J Soc Issues. 1995;51(3):155–75.

Mehta S, Farina A. Is being "Sick" really better? Effect of the disease view of mental disorder on stigma. J Soc Clin Psychol. 1997;16(4):405–19.

Orne MT, Evans FJ. Social control in the psychological experiment: Antisocial behavior and hypnosis. J Pers Soc Psychol. 1965;1(3):189–200.

Popper KR. The open society and its enemies. London: Routledge Classics; 2011.

Proctor RN. The Nazi war on cancer. 1999

Rowland LW. Will hypnotized persons try to harm themselves or others? J Abnorm Soc Psychol. 1939;34(1):114–7.

Sapolsky RM. Determined: Life without free will. London: Bodley Head; 2023.

West SG, Gunn SP, Chernicky P. Ubiquitous Watergate: An attributional analysis. J Pers Soc Psychol. 1975;32(1):55–65.

Yalom ID. Existential psychotherapy. 1st ed. New York: Basic Books; 1980.

Young PC. Antisocial uses of hypnosis. In: LeCron LM, editor. Experimental hypnosis. New York: Macmillan; 1952. p. 376–409.