Neuroethics (2024) 17:1
https://doi.org/10.1007/s12152-023-09535-0

ORIGINAL PAPER

How to Advance the Debate on the Criminal Responsibility of Antisocial Offenders

Marko Jurjako · Luca Malatesti · Inti A. Brazil

Received: 2 November 2022 / Accepted: 25 September 2023 / Published online: 21 October 2023

Abstract Should offenders with psychopathy or those exhibiting extreme forms of antisocial behaviour be considered criminally responsible? The current debate seems to have reached a stalemate. Several scholars have argued that neuropsychological data on individuals with psychopathy might be relevant for determining their criminal responsibility. However, relying on such data has not produced a consensus among legal scholars and philosophers on whether individuals with psychopathy should be excused from responsibility. We offer a diagnosis about why this debate has reached a standoff. We argue that part of the problem is that psychopathy, being a syndrome-based category, is too heterogeneous and thus offers low prospects for being integrated with neuropsychological data that might support significant conclusions about the criminal responsibility of individuals with psychopathy. Moreover, the construct of psychopathy was not originally devised to discriminate criminally accountable from unaccountable antisocial individuals. To overcome these difficulties and advance the debate, we extend a currently less discussed theoretical framework for bridging biopsychology and the law that focuses on neuropsychological constructs, domains, and processes that directly measure capacities of offenders that are relevant for criminal responsibility, without necessarily relying on mediating syndrome-based constructs. The novelty of our contribution is that the biopsychological bases of exculpation can be further developed and used for determining more fine-grained categorisations of antisocial personality types. We show how this framework provides guidelines for interdisciplinary research that can significantly advance our understanding of the preconditions for criminal responsibility and help the legal practice of ascribing or withholding it.

Keywords Antisocial personality disorder · Biopsychological categorisation · Criminal responsibility · Forensic philosophy · Insanity defence · Psychopathy

M. Jurjako · L. Malatesti (✉)
Department of Philosophy and Division of Cognitive Sciences, Faculty of Humanities and Social Sciences, University of Rijeka, Sveučilišna Avenija 4, 51000 Rijeka, Croatia
e-mail: lmalatesti@ffri.uniri.hr
URL: https://portal.uniri.hr/Portfolio/1541; https://cogsci.uniri.hr/people/luca-malatesti

M. Jurjako
e-mail: mjurjako@ffri.uniri.hr

I. A. Brazil
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
e-mail: inti.brazil@donders.ru.nl
URL: https://www.ru.nl/en/people/brazil-I; https://www.intibrazil.com

I. A. Brazil
Forensic Psychiatric Centre Pompestichting, Division Diagnostics Research and Education, Nijmegen, The Netherlands

Introduction

The determination of accountability has been one of the fundamental and challenging issues in forensic psychiatry. The current discussion on whether individuals with psychopathy are fully responsible for their criminal offences offers a significant example of the challenge. This debate highlights the difficulties faced by the project of individuating a group of unaccountable individuals amongst those with severe antisocial personality types [1–3]. Psychopathy is a personality construct typically characterized as involving a deceitful and manipulative interpersonal style, emotional shallowness as exhibited in lack of empathy and guilt, which are strongly associated with antisocial and criminal behaviour [4]. Advancements in the study of the brain and of the psychological, and behavioural correlates of psychopathy have provided novel frameworks that can further the discussion [5, 6]. Nonetheless, thus far, the investigation of whether offenders with psychopathy or other individuals with severe antisocial personality characteristics should be legally exculpated for their misdeeds has run into many obstacles (see, e.g., [7–9]).

There are two main limitations in the ongoing debates. First, these discussions concern categories of antisocial conditions that are based on classification systems, such as the Diagnostic Statistical Manual of Mental Disorders [10]. These classifications offer diagnoses that group individuals based on behavioural and psychological clusters of signs and symptoms for which it is often unclear how they can be effectively integrated with the available legally relevant neuropsychological data and paradigms (e.g., [11, 12]). Second, and more importantly, such categories are not devised for determining criminal responsibility. Thus, we should not expect that empirical results about them would be directly relevant for distinguishing criminally accountable from unaccountable individuals.

To solve these problems, we propose a biopsychology-informed framework aimed at individuating exculpable individuals with antisocial personality profiles. By "biopsychological" we refer to approaches that integrate psychological explanations with subpersonal models and explanations offered by cognitive (neuro)science. In psychiatry, biopsychological research strives to establish valid measures of disorders by combining psychological data with evidence from the genetic, neural, cognitive, and affective systems underlying psychiatric conditions, with the aim of creating new or refining existing psychiatric categorizations (see e.g., [11]). Our aim is to offer programmatic suggestions on how biological, psychological, and behavioural data may be integrated and be used for refining categories that delineate the unaccountable from accountable members of antisocial populations. The overarching goal is to provide a framework for a more effective use of available neuroscientific and psychological data to fuel a novel research agenda at the intersection of law, psychology, and neuroscience.

In this paper, we proceed as follows. In Sect. "Offenders with psychopathy and the insanity defence: an inconclusive debate", we show how the debate on the criminal responsibility of offenders with psychopathy has encountered difficulties that prevent progress. In Sect. "A biopsychology-informed categorisation for exculpation", to solve these difficulties, based on a proposal by Buckholtz, Reyna, and Slobogin [13], we develop a framework for a categorisation that is based on the biopsychological data that should track differences between types of antisocial individuals that are significant for legal exculpation. In the last section, we use the construct of inhibition to illustrate and motivate our framework for the study of the responsibility of antisocial populations. Many legislations recognise the capacity of control as a prerequisite for responsibility and severe deficit in inhibition might compromise such capacity for control. We argue that research on reward processing in individuals with psychopathy, which is pertinent to studying inhibition in them, clearly support the plausibility and fruitfulness of the suggested framework. This evidence suggests that we could individuate within the class of offenders with psychopathy a subclass of those who can be excused based on specific impairments that hinder their ability to exercise control.

Offenders with Psychopathy and the Insanity Defence: An Inconclusive Debate

The goal of the insanity defence is to distinguish between individuals who are responsible from those who are not, and thus do not deserve to be punished for their misdeeds [14]. Across legal systems, the capacities necessary for being criminally responsible are divided into epistemic and volitional capacities

[15]. Epistemic capacities include having a good grasp of the reality of the situation in which a person acts and the legal and/or moral norms that pertain to this action. For instance, if a person commits a crime under a persecutory delusion, she will typically not satisfy the epistemic requirement for responsibility. Volitional capacities concern the ability to act in accordance with the knowledge of the relevant legal and/or moral norms. For instance, if a person knows the relevant legal norms that pertain to a situation (i.e., has sufficient epistemic capacities) but commits a crime due to an irresistible impulse that prevents them from acting in accordance with the relevant legal norms (i.e., lacks sufficient volitional capacities), they will be deemed not accountable for this crime.

Stating the capacities that are required for being held criminally responsible in such abstract terms leaves crucial questions unanswered. It does not offer a sufficiently informative description of the psychological capacities that are preconditions for responsibility and under what circumstances their impairment implies exculpation. For instance, a drunk person may commit a crime while their responsibility-conferring capacities are undermined. However, if they were voluntarily intoxicated, then, despite the impairment in the responsibility-conferring capacities, typically they would be held responsible for the offending act. This is because the relevant incapacities must be beyond the control of the agent for them to grant an excuse.

Focusing on the clinical notion of mental disorder might provide a good starting point for tackling this issue. The presence of a mental disorder usually is a necessary condition for being deemed legally insane or for having diminished accountability. This is because mental disorders are typically understood as conditions that happen to a person and are therefore outside of their control. However, being mentally disordered is not sufficient for judging someone unaccountable for a committed crime. For instance, an offender may be diagnosed with schizophrenia but not exhibit any signs of psychosis at the time of committing a crime. Thus, there would not be any justification for exculpating this person, despite their mental illness.

The category of mental disorder is only relevant when a person committed a crime and the fact that they suffer from a mental disorder played a causal role in their committing it. But even then, not all diagnoses of mental disorder will be sufficient for being considered unaccountable. For instance, hoarding was introduced as a mental disorder in the DSM-5 [10]. But it is far from clear that if a person with a hoarding disorder intentionally commits a crime, and their hoarding behaviour played a causal role in leading to the unlawful action, any court would find them less accountable for the crime. Thus, we should specify a category of individuals whose disorder would be more relevant for exculpating them.

Here, discussions between legal experts tend to fall back on mental disorders involving psychosis and delusions. Most jurisdictions that include a form of insanity or diminished responsibility defences, consider crimes executed under a delusion or a psychotic episode as deserving exculpation or diminished culpability [16, 17]. Indeed, when someone performs a criminal act during a psychotic episode and does not know the meaning of their action, or their reasons for doing it are based on a wildly inaccurate representation of the actual events, this presents a paradigmatic example of a person that should not be held entirely accountable [18]. However, crimes committed under a psychotic delusion capture only cases that violate the cognitive requirement of legal responsibility. Many think that mental disorders can cause diminished responsibility in more subtle ways that do not involve clearly wild misrepresentations of physical reality [5, 19–21]. These ways may depend on more specific disruptions in the understanding of the shared morality, or more general problems in impulsivity and self-control that are thought to underpin pervasive and violent antisocial behaviour [5].

In recent legal discussions, there was some hope that the categories of severe antisocial personality or psychopathy might play this role in delineating at least one class of legally insane individuals [3]. Indeed, antisocial personality and psychopathy are typically defined by reference to pervasive and persistent immoral and criminal behaviour [22, 23]. For instance, typical signs and symptoms of antisocial personality disorder (ASPD) and psychopathy include persistent criminal behaviour, disregard for right and wrong, hostility, aggression, and violence [10]. Although psychopathy should be distinguished from ASPD, because it puts greater emphasis on maladaptive emotional and interpersonal styles that include lack of empathy, guiltlessness, callousness, and pathological lying [24], this construct involves also

antisociality that includes criminal versatility, pervasive violations of other people's rights, and criminal recidivism [4].

Some researchers even maintain that psychopathy can be used as a template for a general theory of crime, due to the potential of the psychopathy construct to capture maladaptive personality traits, inadequate interpersonal style, and their manifestations in antisocial behaviour. Specifically, Matt DeLisi [25] argues that psychopathy provides a unified theory of crime. According to him, scientific research on psychopathy represents our best attempt at providing a unified theory of antisocial behaviour, because psychopathy can be used as a categorical and a dimensional construct that captures personality traits and motivational profiles that are strongly associated with antisocial behaviour [4]. As such, they provide scientifically solid foundations for studying various aspects of antisocial personality and behaviours across populations and life stages [26]. Moreover, psychopathic personality traits are thought to provide well-researched and promising new avenues for developing neurobiological explanations of the associated antisocial behaviours and propensity for committing crime [24].

However, even in the case of psychopathy, there is uncertainty about whether this construct can be used for delineating an exact subgroup of the criminally unaccountable. First, there is uncertainty about whether psychopathy should be considered as a mental disorder. For instance, amongst those who think that mental disorders should be underpinned by biological dysfunctions, there is debate whether it is a disorder (e.g., [27, 28]) or an unusual but adaptive life strategy (e.g., [29–31]. Thus, it is not clear whether psychopathy would satisfy the mental disorder clause of the insanity defence.

Second, it is unlikely that psychopathy falls under the epistemic clause for the insanity defence. This clause is usually captured by the McNaughton rule that states that a person is not guilty of a crime if, due to a disease of the mind, the person did not know that what they were doing is wrong. There is some controversy about how to interpret the concept of knowing that an action is wrong [32], pp. 14–19). It is being debated whether the wrongness includes knowledge about the prevailing legal norms or if the offender should also possess a deeper moral understanding why the norm is in place [33]. If this issue is not settled, it is not clear how to treat individuals with psychopathy with respect to this clause. For instance, some research has shown that individuals with psychopathy may have problems distinguishing moral norms from other types of norms when it comes to justifying why other people should not be harmed (e.g., [34]). But there seems to be an agreement that individuals with psychopathy are aware that their behaviour often violates legal norms [5]. To complicate the situation even further, newer research indicates that individuals with psychopathy possess minimal moral and legal understanding that their antisocial behaviour violates prevailing moral norms [35, 36], for review, see [37].[1]

Third, it is uncertain whether psychopathy would satisfy the volitional component of the insanity defence. This clause is usually captured by the irresistible impulse test. The claim is that if a person cannot control their action due to severe problems in inhibiting behavioural impulses, they should not be held responsible for it. There are well-known problems with this criterion: it is often unclear how to distinguish an irresistible impulse from an impulse that was not resisted (e.g., [20, 40]). This problem is clearly exhibited in the case of psychopathy. Their crimes are often a consequence of calculated behaviours that are aimed at achieving their ends by any means available. Thus, based on behavioural criteria, their antisocial behaviour seems to be intentional and instrumentally rational [41, 42]. On the other hand, psychopathy is characterized by impulsive and irresponsible behaviour with apparent problems in creating and maintaining feasible long-term plans [24, 43, 44]. These behavioural characteristics suggest that individuals with psychopathy may have problems in controlling their behaviour.

To address the difficulties associated with determining the presence of the epistemic and volitional components of the insanity defence, it seems prudent to go beyond strictly behavioural and psychological criteria. Incorporating information on the biopsychological mechanisms that contribute to antisocial behaviour would offer more compelling evidence

[1] This issue is further complicated by studies suggesting that there is no unified cognitive capacity for delivering moral judgments (for discussion, see [38]). This might also explain why empirical research failed to provide sufficient evidence for general deficits in moral understanding associated with the construct of psychopathy [39].

and a more reliable perspective [13]. Indeed, based on the biopsychological research of emotional, interpersonal, and cognitive abnormalities associated with psychopathy, some have argued that psychopathy should be considered as an exculpatory condition in criminal trials (for a critical review, see [45]). For instance, Glenn et al. [5, 19] argue that, due to problems in learning from punishment, individuals with psychopathy typically cannot translate their abstract knowledge of legal and potentially moral norms into reasons for performing or abstaining from certain actions. For instance, some studies suggest that individuals with psychopathy have problems recognizing emotional expressions, such as fear or sadness, when observing other people's faces (see, e.g., [46, 47]). These studies suggest impairments in brain regions that process affective information. Indeed, some studies suggest that psychopathy is associated with functional abnormalities in parts of the amygdala, right lateral prefrontal cortex, and dorsomedial prefrontal cortex [48], cf. [49]. The amygdala is commonly seen as a key structure for emotional processing and affect-based learning, lateral parts of the prefrontal cortex are associated with action execution and ability to inhibit prepotent impulses, while the dorsomedial prefrontal cortex is associated with social cognition and empathy.

Thus, diminished, or aberrant functioning in certain brain areas, and the associated behavioural evidence, could explain the disposition in these individuals to behave unempathically, disregard other people's rights, and exhibit impulsive and aggressive behaviour [50]. Some argue that these deficits might explain why individuals with psychopathy, although have abstract knowledge about what is right and wrong, cannot use it to adapt their behaviour accordingly (for similar arguments, see [21, 51].

However, even in the case of psychopathy, the biopsychological data do not give clear verdicts on the presence of exculpatory deficits. The problem is that many of the studies that indicate functional and neural abnormalities associated with psychopathy are not replicated or indicate subtle impairments for which it is not clear that they should be considered as entirely exculpatory [8, 9]. For instance, studies suggesting that individuals with psychopathy have impairments in fear recognition are contradicted by other studies showing that individuals with psychopathy recognize emotions similarly to control groups (see, e.g., [52]). This incongruence in results might be explained by the fact that, without priming, individuals with psychopathy in such studies tend to focus less on the eye region of human faces, which provides cues about whether someone is fearful or sad [53].

Similarly, the results of the neuroimaging studies indicating functional and structural abnormalities in the amygdala and areas in the prefrontal cortex seem to be modulated by attention, instruction, and the level of intentional engagement with the task (see, e.g., [54, 55]). Such results prompted some researchers to propose the motivational framework for psychopathy, according to which neural, cognitive, and behavioural abnormalities associated with psychopathy are a downstream consequence of their motivational dispositions [56]. In other words, the cognitive and neural peculiarities associated with psychopathy might not be symptoms of an underlying impaired mechanism. Instead, they might be consequences of the differences in how they evaluate rewards and punishments and what they find motivationally appealing (see, also [43, 44]). If that really is the case, then we would have a reason to think that individuals with psychopathy do not suffer from impairments in the legally relevant cognitive and affective capacities. Instead, we would have a reason to believe that they are simply motivated to behave badly and that these motivational structures are reflected in their underlying neural and cognitive mechanisms in an interactive loop. On the one hand, circumstantial neural and cognitive peculiarities would be a downstream consequence of motivations. On the other hand, such peculiarities would provide enabling conditions for such motivations. In other words, we would have a reason to believe that the biopsychological makeup of individuals with psychopathy reflects their bad character, rather than them being insane [57].

However, even setting aside the motivational hypothesis, the inconsistency across biopsychological studies of psychopathy suggests deeper methodological issues in the scientific study of psychopathy. Some of these issues will likely reflect the more general replicability crisis in psychological research [49]. But, more importantly for our line of reasoning, the inconsistencies across biopsychological studies of psychopathy reflect the heterogeneity of the construct itself [26, 58–62]. Like other forensic and psychiatric constructs, the currently dominant approach in forensic psychiatry is to view psychopathy as a collection

of symptoms that form a syndrome. However, when measuring psychopathy, none of the symptoms represent a necessary feature for assigning psychopathy as a diagnostic label. Notable examples are lack of empathy, glibness, and guiltlessness, traits that are typically associated with psychopathic personality. However, when using instruments such as the PCL-R [4], which is taken to be the gold standard for measuring psychopathy in forensic settings [63], getting high scores on such traits is not necessary for receiving the "psychopathy diagnosis". The PCL-R has 20 items that can be scored 0, 1, or 2 (zero indicating that the trait is not present, 2 indicating that the trait is fully present). The maximal score is 40, and the typical cut-off score for assigning the "psychopathy" label is 30 (North America) or 25 (Europe) points. This means that a person who scores zero on lack of empathy, glibness, and lack of remorse/guilt, can still exceed the cut-off score and be diagnosed with psychopathy by scoring sufficiently high on other traits (for discussion, see [62, 64]).

If the construct of psychopathy can designate such a diverse constellation of personality and behavioural traits, it might come as no surprise that we cannot easily find common etiological and biopsychological mechanisms underlying psychopathy. Indeed, some conceptualizations of psychopathy presuppose that this construct involves diverse aetiologies (i.e., acquired vs. inborn psychopathy), others presuppose differential associations with external criteria (e.g., low and high anxious psychopathy), or different life histories (i.e., successful/community-dwelling vs. unsuccessful/criminal psychopathy), and so on (for review, see [59]). Moreover, different conceptualizations of psychopathy are often accompanied by different measuring instruments. For instance, the triarchic model of psychopathy explicitly presupposes that psychopathy is characterized by a constellation of three distinct and loosely correlated phenotypic constructs involving disinhibition, boldness, and meanness [65].

The proliferation of various measuring instruments can further complicate the matter because some studies indicate that incompatible scientific results can be obtained depending on which instrument for measuring psychopathy is used. Most notably, Baskin-Sommers et al. [66] examined associations between executive function scores and psychopathic traits in a population of prisoners. To measure psychopathic traits, they used the PCL-R and the Multidimensional Personality Questionnaire-Brief (MPQ-B). Their results indicated, among other things, that the same population of prisoners showed positive associations with executive function scores when psychopathy was measured with the MPQ-B, but no or even negative associations when psychopathy was measured with the PCL-R. Therefore, relying solely on psychopathy to investigate the accountability of individuals who manifest severe forms of antisocial behaviour is insufficient and not justifiable [7, 8].

In the next section, we highlight the main obstacles for the lack of progress in the debate concerning the criminal responsibility of offenders with psychopathy and offer a theoretical framework that might be used to overcome them.

A Biopsychology-Informed Categorisation for Exculpation

The current debate on the criminal responsibility of offenders with psychopathy has many shortcomings. Several studies on the ethical or legal consequences of scientific research on psychopathy might be affected by simplistic readings of the data, especially due to a lack of appreciation of the statistical subtleties involved in the research, unwarranted extrapolations, and an uncritical acceptance of the authority of specialists (for discussion, see [45]). We are, however, more concerned with two shortcomings in the debate that have a more theoretical nature. First, most of the discussions rely on syndrome-based measures of the construct of psychopathy that some find problematic on general grounds. Specifically, it has been argued that that these measures are not sufficiently specific or sensitive, and often lack biological validity [26, 61]. Second, the contemporary construct of psychopathy that underlies these measures is not devised for delineating criminally responsible from not responsible antisocial individuals [4]. Thus, it should not come as a surprise that not all individuals labelled as psychopaths will belong to the group of people with diminished responsibility. Let us start with the first shortcoming.

Towards More Homogenous Categories

Most diagnostic categories of mental disorders, including the widely used classification system such as the DSM-5 [10], predominantly rely on syndrome-based approaches for classification. Such classifications are mainly based on signs and symptoms of mental disorders that form statistical clusters, and often do not take into consideration the underlying causes. Such diagnostic systems show several problems; they are overly heterogeneous (i.e., mixed groups of people and symptoms clustered under the same label), have delivered relatively small advances in treatment, tend to have low validity, include comorbidity, with low prospects of integration with neuroscience, genetics, and neuropsychological paradigms [67]. These general problems with syndrome-based categories can help explain the lack of clear and consistent results in biopsychological studies with individuals with psychopathy that can be used for determining the presence of cognitive-affective and behavioural disabilities relevant for the notion of criminal responsibility [26, 60, 61].

To overcome these issues underlying mental health research, multidisciplinary initiatives, such as the Research Domain Criteria (RDoC), have been proposed [11]. Such initiatives tend to propose rethinking the standard categories of mental disorders. This involves rebuilding or refining them based on available knowledge of the biopsychological mechanisms that underpin psychopathology. Similarly, in the case of the scientific study of psychopathy and antisocial behaviour, an influential suggestion has been to build data-derived profiles based on individual differences in biological and psychological dimensions, in a way that allows such functions to be linked to, for example, criminogenic variables [26]. Such an approach does not need to be reductive, because, similarly to the biopsychosocial model in medicine [68], these dimensions that enter in the categorisation, in the ideal case, would include genetic, neurobiological, behavioural, self-report, social, and cultural data, without a presupposition that any of these levels have an a priori explanatory priority [61]. The procedure can be carried forward even to the point of creating a unique signature or 'biopsychological fingerprint', that is specific to an individual and would enable us to differentiate the person from other people based on such biopsychological variables [26]. In this way, by utilizing more biopsychological variables we would be able to delineate more homogeneous categories of individuals that predict clinically relevant outcomes.

Here we propose that a similar approach can be extended to the forensic domain, and that the heterogeneity in the syndrome-based classifications can be overcome by using more integrated biopsychological data when devising our forensic classifications.[2] When building profiles of antisocial individuals, in addition to higher up psychological variables, we should also rely on the available bottom-up (neuro)biological data to clearly delineate the offenders with similar neuropsychological impairments. However, increasing the homogeneity of the categories of antisocial individuals, will not directly address the second problem that afflicts the current debate on the responsibility of offenders with psychopathy. This difficulty stems from the fact that these categories are not meant to distinguish accountable from not accountable offenders. We discuss this issue next.

Mapping Criminal Responsibility Onto Biopsychological Constructs

Categories in standard psychiatric manuals, such as the DSM-5, including those pertaining to extreme forms of antisocial behaviour and psychopathy, are not devised to distinguish the legally sane from the insane. This is not surprising because the general aims of forensic and clinical approaches to psychopathology are different. The primary objective of forensic approaches to psychopathology is to give opinions and assessments concerning issues that are relevant to the legal system. This includes, for instance, determining a person's accountability for a crime, their competency to stand trial, or to predict the likelihood of recidivism. In contrast, clinical mental health experts aim to provide care and treatment to their patients. This involves diagnosing mental illness,

[2] In this respect, our approach can be characterized as sharing the general principles of RDoC and applying them to the legal domain. However, it is important to note that we only adopt an "RDoC-style" approach, as we want to remain uncommitted about the specifics of the RDoC model in clinical practice. In fact, there are several RDoC-style proposals in the psychological and psychiatric literature (e.g., biocognitive fingerprinting, computational psychiatry, etc.) that, while sharing the fundamental principles of RDoC, implement them in varying manners (see, e.g., [26, 61]).

selecting appropriate treatment options, and monitoring the effectiveness of the chosen treatments.

The differences in the aims will be reflected in the categories they use. For instance, forensic psychologists may focus on categories that are aimed at reducing risk or improving functioning in a legal or institutional setting, such as anger management or substance abuse treatment, while clinical psychologists may focus on treatments aimed at improving overall well-being, such as psychotherapy. More pertinent to our context, forensic experts will focus on categories that are specific to decision-making capacities of individuals that underpin legal responsibility, that explain and predict violent offending, recidivism, etc. These categories may not play the same role in clinical practice where they may be just another symptom across different illness categories that are more pertinent to treatment.[3] For instance, deficits in inhibitory control capacities are relevant in many forensic contexts. They are also associated, to varying degrees, with clinical categories such as ADHD, PTSD, and autism spectrum disorder. However, it is important to note that within clinical practice, these deficits in inhibitory control would be addressed as part of a broader treatment approach targeting the underlying clinical condition. In the forensic context, on the other hand, the focus would be more on risk assessment and management strategies related to decision-making capacities, which can cut across or even may be unrelated to specific clinical conditions.

Like other clinical categories, psychopathy is primarily used for detecting socially maladjusted individuals, with the goal of finding effective interventions for reducing the associated socially maladaptive behaviour [26, 69]. However, the category of the 'criminally excusable' purports to track those individuals who cannot be held accountable for their misdeeds. Given their divergent purposes, there is no a priori reason for expecting that clinical or research views on antisocial behaviour and psychopathy will correspond to the category of criminally unaccountable [70, 71].

To solve these problems, we suggest extending the use of biopsychology-informed categorisation in forensic research, with an emphasis on the data that are clearly relevant for the issue of criminal responsibility. The key step is to start thinking about how legally relevant capacities could be detached from the standard views of mental disorder and more directly interfaced and assessed using neuropsychological notions and paradigms. This approach can provide valuable biopsychological data to help categorize individuals who cannot be held legally responsible (see Fig. 1).

To clarify how the legally relevant data can inform the biopsychological categorisation of unaccountable individuals with antisocial disorders, we consider a proposal by Buckholtz, Reyna, and Slobogin [13]; see also [72]. The core of their proposal is that legally relevant capacities, such as those involved in criminal responsibility, should be operationalised in a way that can be measured and related to processes studied by science. For example, they take the notion of self-control as a target construct, arguing that it has a significant role in several legal doctrines and practices. Moreover, *domains* that constitute distinct subdivisions of the target constructs need to be individuated. Finally, domains are subserved by mental processes that can be measured by means of specific tasks or paradigms that involve the experimental observation of behaviours in certain specific conditions.[4]

[3] The difference may also be reflected in some of the desiderata for forensic and clinical categories. Generally, we would want clinical and forensic constructs that are valid and reliable. However, in certain contexts, the trade-off between their sensitivity and specificity might differ between clinical and forensic cases. Sensitivity refers to the ability of a construct to accurately identify individuals who have a particular condition or characteristic, while specificity refers to the ability of a construct to accurately identify individuals who do not have the condition or characteristic. In a forensic setting, specificity might be more important, as justice requires that it is better for a guilty person to go free than for an innocent person to be punished. If legal constructs are not specific enough, they run the risk of including innocent individuals within their scope. For instance, if a forensic DNA test used to identify a suspect from a crime scene is not specific enough and produces a false positive result, an innocent person might be falsely accused and wrongly convicted of a crime they did not commit. In the clinical case, sensitivity might be more important, as the primary goal of the clinical system is to accurately identify and diagnose individuals who may need treatment. For instance, if a mammogram lacks sensitivity, it can potentially result in delayed diagnosis and treatment of breast cancer, which may cause a patient to miss the chance for early detection and treatment, significantly decreasing their chances of survival.

[4] There are intriguing philosophical questions regarding the understanding of relations between constructs, domains, and measurable processes. For instance, we will later discuss inhibition as a significant domain of self-control. This raises the question of whether inhibition and its associated mental pro-

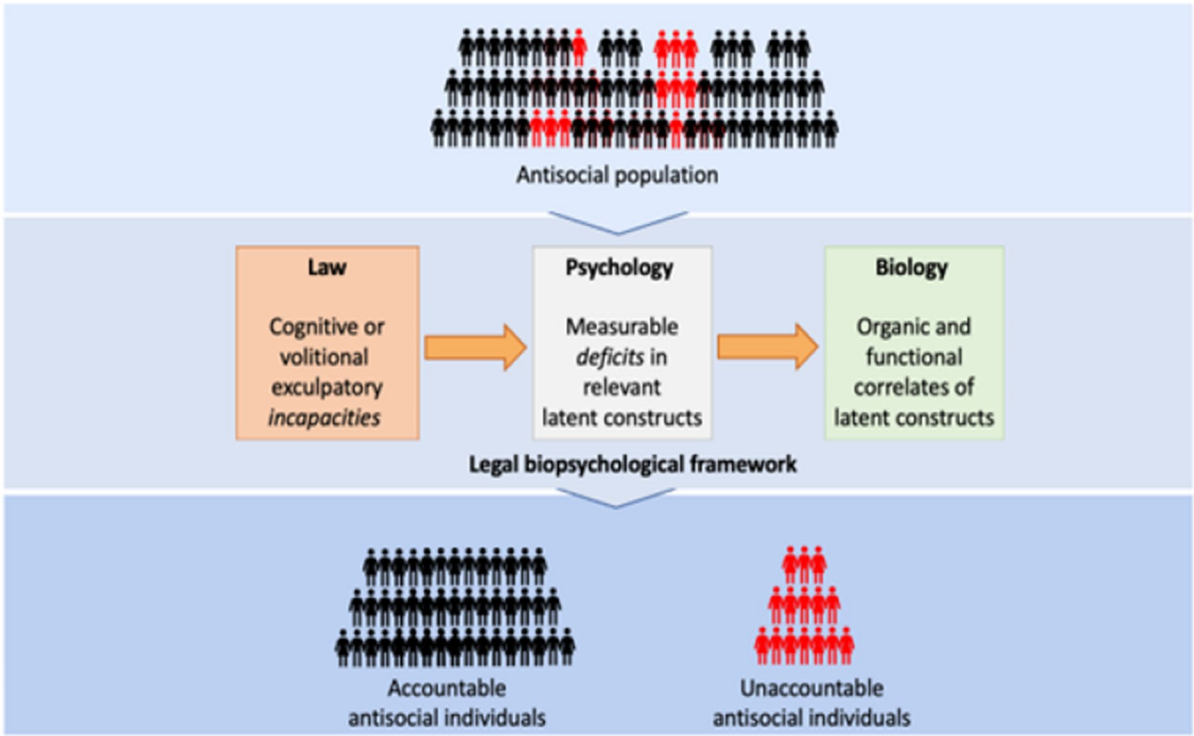

Fig. 1 Framework for the categorisation of unaccountable antisocial individuals. In this approach the target legal notions are first defined. Next, the psychological constructs that can be mapped onto the legal notions are identified, making it possible to operationalize the legal notions in quantifiable ways. Insights from (neuro)biological research are used to anchor the psychological constructs into biologically plausible mechanisms. The multi-domain information can then be used to help determine whether certain individuals should be considered legally accountable

The characterisation of the relevant target constructs is a complex matter that requires interdisciplinary research. As mentioned earlier, the Law characterises the psychological prerequisites for exculpation in very generic terms, such as, for instance, lack of an understanding of the nature of an action or lack of control. Such characterisations do not tell us how to measure the presence of these incapacities and how they might relate to behavioural, psychological, or neural evidence. In fact, devising such operationalisations of exculpatory capacities, far from being simply a scientific task, requires addressing several interrelated normative and scientific problems (e.g., [72, 75, 76]). For instance, in different legal regimes there are different views on how to understand the requirement that for a person to be criminally accountable they should understand the nature of their action. As mentioned earlier when discussing psychopathy, while some consider an essential part of this understanding the capacity to appreciate the moral reasons, embedded in the Law, that prohibit certain course of action, others support the view that such understanding involves the simpler requirement that the person is aware that the Law prohibits the action (for discussion, see [33]). Similarly, there are disputes on whether and how a notion of capacity for control

Footnote 4 (continued)
cesses provide causal underpinnings for self-control or if they are constitutively related (for empirically grounded discussion, see [73]). Answering these questions requires addressing broader issues in the philosophy of psychology/mind, particularly concerning levels of psychological explanation and their ontology (for contrasting perspectives in the legal context, see [72, 74]). In our discussion, we need not commit to the precise metaphysics of these relations. We will minimally presuppose that the relevant domains and associated mental processes (to be determined collectively by philosophers, legal scholars, and mental health practitioners) should provide at least necessary conditions for the proper functioning of legally relevant mental capacities.

should enter in the formulation of criminal responsibility [77]. There are also issues concerning how much someone must be incapacitated, and what is the relevant group of reference, to be declared, for instance, criminally unaccountable [13].

So far, we have offered, in general terms, a framework for the (re)categorization of antisocial individuals for legal exculpation. In the next section we motivate it by showing how it might work in practice. We consider the operationalisation of the legal notion of control by means of the psychological construct of inhibition (for other examples, see [13]). Focussing on deficits in reward processes that undermine inhibition, we show how within the class of individuals with psychopathy we could determine whether there is a subclass that should be exculpated due to lack or diminished capacities for control. Given the programmatic nature of our framework, this illustration is also an opportunity to highlight some of its methodological features and challenges that it must address.

A Case Study: Inhibition as a Domain Within the Construct of Control

To illustrate and motivate our framework for the study of the responsibility of antisocial populations, we focus on the construct of inhibition that is relevant for the legal notion of control. It is important to note that inhibition is not a unitary construct [78].[5] For the present purposes, we understand inhibition as a higher-order behavioural capacity that enables people to refrain from acting based on certain desires and impulses. This ability is typically exercised in situations where there is a conflict between desires (e.g., eating another cake or resume with the diet plan) or situational demands require delay of gratification (e.g., not taking a smaller reward now to get a larger reward later). This capacity is also especially important for social interactions because it constrains people from acting on their immediate impulses (e.g., taking things from other people just because we like them), and promotes socially adaptive behaviours (e.g., restraining us from injuring someone when we are angry). This higher-order capacity can be implemented via various lower-level capacities and behavioural strategies. For instance, the capacity of inhibition is typically required when one seeks to defer gratification, such as refraining from consuming excessive amounts of sweet food while on a diet. However, this objective can be accomplished through different lower-level processes or strategies, including abilities such as contemplating the additional calories present in a chocolate cake, diverting attention away from purchasing cupcakes while inside a bakery, making a verbal commitment to abstain from eating dessert, and so forth (see Fig. 1 in [73].

It seems plausible that deficits in the capacity to deploy inhibition (i.e., disinhibition), might affect the ascription of legally relevant control. In fact, disinhibition is directly implicated in psychopathy research, has legal ramifications, and can be parsed into biopsychological factors that can be measured in terms of the available neuropsychological paradigms [65, 79]. To show how this type of neuropsychological research into inhibition can be efficiently utilized within our framework, consider the following problem besetting the psychopathy research.

As mentioned, at some level of behavioural and psychological functioning, disinhibition is usually considered as an important trait of psychopathy (e.g., [4, 65]). Yet, this behavioural characterization leaves unresolved whether disinhibition associated with psychopathy or antisocial behaviour should diminish or exculpate from responsibility. This is because studies on psychopathy often fail to address the question of whether the disinhibition observed in individuals with psychopathy stems from a lack of inhibitory capacity or simply from their lack of motivation to refrain from inhibiting antisocial behaviours. In other (more metaphorical) words, it remains unclear whether individuals with psychopathy lack effective behavioural "brakes" to reduce their antisocial impulses, or if they possess adequate brakes but do not exert control as effectively as others [26]. Focusing on the neuropsychological studies of disinhibition provides a useful approach to properly (re)categorize and distinguish individuals with psychopathic tendencies and other subgroups of antisocial populations that might lack behavioural "brakes" from those who have them but do not use them (see, also [20, 79]).

[5] It should also be noted that there is some uncertainty about whether inhibition should be understood as a process or an outcome of goal-directed behaviour (for discussion, see [73]). We will remain neutral about these issues by construing attention as a capacity that might be subserved by different processes.

A central step towards the recategorization that we recommend is to individuate deficits in inhibition based on the relevant biopsychological factors. This means that we need to individuate measurable deficits in relevant latent constructs and their organic and functional correlates. Now, recent work suggests that the following cognitive and affective factors are associated with disinhibition: negative emotionality, executive functioning, and reward processing [79], cf. [73]. One important consideration is that, in turn, such factors consist of multiple components that can be investigated using different tools. Negative emotionality, for example, can be divided in component processes that can be measured with behavioural (e.g., by questionnaires and tasks for testing resilience to stress), physiological (e.g., measuring heart rate), and neuroimaging (e.g., the activity in the amygdala and areas of the prefrontal cortex) methods [79]. At the general level, negative emotionality involves disturbances associated with emotions and traits such as anger, fear, and distress. Disinhibition can be a consequence of a more intense or of a reduced experience of negative emotions. For instance, a person with schizophrenia can exhibit excessive anger, frustration, and distress, which may lead to uncontrolled outbursts of aggression and socially maladaptive behaviours. Psychopathy, on the other hand, is associated with reduced experiences of negative emotions. Such deficits might partly explain why such individuals are more prone to exhibit violent and aggressive behaviours. The contrast sketched here illustrates the challenge imposed by equifinality (i.e., the possibility of arriving at the same outcome through different paths or processes), and the importance of finding ways to better accommodate the latent internal correlates of observed behavioural disinhibition into judgements of responsibility [80].

Clearly, to implement our framework we are interested in relevant factors that impact inhibition in a way significant for the ascription of responsibility. In this regard, reduced negative emotionality associated with psychopathy seems less pertinent (cf. [21]). It has been argued that reduced experience of negative emotions will not necessarily affect our attributions of responsibility for a criminal act [81]. This might be because reduced negative emotionality does not directly affect the ability to act deliberately, as it is not conceptualized as a deficit in the mechanism for self-control itself. Thus, low negative emotionality alone cannot be used for distinguishing offenders who act based on bad motives and are responsible for their actions from those who cannot control their behaviour in the light of their motives and are not responsible.

In contrast, impairments in executive functioning may affect inhibition in a way that is more relevant for withdrawal of responsibility. Executive functions (EF) refer to higher-order cognitive abilities for planning and goal-directed behaviour. The components of EF include processes such as focusing of attention, inhibition of prepotent actions, and working memory. EF are implemented broadly in the prefrontal cortex, involving areas such as the dorsolateral prefrontal cortex, orbitofrontal cortex, and their functional connections with the mesocorticolimbic reward systems [72]. Problems with disinhibition can be associated with deficits in executive functions, since impaired EF can make it challenging for individuals to control impulsive and inappropriate behaviours [79]. Such individuals may have trouble in refraining from their actions, managing their impulses, or considering the potential consequences of their behaviour before acting. Given that EF provide a set of cognitive processes that underpin our ability for conscious self-control, some philosophers and legal scholars maintain that EF provide some of the crucial cognitive abilities that underpin the notion of criminal responsibility (for a seminal discussion, see [72]). Importantly, self-control is also influenced by motivational factors, with reward processing being one of the most studied ones.

Impairments in reward processing might be significant for establishing exculpation. Reward processing, sometimes referred to as “hot” executive function, is associated with the limbic areas of the brain, such as the amygdala, that regulate processing of emotional valence of information and are functionally connected to the areas in the prefrontal cortex that encode and update the expected value of rewards [43]. Severe deficits in these processes are plausibly associated with disinhibition, wherein they cause a failure to take into consideration the fact that a certain action will lead to adverse consequences (e.g., robbing a store leads to imprisonment) and thus to a failure to refrain from executing it. Furthermore, a person may be especially sensitive to a certain type of reward, e.g., money, that might lead to an overactive response from the reward system that interferes with the proper function of the more deliberate executive processes.

For the purposes of our discussion, it may be beneficial to narrow our focus from the broad construct of EF and place greater emphasis on reward processing. This is for several reasons. First, it has been argued that EF tends to disguise responsibility relevant capacities in terms of neurocognitive constructs that do not manifest greater utility than other more familiar psychological and behavioural constructs underpinning our notion of criminal responsibility [82]. Second, although EF dysfunction is associated with antisocial populations, standardized tests of EF show mixed results with psychopathy (for references, see [83]). In this regard, focusing on EF does not seem to provide a template for thinking about how to parse out the accountable and unaccountable groups within the antisocial populations associated with psychopathy. Third, reward processing has received less attention than EF in legal discussions, despite its significant role in underpinning inhibitory and thus responsibility relevant capacities (for instance, see [82, 84]). Fourth, there is extensive neuropsychological research on reward processing in psychopathy and extreme forms of antisocial behaviour [43, 44]. By considering reward processing, we can effectively show how biopsychological factors might be used to think more effectively about the domains and measurable processes that are relevant for distinguishing, within antisocial populations, responsible from non-responsible individuals.

To illustrate how reward processing studies can address the initial question about distinguishing between individuals high on psychopathic tendencies with and without inhibitory deficits, let us examine the typical measurement of reward processing. Reward processing is typically investigated via studies on associative learning. Associative learning includes classical Pavlovian learning (such as associating a bell ring with food) and different forms of operant learning (that include associations of own behaviour with rewards or punishments). Reversal learning tasks can assess the behavioural adaptation during associative learning. Such tasks usually have two phases. First, subjects are supposed to learn to associate one type of stimuli with a rewarding outcome and another type of stimuli with a punishing outcome. In the reversal phase of the task, the contingencies change so that the outcomes that were previously rewarded are punished and vice versa.

Some studies suggest that individuals with psychopathy have impairments in learning to respond to negative feedback and that they fail to notice changes in contingencies during reversal tasks [50]. Such a failure to learn from negative feedback and environmental changes in how punishing and rewarding stimuli are distributed, may (partly) explain why a person would have problems inhibiting impulsive behaviours. If a person is unable to learn from negative experiences and cannot learn the associations between rewards, punishments, and action outcomes, then they might fail to inhibit prepotent behavioural dispositions, including those that lead to antisocial outcomes. However, other studies failed to replicate these results, or they showed less extensive abnormalities in how individuals with psychopathy perform on associative learning tasks (for an overview, see [43]. This prompted Glimmerveen et al. [43] to test whether individuals with psychopathy perform differently from others as a function of how different rewards are subjectively valued. In what follows, we indicate how this type of studies can be used to further inform an integrative biopsychology-based (re)categorization of legally relevant deficits associated with psychopathy.

Glimmerveen et al. [44] relied on the assumption that individuals with elevated levels of psychopathy may not be sufficiently motivated to monitor contingencies and adapt their behaviour when subjectively uninteresting rewards (i.e., points) are used in a task (see, also [56]). Thus, they first determined the extent to which each of 25 different reinforcers were subjectively rewarding for each individual participant. The subjective rewards were identified in a previous study [85]. The reward items participants could choose from included "material rewards, food-related rewards, and rewards related to personal development (e.g., attending workshops)" [44, p. 48], see, also [82]. The participants were asked to rank the reinforcers, and the reinforcers with the lowest, median (or 'neutral'), and highest subjective values were incorporated as outcome stimuli in an associative learning task. Moreover, they also used electroencephalography (EEG) to measure an event-related potential (ERP) known as the feedback-related negativity (FRN). The magnitude of the FRN is known to vary as a function of how subjectively valuable the individual finds the reinforcers in such tasks [86, 87]. This enabled them to explore the possibility that associative learning deficits in psychopathy are underpinned by impairments in the coding of the value of different reinforcers.

Interestingly, the results of this study indicated that offenders with psychopathy mostly perform differently from healthy subjects in the condition where the "neutral reward" is used. This indicates that some of these offenders perform equally to healthy participants in the conditions with rewards that they experience as having high (and low) subjective value. The EEG results showed that the subjective value of rewards affected the ERPs equally in each group, ruling out the possibility that differences in the processing of outcome-related feedback can account for the learning effects. Moreover, when controlling for the effects of subjective rewards, the results indicate that the groups did not consistently differ on how they perform on associative learning tasks employed. This study might suggest that a subpopulation of individuals with psychopathy possesses the capacity to learn and adapt, but motivational and other contextual factors might impact the degree to which these abilities for learning will be exercised in real life circumstances. Alternatively, this study might also suggest that there are other antisocial individuals whose reward processing is disrupted by inefficient coding of value or more downstream deficits in translating the perceived value into motivations and motor commands that control behaviour (see [88]). Importantly for the context of our discussion, this type of study enables the development and refinement of parameters and variables that track relevant biocognitive processes that are associated with legally relevant capacities for control.

More generally, the discussion of the case of reward processing in individuals with psychopathy offers several insights for the elaboration of the framework we are proposing. First, empirical evidence and theoretical assumptions about the legally relevant constructs and associated domains might offer more precise and relevant tasks to gather relevant data for distinguishing between agents that possess and those who do not possess the legally relevant psychological capacities. In this regard, associative learning tasks can be used to operationalise and test an aspect of the psychological/behavioural construct of impulsivity and disinhibition (for other examples of tasks that might be used to test for impulse control and self-regulation, see [13]. However, just administering such a task may not be enough, because at the behavioural level a person being assessed might perform worse than an average healthy control, and thus, these behavioural results could be mistaken for evidence that the person is unable to inhibit prepotent behaviours. This leads us to a further lesson.

The second lesson is that our framework recommends relying on multiple explanatory domains and data types. For example, the problem of determining the level of impulsivity or relevant forms of disinhibition can be approached by relying on insights on the neural underpinnings of the motivational structure of the person (such as checking the patterns of the relevant ERPs) or using other approaches that incorporate multiple levels of explanation, in a way that each level is complementary to the others and yields unique insights [43, 44, 85]. In other words, without multi-domain data about the mechanisms underlying behavioural control, we would not be able to distinguish those individuals with psychopathy who suffer from responsibility diminishing or exculpatory level of disinhibition, from those who have the relevant capacity but are not sufficiently motivated to use it [56]. Thus, the relevance of (experimental) behavioural observations for deciding on the impulsivity of a person, or whether they suffer from other legally relevant incapacity, in the case of individuals with psychopathy will often be possible to properly decide only if we consider additional measures that account for their motivations for solving the task.

Our discussion of the case of reward processing and inhibition also shows how our framework makes perspicuous some fundamental challenges that need to be addressed when bringing scientific evidence about the mind, brain, and behaviour to bear into the practice of ascribing or withdrawing responsibility. Our framework, for instance, highlights issues of heterogeneity at multiple levels. This is clear in the case of inhibition, which as we have seen can be divided into legally significant constructs such as negative emotionality, reward processing, and executive functions. Analysing further these processes indicates that they are themselves heterogeneous and can have many causes and effects on a person's control capacities. Furthermore, these processes are not devised for the purposes of determining legal responsibility, so the concern might be that replacing one heterogeneous construct with others would only result in creating new problems instead of resolving the original one.

Given the programmatic nature of our proposal, here we can only sketch why these issues do not present insurmountable obstacles. As we have seen, exactly the advantage of our biopsychology-informed approach is that it directs attention towards the pertinent issues and

provides a framework for addressing them. Specifically, we have argued that in the case of inhibition not all processes underlying this construct, such as negative emotionality, will be equally relevant for determining whether a person should be exculpated. Moreover, even though processes such as reward processing and executive function were not created for legal purposes, we have argued that they are parsed at the appropriate level for identifying and categorizing individuals who possess the necessary psychological prerequisites for being held responsible. In this regard, adopting the integrative biopsychological approach should enable relatively easier identification and differentiation of the decision-making processes that have legal significance than solely relying on syndrome-based categories such as psychopathy.

Overall, the considerations about inhibition motivate and show in more detail our proposed framework for refining the categorization of individuals with psychopathy and extreme antisocial personality profiles. The aim of this approach is to integrate traditional, behavioural, and symptom-based views on severe antisocial personality and start thinking about testing individuals that we suspect may have significant disabilities in the capacities underpinning criminal responsibility. Regarding the abilities to control behaviour, this crucially includes relying on biopsychological data concerning the individual's performance across self-report measures, behavioural tasks, physiological and neuroimaging tests that measure the processes underpinning inhibitory capacities. In this way, we would avoid validity issues stemming from reliance on mostly syndrome-based categories and utilize more directly the neuroscientific data measuring the presence and functionality of capacities relevant for distinguishing between accountable and unaccountable offenders.[6]

[6] A reviewer has noted that our method may be relevant to other categories of mental disorders, which aligns with our RDoC-style approach for reconsidering traditional psychiatric categories in forensic sciences. For instance, the specific type of inhibitory control dysfunction that may excuse individuals from accountability could be present across various conventional categories, such as bipolar disorder, schizophrenia, and autism spectrum disorder, in addition to antisocial personality disorder. Although this has implications for our approach, exploring these other categories is beyond the scope of this paper. It is worth noting, though, that where these other categories are associated with severe forms of antisocial behaviour, our general approach can be straightforwardly applied to them as well (see, e.g., [89]).

Conclusion

In the paper, we have discussed how we could overcome the impasse regarding the problem of distinguishing antisocial individuals who are criminally accountable from those who are not. In recent decades, significant emphasis has been put on the idea that the construct of psychopathy could play this role. However, we have indicated that recent biopsychological and conceptual studies of psychopathy do not allow for clear verdicts on whether psychopathy could be used for this purpose in forensic settings. The main problem is that we still lack a good approach for aligning biopsychological knowledge with constructs used in legal settings.

To solve this problem, we have advanced reasons for implementing a framework for biopsychology-informed legal exculpation that would integrate neuroscientific and cognitive/behavioural insights with the law, using philosophical reasoning as a medium for mapping the different fields. Our programmatic proposal is also meant to highlight the necessity and fruitfulness of theoretical investigations that might fall under the heading of "forensic philosophy".

Funding This paper is an outcome of the project *Responding to antisocial personalities in a democratic society* (RAD), http://rad.ffri.hr/en/, that is funded by the *Croatian Science Foundation* (CSF – HRZZ) (PI: Luca Malatesti, grant IP-2018-01-3518).

Data Availability The data availability statement is not applicable.

Declarations

Originality and Exclusivity All authors have seen and agreed with the contents of the manuscript. They certify that the submission is original work and is not under review at any other publication.

Competing Interest The authors have no conflicts of interest to declare.

References

1. Kiehl, K.A., and W.P. Sinnott-Armstrong, eds. 2013. *Handbook on psychopathy and law*. Oxford University Press.
2. Malatesti, L., J. McMillan, and P. Šustar, eds. 2022. *Psychopathy: Its uses, validity and status*, vol. 27. Springer.

3. Malatesti, L., and J. McMillan, eds. 2010. *Responsibility and psychopathy: Interfacing law, psychiatry, and philosophy*. Oxford University Press.
4. Hare, R.D. 2003. *The Hare psychopathy checklist revised* (2nd ed.). Multi-Health Systems.
5. Glenn, A.L., A. Raine, and W.S. Laufer. 2011. Is it wrong to criminalize and punish psychopaths? *Emotion Review* 3 (3): 302–304.
6. Patrick, C.J., ed. 2018. *Handbook of psychopathy*, 2nd ed. The Guilford Press.
7. Jalava, J., and S. Griffiths. 2017. Philosophers on psychopaths: A cautionary tale in interdisciplinarity. *Philosophy, Psychiatry, and Psychology* 24 (1): 1–12. https://doi.org/10.1353/ppp.2017.0000.
8. Jefferson, A., and K. Sifferd. 2018. Are psychopaths legally insane? *European Journal of Analytic Philosophy* 14 (1): 79–96. https://doi.org/10.31820/ejap.14.1.5.
9. Jurjako, M., and L. Malatesti. 2018. Neuropsychology and the criminal responsibility of psychopaths: Reconsidering the evidence. *Erkenntnis* 83 (5): 1003–1025. https://doi.org/10.1007/s10670-017-9924-0.
10. American Psychiatric Association & American Psychiatric Association, eds. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association.
11. Cuthbert, B.N. 2022. Research Domain Criteria (RDoC): Progress and potential. *Current Directions in Psychological Science* 31 (2): 107–114. https://doi.org/10.1177/09637214211051363.
12. Murphy, D. 2017. Can psychiatry refurnish the mind? *Philosophical Explorations* 20 (2): 160–174. https://doi.org/10.1080/13869795.2017.1312499.
13. Buckholtz, J.W., V.F. Reyna, and C. Slobogin. 2016. A neuro-legal lingua franca: Bridging law and neuroscience on the issue of self-control. *Mental Health Law & Policy Journal*. https://ssrn.com/abstract=2788178. Accessed 18 Aug 2022
14. Meynen, G. 2016. *Legal insanity: Explorations in psychiatry, law, and ethics*. Springer. https://doi.org/10.1007/978-3-319-44721-6.
15. Simon, R.J., and H. Ahn-Redding. 2006. *The insanity defense, the world over*. Rowman & Littlefield.
16. Gröning, L., U.K. Haukvik, S.J. Morse, and S. Radovic. 2022. Remodelling criminal insanity: Exploring philosophical, legal, and medical premises of the medical model used in Norwegian law. *International Journal of Law and Psychiatry* 81: 101776. https://doi.org/10.1016/j.ijlp.2022.101776.
17. Moore, M.S. 2015. The quest for a responsible responsibility test: Norwegian insanity law after Breivik. *Criminal Law and Philosophy* 9 (4): 645–693. https://doi.org/10.1007/s11572-014-9305-6.
18. Robinson, D.N. 1996. *Wild beasts & idle humours: The insanity defense from antiquity to the present*. Harvard University Press.
19. Glenn, A.L., W.S. Laufer, and A. Raine. 2013. Author Reply: Vitacco, Erickson, and Lishner: Holding psychopaths morally and criminally culpable. *Emotion Review* 5 (4): 426–427.
20. Moore, M.S. 2016. The neuroscience of volitional excuse. In *Philosophical foundations of law and neuroscience*, ed. D. Patterson and M.S. Pardo, 179–230. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198743095.003.0010.
21. Morse, S.J. 2008. Psychopathy and criminal responsibility. *Neuroethics* 1 (3): 205–212. https://doi.org/10.1007/s12152-008-9021-9.
22. Jurjako, M., and L. Malatesti. 2022. The value-ladenness of psychopathy. In *Psychopathy: Its uses, validity and status*, vol. 27, ed. L. Malatesti, J. McMillan, and P. Šustar, 215–233. Springer. https://doi.org/10.1007/978-3-030-82454-9_12.
23. Sadler, J.Z. 2008. Vice and the diagnostic classification of mental disorders: A philosophical case conference. *Philosophy, Psychiatry, & Psychology* 15 (1): 1–17. https://doi.org/10.1353/ppp.0.0152.
24. De Brito, S.A., A.E. Forth, A.R. Baskin-Sommers, I.A. Brazil, E.R. Kimonis, D. Pardini, P.J. Frick, R.J.R. Blair, and E. Viding. 2021. Psychopathy. *Nature Reviews Disease Primers* 7 (1): 49. https://doi.org/10.1038/s41572-021-00282-1.
25. DeLisi, M. 2016. *Psychopathy as unified theory of crime*. Palgrave Macmillan.
26. Brazil, I.A., J.D.M. van Dongen, J.H.R. Maes, R.B. Mars, and A.R. Baskin-Sommers. 2018. Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews* 91: 259–277. https://doi.org/10.1016/j.neubiorev.2016.10.010.
27. Malatesti, L., and E. Baccarini. 2022. The disorder status of psychopathy. In *Psychopathy. Its uses, validity, and status*, ed. L. Malatesti, J. McMillan, and P. Šustar, 291–309. Springer.
28. Nadelhoffer, T., and W.P. Sinnott-Armstrong. 2013. Is psychopathy a mental disease? In *Neuroscience and legal responsibility*, ed. Nicole A. Vincent, 229–255. Oxford University Press.
29. Ene, I., K.K.-Y. Wong, and G.D. Salali. 2022. Is it good to be bad? An evolutionary analysis of the adaptive potential of psychopathic traits. *Evolutionary Human Sciences* 4: e37. https://doi.org/10.1017/ehs.2022.36.
30. Jurjako, M. 2019. Is psychopathy a harmful dysfunction? *Biology & Philosophy*, *34*(5). https://doi.org/10.1007/s10539-018-9668-5.
31. Međedović, J., B. Petrović, J. Želeskov-Đorić, and M. Savić. 2017. Interpersonal and affective psychopathy traits can enhance human fitness. *Evolutionary Psychological Science* 3 (4): 306–315. https://doi.org/10.1007/s40806-017-0097-5.
32. Yannoulidis, S. 2012. *Mental state defences in criminal law*. Ashgate.
33. Sifferd, K.L. 2022. Legal insanity and moral knowledge: Why is a lack of moral knowledge related to a mental illness exculpatory? In *Agency in mental disorder*, ed. M. King and J. May, 113–135. Oxford University Press. https://doi.org/10.1093/oso/9780198868811.003.0006.
34. Blair, R.J.R. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition* 57 (1): 1–29. https://doi.org/10.1016/0010-0277(95)00676-P.
35. Aharoni, E., W. Sinnott-Armstrong, and K.A. Kiehl. 2012. Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. *Journal of*

Abnormal Psychology 121 (2): 484–497. https://doi.org/10.1037/a0024796.
36. Aharoni, E., W. Sinnott-Armstrong, and K.A. Kiehl. 2014. What's wrong? Moral understanding in psychopathic offenders. *Journal of Research in Personality* 53: 175–181. https://doi.org/10.1016/j.jrp.2014.10.002.
37. Marshall, J., A.L. Watts, and S.O. Lilienfeld. 2018. Do psychopathic individuals possess a misaligned moral compass? A meta-analytic examination of psychopathy's relations with moral judgment. *Personality Disorders: Theory, Research, and Treatment* 9 (1): 40–50. https://doi.org/10.1037/per0000226.
38. Rosenberg Larsen, R., and D. Sackris. 2023. Are there "moral" judgments? *European Journal of Analytic Philosophy* 19 (2): (S1)1-23. https://doi.org/10.31820/ejap.19.2.1.
39. Sackris, D. 2022. The disunity of moral judgment: Implications for the study of psychopathy. *Philosophical Psychology*, 1–26. https://doi.org/10.1080/09515089.2022.2155125.
40. de Sio, F.S. 2011. Irresistible desires as an excuse. *King's Law Journal* 22 (3): 289–307. https://doi.org/10.5235/096157611798456753.
41. Jurjako, M., and L. Malatesti. 2016. Instrumental rationality in psychopathy: Implications from learning tasks. *Philosophical Psychology* 29 (5): 717–731. https://doi.org/10.1080/09515089.2016.1144876.
42. Meloy, J.R., A. Book, A. Hosker-Field, T. Methot-Jones, and J. Roters. 2018. Social, sexual, and violent predation: Are psychopathic traits evolutionarily adaptive? *Violence and Gender* 5 (3): 153–165. https://doi.org/10.1089/vio.2018.0012.
43. Glimmerveen, J.C., J.H.R. Maes, and I.A. Brazil. 2022. Psychopathy, maladaptive learning and risk-taking. In *Psychopathy. Its uses, validity, and status*, eds. L. Malatesti, J. McMillan, and P. Šustar, 189–211. Springer.
44. Glimmerveen, J.C., J.H.R. Maes, E. Bulten, I. Scheper, and I.A. Brazil. 2022. So what'cha want? The impact of individualised rewards on associative learning in psychopathic offenders. *Cortex* 149: 44–58. https://doi.org/10.1016/j.cortex.2022.01.006.
45. Jalava, J., and S. Griffiths. 2022. Psychopathy: Neurohype and its consequences. In *Psychopathy. Its uses, validity, and status*, ed. L. Malatesti, J. McMillan, and P. Šustar, 79–98. Springer. https://doi.org/10.1007/978-3-030-82454-9_6.
46. Blair, R.J.R., D.G.V. Mitchell, K.S. Peschardt, E. Colledge, R.A. Leonard, J.H. Shine, L.K. Murray, and D.I. Perrett. 2004. Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences* 37 (6): 1111–1122. https://doi.org/10.1016/j.paid.2003.10.008.
47. Hastings, M.E., J.P. Tangney, and J. Stuewig. 2008. Psychopathy and identification of facial expressions of emotion. *Personality and Individual Differences* 44 (7): 1474–1483. https://doi.org/10.1016/j.paid.2008.01.004.
48. Poeppl, T.B., M.R. Donges, A. Mokros, R. Rupprecht, P.T. Fox, A.R. Laird, D. Bzdok, B. Langguth, and S.B. Eickhoff. 2019. A view behind the mask of sanity: Meta-analysis of aberrant brain activity in psychopaths. *Molecular Psychiatry* 24 (3): 463–470. https://doi.org/10.1038/s41380-018-0122-5.
49. Deming, P., M. Heilicher, and M. Koenigs. 2022. How reliable are amygdala findings in psychopathy? A systematic review of MRI studies. *Neuroscience & Biobehavioral Reviews* 142: 104875. https://doi.org/10.1016/j.neubiorev.2022.104875.
50. Blair, R.J.R., D. Mitchell, and K. Blair. 2005. *The psychopath: Emotion and the brain*. Blackwell.
51. Greenspan, P. 2016. Responsible psychopaths revisited. *The Journal of Ethics* 20 (1–3): 265–278. https://doi.org/10.1007/s10892-016-9231-z.
52. Glass, S.J., and J.P. Newman. 2006. Recognition of facial affect in psychopathic offenders. *Journal of Abnormal Psychology* 115 (4): 815–820. https://doi.org/10.1037/0021-843X.115.4.815.
53. Dargis, M., R.C. Wolf, and M. Koenigs. 2018. Psychopathic traits are associated with reduced fixations to the eye region of fearful faces. *Journal of Abnormal Psychology* 127 (1): 43–50. https://doi.org/10.1037/abn0000322.
54. Larson, C.L., A.R. Baskin-Sommers, D.M. Stout, N.L. Balderston, J.J. Curtin, D.H. Schultz, K.A. Kiehl, and J.P. Newman. 2013. The interplay of attention and emotion: Top-down attention modulates amygdala activation in psychopathy. *Cognitive, Affective & Behavioral Neuroscience* 13 (4): 757–770. https://doi.org/10.3758/s13415-013-0172-8.
55. Shane, M.S., and L.L. Groat. 2018. Capacity for upregulation of emotional processing in psychopathy: All you have to do is ask. *Social Cognitive and Affective Neuroscience* 13 (11): 1163–1176. https://doi.org/10.1093/scan/nsy088.
56. Groat, L.L., and M.S. Shane. 2020. A motivational framework for psychopathy: Toward a reconceptualization of the disorder. *European Psychologist* 25 (2): 92–103. https://doi.org/10.1027/1016-9040/a000394.
57. Maibom, H.L. 2008. The mad, the bad, and the psychopath. *Neuroethics* 1 (3): 167–184. https://doi.org/10.1007/s12152-008-9013-9.
58. Aono, D., G. Yaffe, and H. Kober. 2019. Neuroscientific evidence in the courtroom: A review. *Cognitive Research: Principles and Implications* 4 (1): 40. https://doi.org/10.1186/s41235-019-0179-y.
59. Brzović, Z., M. Jurjako, and P. Šustar. 2017. The kindness of psychopaths. *International Studies in the Philosophy of Science* 31 (2): 189–211. https://doi.org/10.1080/02698595.2018.1424761.
60. Griem, J., N.J. Kolla, and J. Tully. 2022. Key challenges in neurocognitive assessment of individuals with antisocial personality disorder and psychopathy. *Frontiers in Behavioral Neuroscience* 16: 1007121. https://doi.org/10.3389/fnbeh.2022.1007121.
61. Jurjako, M., L. Malatesti, and I.A. Brazil. 2020. Biocognitive classification of antisocial individuals without explanatory reductionism. *Perspectives on Psychological Science* 15 (4): 957–972. https://doi.org/10.1177/1745691620904160.
62. Maibom, H.L. 2018. What can philosophers learn from psychopathy? *European Journal of Analytic Philosophy* 14 (1): 63–78.
63. Cooke, D.J. 2018. Psychopathic personality disorder: Capturing an elusive concept. *European Journal of Analytic Philosophy* 14 (1): 15–32. https://doi.org/10.31820/ejap.14.1.1.

64. Rosenberg Larsen, R. 2018. False-positives in psychopathy assessment: Proposing theory-driven exclusion criteria in research sampling. *European Journal of Analytic Philosophy* 14 (1): 33–52.
65. Patrick, C.J., D.C. Fowles, and R.F. Krueger. 2009. Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology* 21 (3): 913–938.
66. Baskin-Sommers, A.R., I.A. Brazil, J. Ryan, N.J. Kohlenberg, C.S. Neumann, and J.P. Newman. 2015. Mapping the association of global executive functioning onto diverse measures of psychopathic traits. *Personality Disorders* 6 (4): 336–346. https://doi.org/10.1037/per0000125.
67. Lilienfeld, S.O., S.F. Smith, and A.L. Watts. 2013. Issues in diagnosis: Conceptual issues and controversies. In *Psychopathology: History, diagnosis, and empirical foundations*, ed. W.E. Craighead, D.J. Mikllowitz, and L.W. Craighead, 1–35. Wiley.
68. Bolton, D., and G. Gillett. 2019. *The biopsychosocial model of health and disease: New philosophical and scientific developments*. Palgrave Macmillan.
69. Pickersgill, M. 2012. Standardising antisocial personality disorder: The social shaping of a psychiatric technology. *Sociology of Health & Illness* 34 (4): 544–559. https://doi.org/10.1111/j.1467-9566.2011.01404.x.
70. Campbell, C., and N. Eastman. 2014. The limits of legal use of neuroscience. In *Bioprediction, biomarkers, and bad behavior*, ed. I. Singh, W.P. Sinnott-Armstrong, and J. Savulescu, 91–117. Oxford University Press.
71. Glannon, W. 2014. The limitations and potential of neuroimaging in the criminal law. *The Journal of Ethics* 18 (2): 153–170. https://doi.org/10.1007/s10892-014-9169-y.
72. Hirstein, W., K. Sifferd, and T. Fagan. 2018. *Responsible brains: Neuroscience, law, and human culpability*. The MIT Press.
73. Werner, K.M., M. Inzlicht, and B.Q. Ford. 2022. Whither inhibition? *Current Directions in Psychological Science* 31 (4): 333–339. https://doi.org/10.1177/09637214221095848.
74. Pardo, M. S., & Patterson, D. M. 2013. *Minds, brains, and law: The conceptual foundations of law and neuroscience*. Oxford University Press.
75. Buckholtz, J.W., and D.L. Faigman. 2014. Promises, promises for neuroscience and law. *Current Biology: CB* 24 (18): R861-867. https://doi.org/10.1016/j.cub.2014.07.057.
76. Pernu, T.K., and N. Elzein. 2020. From neuroscience to law: Bridging the gap. *Frontiers in Psychology* 11: 1862. https://doi.org/10.3389/fpsyg.2020.01862.
77. Schopp, R.F. 2008. *Automatism, insanity, and the psychology of criminal responsibility: A philosophical inquiry*. Cambridge Univ. Press.
78. Dillon, D.G., and D.A. Pizzagalli. 2007. Inhibition of action, thought, and emotion: A selective neurobiological review. *Applied and Preventive Psychology* 12 (3): 99–114. https://doi.org/10.1016/j.appsy.2007.09.004.
79. Baskin-Sommers, A.R., S. Ruiz, B. Sarcos, and C. Simmons. 2022. Cognitive–affective factors underlying disinhibitory disorders and legal implications. *Nature Reviews Psychology* 1 (3): 145–160. https://doi.org/10.1038/s44159-022-00020-8.
80. Baskin-Sommers, Ae. R. 2016. Dissecting antisocial behavior: The impact of neural, genetic, and environmental factors. *Clinical Psychological Science* 4 (3): 500–510. https://doi.org/10.1177/2167702615626904.
81. Aspinwall, L.G., T.R. Brown, and J. Tabery. 2012. The double-edged sword: Does biomechanism Increase or decrease judges' sentencing of psychopaths? *Science* 337 (6096): 846–849. https://doi.org/10.1126/science.1219569.
82. Morse, S.J. 2022. Is executive function the universal acid? *Criminal Law and Philosophy* 16 (2): 299–318. https://doi.org/10.1007/s11572-021-09607-3.
83. Burghart, M., Schmidt, S., & Mier, D. (2023). Executive functions in psychopathy: A meta-analysis of inhibition, planning, shifting, and working memory performance [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/naxvg
84. Hirstein, W. 2022. Neuroscience and normativity: How knowledge of the brain offers a deeper understanding of moral and legal responsibility. *Criminal Law and Philosophy* 16 (2): 327–351. https://doi.org/10.1007/s11572-021-09600-w.
85. Glimmerveen, J.C., I.A. Brazil, B.H.(Erik) Bulten, and J.H.R. Maes. 2018. Uncovering naturalistic rewards and their subjective value in forensic psychiatric patients. *International Journal of Forensic Mental Health*, *17*(2), 154–166. https://doi.org/10.1080/14999013.2018.1452081.
86. Ichikawa, N., G.J. Siegle, A. Dombrovski, and H. Ohira. 2010. Subjective and model-estimated reward prediction: Association with the feedback-related negativity (FRN) and reward prediction error in a reinforcement learning task. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology* 78 (3): 273–283. https://doi.org/10.1016/j.ijpsycho.2010.09.001.
87. Loehr, J.D., D. Kourtis, and I.A. Brazil. 2015. It's not just my fault: Neural correlates of feedback processing in solo and joint action. *Biological Psychology* 111: 1–7. https://doi.org/10.1016/j.biopsycho.2015.08.004.
88. Baskin-Sommers, A.R., and I.A. Brazil. 2022. The importance of an exaggerated attention bottleneck for understanding psychopathy. *Trends in Cognitive Sciences* 26 (4): 325–336. https://doi.org/10.1016/j.tics.2022.01.001.
89. Bošnjak, M., M. Jurjako, and L. Malatesti. 2022. The criminal responsibility of high-functioning autistic offenders in Croatia. *Balkan Journal of Philosophy* 14 (2): 137–148. https://doi.org/10.5840/bjp202214217.