Agency, and its relationship to responsibility, are issues over which philosophers have yet to find entire agreement. The above passage by Timothy Bayne was written some time ago and preceded his own valuable clarification. Yet the subjective experience (or phenomenology) of the personal action that philosophers call agency (the term is explained more fully below) remains, as Bayne says, unmapped, as do related value-based (or normative) questions about responsibility. Nowhere is this more evident than at the juncture of agency, responsibility and mental disorder. And it is in the spirit of sketching some corners of Bayne’s map that I consider what can be learned from atypical states and behaviour. A reflection on images – here, Goya’s great Madhouse scene, and photographs of Mad Pride events since the 1980s – is shown to provide assistance towards this end.

Keywords Philosophy; mental health; agency; Mad Pride; Goya.
examines the philosophical concepts grounding mental health activism. The first analysis of its kind, his is a respectful but stringent account of the recognition sought by activists, locating their demands within contemporary political philosophy, and providing guidance as to how those demands might be met. In the present essay, I suggest that some of the ambiguity and complexity entailed in the categories of madness, agency and responsibility can be approached and perhaps illuminated through a reflection on images. I use a familiar painting by Goya, on the one hand, and on the other, photographs of Mad Pride marches and public events since the 1980s. The message from what follows is simple. Even applied to normal states and behaviour, conceptions of agency and responsibility may be incompletely understood. We must remain open to revising and expanding them in light of what can be learned from atypical states and behaviour. And for this purpose, images seem able to offer powerful assistance.

Both agency, and responsibility in relation to it, are contested topics. But few would dispute that personal responsibility can be appropriately assigned only when there is some degree, or form, of (personal) agency. The converse may not hold; arguably, there can be agency in the absence of full responsibility. Philosophical accounts of agency and responsibility differ, but as is indicated by legal exculpating factors such as ignorance, compulsion and duress, ascriptions of responsibility will sometimes be withheld, even from what appear to be voluntary actions. In keeping with much discussion about agency and responsibility, the two terms are sometimes merged; in other places, as will become apparent, they need to be separated.

Some basic distinctions

A group of contrasts drawn from philosophical discussions of agency are introduced here: actions and movements; the proprietary phenomenal character of agency (how it feels subjectively) in contrast to interpretations about, or judgements over it; self- and other-assigned agency; mental and overt agency; agency emanating from more and less proximal and distal intentions; and more and less deliberative and absorbed agency.

Actions and movements

The ‘sense’ of agency, as we seem to experience it, is sometimes used by philosophers to mark the contrast between true actions (those we voluntarily undertake ourselves) and mere movements. This is the difference between lying down and unintentionally falling, for example.

Phenomenal character of agency

When agency is involved (in lying down), two distinct parts have been recognised: the phenomenological feeling of doing it oneself (known as its proprietary phenomenal character) and the interpretation or judgement we make about that feeling (the awareness that, rather than either falling, or having been pushed to the ground, one lies down). Both phenomenal feelings of agency, and the interpretation or judgement we – and others – make about our agency, enter into the following discussion.

Self- and other-assigned agency

Complicating questions about self-assigned agency and responsibility arise when we speak of agency and responsibility, whether as felt agency or as judgements about one’s agency, and these are ostensibly magnified when mental health is part of the mix. Self-assigned agency and responsibility, whether as feelings or judgements, will not be entirely reliable guides if they are distorted by disorder. Examples of that distortion are the self-blaming depressive, who exaggerates their personal responsibility (‘It’s all my fault’), or the grandiosity of the narcissistic personality’s boast (‘All thanks to me’) illustrate. Similarly, the proprietary phenomenal character of agency appears to be absent in some psychotic experience, as suggested by what are known as ‘inserted thoughts’.

When agency (and responsibility) is judged to be present, by others or by oneself, may thus depend on the opinion of psychiatric experts. And asked whether those with psychiatric diagnoses possess agency, the expert’s answer will usually be, it depends – on the situation, the nature of their disorder, the particularities of the thoughts and actions in question, and so on – and thus on case-by-case clinical observation and judgement.

Mental and overt agency

Discrepancies between felt agency (and responsibility) and agency (and responsibility) with normative force imply that some judgements about agency (and responsibility) may be inaccurate, or unreliable. (Philosophers speak of ascriptions of agency and responsibility as possessing normative force when they are accurate (as the moral realist might put it) or appropriate). Yet such inaccuracy or unreliability will arise from multiple factors. Mistaken or distorted self-assigned agency and responsibility are also associated with varying cognitive styles and socialised habits in normal populations; for example, the widespread tendency to deflect blame onto others, or take sole credit for achievements more accurately attributed to many. They will then be culturally local; other times, places and cultures might rely on different substantive norms. Not only disorder, and non-disordered personality bias, but also any number of social and cultural expectations can thus shape judgements of agency and responsibility, self- and other-assigned. And given that disorder, personality and culture are inextricably entwined in any given instance, these distorting factors may not be separated easily or, in many instances, helpfully.

With regard to judgements about agency, whether made by oneself or another, it will be worth noting that there are mental as well as more outward bodily, actions. When Thomas Jefferson spoke of the ‘illimitable freedom of the human mind’, he referred to our immediate ability to recall the past, envision the future, rehearse what might have been as well as what was – indeed our ability to direct our thoughts at all, to form intentions, revise and renounce earlier attitudes, to call up memories and imaginings. This kind of agency, associated with the notion of rational capacity, has sometimes been denied those with mental disorder, even
while it was said to be possessed by prisoners in chains, and those trapped in ill and unresponsive bodies.

Ascriptions of responsibility also sometimes extend to these inner efforts of mental agency. For some mental actions, we hold ourselves responsible and are also held responsible by others (‘This is an uncharitable thought, but . . .’ it might be said, or ‘You mustn’t think like that . . .’). That said, whether such responsibility is ever rightly ascribed to one’s own mental ‘actions’ is a contested matter, as is the applicability of the term ‘agency’ to them. (I am grateful to David Foreman for pointing out that ethical systems differ over this point. Only some Christian theology acknowledges the sinfulness of ‘immoral thoughts,’ for example.) And the exactness of the parallels between mental agency and the overt agency we exhibit to the world are similarly debated. The structure of mental action has been seen by some to differ in not involving intentions, or initiated by decisions, or reasons, for example, whereas others insist that mental agency is the originating source of all agency, outer as much as mental.34

The details of these additional differences need not detain us here, as long as it is acknowledged that some of what we humans do involves our inner as well as our outer lives, and that the coherence of mental agency must be understood within the whole, including personal-level intentions and more overarching goals, that may not manifest in any way that is outwardly evident. Examples of such expansive goals over mental health might include the aim to maintain psychic stability and avoid or reduce personal suffering.

Proximal and distal agency

Another feature of agency, occurring with both mental and overt agency, are the less and more direct and immediate forms it takes, for each of which we are also, sometimes, held responsible, both by others and perhaps ourselves. Examples of less direct and immediate agency include Elster’s ‘imperfect rationality,’ when we outwit ourselves, anticipating and engineering outcomes we are unable to produce directly and immediately.6 Often, we do so aware (or not unaware) of the process as we manipulate our beliefs and feelings to achieve desired ends. The way cognitive-behavioural therapy is widely supposed to work employs this approach, both for mental agency, where the power a person has to change beliefs and attitudes is used to bring about changed feelings, as well as directing what is said and done. The effectiveness of such imperfect rationality cannot be guaranteed, of course. Cognitive–behavioural therapy has now quite self-consciously adopted its classical ancestor in the imperfect rationality practises of those who recommended care of the soul that aimed to expunge all negative and unproductive feelings (debatably, all feelings of any kind), through a programme of self-analysis and cognitive therapy. We are free to, can and should adjust the beliefs on which our feelings rest, it is supposed. (Even in their own time, and almost ever since, owing to the influence of Aristotelianism, the excessive responsibility for controlling thoughts and feelings accepted by the Stoics were derided as unrealistic and undesirable.) Outcomes of agency that are more and less immediate and direct have been described as emanating from intentions that are ‘proximal’ and ‘distal’, respectively, with imperfect rationality exhibiting the latter.67

Deliberative and absorbed agency

As examples of more planful and less immediate (or proximal) agency illustrate, particular instances of agency, whether mental or overt, fall within complexes of goals, deliberations, purposes and plans – even very long-term aims and dreams. This nesting of agency within its broader setting has also been recognised to involve a consequence in terms of phenomenology: although recognisable feelings may be associated with agency, not all agency is immediately felt. Agency often occurs as part of a seemingly effortless, automatic and non-conscious flow of engagement with the world. In that case, it leaves little space for, and bears little resemblance to, conscious deliberation or, the felt sense of actively doing, rather than passively experiencing. Typically, indeed, felt agency is ‘recessive’, in being confined to the margins of attention. Much of everyday life, and more than philosophers always acknowledge, involves this kind of agency, as the case of habitual action illustrates. With skills like driving, we usually proceed with little or no conscious awareness of our complex responses. When we engage in practical or creative endeavours with full, undistracted absorption, we have been described as undertaking ‘skilled coping’,72 and awareness in the latter experience is said to be immersed rather than detached.9

Images: ‘dark freedom’ in Goya’s ‘The Madhouse’ and Mad Pride parades

The above contrasts have been drawn from within disparate theories, and only fuller accounts of each can determine their compatibility, and the final plausibility of the theses put forward here. Nonetheless, felt agency and ‘freedom’; self- and other-assigned agency (and responsibility); the effects on agency and responsibility wrought by personality style and cultural norms as well as by disorder; mental agency and agency in the outside, shared world; and more and less direct, and more and less ‘online’ or conscious, and deliberative, forms of agency, all seem likely part of a rough sketch of the territory, although each element still wants for a fuller analysis.

In what follows, I try to elucidate some of the features of mad agency (and responsibility) with the help of images: a much-interpreted painting, showing the inside of a madhouse at the beginning of the 19th century; and photographs of Mad Pride parades. Why turn to images? Arguably, imagery can communicate qualities of agency that are hard to capture in language alone. And pictorial commentary, as we will see, offers suggestive associations worth exploring. Moreover, in addition to written descriptions, and long before the present-day use of images, pictorial traditions shaped how mental health and ill health, disorder and madness were apprehended and understood.10 This ancient, long-lived, visual record may have something more to teach us.

‘The Madhouse’, Goya’s magnificent depiction of the madhouse in his hometown of Zaragoza, was painted between 1812 and 1813. Unlike the more positive images from the Reform era that followed, Goya’s several pictures of that institution have been judged to depict unutterable suffering and awfulness, even depravity. In a typical description, the painting’s only light source is said to be:
With its striking use of darkness and light, the picture shows ‘a world of darkness opposed to the world of light’, in keeping with Goya’s ‘emphasis on the soul as the dark hiding place of terror’, for example.10 

Echoing this negative interpretation, the actions of the inmates have been described by Robert Hughes as ‘delusory gestures of power’.11 However, free, responsible and agentic these transgressive gestures appear to be, this ‘power’ (to use Hughes’ word) will not guarantee that full, or perhaps any, responsibility could be appropriately ascribed, by the inmates themselves or by observers. There are bars on the window. The inmates’ agency in the outer world is limited. Yet delusory or not, each seems to reframe his birthright as a world of darkness opposed to the world of light. (Hughes attributes to Goya a well-known medieval saying, attributed to Terence, to this effect.) And few subsequent depictions of madness behind bars are as sympathetic, at least until we reach 20th century pictures such as Bellows’ ‘Dance in a Madhouse’ (1917) and other images showing asylum entertainments.

Arguably, then, Goya’s attitude toward his subject matter was expressive of sympathy, not revulsion. Conforming to that positive conception is a revealing passage from another 20th century commentator. Presenting mad agency as a certain, transgressive ‘dark liberty’, Michel Foucault observes of the madman in the hat in Goya’s painting.

[the figure] leaps out, by virtue of the silent language of his well-muscled form and the wild, marvelous freedom of his youth, a free human presence who affirms his birthright as though this were the beginning of new era. “The Madhouse” speaks…[of] those new bodies, brought into the light in all their vigour, and whose gestures, if they call up their dreams, sing above all of their dark liberty. (Foucault: pp. 530–1).12

What dark liberty is this? Where does the mad agency lauded by interpretations like Foucault’s fit among the forms of agency sketched earlier? I propose that the ‘freedom’ expressed in the painting anticipates something that we only entirely recognise now as a consequence of political events: it is a demonstration of the powerfully ‘freeing’ and healing transgressiveness and ludic exhalation celebrated in today’s Mad Pride activism, rhetoric and writing.

We are tempted to agree with Hughes that the gesture of power is merely delusory – a mistaken parody of real agency – or even a feeling of agency distorted by disorder-wrought cognitive error. Yet, in light of writing by mad activists, we also perhaps begin to see a quality inviting attitudes of hope and delight, rather than abject despair, in the gestures of Goya’s inmates. To cite just one example of such writing:

‘Madness is the new rock “n” roll!…All of us who’ve experienced “deep sea fishing” will know the sensation of heightened awareness, of consciousness enhanced…of feelings of wonder and terror that can’t be verbalized…’ (Morris: pp. 207).13

Generally, what we see in images is limited by what we expect to see, and what we suppose the artist wanted us to see. We might guess that the inmates in the asylum were performing for an audience, perhaps even engaging in such display for monetary reward, as did the inmates at English asylums during the same era. (I am grateful to David Foreman for drawing my attention to this possibility.) Were this so, the actions depicted by Goya are replete with the features of classic rational agency, and a pretence. But whether or not it is a pretence, later events and subsequent understanding allow us to go beyond the constraints imposed by what we would expect, and what, intentionally or not, Goya may have conveyed. From today’s perspective and understanding, we can see it differently. To illustrate this interpretive shift, we might turn to Bellows’ ‘Dance in a Madhouse’, where movements have been dismissed as ‘wild and uncontrollable’, and reflecting passive (‘melancholic’) postures. Yet here, too, and contrary to Gilman, the central female figure in Bellows’ picture also reveals something joyous, triumphant, freeing, enlivening and, perhaps, empowering.

Goya’s madhouse has parallels and echoes in today’s Mad Pride activism, some of them quite evidently self-conscious. Here we find not only resistance to the mental health system, and solidarity, but repeated emphasis on ‘the celebration of difference’. ‘Celebration’ is a useful encapsulation, suggesting exhilaration and joy. Its object is more than mere difference, however. What is being celebrated? Minimally, the ‘creativity of mad people, pride in a unique way of looking at life, the validity of such a distinct way of life…’ (Sen: pp. 5) are grounds for pride and reason for celebration. Enumerating elements of the demand for recognition of mad identity, Rashed emphasises that Mad Pride discourse includes a range of attitudes and interpretations. It emphasises mad identity, creativity, spirituality, suffering and the gifts that, although valuable, are dangerous. Each of those aspects (identity, creativity, spirituality, suffering and dangerous gifts) offers grounds to applaud and celebrate.

Images of broadly ‘celebratory’ Mad Pride parades (in Canada, the USA, the UK, Ireland, Belgium, France and Australia, for example), provide us with readily recognisable descendants of our madhouse scene. Here are flamboyant and excessive dress, gesture and performance; here are the seeming freedoms and agency associated with the transgressive, the parodic and ludic. Unlike the barred madhouse, the street now forms the stage for performance. The demands for political recognition, I suggest, provide examples of agentic forms identified earlier, and perhaps anticipated by Goya.

By recognising that agency works in many ways, indirect as well as direct, and through imperfect as well as perfect rationality, we may regard the mad gestures in Goya’s painting as not only exhilarating, and felt, although ultimately delusional agency and freedom, but also as healing.
consoling, emboldening, strengthening and even, yes, freeing. Like the antics of the marchers celebrating Mad Pride, these gestures may convey the shoring up of what is depleted, diminished and misunderstood by the surrounding culture, or concealed through prejudicial expectations.

Viewed as a kind of self-help or self-care, the gestures depicted in both Goya’s and the later celebratory, Mad Pride images, can also be likened to the indirect agency of imperfect rationality. Arts-based healing rituals, healing effects and the mental health benefits of expressive therapies have been acknowledged since ancient times, and are well documented.\(^\text{17-19}\) Their proponents speak of the way such activities integrate mind and body in a unified, healing whole; the ‘act of responding to what is given, imagining its possibilities and reshaping it in accordance with what is emerging’, is described as akin to that found in all imaginative play, and improvisation (Levine: pp. 71).\(^\text{16}\)

The agent may not seek anything further through their exuberant gesture, or even be alert to broader intentional or purposive context within which it occurs. Their agency may reflect the obliviousness of intense absorption. Moreover, it may not be agency with outer effects. In contrast to the powerless inhabitants of the madhouse, today’s activists bring about real world, consequential political outcomes (changed cultural attitudes, for example, and more enlightened mental health policies). Yet even without doing so, the gestures of Goya’s figures may be seen to intimate certain forms of distal, immersed agency, as Foucault tried to suggest that none quite convey the healing and freeing effects of expressive therapies.

Reviewing the discourse and tenets of Mad Pride activism, Rashed notes four main elements, as we saw: demands around disability, and the gestures of Goya’s figures may be seen to intimate certain forms of distal, immersed agency, as Foucault’s analysis may be taken to suggest.

Among the demands identified by Rashed and these activists, I have tried to suggest that none quite convey the healing and freeing power captured in the ludic, the parodic and the transgressive that are so evident in today’s Mad Pride parades and celebrations and, in light of them, we are now able to recognise as foreseen in the dark liberty of Goya’s madhouse.

About the dangerous gifts of madness, it has been explained that ‘...we are members of a group that has been misunderstood and persecuted throughout history but has also been responsible for some of its most brilliant creations’ (Du Bru: pp. 259).\(^\text{20}\) Any incautious ability to violate convention and social norms will be dangerous, including such gifts, it must be conceded, and will likely bring its own exhilarating phenomenology, as well as fateful and often self-defeating consequences. The demands of mad activism can be met, perhaps, only with the openness, imagination and negotiated agreement of the larger culture.

Such negotiated agreement requires a preparedness to revise and expand accepted ideas about social norms and concepts of mental health, as I have pointed out elsewhere.\(^\text{21}\)

Among those social norms and concepts of mental health, it has been proposed here, are ideas about agency.

**About the author**

Jennifer Radden is a Professor Emerita at the Philosophy Department of University of Massachusetts Boston, Massachusetts, USA.

**Supplementary material**

To view supplementary material for this article, please visit https://doi.org/10.1192/bjb.2020.119.

**Acknowledgements**

I am grateful for audience questions at the Royal College of Psychiatrists’ Philosophy of Psychiatry Special Interest Group Biennial Conference in September 2019, ‘Madness and Society: Pathways to Reconciliation’. My greatest thanks go to Dr David Foreman for his careful reading and insightful suggestions on an earlier draft of this paper.

**References**

1 Bayne T. The phenomenology of agency. Philos Compass 2008; 3: 182–202.
2 Rashed M. Madness and the Demand for Recognition: A Philosophical Inquiry into Identity and Mental Health Activism. Oxford University Press, 2019.
3 Hieronymi P. Two kinds of agency. In Mental Actions (eds L O’Brien, M Soteriou): 138-62. Oxford University Press, 2009.
4 Schlosser M. Agency. In The Stanford Encyclopedia of Philosophy (Winter 2019 edn) (ed. EN Zalta). Stanford University, 2015 (https://plato.stanford.edu/archives/win2019/entries/agency/).
5 Elster J. Ulysses and the Sirens: A Theory of Imperfect Rationality. Cambridge University Press, 1977.
6 Mele A. Recent work on intentional action. Am Philos Q 1992; 29: 199–217.
7 Searle J. Intentionality: An Essay in the Philosophy of Mind. Cambridge University Press, 1983.
8 Dreyfus H. Overcoming the myth of the mental: how philosophers can profit from the phenomenology of everyday expertise. Proc Addresses Am Philos Assoc 2005; 79: 47–65.
9 Pacherie E. The phenomenology of action. Cognition 2008; 107: 179–217.
10 Gilman S. Seeing the Insane. University of Nebraska Press, 1996.
11 Hughes R. Goya. Knopf, 2003.
12 Foucault M. History of Madness. Routledge, 1972.
13 Morris S. Heaven is a mad place on earth. In Mad Pride: A Celebration of Mad Culture (eds T Curtis, R Dellar): 207–8. Spare Change Books, 2000.
14 Porter R. Bodies Politic: Disease, Death and Doctors in Britain, 1650–1900. Cornell University Press, 2001.
15 Sen D. What is mad culture? Asylum: Mag Democratic Psychiatry 2011; 18(1): 5.
16 Levine S. Poiesis: The Language of Psychology and the Speech of the Soul. Jessica Kingsley, 1997.
17 Winnicott D. Playing and Reality. Tavistock, 1971.
18 Jones P. The Arts Therapies: A Revolution in Healthcare. Taylor and Francis, 2004.
19 Malchiodi CA (ed.). Expressive Therapies. The Guilford Press, 2003.
20 Du Brul SA. The icarus project: a counter narrative for psychic diversity. J Med Humanit 2014; 3: 257–71.
21 Radden JH. Recognition rights, mental health consumers and reconstructive cultural semantics. Philos Ethics Humanit Med 2012; 7: 6.