REGISTERED REPORT PROTOCOL

Morally excused but socially excluded: Denying agency through the defense of mental impairment
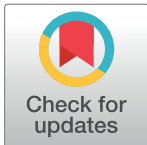
Melissa de Vel-Palumbo[1]*, **Chelsea Schein**[2¤a], **Rose Ferguson**[3¤b], **Melissa Xue-Ling Chang**[4], **Brock Bastian**[5]

1 College of Business, Government and Law, Flinders University, Adelaide, Australia, **2** Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Victoria, Australia, **4** School of Psychology, University of Queensland, Brisbane, Queensland, Australia, **5** School of Psychological Sciences, University of Melbourne, Melbourne, Victoria, Australia

¤a Current address: Wharton School of Business, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America
¤b Current address: Science, Psychology and Sport, Federation University, Melbourne, Victoria, Australia
* melissa.devel-palumbo@flinders.edu.au

Abstract

Defendants can deny they have agency, and thus responsibility, for a crime by using a defense of mental impairment. We argue that although this strategy may help defendants evade blame, it may carry longer-term social costs, as lay people's perceptions of a person's agency might determine some of the moral rights they grant them. In this registered report protocol, we seek to expand upon preliminary findings from two pilot studies to examine how and why those using the defense of mental impairment are seen as less deserving of certain rights. The proposed study uses a hypothetical vignette design, varying the type of mental impairment, type of crime, and type of sentence. Our design for the registered study improves on various aspects of our pilot studies and aims to rigorously test the reliability and credibility of our model. The findings have implications for defendants claiming reduced agency through legal defenses, as well as for the broader study of moral rights and mind perception.

Introduction

A defense of mental impairment—available in most Western democratic legal systems [1]—recognizes the principle that people lacking full mental capacity should be excused from criminal responsibility. But using this defense may come with a hidden cost. In denying their *agency* (the capacity to engage in intentional action [2]) mentally impaired defendants may initially absolve themselves of blame, but they may also lose certain rights and freedoms conferred to moral agents. The current research examines the potential negative consequences of lost agency for those using the defense of mental impairment, in terms of others' moral regard of such defendants.

Mental impairment excuses criminal responsibility through denial of agency

Defenses of mental impairment (defined differently across jurisdictions) typically involve the denial of one's ability to reason and/or exercise full control over one's actions at the time of the alleged crime, and thus constitute a denial of *mens rea* (a "guilty mind")—a necessary requirement for establishing guilt [3, 4]. While major mental illnesses represent the classic case for a defense of mental impairment, brain injury and physical diseases affecting brain function may also constitute impairment. Across different jurisdictions, criteria for an mental impairment defenses typically involving either a cognitive (not knowing the nature of one's act) or volitional (not being able to control their behavior) element [3, 5, 6]. In the absence of these capacities, the law deems criminal responsibility and punishment to be inappropriate and unjustified.

In addition to a full defense, mental impairment is factored into judicial decision-making through a range of legal provisions that can reduce blame for a crime [7–9]. A partial defense (i.e., diminished capacity) due to mental impairment is available in many jurisdictions. Alternatively, a different plea, often termed "guilty but mentally ill" may impose a less stringent test of mental impairment, acknowledging that a defendant's mental illness substantially contributed to his/her criminal conduct to some degree (for which he/she is nonetheless accountable). Last, mental impairment may be considered a mitigating factor in sentencing. All of these provisions seek to account for the reduced responsibility of mentally ill offenders even when they fall short of the criteria for a full defense of mental impairment; a full defense is the most complete abdication of agency and thus a useful starting point for exploration.

Consistent with the legal rationale for the defense of mental impairment, empirical research suggests that people attribute less blame and punishment to defendants with a mental illness [10–15] (but see [16, 17]). For example, Monahan and Hood [18] presented participants with case descriptions of a violent crime in which the offender either had or did not have a history of psychological disorder. Offenders with a history of psychological disorder were attributed with less free will and were considered less blameworthy and deserving of punishment. Similarly, Kleinke and Baldwin [19] demonstrated that persons supplying "crazy" explanations for criminal acts were ascribed less responsibility and intent, and were given shorter prison sentences than those providing "sane" explanations.

If mental illness reduces perceived blameworthiness, it is likely through reduced perceptions of *moral agency*—a prerequisite for moral responsibility. Agency requires first and foremost the ability to regulate one's actions (e.g., through self-control, planning, and intention; this constitutes a volitional element [20]). In addition, agency also includes an understanding of what one is doing (e.g., perception, knowledge of norms, reason and thought; constituting a cognitive element typically referred to as moral knowledge [21]). Both correlational and experimental studies have shown that perceptions of agency (e.g., controllability of behavior) are related to perceptions of moral responsibility, deservingness of blame, and punishment for wrongdoing [20, 22–28].

Accordingly, insofar as mental impairment leads to perceived lack of agency, those invoking the defense should be perceived as less blameworthy for their criminal actions. Surveys reveal that mentally ill defendants are judged by community members as lacking free will [18] —mapping onto the volition element of agency. In addition, people believe that mentally ill offenders feel less guilt about their crimes than healthy offenders [29]. This implies that mentally ill offenders are perceived as less able to feel remorse for their crime, potentially because they lack moral knowledge to understand the wrongness of actions.

It is important to note that the link between mental impairment and responsibility need not translate into reduced length of sentencing. A study of over 8,000 criminal defendants who entered a defense of mental impairment from 1976 and 1985 indicated that the defense might actually be associated with a greater likelihood of detainment [30], potentially because people are concerned that these defendants are just finding "loopholes" [31–33]. Alternatively, other people report that mentally ill or impaired defendants should receive a longer term of detainment in order to receive appropriate healthcare to aid recovery [34]. In other words, mentally impaired defendants may be seen as less blameworthy and less deserving of punitive responses, but more in need of help, including potentially detainment.

Situating legal cases within the broader moral psychology literature is helpful not only for understanding the broader mental processes underlying criminal judgments about those using specific legal defenses, but also for making predictions beyond the context of criminal responsibility—here, as will be shortly seen, to lay judgements about offenders' ongoing societal rights.

Implications of reduced agency for rights

Not only does mental impairment challenge lay perceptions of responsibility, it may also have downstream consequences for another moral domain: *rights*. Psychological research has shown that perceived mental capacities play a role in lay judgments of rights. The dominant perspective is that perceived agency should have little impact on rights, as moral rights depend on perceptions of the capacity to experience pain (*patiency)*, which is inversely related to agency [20, 35–37]. Under this account, those considered *moral patients* (i.e., entities vulnerable to pain and harm) such as those with a mental impairment, might be seen to deserve *more* rights; in particular, rights pertaining to protection from harm (we expand on this point further below). On the other hand, some recent research suggests that perceptions of perceived agency also contribute to judgments of moral status [38]. Here, we advance this debate by examining a more expansive notion of moral rights. We propose that a denial of agency might help mitigate blame, but it may also undermine provision of certain rights.

The aforementioned debate in mind perception and morality has explored a very narrow set of rights—the right to not be harmed [20], or the right to be helped [39]. Following previous nomenclature [40], we refer to those types of rights as *moral standing*. Yet rights are a broad category, including the right to access public goods (e.g., social support services), as well as civil and political rights, such as freedom of speech, the right to privacy, bodily rights, and property rights [41]. Criminal offenders are often subject to diminution of civil rights, such as revocation of one's right to vote, or prohibitions against hiring people with criminal convictions, even though doing such is not necessarily ordained or codified in sentencing policies [42]. Revocation of these types of rights from criminal offenders has thus been referred to as "invisible punishments" that work as tools of social exclusion [43] by diminishing the rights and privileges normally granted by citizenship and/or residency. Therefore, these types of rights are particularly relevant to lay judgments in the criminal justice context.

When we shift how we conceptualize moral rights beyond the right not to be harmed, the role of moral agency in determining rights becomes more nuanced. All people, from newborn children to the elderly, are typically viewed as deserving certain moral standing [39], such as the right to be treated with care and compassion, or the right not to be harmed. The right to social support services might also be open to anyone, or arguably even more deserving for people who are low in agency and cannot care for themselves. However, only adults capable of higher-order cognition, such as rationality, are given the full legal and political rights to participate in society, including liberties such as the right to vote, to get married, and own property

[44]. Such rights are explicitly justified on the basis of agency [45]. The denial of agency in the defense of mental impairment may therefore directly undermine these rights.

Perceived dangerousness and rights

We predict that a specific aspect of agency, *inhibitive agency*, may play a particularly important role in affordance of rights for defendants with mental impairment. Mentally ill offenders are often perceived as lacking the capacity to inhibit undesirable behaviors (i.e., lacking in inhibitive agency), constituting a threat to community safety [46–48]. Recent research has highlighted that perceived dangerousness is associated with moral standing: people offer less moral protection to entities that is potentially harmful to society [40]. Along these lines, people who are viewed as dangerous due to their lack of agency may thus be attributed less moral standing.

Moreover, treatment in psychiatric facilities (conjuring images of straightjackets and forced medication) may not be seen as rehabilitative by lay audiences, but rather, further stigmatizing [49], potentially due to reinforcement of the belief that a mentally impaired person is unable to control him/herself. The accuracy of these lay perceptions is questionable, as research indicates that perceived recidivism rates for mentally disordered criminals are often overestimated [50]. In fact, defendants who use a mental impairment defense or suffer from psychosis or mood disorders are no more likely—slightly less likely, in some cases—to reoffend than those without these characteristics [51]. Regardless of the accuracy of these judgments, however, people perceived as lacking agency may be stripped of their rights when these judgments are based on perceptions of dangerousness.

Furthermore, lay perceptions of the dangerousness of such defendants are likely to be compounded by the way mental illness is perceived. People tend to view mental illness as biological and constituting a deep part of people's identity—and therefore as manageable but unlikely to fundamentally change [52]. In line with this, Piazza and colleagues [40] found that it was perceptions of another's dangerous disposition, rather than another's actual capacity to act out on that disposition, that was associated with reduced moral standing. That is, mentally impaired people might not only pose an immediate threat to one's physical safety, but also be perceived to present a more generalized threat to societal order and integrity because of who they are. Thus, the effects of mental impairment on one's rights may be particularly pernicious.

This topic has significant practical import, because views about mentally ill offenders' moral status and rights shape public policy. For example, public outrage over a 2016 murder by a man who had previously been under state supervision (following a defense of mental impairment for an earlier crime) led to significant pressure on policy makers to introduce stricter restrictions on those using the defense. While the board responsible for supervisory decisions pushed back on criticisms, stating that such changes "could lead to violations of the rights" of the mentally ill [53], public lobby to diminish rights have in other cases resulted in legislative action [54]. Critically, restoration of the civil rights in particular is essential for those exiting the criminal justice system, given resumption of citizenship roles and inclusion in activities that build social capital are key features of effective rehabilitation [55–57].

Overview

The present research examines the effects of denial of agency through the defense of mental impairment on judgments of moral responsibility, as well as downstream effects on rights for criminal offenders. Two key models are proposed (see Fig 1): (a) The first model states that a defense of mental impairment will reduce perceptions of the defendant's agency, which causes reduced attributions of moral responsibility for a criminal act compared to a defendant who plead guilty (Studies 1 and 2); and (b) The second model states that a defense of mental
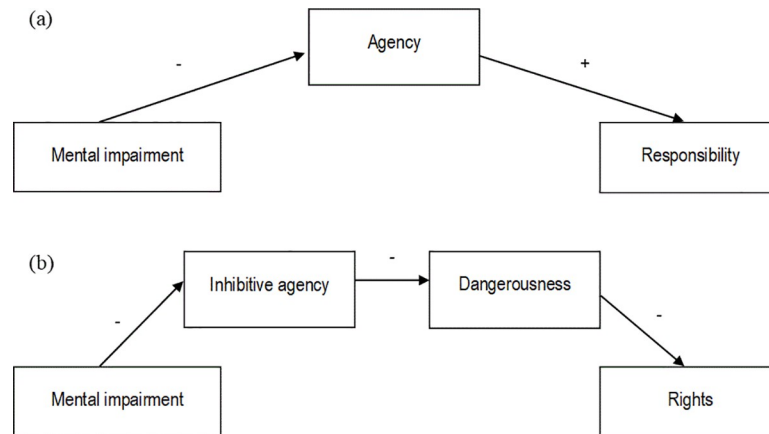
Fig 1. Models of the effects of using the defense of mental impairment on (a) responsibility; and (b) rights.

https://doi.org/10.1371/journal.pone.0252586.g001

impairment will reduce perceptions of the defendant's (inhibitive) agency, which increases perceptions of dangerousness and through this, ascription of rights (Studies 1 and 3). Re-establishing agency should counteract these negative effects—leading to less perceived dangerousness and more rights (Study 2). Studies 1 and 2 demonstrate preliminary support for the models. A pre-registered third study improves on various aspects of the first two studies' design and aims to more rigorously test the reliability and credibility of the model pertaining to rights (b).

While mediation analyses cannot establish causality, we contend that attributions of agency are likely to precede judgments of moral responsibility and rights. The causal effects of attributions of mental capacities (including agency and harmful dispositions) on responsibility and moral standing has been established in experimental studies [58], and we manipulate agency directly in Study 2. While we cannot completely rule out the possibility that motivated cognition may account for at least one of the proposed paths (i.e., perceiving entities as more dangerousness as a justification for granting them fewer rights; see [58]), we consider this causal direction less plausible. Nevertheless, the indirect effects reported in this paper should not be interpreted as definitive proof of causal relationships.

Study 1

Study 1 provided a preliminary test of our model. We manipulated whether a defendant used the defense of mental impairment (or not) and explored effects on agency, moral responsibility, and rights. The study has two distinct parts, corresponding to the two arms of our model. Specifically, we predicted that (a) using the defense of mental impairment would lead to more lenient crime judgments (less responsibility, blame, and punishment) at trial and that these effects would be mediated by perceived agency. However, after the defendants serve their respective sentences, this mitigation of initial responsibility would come at a cost. Specifically, we predicted that (b) after serving their sentences, non-mentally impaired criminals would be ascribed more rights relative to the newly released mentally impaired patients, and that these differences would be mediated by perceptions of lower agency and increased dangerousness.

Method

Study materials for all studies are presented in the S1 File for this article, and are also publicly available (along with datasets) at https://osf.io/w8p4t/?view_only=ecf04457ffd14f2681719 40ea887df6a. All three studies reported in this paper were granted ethics approval from

institutional review board of the University of North Carolina at Chapel Hill, and all participants provided informed consent to participate via the online survey. All measures, manipulations, and exclusions in the studies are disclosed here. Aside from the measures reported in the Materials section below, we collected additional measures in Studies 1 and 2 that are not central to our analyses. Detail about these measures are provided in the S1 File.

Participants. The sample size for Study 1 was primarily determined by budgetary constraints. G*Power 3 [59] indicated that cell sizes of 90 would yield 80% power to detect a medium effect size (d = .42) for a two-tailed t-test at $p < .05$. We deemed this level of sensitivity acceptable for a preliminary study, noting that past research found medium to large effects of mental disorder manipulations on blame and punishment of offenders [18]. Sample sizes were determined prior to data collection (true for all studies).

Four hundred and twenty-four United States residents completed the survey titled "perceptions of a person" through Amazon Mechanical Turk (MTurk) in exchange for payment. Data integrity measures were used, such as requiring HIT approval ratings over 95%, keeping the task visibility private and including two attention checks (see Materials). Eighty-eight respondents failed attention check items, leaving a final sample of 336 participants (132 males, 200 females, 4 did not indicate gender, M_{age} = 37.44 years, SD = 12.55). We recognize that 88 participants seems like a high number for exclusion, but excluding participants who fail to adequately read instructions is standard practice to ensure data integrity in online research [60]. The number of exclusions did not vary significantly by condition, χ^2 (3) = 1.16, p = .763. Including attention check fails produces results consistent with the reported analysis.

Procedure. The design was a full between-subjects vignette study design with random allocation to one of four conditions (control at trial, mental impairment at trial, control after release, mental impairment after release). To examine the two different sides of the defense of mental impairment, the study design can be understood as having two distinct separate parts. The first part related to Hypothesis 1 (i.e., that using the defense would lead to more lenient crime judgments). Participants allocated to this part of the study (the trial conditions) received a vignette describing a 25-year-old man ("Jeff") who was on trial for pushing a stranger off a pedestrian bridge, causing permanent nerve damage and paralysis. Participants read that (a) the defendant pleads guilty (control condition), or (b) the judge rules that Jeff is not guilty by rule of insanity (mental impairment [MI] condition) due to psychotic delusions caused by schizophrenia. To control perceptions of the defendant's remorse, all participants read that "Jeff is distressed and remorseful about what he has done." Participants rated the protagonist Jeff on moral responsibility for the offence, along with other judgments related to the crime, as well as agency and dangerousness.

To test Hypothesis 2 (i.e., the long-term negative effects of using the defense on rights), the other half of the sample read a similar vignette about the character several years later, after release from detainment. These participants read the same vignette to begin with, but were given additional information about the defendant's sentence, in which Jeff was described as having "done his time" and was ready for release, following (a) two years' detainment in a prison (control condition), or (b) a psychiatric ward (MI condition). Participants then rated the protagonist on agency, dangerousness, and rights.

We used generic vignettes rather than courtroom stimuli for our manipulation because we sought to gauge *lay* attitudes about the psychological and moral dimensions of mentally ill defendants, rather than juror decision-making processes. Structured juror scenarios invoke a legal frame, which introduces cognitive biases in the processing of scientific information [61, 62], making this approach unsuitable for our research question. Here, we are less concerned with mimicking jury processes and more concerned with providing simple information that exposes peoples' underlying intuitions and beliefs [63].

Materials. Unless otherwise indicated, all items were rated on a 5-point rating scale, with scale anchors "Not at all" to "Extremely" (or similar). We report the mean inter-item correlation as a measure of reliability for 3-item scales alongside Cronbach's alpha, as scales with few items can bias Cronbach's alpha downwards [64]. A mean correlation above .20 indicates good reliability [65].

Agency. We included two measures of agency in this study. First, we constructed a scale of *general moral agency* using traditional concepts of legal responsibility: volition and moral knowledge (scale α = .80). Since the two components of general moral agency were designed to tap into notions underlying legal responsibility for the criminal act itself, we predicted that this type of agency would be most relevant to judgments about the crime (i.e., at trial). We tapped into the volition component with three items ("In general, to what extent is Jeff [willful]; [intentional]"; "To what extent do you think Jeff . . . can freely choose his own actions"), which were adapted from the battery used in Gray et al. [20]. The scale also included two items relating to moral knowledge ("To what extent do you think Jeff . . . [knows the difference between right and wrong]; [understands the effects of his behaviour on others]").

We also measured a second type of agency, *inhibitive agency*, in order to capture agency judgments most relevant to conferral of rights after release form detainment (which we anticipated would function via dangerousness). Three items made up the inhibitive agency scale ("Jeff is able to inhibit negative impulses"; "Jeff would normally refuse to participate in immoral behavior"; "Jeff has the power to refrain from hurting others"), mean inter-item correlation of .24 (α = .50). Inhibitive agency and general moral agency were positively correlated (r = .30, $p <$ .001).

Crime judgments. Participants judging defendants at the trial stage rated the defendant on single-item measures of responsibility, blame, and punitiveness (e.g., "How much punishment does Jeff deserve for what he did?"). Participants were also asked to indicate how many years Jeff should be detained for (in a prison or psychiatric ward, depending on experimental condition).

Dangerousness. A three-item scale assessed dangerousness. We included an item assessing risk of future harm ("How likely do you think it is that Jeff will commit another crime in his lifetime?"), as well as capturing broader judgments of dangerousness ("How dangerous is Jeff currently?"; "How comfortable would you feel attending the same public function as this man?"), mean item intercorrelation of .53 (α = .76).

Rights. We included three measures of rights: (1) moral standing, a measure of negative rights consistent with previous research in moral psychology, (2) rights to social services, a measure of an individual's positive rights to public services, and (3) rights as a citizen. Importantly, these are dimensions that reflect judgments of moral worth and social inclusion, and are not rights that are stripped from those with a mental impairment by sentencing policies.

Moral standing. Four items assessed moral standing (α = .86), adapted from Piazza et al. [40] and Goodwin and Landy [39], for example, "Harming Jeff would be morally wrong;" "Jeff deserves to be treated with care and compassion."

Social services. Four items measured access to different social services (α = .90), namely whether the state should provide welfare payments, education training, career services and financial advice to Jeff.

Right as citizen. Three items measured rights as a citizen (i.e., Jeff's right to vote, marry, and own property), mean item intercorrelation of .56 (α = .79). This measure of rights is consistent with notions of liberty and closely tied to many of the invisible punishments historically levied against offenders.

Attention checks. To assess whether participants read the vignette properly, a reading check asked participants to select whether the vignette described that Jeff (a) admitted guilt, or (b) was not guilty by reason of mental impairment (though we used the term "insanity defense"

throughout the survey to fit the U.S. context). An instructional attention check was also used [60]. Respondents failing these checks (placed at the end of the study) were excluded from the sample.

Results

In this study, we test for two main predictions: that the defense of mental impairment will reduce responsibility at the time of the trial, and that it will reduce rights after release. Both of these predictions depend on perceptions of agency, which we measured in two different ways. Before turning to each hypothesis, we first report the agency scores and examine how the defense of mental impairment impacts perceptions of moral agency and inhibitive agency both at the time of the trial, and after release.

Judgments of agency.   Mean scores of the protagonist's perceived agency across all four conditions are presented in Fig 2. For the analysis of part 1 of the study (comparing control and mental impairment at trial), a t-test indicated defendants with mental impairment were perceived to have significantly less moral agency than those in the control condition $t(166) =$ 9.33, $p < .001$. There were no differences between the two conditions on inhibitive agency, $t(166) = -0.15$, $p = .879$.

For the analysis of part 2 of the study (comparing control and mental impairment post sentence), mentally impaired defendants were perceived to have less moral agency than control after release from detention, $t(166) = 8.12$, $p < .001$. MI defendants were also perceived as having significantly less inhibitive agency than control after release $t(166) = 3.57$, $p < .001$.

These results indicate that judgments of general moral agency (rather than inhibitive agency) distinguish mentally impaired defendants from sane defendants at trial. Although not expected, a lack of difference in inhibitive agency at the time of the trial makes sense. Each defendant has just committed a crime that suggests a lack of personal restraint in committing negative acts. Arguably, a more general measure of agency is better able to reliably distinguish responsibility judgments about mentally impaired defendants. After release, however, differences in inhibitive agency were apparent. Mediation analyses conducted in the following sections will examine the critical role of agency in determining crime judgments and in conferring rights to defendants after their release from detainment.

Effects of the defense of mental impairment on responsibility at the time of trial.
Descriptive statistics for between-group effects for those in the trial groups are presented in Table 1. As predicted, at the time of the trial, using the defense of mental impairment led to less perceived moral responsibility, blame, and deserved punishment for the crime. There were no significant differences between the conditions on length of detainment.
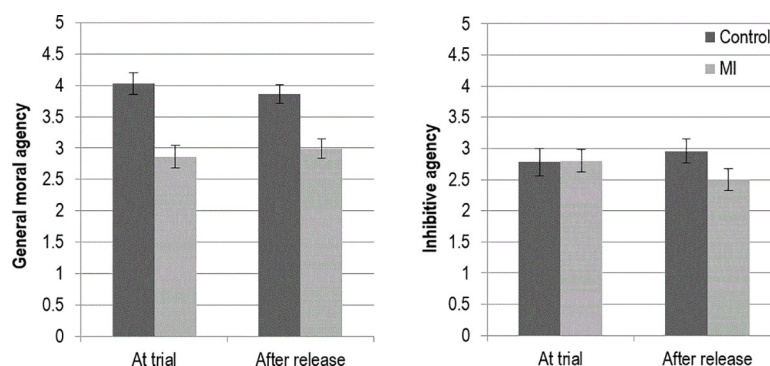

Fig 2. Perceived agency of defendant across experimental conditions (+/- 2 SE).

https://doi.org/10.1371/journal.pone.0252586.g002

Table 1. Descriptive and t-test statistics for dependent variables at trial (Study 1).

| | M (SD) | | | | | |
Dependent variable	Control (n = 86)	MI (n = 82)	t	df	p	d
Responsibility	4.86 (0.44)	3.56 (1.23)	9.05	100.45	< .001	1.41
Blame	4.76 (0.59)	3.23 (1.02)	11.75	128.69	< .001	1.84
Punitiveness	4.13 (0.82)	3.11 (0.92)	7.58	166	< .001	1.17
Years detained	11.19 (13.53)	8.26 (10.94)	1.52	164	.129	0.24
General moral agency	4.03 (0.81)	2.86 (0.81)	9.33	166	< .001	1.44
Inhibitive agency	2.78 (1.00)	2.80 (0.82)	0.15	166	.879	0.02

To test whether perceptions of reduced moral agency corresponded to decreases in blame and punishment, we conducted a mediation analysis using general moral agency as a mediator. In order to determine the total, direct, and indirect effects, the PROCESS macro for SPSS (Hayes, 2013; model 4) was used with 10,000 bootstraps and bias-corrected confidence intervals (conservatively adjusted to 99% CIs for multiple tests). As predicted, general moral agency mediated the effect of the defense of mental impairment on judgments of responsibility, blame, and punitiveness (see Table 2). We also tested models using inhibitive agency instead of general moral agency as the mediator. Unsurprisingly, since there was no effect of condition on inhibitive agency, none of the models were significant. Taken overall, the results indicate that participants' lenient judgments of the mentally ill defendants can be explained—at least partially—by the perception that they lacked moral agency.

Long-term implications of using the defense of mental impairment on rights. Providing some support for our predictions, when Jeff used the defense of mental impairment and had spent time in a psychiatric institution, participants judged him as more dangerous and granted him fewer rights as a citizen, relative to sane Jeff (see Table 3). Participants were, however, *more* likely to endorse access to social services for the mentally ill character, while moral standing did not differ significantly by condition. Thus, results suggest that after serving their time, mentally ill offenders were advantaged in terms of receiving services but they suffered more disadvantage in terms of rights as a citizen.

We then examined whether mentally ill Jeff's reduced rights were related to lower perceived inhibitive agency, and higher perceived dangerousness. To test this proposed model (see Fig 3), we conducted a serial mediation analysis (as per Hayes [66] Model 6; 10,000 bootstraps and 99% bias-corrected confidence intervals). The analysis revealed that claiming a defense of mental impairment was associated with fewer rights as a citizen, and that this was mediated by lower levels of inhibitive agency and higher levels of perceived dangerousness, $B_{indirect}$ = -.12, SE = .05, 99% CI [-.28, -.03]. This indirect pathway completely mediated the effect of the defense on rights as a citizen (B_{direct} = -.07, SE = .14, p = .620).

Table 2. Mediation model statistics: Effects of defense of mental impairment on crime judgments via agency (Study 1).

| | M→DV | | | IV → DV (total effect) | | | IV → DV (direct effect) | | | Indirect effect | | |
Outcome variable	B	SE	CI99%	B	SE	CI99%	B	SE	CI99%	B	SE	CI99%
Responsibility	0.35	.08	0.14, 0.57	-1.30	0.14	-1.67, -0.93	-0.89	0.17	-1.32, -0.46	-0.41	0.11	-0.76, -0.16
Blame	0.42	0.07	0.23, 0.61	-1.52	0.13	-1.86, -1.19	-1.04	0.15	-1.42, -0.66	-0.49	0.10	-0.80, -0.25
Punitiveness	0.29	0.08	0.08, 0.50	-1.02	0.13	-1.37, -0.67	-0.67	0.16	-1.09, -0.26	-0.34	0.10	-0.64, -0.11
Years detained	0.89	1.20	-2.24, 4.03	-2.92	1.92	-7.92, 2.08	-1.88	2.38	-8.09, 4.33	-1.04	1.61	-5.38, 3.25

Condition variable coded as control = 0, MI = 1.

Table 3. Descriptive and t-test statistics for dependent variables after release (Study 1).

Dependent variable	*M* (*SD*)		*t*	df	*p*	*d*
	Control (*n* = 87)	MI (*n* = 81)				
General moral agency	3.86 (0.70)	2.99 (0.69)	8.12	166	< .001	1.25
Inhibitive agency	2.96 (0.89)	2.50 (0.78)	3.56	166	< .001	0.55
Dangerousness	3.83 (0.91)	4.40 (0.88)	-4.11	166	< .001	-0.64
Moral standing	3.81 (1.02)	4.06 (0.87)	-1.66	166	.098	-0.26
Services	2.64 (1.18)	3.05 (1.20)	-2.24	166	.026	-0.34
Rights as citizen	3.89 (0.99)	3.39 (1.14)	3.02	166	.003	0.49

We also tested whether indirect effects via inhibitive agency and dangerousness could predict scores on the other measures of rights. Despite the direction of the total effect on services being positive (i.e., defense of mental impairment led to more services), the indirect effect was negative, indicating that using the defense had negative consequences on access to social services insofar as participants perceived a loss of inhibitive agency and a subsequent increase in dangerousness ($B_{indirect}$ = -.09, SE = .04, 99% CI [-.23, -.02]). In addition, we examined the indirect effect on moral standing, since mediation does not require a total effect of the independent variable on the dependent variable [67]. The indirect effect on moral standing was significant ($B_{indirect}$ = -.08, SE = .04, 99% CI [-.22, -.02]) through inhibitive agency and dangerousness. Last, we ran the same mediation models but substituting general moral agency for inhibitive agency; none of the models were significant (ps>.05). This indicates that post release, it is not volition or moral knowledge—general moral agency—that matters for rights (indeed, lacking these capacities, such persons may simply be passive and accepted members of society), but rather, the ability to control the harm one may inflict on society in the future.

Discussion

Consistent with predictions, the defense of mental impairment effectively mitigated blame for a crime, though it did not affect total time detained (in line with Silver [30]). The effectiveness of this defense occurred in part by reducing perceived moral agency. That is, without the ability to discern right from wrong or act freely, the mentally impaired defendant was held less accountable for his crime. However, there were long-term ramifications to being perceived as having a deficit in agency. In invoking a defense of mental impairment, the perpetrator also had fewer rights as a citizen in the eyes of participants. This cost came from being perceived as less able to control his negative impulses and thus being a danger to society. The defense also reduced other types of rights indirectly through this pathway, though the total effects were
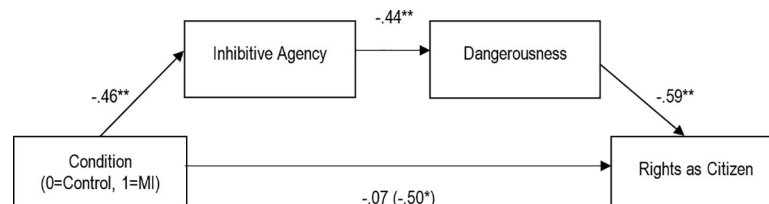
** *p*<.001, * *p*<.05

Fig 3. Indirect effect model of using the defense of mental impairment on rights as citizen, via inhibitive agency and dangerousness. Path values are expressed in unstandardized coefficients.

positive (but nonsignificant). At this stage it is unclear whether the positive effects of mental impairment on moral standing and services reflect noise in the data or if these types of rights are distinct from rights as a citizen. Moreover, the inhibitive agency scale had low internal reliability. Further testing was needed to confirm initial findings and improve the study measures.

Study 2

Study 1 demonstrated that while the defense of mental impairment can reduce perceived responsibility for a crime, the lack of agency implied by one's mental illness also reduces certain rights in the long term. Yet, it is possible that the effects observed in Study 1 were not due to reduced agency per se, but due to other variables commonly associated with mental illness, such as generalised stigma and disgust [68, 69].

Study 2 was designed to provide a more robust test of the causal and critical role of agency in reduced moral status for those using a defense of mental impairment. We replicated and extended the experimental design in Study 1 with a key change: we included *two* mentally impaired conditions at the post-sentence (after release) level. Here, we directly manipulated inhibitive agency (the proposed mediator for the effects of the defense of mental impairment on rights). This design thus employs a *blockage* manipulation [70] to block the proposed mediation process. Though in both of the mental impairment conditions, Jeff is still mentally ill—thereby controlling for mental illness stigma (and stereotypes pertaining to his mental illness in particular)—we varied only the extent to which he had learned to manage his illness and control his behavior (i.e., the extent to which he had regained inhibitive agency). This is consistent with law in many jurisdictions, where forensic patients can be held indefinitely until they have demonstrated such agency (at which time they may be released). If agency is a mediator of the defense, then reduced levels of rights should be observed when Jeff has not learned to control his impulses (MI-unrecovered condition), but not when he has learned to do so (i.e., when he has regained his agency; MI-recovered condition).

We expected to replicate results of Study 1 in relation to implications at trial: that mental impairment would reduce perceived responsibility, blame, and punitiveness, and that this would be mediated by moral agency. With regard to the second arm of the model, we expected those released from a psychiatric ward and unable to control their impulses would be viewed as lacking in inhibitive agency, more dangerous, and having fewer rights relative to those released from prison. We predicted that this diminution in rights would be across the board; that is, we would observe a consistent negative effect across all three rights measures. While there was some evidence in Study 1 that the defense might increase access to services, the lack of inhibitive agency (and therefore dangerousness) of the mentally impaired (unrecovered) character will be more salient, and therefore we expected negative effects of agency on rights to be stronger. Further, we predicted that regaining inhibitive agency would nullify any negative effects of the defense of mental impairment on later rights, such that using the defense is not a barrier to social reintegration so long as one's agency is regained (mediated by a reduction in perceived dangerousness).

Method

Participants. We slightly increased our desired sample size relative to Study 1, as we reasoned that mean differences between the two post-sentence mental impairment conditions (i.e., recovered versus unrecovered agency) could be slightly smaller than those between the control condition and the unrecovered agency condition (i.e., equivalent to the control–MI contrast in Study 1). Working from the main effect of condition on rights as citizen of $d = 0.49$ in Study 1, we chose a d of 0.40 as the smallest effect size of interest in Study 2 if the recovered

agency condition is to be meaningfully distinguished from the unrecovered agency condition. G*Power indicated a sample of 100 per condition was required for 80% power to detect one-tailed contrasts in the predicted directions between two independent groups of $d = 0.40$ at $p = .025$ (adjusted for multiple comparisons between the three post-sentence conditions).

Seven hundred and sixty-seven United States residents completed the survey through MTurk. One hundred and eighty-six respondents were excluded for failing the attention checks. Including attention check fails produces results consistent with the reported analysis. The final sample was 581 participants ($M_{age} = 39.05$ years, $SD = 13.29$; 57% female, 43% male).

Procedure. Study 2 used a design nearly identical to Study 1, with the exception that after release, we included two mental impairment conditions: one in which the mentally impaired offender was described as having regained inhibitive agency, and the other where he had not. This manipulation allowed us to empirically test whether it was an ongoing reduction in inhibitive agency, and a corresponding increase in perceived dangerousness, that accounted for any reduction in rights upon release from detainment. Participants in all three post-sentence conditions (including the control condition) read that before release, Jeff had a hearing to review his progress, in which the wardens (or prison officers, in the case of the control condition) and psychologists at the facility in which Jeff was detained testify that "Jeff is eligible for release" and is therefore returned to society, having "done his time." The two MI conditions received some additional information. In the MI-recovered condition, participants read that "they state that Jeff is in control of his illness and has learned to manage his illness independently. They also testify that Jeff has developed the capacity to control his emotions and impulsive behaviors." In contrast, in the MI-unrecovered condition, participants read, "he could, however, learn to be more in control of his illness and to better manage his illness independently. They also testify that Jeff has struggled at times to control his emotions and impulsive behaviors."

Materials. Measures were unchanged from Study 1, with the following exceptions. We sought to improve the internal reliability of the three-item inhibitive agency scale by amending one item (that had weak correlations with the other items) to be more consistent with the others: "Jeff has the capacity to resist doing the wrong thing" (scale mean item intercorrelation = .81, $\alpha = .93$). In addition, we modified the six-item general agency scale. Two of the scale items were changed to increase the specificity and face validity of the terms ("intentional" and "willful" were replaced with "mental capacity to think about and plan for the future" and "exercise self-control"), and an extra item measuring the moral knowledge dimension of the scale was added ("is aware of social norms") to balance it with the volition dimension. Internal consistency for the general agency scale was excellent ($\alpha = .91$).

We focused on different measures of agency across the conditions. Since Study 1 results confirmed that general moral agency was relevant to judgments about the crime (responsibility, blame, and punitiveness), the general moral agency scale was only presented to participants in the trial condition. For the post-sentence analysis we only measured judgments of inhibitive agency, since Study 1 indicated that this measure was more relevant for conferring rights.

Remaining scales showed good internal reliability: dangerousness (mean item intercorrelation = .56, $\alpha = .79$), moral standing ($\alpha = .82$), access to social services ($\alpha = .89$), rights as citizen (mean item intercorrelation = .62, $\alpha = .82$).

Along with the two attention check items from Study 1 (presented to all participants), we presented an additional attention check item to those in the MI conditions post sentence, in order to eliminate those who did not read the recovery manipulation vignette properly. Participants were asked whether in the vignette, wardens and psychologists testified at Jeff's hearing that Jeff had: (a) learned to control his illness and impulses, or (b) struggled to control his illness and impulses. Those who answered incorrectly were excluded from the analyses.

Table 4. Descriptive and t-test statistics for dependent variables at trial (Study 2).

| | M (SD) | | | | | |
Dependent variable	Control (n = 125)	MI (n = 123)	t	df	p	d
Responsibility	4.71 (0.66)	3.52 (1.22)	9.57	186.99	< .001	1.21
Blame	4.63 (0.76)	3.42 (1.17)	9.63	208.06	< .001	1.23
Punitiveness	4.15 (0.80)	3.24 (1.10)	7.43	223.61	< .001	0.94
Years detained	13.47 (13.61)	15.55 (22.99)	-0.85	185.91	.391	-0.11
General moral agency	3.56 (1.05)	2.75 (0.99)	6.23	246	< .001	0.79

Results

Effects of the defense of mental impairment on responsibility at the time of trial. Replicating Study 1, at the time of trial, using the defense of mental impairment led to less moral responsibility, blame, and punishment for harming the victim (see Table 4). There were no differences in the length of time participants believed the perpetrator should be detained.

Next, we examined the predicted model of effects on crime judgments via moral agency. Mediation models were tested as per the procedure detailed in Study 1. Results indicated that using the defense of mental impairment led to less perceived responsibility, blame, and punitiveness (but not years detained), partially through a reduction in perceived moral agency (see Table 5).

Long-term implications of using the defense of mental impairment on rights: Mentally ill offenders can be redeemed—if they regain agency. Next we explored the implications for perceived rights: whether lack of agency hindered mentally ill offenders from being accepted back into society post-release, and whether these effects could be mitigated by recovered agency. See Table 6 for group means on all dependent variables after perpetrator release. The below analyses draw on one-way ANOVA models, examining contrasts between the three post-sentence conditions.

Analyses revealed that after release, participants attributed more inhibitive agency to sane Jeff (control) than to unrecovered Jeff. However, attributions of inhibitive agency did not differ between control and the recovered agency condition. This suggests that a defendant described as having learned to regain control of their behavior was perceived as having equal inhibitive agency to a defendant without a mental illness. Comparing the two mental impairment conditions confirmed that recovered Jeff was perceived as possessing significantly more inhibitive agency than unrecovered Jeff—as would be expected if the manipulation were successful.

In line with predictions, when Jeff had used the defense of mental impairment and had not recovered any agency, participants found him to be more dangerous, and granted him fewer rights as a citizen, relative to sane Jeff. They did not, however, grant him fewer services or less moral standing. In contrast, when Jeff had used the defense and had recovered his agency, no

Table 5. Mediation model statistics: Effects of defense of mental impairment on crime judgments via agency (Study 2).

| | M→DV | | | IV → DV (total effect) | | | IV → DV (direct effect) | | | Indirect effect | | |
Outcome variable	B	SE	CI99%	B	SE	CI99%	B	SE	CI99%	B	SE	CI99%
Responsibility	0.22	0.06	0.07, 0.38	-1.19	0.12	-1.51, -0.87	-1.01	0.13	-1.35, -0.67	-0.18	0.06	-0.37, -0.06
Blame	0.21	0.06	0.05, 0.37	-1.21	0.13	-1.53, -0.88	-1.04	0.13	-1.38, -0.70	-0.17	0.06	-0.36, -0.04
Punitiveness	0.21	0.06	0.06, 0.36	-0.91	0.12	-1.22, -0.59	-0.74	0.13	-1.07, -0.41	-0.17	0.06	-0.36, -0.04
Years detained	0.86	1.17	-2.19, 3.91	2.07	2.41	-4.19, 8.33	2.76	2.59	-3.97, 9.49	-0.69	1.03	-3.36, 2.29

Condition variable coded as control = 0, MI = 1.

Table 6. Descriptive and inferential statistics for dependent variables after release (Study 2).

Dependent variable	M (SD)			F	df	p
	Control (n = 115)	MI recovered (n = 115)	MI unrecovered (n = 103)			
Inhibitive agency	3.55 (1.02)[a]	3.42 (0.93)[b]	2.70 (0.95)	24.22	2, 330	< .001
Dangerousness	3.62 (0.89)[a]	3.78 (0.84)[b]	4.37 (0.83)	22.95	2, 330	< .001
Moral standing	4.01 (0.76)	4.19 (0.74)	3.97 (0.86)	2.61	2, 330	.08
Services	2.76 (1.27)	3.01 (1.12)	2.79 (1.21)	1.42	2, 330	.243
Rights as citizen	3.93 (0.99)[a]	3.89 (0.97)[b]	3.48 (1.14)	6.20	2, 330	.002

[a]Control different to MI unrecovered (p < .05).

[b]MI recovered different to MI unrecovered (p < .05).

differences existed at all; he was treated in the same manner as his sane counterpart across all outcome variables. In addition, recovered Jeff fared significantly better than unrecovered Jeff in terms of perceived dangerousness and rights as a citizen (e.g., see Fig 4). Overall, the results supported our hypothesis: Recovering one's agency nullified the deleterious effect of using the defense of mental impairment on dangerousness and rights.

Finally, we tested whether the restoration of inhibitive agency was responsible for the differences observed between the MI-recovered and MI-unrecovered conditions. Since inhibitive agency was directly manipulated in Study 2, using the same serial mediation model as in Study 1 (condition → inhibitive agency → dangerousness → Y) would be problematic—the exogenous variable (restoration of inhibitive agency) and the mediator (inhibitive agency) are too proximal and would produce multicollinearity [71]. Instead, we omitted the measured inhibitive agency scale and conducted a simple mediation model (condition → dangerousness → Y), with the contrast between unrecovered and recovered conditions set as the exogenous variable (unrecovered = 0, recovered = 1) [72].
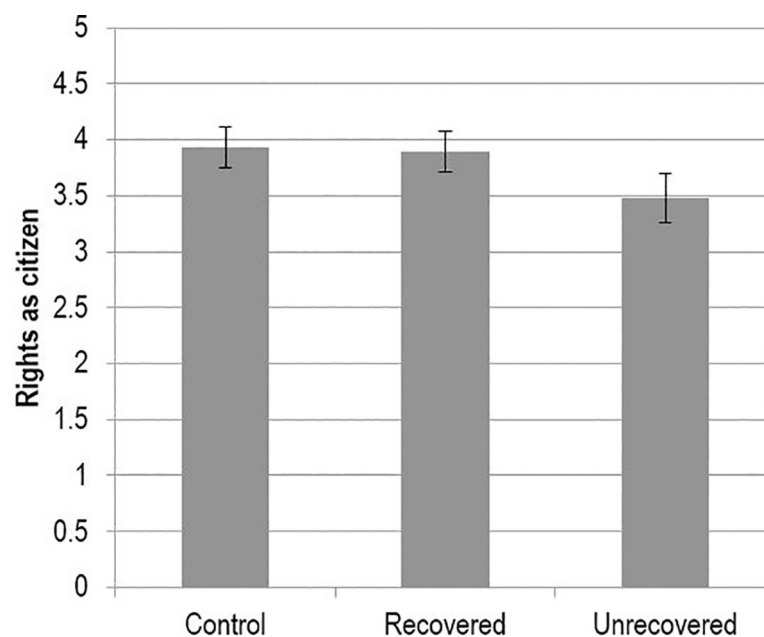

Fig 4. Means scores for rights as citizen measure (+/- 2 SE).

Table 7. Mediation model statistics: Effects of recovered agency on rights via dangerousness (Study 2).

Outcome variable	M→DV			IV → DV (total effect)			IV → DV (direct effect)			Indirect effect		
	B	SE	CI$_{99\%}$	B	SE	CI$_{99\%}$	B	SE	CI$_{99\%}$	B	SE	CI$_{99\%}$
Moral standing	-0.36	0.06	-0.51, -0.20	0.23	0.11	-0.05, 0.51	0.02	0.11	-0.26, 0.29	0.21	0.06	0.09, 0.40
Services	-0.45	0.09	-0.68, -0.21	0.21	0.16	-0.20, 0.62	-0.05	0.16	-0.47, 0.36	0.26	0.08	0.10, 0.53
Rights as citizen	-0.63	0.07	-0.82, -0.43	0.41	0.14	0.04, 0.78	0.04	0.13	-0.30, 0.38	0.37	0.09	0.17, 0.65

Note. Condition variable coded as MI unrecovered = 0, MI recovered = 1.

The analysis revealed that dangerousness completely mediated the effect of the agency manipulation on rights as citizen. In addition, there were significant indirect effects on services and moral standing (see Table 7). These results indicate that when those using the defense of mental impairment are seen as less in control of their actions, they are perceived as more dangerous, and this corresponds to being granted fewer rights of all kinds.

Discussion

The reduction of perceived moral agency through the defense of mental impairment once again resulted in immediate benefits to defendants in terms of reduced responsibility for one's crime; however, the loss of inhibitive agency resulted in some long-term costs insofar as invoking the defense signaled that the offender was dangerous. However, when the mentally impaired defendant was able to demonstrate he had regained his inhibitive agency, he was able to avoid the deleterious effects of the defense of mental impairment. This suggests that inhibitive agency plays a causal role in conferring rights, especially as it relates to dangerousness.

Study 3 (registered study)

Taken together, the results of the first two studies suggest that laypeople judge those using a mental impairment defense as not only less responsible for their actions (and thus their crime), but possessing less inhibitive agency—and consequently being less deserving of certain rights. When we examined the most traditional definition of rights (moral standing), we found mixed findings. Study 1 suggested that the defense of mental impairment increases moral standing at release relative to a guilty plea, though this effect did not replicate in Study 2, and in fact the indirect effect of reduced agency on moral standing via dangerousness was negative in both studies. It is possible that the defense exerts heterogenous effects on rights, involving both positive and negative indirect effects. Specifically, the mental impairment defense might signal increased *patiency* as well as diminished agency—which would theoretically increase rights, particularly those that are strongly tied to notions of vulnerability. Moral standing represents rights that have typically been associated with entities requiring protection (moral patients), so it is not surprising that those with mental impairment could be granted these rights (as in Study 1). Where one's lack of agency is made particularly salient (as in the unrecovered MI condition in Study 2), the negative effects via diminished agency may override any positive effects via patiency on rights.

Moving beyond the basic negative rights definition involving protection from harm, we also examined *positive* liberties, such as the right to marry, as well as the defendant's rights to social services. Access to services followed the pattern seen for moral standing—the mental illness defense either increases access to services (Study 1) or does little to affect it (Study 2), though there is a consistent negative indirect effect of the defense (or manipulated agency in Study 2) via dangerousness. Again, access to services might imply vulnerability to some extent, possibly leading people to grant those using the defense more of these rights.

Rights as citizens showed a more consistent pattern in relation to agency—both total effects and indirect effects across both studies—likely because these rights more naturally align with moral agency, as rights of citizenship assume adult agency. This may also be because citizenship rights are most pertinent to the criminal justice realm, reflecting our historical tendency to socially exclude offenders through diminution of such rights [42, 43]. Crucially, these rights are "invisible punishments" [43] that are not strictly demanded by sentencing policy (i.e., unrelated to one being found guilty or not guilty by reason of mental impairment) but rather, subtle and informal tools of social exclusion. In fact, since many states in the United States and other jurisdictions withhold some citizenship rights from ("sane") prisoners (at least, during their sentence), it is conceivable that the results could have been the other way around—"sane" defendants could have suffered more extreme revocation of rights relative to those using the defense of mental illness. This was not so. Our research thus goes beyond previous literature by showing a clear link between mental impairment and the determination of certain rights— one that cannot be entirely explained by generalized stigma towards those with mental illness (as demonstrated in Study 2) but, rather, is underlain by judgements of agency.

It is worth nothing that we did include an exploratory patiency measure in Study 1, which allow us to test the idea about a second, opposing, indirect effect through patiency. These data do provide support for the indirect effects (results are presented in the S1 File). Of course, since this analysis was based on post-hoc reasoning, a confirmatory, pre-registered study is necessary to test for these two indirect effects in a more robust fashion.

Our previous two studies have some limitations. First, the link between mental impairment and reduced rights is novel, and requires further validation with concrete predictions about the role of agency in conferring various rights. A second, significant, limitation of our existing work is that we only used one example of a crime and mental impairment. The effect may not generalize to other circumstances, even though we theoretically expect it to. For instance, according to our framework, any reduction in agency should yield the same effects, regardless of type of mental illness.

The primary aim of the pre-registered Study 3 is to conduct a conceptual replication of Study 1, focusing on the more novel arm of the research—Hypothesis 2: that the defense of mental impairment corresponds with reduced rights at the end of a sentence. That is, the study will be a 1-factor, 2-level (defense vs. control) experimental vignette design. We will test for heterogenous effects of the defense on rights, exploring two indirect effects of the defense on rights (that is, one via agency/dangerousness and one via patiency). Study 3 thus offers the opportunity to better explain positive (or null) main effects of the manipulation on rights. In addition, the pre-registered study aims to address concerns regarding the generalizability of the research by using a stimulus sampling approach. In this design, participants are exposed to multiple vignettes varying on other (theoretically irrelevant) features. If the proposed effect is robust, it should hold across other stimuli. The registered study thus aims to increase confidence in the model and earlier findings.

Hypotheses

The study will test five hypotheses:

H1: Mentally impaired defendants will have less inhibitive agency, and will be perceived as more dangerous than control.

We predict differentiated main effects of the defense on rights. Since rights as a citizen are more closely tied to agency, we still expect to observe a negative main effect of the defense on this measure. That is, H2: Those using the defense will be deemed less deserving of rights as a citizen relative to control. In contrast, we predict that the defense will result in *more* (or no

change in) moral standing and access to services. That is, H3: Those using the defense will be granted more (or equal) moral standing and right to social services relative to control.

Based on the results of the first two studies, there is reason to suspect that an alternate mediator is exerting a counteracting influence on rights. We therefore expect to find two indirect effects of the defense on these all three rights measures. This constitutes the fourth hypothesis:

H4: We will observe a negative indirect effect on rights via inhibitive agency and dangerousness; and a positive indirect effect via patiency.

We make further predictions about the nature of these indirect effects in the final hypothesis:

H5: We expect the indirect effect on rights as citizen via the inhibitive agency and dangerousness route to be stronger than the indirect effect through the patiency route. In contrast, in the models for access to services and moral standing, we expect the indirect effect via patiency to be equal in magnitude to the indirect effect via inhibitive agency and dangerousness.

Method

Participants. Participants (United States residents) will be recruited from MTurk. As per our previous protocol, exclusion criteria will be limited to MTurk users with a HIT approval rating below 95% and respondents failing attention checks. No interim analyses or stopping rules will be applied to data collection. This study will be covered by existing ethics approval from the institutional review board of the University of North Carolina at Chapel Hill.

Exclusion criteria. Data will be screened to exclude participants who fail the attention check, which will be a reading check placed at the end of the survey, identical to the one in Study 1 (asking whether the character was found guilty or not guilty by reason of insanity). Any respondent who does not check the correct response will be excluded from further analysis. In addition, the instructional attention check will be placed at the beginning of the survey, filtering out respondents before they can complete the study. This latter group will not be considered dropouts.

Power analysis. Building on the findings from the previous studies, we conducted a robust power analysis for the registered study. We determined that the sample should have enough statistical power not only to detect medium and large (main) effects, but that smaller differences between control and MI—such as those consistent with effects observed on moral standing ($d = 0.26$) and access to services ($d = 0.34$) in Study 1—would also be meaningful. We therefore set our smallest effect size of interest to a small-medium d of 0.3. According to G*Power, in order to obtain 90% power to detect an effect of this magnitude using a one-tailed independent samples t-test (with alpha set at .0167 to adjust for testing of three rights measures) we require 260 participants per condition.

We followed the method and software described in Schoemann et al. [73] to calculate statistical power required for the indirect effects. Using Monte Carlo simulations and the exact parameters from Study 1, the analysis indicated that an N of 520 would achieve a power >99% (at $\alpha = .0167$ to correct for mediation analyses on three rights measures) to detect the indirect effect of condition on moral standing via inhibitive agency and dangerousness—the smallest indirect effect of condition on any rights measure in that study. Since any counteracting mediation via patiency is expected to be similar (or larger) in magnitude to the negative effect via inhibitive agency and dangerousness, this power analysis is suitable for testing both pathways.

Materials. The experimental manipulation of the defense of mental impairment will be conceptually similar to the one used in Study 1, except for varying features of the vignette. Participants will read one of several vignettes describing a person (Liam) who has recently completed the terms of their court order. The character will be described as having committed a
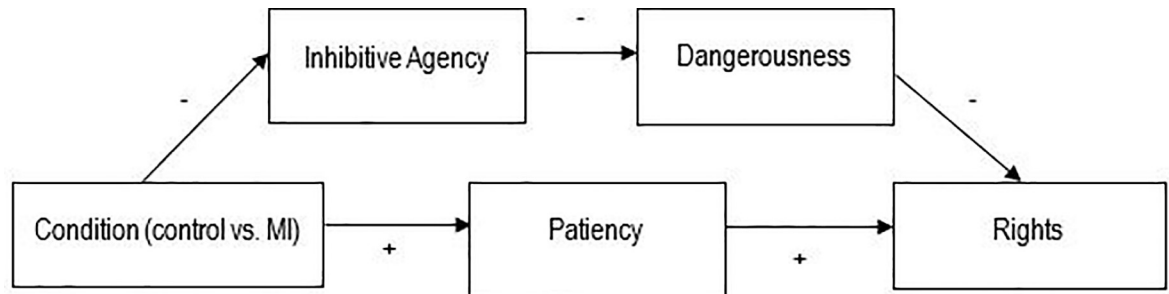
Fig 5. Indirect effect model that will be tested in the registered study.

https://doi.org/10.1371/journal.pone.0252586.g005

crime several years early and was either: found guilty of the crime (control condition); or found not guilty by mental impairment (mental impairment condition). Vignette features will vary by: type of crime (robbery or sexual offending); type of sentence (not crossed but matched to the crime; custodial- or community-based sentence/treatment); and type of mental impairment (bipolar disorder or brain injury). Crossing of vignette features will result in six unique vignettes. These are presented in the S1 File along with outcome measures.

Outcome measures will comprise inhibitive agency, dangerousness, and rights measures (moral standing, rights to social services, and rights as a citizen). These will be unchanged from Study 2. We will also introduce a three-item patiency measure (Goya-Tocchetto, Gray, Vaisey, & Kapsaskis, 2019, unpublished manuscript).

Analysis plan

Independent samples t-tests will be used to test for differences between the control and MI conditions on inhibitive agency, dangerousness, and rights measures (H1, H2, and H3). We will test H1 (mentally impaired defendants will have less inhibitive agency, and will be perceived as more dangerous than control) and H2 (mentally impaired defendants will be deemed less deserving of rights as a citizen relative to control) using a one-tailed test. We will test H3 (mentally impaired defendants will be granted more, or equal, moral standing and right to social services) using a two-tailed test. For the tests on rights, an alpha correction will be applied (α = .017) to account for the inclusion of three rights measures.

To examine indirect effects (H4 and H5), three mediation analyses will be conducted—one for each rights measure as a dependent variable—using SEM. While we will report model fit indices, they are not important for testing our predictions, which pertain to the indirect pathways. Fig 5 below illustrates the pathways that will be tested.

Specifically, use of the mental impairment defense will lead to less inhibitive agency, which will be associated with an increase in dangerousness and a corresponding diminution of rights. At the same time, a counteracting pathway to rights will be observed whereby the defense increases patiency and rights. The alpha level for each indirect effect will be set to .017 to adjust for the use of three rights measures.

Supporting information

S1 File. Supporting information file.
(DOCX)

Author Contributions

References

1. Simon RJ, Ahn-Redding H. The insanity defense, the world over. Lanham: Lexington Books; 2006.

2. Bastian B, Laham SM, Wilson S, Haslam N, Koval P. Blaming, praising, and protecting our humanity: The implications of everyday dehumanization for judgments of moral status. British Journal of Social Psychology. 2011; 50(3):469–83.

3. Allnutt S, Samuels A, O'Driscoll C. The insanity defence: From wild beasts to M'Naghten. Australasian Psychiatry. 2007; 15(4):292–8. https://doi.org/10.1080/10398560701352181 PMID: 17612881

4. Hart HLA, Honoré T. Causation in the law. New York: Oxford University Press; 1985.

5. Platt AM, Diamond BL. The origins and development of the "wild beast" concept of mental illness and its relation to theories of criminal responsibility. Journal of the History of the Behavioral Sciences. 1965; 1 (4):355–67.

6. Robinson PH, Dubber MD. The American model penal code: A brief overview. New Criminal Law Review. 2007; 10(3):319–41.

7. Bagaric M. A rational (unapologetically pragmatic) approach to dealing with the irrational-The sentencing of offenders with mental disorders. Harv Hum Rts J. 2016; 29:1–56.

8. Desmond BC, Lenz PJ. Mental health courts: An effective way for treating offenders with serious mental illness. Mental and Physical Disability Law Reporter. 2010; 34(4):525–30. PMID: 20960980

9. Traynor S. Sentencing trends no 23—Sentencing mentally disordered offenders: The causal link. Sydney, Australia: Judicial Commission of New South Wales; 2002.

10. Barnett ME, Brodsky SL, Davis CM. When mitigation evidence makes a difference: effects of psychological mitigating evidence on sentencing decisions in capital trials. Behavioral Sciences & the Law. 2004; 22(6):751–70. https://doi.org/10.1002/bsl.591 PMID: 15386561

11. Barnett ME, Brodsky SL, Price JR. Differential impact of mitigating evidence in capital case sentencing. Journal of Forensic Psychology Practice. 2007; 7(1):39–45.

12. Berryessa CM. The effects of psychiatric and "biological" labels on lay sentencing and punishment decisions. Journal of Experimental Criminology. 2018; 14(2):241–56.

13. Berryessa CM, Wohlstetter B. The psychopathic "label" and effects on punishment outcomes: A meta-analysis. Law and Human Behavior. 2019; 43(1):9–25. https://doi.org/10.1037/lhb0000317 PMID: 30570278

14. Garvey SP. Aggravation and mitigation in capital cases: What do jurors think? Columbia Law Review. 1998; 98(6):1538–76.

15. Gillespie LK, Smith MD, Bjerregaard B, Fogel SJ. Examining the impact of proximate culpability mitigation in capital punishment sentencing recommendations: The influence of mental health mitigators. American Journal of Criminal Justice. 2014; 39(4):698–715.

16. Mossière A, Maeder EM. Defendant mental illness and juror decision-making: A comparison of sample types. International Journal of Law and Psychiatry. 2015;42–43:58–66. https://doi.org/10.1016/j.ijlp.2015.08.008 PMID: 26314888

17. Roberts CF, Golding SL, Fincham FD. Implicit theories of criminal responsibility. Law and Human Behavior. 1987; 11(3):207–32.

18. Monahan J, Hood GL. Psychologically disordered and criminal offenders: Perceptions of their volition and responsibility. Correctional Psychologist. 1976; 3(2):123–34.

19. Kleinke CL, Baldwin MR. Responsibility attributions for men and women giving sane versus crazy explanations for good and bad deeds. The Journal of Psychology. 1993; 127(1):37–50. https://doi.org/10.1080/00223980.1993.9915541 PMID: 8510051

20. Gray HM, Gray K, Wegner DM. Dimensions of mind perception. Science. 2007; 315(5812):619. https://doi.org/10.1126/science.1134475 PMID: 17272713

21. Meyer SS. Moral responsibility: Aristotle and after. In: Everson S, editor. Ethics. Companions to Ancient Thought. Cambridge: Cambridge University Press; 1998. p. 221–40.

22. Alicke MD. Culpable control and the psychology of blame. Psychological bulletin. 2000; 126(4):556. https://doi.org/10.1037/0033-2909.126.4.556 PMID: 10900996

23. Cushman F. Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. Cognition. 2008; 108(2):353–80. https://doi.org/10.1016/j.cognition.2008.03.006 PMID: 18439575

24. Fincham FD, Roberts C. Intervening causation and the mitigation of responsibility for harm doing II. The role of limited mental capacities. Journal of Experimental Social Psychology. 1985; 21(2):178–94.

25. Gray K, Schein C. When your life is on the line, be a victim, not a hero. Jury Expert. 2011; 23(6):8–14.

26. Guglielmo S, Monroe AE, Malle BF. At the heart of morality lies folk psychology. Inquiry. 2009; 52 (5):449–66.

27. Monterosso J, Royzman EB, Schwartz B. Explaining away responsibility: Effects of scientific explanation on perceived culpability. Ethics & Behavior. 2005; 15(2):139–58.

28. Pizarro D, Uhlmann E, Salovey P. Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. Psychological Science. 2003; 14(3):267–72. https://doi.org/10.1111/1467-9280.03433 PMID: 12741752

29. Patricny N, & Woodworth M. Examining guilt and blame attribution for offence: Comparing a public and forensic psychiatric sample Okanagan, Canada: University of British Columbia; 2010.

30. Silver E. Punishment or treatment? Law and Human Behavior. 1995; 19(4):375–88.

31. Hans VP. An analysis of public attitudes toward the insanity defense. Criminology. 1986; 24(2):393–414.

32. Perlin ML. Myths, realities, and the political world: the anthropology of insanity defense attitudes. Bull Am Acad Psychiatry Law. 1996; 24(1):5–26. PMID: 8891319

33. Silver E, Cirincione C, Steadman HJ. Demythologizing inaccurate perceptions of the insanity defense. Law and Human Behavior. 1994; 18(1):63–70.

34. Allen CH, Vold K, Felsen G, Blumenthal-Barby JS, Aharoni E. Reconciling the opposing effects of neurobiological evidence on criminal sentencing judgments. PLOS ONE. 2019; 14(1):e0210584. https://doi.org/10.1371/journal.pone.0210584 PMID: 30657761

35. Gray K, Wegner DM. Moral typecasting: divergent perceptions of moral agents and moral patients. Journal of Personality and Social Psychology. 2009; 96(3):505. https://doi.org/10.1037/a0013748 PMID: 19254100

36. Knobe J, Prinz J. Intuitions about consciousness: Experimental studies. Phenomenology and the Cognitive Sciences. 2008; 7(1):67–83.

37. Robbins P, Jack AI. The phenomenal stance. Philosophical Studies. 2006; 127(1):59–85.

38. Sytsma J, Machery E. The two sources of moral standing. Review of Philosophy and Psychology. 2012; 3(3):303–24.

39. Goodwin GP, Landy JF. Valuing different human lives. Journal of Experimental Psychology: General. 2014; 143(2):778–803. https://doi.org/10.1037/a0032796 PMID: 23647311

40. Piazza J, Landy JF, Goodwin GP. Cruel nature: Harmfulness as an important, overlooked dimension in judgments of moral standing. Cognition. 2014; 131(1):108–24. https://doi.org/10.1016/j.cognition.2013. 12.013 PMID: 24462925

41. Wenar L. Rights. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Fall 2015 ed. Stanford, CA: Metaphysics Research Lab, Stanford University; 2015.

42. Newheiser A-K, DeMarco TC. Who deserves basic rights? People condone violations of procedural and physical rights in the treatment of terrorist suspects. Law and Human Behavior. 2018; 42(1):50–6. https://doi.org/10.1037/lhb0000275 PMID: 29461077

43. Travis J. Invisible punishment: An instrument of social exclusion. In: Chesney-Lind MMM, editor. Invisible punishment: The collateral consequences of mass imprisonment. New York, NY: The New Press; 2002. p. 15–36.

44. Carter I. Positive and negative liberty. The Stanford Encyclopedia of Philosophy. Summer 2018 ed. Stanford, CA: Metaphysics Research Lab, Stanford University; 2003.

45. Kant I. Groundwork for the metaphysics of morals. Reprint (original work 1785), Binghampton, NY: Yale University Press; 2002.

46. Angermeyer MC, Dietrich S. Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatrica Scandinavica. 2006; 113(3):163–79. https://doi.org/10. 1111/j.1600-0447.2005.00699.x PMID: 16466402

47. Martinez AG, Piff PK, Mendoza-Denton R, Hinshaw SP. The power of a label: Mental illness diagnoses, ascribed humanity, and social rejection. Journal of Social and Clinical Psychology. 2011; 30(1):1–23.

48. Perlin ML. On sanism. SMUL Rev. 1992; 46:373.

49. Sowislo JF, Lange C, Euler S, Hachtel H, Walter M, Borgwardt S, et al. Stigmatization of psychiatric symptoms and psychiatric service use: a vignette-based representative population survey. European Archives of Psychiatry and Clinical Neuroscience. 2017; 267(4):351–7. https://doi.org/10.1007/s00406-016-0729-y PMID: 27761652

50. Stuart HL, Arboleda-Flórez JE. A public health perspective on violent offenses among persons With mental illness. Psychiatric Services. 2001; 52(5):654–9. https://doi.org/10.1176/appi.ps.52.5.654 PMID: 11331801

51. Bonta J, Law M, Hanson K. The prediction of criminal and violent recidivism among mentally disordered offenders: a meta-analysis. Psychological Bulletin. 1998; 123(2):123. https://doi.org/10.1037/0033-2909.123.2.123 PMID: 9522681

52. Haslam N, Ernst D. Essentialist beliefs about mental disorders. Journal of Social and Clinical Psychology. 2002; 21(6):628–44.

53. Fraser J. Oregon board explains why it repeatedly released killer from psychiatric hospital. ProPublica. 2018 Apr 25.

54. Baillie R. Not guilty by reason of mental illness: 'It's a slap in our faces' 2020 Jun 3. Available from: https://www.sbs.com.au/news/insight/not-guilty-by-reason-of-mental-illness-it-s-a-slap-in-our-faces.

55. Moore KE, Stuewig JB, Tangney JP. The effect of stigma on criminal offenders' functioning: A longitudinal mediational model. Deviant Behavior. 2016; 37(2):196–218. https://doi.org/10.1080/01639625. 2014.1004035 PMID: 26973364

56. Ward T, Maruna S. Rehabilitation. London: Routledge; 2007.

57. Winnick TA, Bodkin M. Anticipated stigma and stigma management among those to be labeled "ex-con". Deviant Behavior. 2008; 29(4):295–333.

58. Goodwin GP. Experimental approaches to moral standing. Philosophy Compass. 2015; 10(12):914–26.

59. Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007; 39(2):175–91. https://doi.org/10.3758/bf03193146 PMID: 17695343

60. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology. 2009; 45(4):867–72.

61. Berryessa CM. Jury-eligible public attitudes toward biological risk factors for the development of criminal behavior and implications for capital sentencing. Criminal Justice and Behavior. 2017; 44(8):1073–100.

62. Devine DJ, Clayton LD, Dunford BB, Seying R, Pryce J. Jury decision making: 45 years of empirical research on deliberating groups. Psychology, Public Policy, and Law. 2001; 7(3):622.

63. Bieneck S. How adequate is the vignette technique as a research tool for psycho-legal research? In: Oswald SB M. E., & Hupfeld-Heinemann J., editor. Social psychology of punishment of crime. New York, NY, US: John Wiley & Sons Ltd; 2009. p. 255–71. https://doi.org/10.1007/s00132-009-1474-y PMID: 19609773

64. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology. 1993; 78(1):98.

65. Briggs SR, Cheek JM. The role of factor analysis in the development and evaluation of personality scales. Journal of Personality. 1986; 54(1):106–48.

66. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. New York, NY: Guilford; 2013.

67. Zhao X, Lynch JGJ, Chen Q. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. Journal of Consumer Research. 2010; 37(2):197–206.

68. Link BG, Yang LH, Phelan JC, Collins PY. Measuring mental illness stigma. Schizophrenia Bulletin. 2004; 30(3):511–41. https://doi.org/10.1093/oxfordjournals.schbul.a007098 PMID: 15631243

69. Kvaale EP, Gottdiener WH, Haslam N. Biogenetic explanations and stigma: a meta-analytic review of associations among laypeople. Soc Sci Med. 2013; 96:95–103. https://doi.org/10.1016/j.socscimed.2013.07.017 PMID: 24034956

70. MacKinnon DP. Introduction to statistical mediation analysis. Mahwah: Erlbaum; 2008.

71. Kenny DA. Mediation 2018. Available from: http://www.davidakenny.net/cm/mediate.htm

72. Running the analysis using the full serial mediation models produces results consistent with the simple mediation models: Significant indirect effects are found on all four outcome variables, though as expected, the unique effects contributed by the more direct route through a single mediator are slightly larger than the longer route through both mediators.

73. Schoemann AM, Boulton AJ, Short SD. Determining power and sample size for simple and complex mediation models. Social Psychological and Personality Science. 2017; 8(4):379–86.