Prof. dr hab. Wojciech Załuski
Department of the Philosophy of Law and Legal Ethics
Faculty of Law and Administration
Jagiellonian University, Krakow, Poland

Criminal Responsibility and Mental Incompetence

Abstract. The purpose of this paper is to distinguish and examine the philosophical underpinnings of three models of criminal responsibility of the mentally incompetent. The models are determined by the assumed conceptions of criminal responsibility and of mental illness. They include: the Standard Model (admitting only responsibility of the mentally competent), the Positive Equalization Model (admitting criminal responsibility of both the mentally competent and the mentally incompetent); and the Negative Equalization Model (abolishing criminal responsibility of both the mentally competent and the mentally incompetent). In the evaluative part several argument are proposed in favour of the Standard Model.
Key words: responsibility, mental illness, mental incompetence, anti-psychiatry

1. Introduction

There is no doubt that one speak about the existence of the standard (adopted in nearly all legal systems) model of the regulation of criminal responsibility of the mentally incompetent; the model presupposes that: (a) freedom is a condition of criminal responsibility, (b) the concept of mental illness/disorder is fully legitimate; (c) mental illness/disorder *in conjunction with* its psychological effects (which are certain mental – cognitive and/or volitional – incapacities or at least serious impairments of certain mental capacities) constitute an exculpatory circumstance or at least a circumstance mitigating the penalty for committing an offence. But, needless to say, the very fact that a given model is dominant/standard does not mean that it is uncontroversial – that none of its presuppositions cannot or have not already been called into doubt. Furthermore, as we shall see, given the ambiguity of the concept of freedom, the Standard Model can itself be understood in several different ways. For these reasons in this paper an attempt will be made at presenting the standard model in a broader context depicting other possible models of criminal responsibility of the mentally incompetent.

As can be inferred from the above description of the Standard Model, models of criminal responsibility of the mentally incompetent are determined by the conceptions of criminal responsibility and of mental illness. Accordingly, I shall present the former in section

2, and the latter in section 3; the section 4 will be devoted to the presentation of the models of criminal responsibility of the mentally incompetent. At the outset, however, three important *caveats* must be made. *Firstly*, as we shall see in section 2, not all models of criminal responsibility of the mentally incompetent are determined by both conceptions (of criminal responsibility and mental illness); some of the conceptions of criminal responsibility can *alone/by themselves* justify a given model. *Secondly*, the second dimensions (an account of mental illness) is relevant only for the so-called 'double' or 'mixed' – *psychiatric-psychological* – account of mental incompetence (implied in the above description of the Standard Model) and within 'single', *purely psychiatric* account of mental incompetence.[1] It is clear that if one accepted a 'single', *purely psychological* account (which dispenses with the very concept of mental illness/disorder), then it would be of minor importance what conception of mental illness one assumes. Additionally, it should be noticed that the *strength* of the relevance of a conception of mental illness is greater in the 'single', purely psychiatric model than in the 'double' one, and how great this strength is in the 'double' one will depend on how the relationship between mental illness/disorder and the mental incapacities (psychologically defined) is exactly conceived: whether conceptually (in which case it is part of the definition of a mentally incompetent person that he has suffered, at least at the time of criminal act, from mental illness/disorder), or, as one may call it, 'diagnostically' (in which cases the fact that a person has suffered from a mental disease/disorder is not part of the concept of mental incompetence but makes the hypothesis of the person's being mentally incompetent more probable). In the former case (which implies that a mentally healthy person cannot be mentally incompetent) the relevance is stronger than in the latter case (which implies that a mentally healthy person can be mentally incompetent).[2] Arguably, the Polish legal doctrine assumes the former interpretation (postulating the conceptual relation between mental health and mental incompetence). *Thirdly*, one might take into account an additional dimension – that of the

[1] The 'single' models are much less frequent than the 'double' one; for instance, the purely psychiatric was accepted only for a short time in the American legal doctrine (in the wake of the case *Durham v United States* in 1954). The so-called Durham rule provided that a defendant is non-responsible if his unlawful act was "the product of mental disease or mental defect".

[2] I discount here the complication which arises from the fact that, in the Polish Penal Code, mental incompetence can also be caused by *mental deficiency/retardation*. When we take into account this complication, we cannot, strictly speaking, say that there is a conceptual/definitional connection between mental illness/disorder/disturbance and mental incompetence, though, of course, we can say that there is such a connection between mental illness/disorder/disturbance *or* mental deficiency/retardation on the one hand and mental incompetence on the other.

psychological *content* of mental incompetence – for distinguishing models of criminal responsibility of the mentally incompetent: depending on how precisely the dimension is conceived, viz. as concerning only the cognitive or also a volitional side, and how these two components are interpreted, one could distinguish additional models (or, rather, sub-models) of criminal responsibility of the mentally incompetent. This problem, however, leads to a number of new questions (for instance, about the legitimacy of the so called volitional component of the mental incompetence), which are not directly connected with the subject matter of this paper and, deserving, owing to their importance and intricacy, a separate, in-depth treatment

2. Conceptions of criminal responsibility

I shall use in my considerations the term 'criminal responsibility' in the general sense of accountability for failure to comply with the norms of criminal law. By 'conceptions of criminal responsibility' I shall mean, in turn, theories which determine the conditions under which an agent may be held accountable for this failure. I assume that there is a strict (that of a necessary condition) connection between the attribution of responsibility and punishability, that is: an agent can be punished for an act *a* only if he is (criminally) responsible for it. Since criminal law is, of all the branches of law, the one which is most closely related to morality, each of the conceptions of criminal responsibility discussed below could also be a conception of moral responsibility (with due allowance for the obvious and philosophically unimportant differences, for instance the type of norms for the violation of which one is responsible: moral or social norms in the case of moral responsibility and norms of criminal law in the case of criminal responsibility). In fact, some of these conceptions were initially proposed as conceptions of moral responsibility. One could perhaps also say that a conception of criminal responsibility in a given legal system is usually a reflection of a conception of moral responsibility presupposed in the morality of the society in which the legal system is intended to operate (though the truth or falsity of this sociological claim has no consequences whatsoever for my further analysis).

Each plausible conception of criminal responsibility must have the following two features: it must provide that a person *P* can be held criminally responsible for committing an act *a* if, and only if, *P* did *a*, and that *a* must be prohibited by the criminal law.

Conceptions which fail to satisfy these obvious requirements (e.g., those which *openly* – not just as an un-intended and not very likely consequence of their assumptions – declare that the innocent persons can be held criminally responsible) are omitted in my analysis. The plausible conceptions differ, however, in how they describe the 'state of the mind' in which *P* must have been at the time of committing *a* (or at the time of inflicting the penalty) if *P* is to be held responsible for *a*. The most general criterion for distinguishing conceptions of criminal responsibility is whether they assume determinism and criminal responsibility to be incompatible (the *incompatibilist conceptions*) or compatible (the *compatibilist conceptions*). From among the eight conceptions presented below Conceptions 1-3 are incompatibilist, and conceptions 4-8 are compatibilist.

2.1. The incompatibilist conceptions

According to *Conception 1*, *P* can be held criminally responsible for committing an offence *a* (or any other offence) if, and only if, *P* is endowed with free will, i.e., "the ability to do otherwise" (*a* or mental intention to commit *a* must not therefore be causally determined if *P* is to be held responsible for it). The phrase "any other offence" is justified because, on *this* conception, free will is treated as a durable feature/faculty: if human beings possess this faculty, they cannot lose it. According to this conception, *P* will be held responsible for *a* even if his action, though based on free will, was not voluntary, i.e., was not free from internal and external coercion (from *vis compulsiva*, though not *vis absoluta*[3]). Since this conception implies that free will is exhibited also in a non-voluntary action, it leads to the conclusion that criminal responsibility of the mentally competent and the mentally incompetent should be equalized (should be *none*, if the existence of free will is denied, and should be *full*, if the existence of free will is affirmed). This conception, therefore, determines *by itself* (without the necessity of its supplementation by a conception of mental illness) the model of criminal responsibility of the mentally incompetent. The main reason why this conception is highly counterintuitive is that it implies that human beings exhibit free will (assuming that it exists) also in non-voluntary actions, whereas most of those who believe that free will exists seem to assume (though they seldom make this assumption

[3] Because 'an act' done as a result of *vis absoluta*, in which the decisional component is entirely absent, is not an act in the strict sense.

explicit) that it is manifested only in voluntary actions. This leads us to *Conception 2*, which says that *P* can be held criminally responsible for committing an offence *a* if, and only if, *P* is endowed with free will *and* his action *a* was voluntary (free from internal and external coercion). Unlike *Conception 1*, this conception does not by itself (but only in conjunction with a conception of mental illness) decides the problem of criminal responsibility of the mentally incompetent.

A different incompatibilist conception (*Conception 3* in my list), based on Galen Strawson's account of moral responsibility, asserts that *P* can be held criminally responsible for committing an offence *a* (or any other offence) if he is *causa sui*, "at least in certain mental respects", such as "preferences, values, ideals, pro-attitudes", which determine the way in which we react to various reasons for action (cf. Strawson 1994: 5-7).[4] There arise two particularly interesting problems in this context. The first one concerns the relations between the concepts of free will and *causa sui*. Strawson claims that they are different, and he may be right in the sense that the 'object of the activity' (as one may put it) of *causa sui* is the agent himself (his character, his hierarchy of values), whereas the 'the object of the activity' of free will are actions (free will is the capacity which manifests itself in decision-making). Accordingly, as it seems, an agent does not have to be *causa sui* to have free will: in order to have free will, he must be, as one might put it, *causa suae actionis.* However, I shall not be examining this problem further; I shall focus on the second one, which seems to be more important, viz. Strawson's argument that no one can be *causa sui* and thereby no one (irrespective of whether he is mentally competent or incompetent) can be held morally and (if we treat his conception also as a conception of criminal responsibility[5]) also criminally responsible (the conception of Strawson therefore, *by itself* – without the need of being supplemented by a conception of mental illness – resolves the problem of the responsibility of the mentally incompetent) The argument of Strawson is simple (he calls it 'the Basic Argument'): an agent always chooses an action in accordance with some previous criteria ("principles of choice", as he calls them); however, these criteria either were not chosen by him in a "reasoned, conscious, intentional fashion" but imposed on him (by his upbringing,

[4] Galen Strawson ascribed this conception rather controversially, to Kant and Sartre, and also (less controversially) to Nietzsche (as noticed by Strawson, Nietzsche rejects the very concept of *causa sui*, for instance, in par. 21 of his *Jenseits von Gut und Böse*, and in consequence rejects the very concept of responsibility).

[5] This not Strawson's intention, as can be inferred from the quotation below.

genes, etc), in which case he is not *causa sui*, or they were chosen by him; but if they were chosen by him, they must have been chosen in accordance with some earlier criteria, which means either that he is not *causa sui* (because these second-order criteria were in some way imposed on him) or that there occurs *regressus ad infinitum* (an infinite series of the agent's choices, which is impossible). Strawson summarizes his argumentation in the following way:

> True self-determination is impossible because it requires an actual completion of an infinite series of principles of choice. So true moral responsibility is impossible, because it requires true self-determination (…) We are what we are, and we cannot be thought to have made ourselves *in such a way* that we can be held to be free in our actions *in such a way* that we can be held to be morally responsible for our actions *in such a way* that any punishment or reward for our actions is ultimately just or fair. Punishments and rewards may seem deeply appropriate or intrinsically 'fitting' to us in spite of this argument, and many of the various institutions of punishment and reward in human society appear to be practically indispensable in both their legal and non-legal forms. But if one takes the notion of justice that is central to our intellectual and cultural tradition seriously, then the evident consequences of the Basic Argument is that there is a fundamental sense in which no punishment or reward is ever ultimately just (Strawson 1994: 7, 15-16).

There are some problems with Strawson's 'Basic Argument' The gist of this argument boils down to the claim that we must be responsible for who we are in order to be responsible for what we do. But this claim rests on two dubious assumptions: that the very notion of "who we are" is tenable; and that who we are ("our principles of choice") determines our choices. The first assumptions seems dubious because, arguably, only with respect to some people one can say about "the way they are", i.e., that they have a rigid set of principles (values, preferences, etc.) which determine their choices; most people have a rather fluid set of principles, shaped in the course of the very process of their decision-making. The second assumption seems dubious because free will is precisely the capacity to decide even against one's own customary 'principles of choice' in such a way that the 'new principles' of choice are not *presupposed* but *discovered* or *created*. Strawson does not consider these objections. He deals only with the objection that there exists 'Self' which is independent of character or motives. He points out that such a Self would have to decide "as it decides because of the way it is. And this returns us to where we started. To be a source of true or ultimate responsibility, Self must be responsible for being the way it is. But this is impossible (Strawson 1994: 20-21)". This counter-objection of Strawson, however, can be criticized on the same two grounds indicated above: the Self may not be identical with a rigid set of principles, and the Self may discover or create principles in the course of making a choice. In general, it seems that Strawson is wrong when he claims that "however self-consciously aware we are, as we deliberate and reason, every act and operation of our mind happens as

it does as a result of features for which are ultimately in no way responsible (Strawson 1994: 22)".

2.2. The compatibilist conceptions

Conception 4, defended, for instance by Barbara Wootton (1963), states that *P* can be held criminally responsible for committing an offence *a* if, and only if, his action caused *a* (or is identical with *a*); the state of the mind of *P* at the time of committing *a* (or at any other time) is irrelevant for ascribing him responsibility for *a*. Thus, only *actus reus*, not *mens rea*, should be taken into account in the ascriptions of criminal responsibility: the prosecution does not have either to prove intent or recklessness or to query whether the defendant was mentally competent. On this conception, all crimes become crimes of strict liability. Clearly, this conception *by itself* implies that both the mentally competent and mentally incompetent can be held criminally responsible. Needless to say, this conception is extremely controversial, as it rejects the basic assumption of the modern criminal law, viz. that the 'subjective' side of a crime is equally important as its objective side (the outward act).

Conception 5, defended by many thinkers (to mention only Aristotle, David Hume, John Stuart Mill, or Kazimierz Twardowski), implies that *P* can be held criminally responsible for committing an offence *a* if, and only if, his action *a* was voluntary (free from internal and external coercion). It might be (rightly) objected, however, that my description of this conception is too sketchy, as it does not explain satisfactorily why its adherents could assert that an agent who is not endowed with free will (or is not *causa sui*) but is 'merely' free from internal and external coercion may be held criminally responsible for his (criminal) acts. One must therefore show that the lack of coercion allows some other ground of criminal responsibility (other than free will which, *ex hypothesi*, is non-existent) to manifest itself. The ground may be called generally 'an agent's individuality'. This concept was interpreted in various (though not mutually exclusive) ways by various philosophers: as an agent's character, as an agent's 'essence'[6], as his 'decisional centre', as his capacity for self-control,

[6] It seems that Aristotle's term '*hekousion*', usually translated as 'voluntary', should be translated 'flowing from the substance' (the substance – *ousia* – is the agent himself); this translation shows immediately why an agent can be held responsible for act done non-coercively: this is his *own* act*, flowing from him* as a separate entity/agent (substance).

or as a set of his *own* beliefs and desires. Depending on how this 'ground' is explained, one obtains different variants of this conception.

According to *Conception 6*, *P* can be held criminally responsible for committing an offence *a* if, and only if, the penalty brought upon him is likely to cause a desired change in his behavior. As Andrew Eshleman put it, calling this conception 'consequentialist', the agent can be held responsible if he "exercised a form of control that could be influenced through outward expression of praise and blame in order to promote or curb certain behaviours (Eshleman 2016)". It is worth noticing that *Conception 6* seems less restrictive that *Conception 5*, since, arguably, the group of people whose actions can be influenced by penalty will be broader than the group of people whose actions were voluntary. This conception can also be found in the work of Barbara Wootton (1963), especially in the context of her analysis of the function of punishment (which is – or, rather is to be – purely preventive, in her view). The Danish legal philosopher Alf Ross summarized Wootton's view in the following way:

> the criterion of mental responsibility (imputability) should simply be dispensed with as a condition for conviction, and the criminological reaction to crime should be arrived at in each individual case without regard to guilt and only with a view to what in the particular instance will offer the best chance of preventing recidivism. The traditional system, which bases punishment on retribution for guilt, should be replaced, according to this view, by a system designed as a means of preventive hygiene (…). Wootton's doctrine amounts to no less than the view that questions of both imputation and imputability be discounted as conditions for convicting a person of an offence, while at the same time they are to be taken into account as circumstances which partially determine the nature of legal reaction to be applied in particular case (Ross 1975: 73)[7].

This is not surprising that Wootton defended at the same time *Conceptions 4* and *6*, because they can be reconciled on the assumption that each agent who is causally responsible is (more or less) responsive to the sanctions. It should be noticed, however, that it is certainly controversial to treat *Conception 6* as a conception of criminal responsibility. Some might argue that conception of criminal responsibility should be 'backward-looking', not 'forward-looking'. Furthermore, if we accept *Conception 6*, then the distinction between penalty as a just retribution and preventive measures becomes blurred. The conception also justifies the punishing of children, since their conduct can assuredly be modified by sanctions. But, as it seems, the adherents of *Conception 6* are inclined to accepts these consequences.

[7] In Ross's terminology, imputation is related to the question of the form of mental attitude of the agent to his act (whether the act was negligent or intentional), whereas imputability concerns the question of mental responsibility, viz. whether the act can be 'imputed' to him.

Conception 7, in turn, defended by Herbert L. A. Hart (2008, 1968), assumes that *mens rea* and *actus reus* of an agent *P* are sufficient for holding him/her criminally responsible even if his/her (criminal) act was not voluntary (in the sense of freedom from internal coercion).

On *Conception 8*, proposed by Peter Strawson (1962), *P* can be held morally responsible for committing an offence *a* if, and only if, the type of act he did (the type including not only outward features of the act but also the mental attitude of the agent) elicits a proper – impersonal reactive – attitude (moral indignation, moral disapproval) in most other agents (so that we can speak about social practice of reacting in a given way to a given type of offence). *R*'s reactive attitude is personal when *P*'s immoral act endangers his personal interests; if it does not (i.e., is harmful for people whose interests are not the object of concern for *R*), *R*'s reactive attitude is impersonal/moral. This is a conception of *moral* responsibility which cannot (unlike the previous conceptions) be easily transformed into a conception of criminal responsibility. But, as we shall see, it has several implications for the latter. Let me first notice that, on Strawson's conception, it would be improper to say that because the agent is responsible (deserves punishment), he elicits certain negative attitudes; the proper way of saying would be: an agent can be held responsible because he elicits certain – negative – attitudes, i.e., is *held* responsible. Thus, paradoxically, as it may sound, the idea of holding responsible become primary: an agent is responsible for an act *a* if, and only if he is held responsible (morally disapproved) for this act. Conversely, an agent is not responsible for an act he did if, and only if, he triggers off an *objective* attitude of excuse/understanding. Here we reach a point of Strawson's account which is especially important for our analyses, viz. that reactive attitudes (personal and impersonal) are sensitive not only to the positive or negative attitudes manifested in the behavior of the others towards us/others but also to whether these attitudes are of the mentally sane or of the insane. In the former case we take, as Strawson calls it, a 'participant's view' of their attitudes, in the latter – an 'objective' view, which is directed at understanding rather than evaluation. This implies that we cannot forbear distinguishing between the responsibility of the mentally competent and the mentally incompetent. The distinction, as Strawson argues, is deeply rooted in the way our attitudes function, though Strawson does not explain *how* deeply rooted they are – whether they are changeable or not. He seems to maintain that the rootedness is really deep, since he says they the way our attitudes function is part of "the

general framework of human life (Strawson 1962: 6)". Furthermore, according to Strawson, our personal reactive attitudes would not change even if it turned out that determinism is true and we knew that it is true; Strawson deems it "practically inconceivable" that "a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships as we normally understand them (Strawson 1962: 4)". The same applies to impersonal reactive attitudes, though Strawson's argument for this claim is indirect: he claims that it is harder to imagine the situation in which personal reactive attitudes remain while impersonal reactive attitudes disappear than the situation in which both types of reactive attitudes disappear, because, they "stand or lapse together (Strawson 1962: 10)". Thus, since personal reactive attitudes cannot disappear, impersonal reactive attitudes cannot disappear either. To the question whether it would not be *rational* to suspend our moral reactions if determinism were true, Strawson replies that "it is useless to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do (Strawson 1962: 10)". On a superficial reading of Strawson's paper, one might assert that Strawson's conception does not have a normative dimension, as it takes the social practices of blaming/praising *as they are*, and does not say what they should be like. Accordingly, if the practice were different, e.g., sensitive only to the 'outward' action, or if we were mentally constructed in such a way that we would regard the mentally ill as capable of participating in the normal relationships, then there would be no reason to introduce the 'insanity defence' to a legal system. But the crucial point is that Strawson seems to deny that it could be different. So his descriptive analysis become normative if one assumes a plausible thesis that one cannot require the change of what cannot be changed.[8] The consequence of this conception for the problem of criminal responsibility of the mentally incompetent are straightforward: given the depth of rootedness of the Standard Model in our social practices, it makes little sense to discuss other models; if they were adopted, they could not effectively function because they are too much at odds with those practices. But this consequence, by

[8] One must notice, however, that the above interpretation of Strawson complex and many-layered paper is not the only possible one. His conception would turn out to be much closer to the more traditional ones if it were interpreted as saying that an agent is responsible for committing an offence *a* if, and only if, the offence elicited negative emotions *warranted by the existing social norms* – the norms which decide, for instance, what circumstances justify the adoption of an 'objective' – exculpatory – attitude to the agent. On this interpretation, the crucial factor for responsibility ascription would be social norms rather than emotions themselves, and if the norms invoked such factors as freedom or voluntariness, then the conception of Strawson would not be much different from those which are based on the conception of free will or freedom from coercion.

itself, may be treated as an argument against Strawson's conception, since as we shall see in section 4, non-standards models were implemented in legal systems (though, on the other hand, the fact that they were very *rarely* implemented weakens this argument).

At the end of this section, it is worth emphasizing that the incompatibilist conceptions are stronger than the compatibilist ones in the sense that they make criminal responsibility contingent on the truthfulness of strong metaphysical assumptions: about the existence of free will or about an agent's being *causa sui* (in consequence, they are arguably also stronger in a different sense: if the assumptions are true, they provide a stronger ground for holding an agent responsible for his actions; I shall return to this point in the last section). If a metaphysical assumption of a given conception is rejected, then, on this conception, criminal responsibility proves to be an empty concept: it cannot be ascribed to anyone. For instance, as already mentioned, Galen Strawson claimed that „nothing can be *causa sui*, and in order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects (Strawson 1994: 21)". The compatibilist conceptions are weaker: since they do not make any questionable metaphysical assumptions, they do not create the risk that criminal responsibility will turn out to be an empty concept.

3. Accounts of mental illness

Let me now turn to the second dimension of the models of criminal responsibility of the mentally incompetent, viz. the accounts of mental illness. I shall distinguish two such accounts, which I shall dub the 'affirmative' and the 'negationist'.

3.1. The affirmative account

This account implies that the distinction between mental illness/disorder and mental health is fully legitimate, because, even though it may be difficult or impossible to provide a value-neutral conception of mental illness/disorder/health, it is possible to construct a mixed one, combining a value and scientific element. One such conception was proposed by

James Wakefield.[9] On his account, mental disorder is an objective phenomenon: it refers to some *real* dysfunction of a mental mechanism. By a 'real' dysfunction he means the one whose recognition does not depend on specific cultural assumptions of the 'observer' but is defined without reference to the 'observer', viz. as "the failure of a mental mechanism to perform a natural function for which it was designed by evolution (Wakefield 1992: 373)". The value-neutrality of his account, however, is not complete, because, as he notices, not each kind of dysfunction must be viewed as a disorder. For a dysfunction to be a disorder, it must also be harmful for the agent, and what counts as being harmful depends on the norms of the agent's culture. However, the value component is secondary in the sense that it 'selects' disorders from the set of dysfunctions determined by the scientific component. Wakefield's final definition of a disorder as a harmful dysfunction is as follows:

> A condition is a disorder if and only if (a) the condition causes some harm or deprivation of benefit to the person as judged by the standards of the person's culture (the value criterion), and (b) the condition results from the inability of some internal mechanism to perform its natural function, wherein a natural function is an effect that is part of the evolutionary explanation of the existence and structure of the mechanism (the explanatory criterion) (Wakefield 1992: 384).

Of course, this definition is not uncontroversial. The obvious objection that may be raised to it is that it is not always clear what function of a given mechanism is supposed to perform. We may therefore not know whether a mechanism has a dysfunction until we know its evolutionary function, and such knowledge, especially in the case of mental mechanisms, may be often difficult to gain. But I would like to formulate a different objection to this account, viz. that it does now allow one to distinguish between a mental disorder and mental illness (it is clear that not every mental disorder, e.g., dyslexia is a mental illness which has to be cured psychiatrically). This problem is not insoluble, of course. It seems that it could be solved by supplementing Wakefield's definition with the common-sense criteria of (full) mental incompetence encoded in most penal systems (and treated as exculpatory circumstances). The Polish Penal Code (in Art. 31) provides two such criteria: the incapacity to recognize the significance of the committed act (the cognitive component) and to control one's conduct (the volitional component). Thus, on this account, a mental illness would be a harmful dysfunction which consists in the mental incompetence. This account does not imply that the concept of mental illness is clear and distinct; it recognizes its vagueness – a number of 'borderline' cases, with regard to which it is not clear whether they can be legitimately

[9] For an in-depth, critical review of various conceptions of illness/disease/health cf. Dryla 2018, pp.15-89.

considered as mental illnesses or not. For instance, while it is clear that endogenous psychoses, such as schizophrenia, paranoia, bipolar disorder (manic depression) are mental illnesses, it is less clear, for instance whether borderline personality disorder, obsessive-compulsive disorder, psychopathy or pedophilia are such illnesses (the dominant opinion seems to be that they are not). There is no place here to analyze in depth the sources of its vagueness, but it may be illuminating to notice that the uncontroversial cases of mental illnesses seem to be connected with the cognitive defect (being 'out of touch with reality', having delusions), whereas controversies tend to arise mainly in connection with the disorders which affect the volitional side. This is interesting because from among the two component of the commonsense definition of mental incompetence the volitional one is by far more controversial.[10]

3.2. The negationist account

The negationist account is connected with the so called 'anti-psychiatry', the term coined by one of its proponents David Cooper in his book *Psychiatry* and *Anti-Psychiatry* from 1967. There are several different claims characteristic for the current, though it must be stressed that not all unorthodox psychiatrists, i.e., anti-psychiatrists, defended all of them.

The first claim, which may be called 'sociological', is least controversial: it says that *the practices* of orthodox psychiatry, i.e., the typical attitude and methods of treatment of the mentally ill within the psychiatric system (e.g., electroshocks, lobotomy), cannot be morally accepted as they are inhumane, lacking in respect for the patients' dignity and rights, based on violence, oppression, the rigidly hierarchical relation between the patient and the doctor (this inhumanity of the institutional psychiatry was especially strongly emphasized by Erving Goffman in his classical book *Asylum* from 1961, which profoundly influenced the 'anti-psychiatric' movement[11]). In the more radical versions of this claim, the critique of the practice of orthodox psychiatry becomes part of a more general critique of

[10] They are two main problems here: whether there are any irresistible impulses at all, and (assuming that there are such impulses) how to distinguish a situation in which a criminal impulse was irresistible form the one in which it was simply not resisted. The very idea of irresistible impulse is convincingly criticized, e.g., by Roy Baumeister (2011).

[11] For instance, the works of the Italian psychiatrists Franco Basaglia (1973) have been much influenced by Goffman's analyses.

the system of (capitalist) power, the 'doctor' being seen as a representative of this power. The postulate flowing from this claim is that the psychiatrist system be changed so that that the dignity and human rights of the patients should be recognized and respected

The second claim, more important in the context of my considerations, is that the mental illness is a social construct or a 'myth': the mentally 'ill' are not ill in the strict – and proper – sense of an objective dysfunction (only physical illness is an illness in the strict and proper sense); they just behave in a way which diverges from what the society regards as 'normal'. To put on them the etiquette of mental illness is an insidious attempt at disciplining them and imposing on them the prevailing social norms. On this account, the concept of mental illness is an instrument of stigmatizing socially undesirable conduct, and thereby is *intentionally* used in bad faith by those in power, though anti-psychiatrists sometimes interpreted it also in a less 'unmasking' way, viz. as an expression of sexual, racial or social prejudices (which may be cherished, or rather, are usually cherished, in an unconscious way). The second claim may, however, take a stronger form: viz. that mental 'illness', besides being no illness at all (a weaker form of the claim), is also or at least may be a privileged form of cognition, offering an access to some other dimension of reality (a stronger form of the claim). The second claim is, obviously, fully compatible with the first one, or even implies it: if a mental illness is a myth, a way of disciplining those who deviate from what the society regards as normal, one cannot be uncritical towards the practice of and perhaps even the very existence of the institutions of the orthodox psychiatry (which, on this view, is based on a fundamental mistake). The second claim in its weaker, but occasionally also stronger, form was defended, for instance, by Michel Foucault, whereas Thomas Szasz, perhaps the most famous representative of 'anti-psychiatry', defended its weaker form. A short presentation of their views may be in order here, as it will help better understand their reasons for adopting the critical stance on the traditional account of mental illness.

Michel Foucault in his classical book *Madness and Civilization* from 1961 maintained that the concept of mental illness is simply one of many ways in which the state is trying to control its citizens. According to this account, mental illness, unlike physical illness, is not an objective phenomenon: the deficiencies to which it refers are *always* "socially constructed", distinguished on the basis of subjective ideas and evaluations of those who ascertain them. There exists in a given society a certain standard view of happy/healthy/adaptive life, and

anyone who deviates from this standard is considered to be mentally ill. Thus, there exist no strictly scientific, objective grounds for introducing the concept of mental illness. In Foucault's view, the modern understanding of mental illness – precisely as an *illness* – is a consequence of the reinterpretation of madness that took place in the Enlightenment: the voice of madness – the voice of Unreason – was no longer considered as legitimate; it was smothered by the Reason, which, in Foucault's radical view, is nothing more than the face of power. It may be interesting to notice that Foucault's interpretation of the 'history of madness' was much influenced by his reading of an essay 'Van Gogh, The Man Suicided by Society' written by the French writer Antonin Artaud, who after a nervous breakdown, spent nine years in a mental hospital, where he was subjected to severe treatment, including, among others, sixty electroshock (cf. Kotowicz 1997: 61). In this essay, which was an scathing critique of the orthodox psychiatry, Artaud defended the view that mental illness is not only no illness at all; it is also a privileged way of knowing:

Things are going badly because sick consciousness has a vested interest right now in not recovering from its sickness. This is why a tainted society has invented psychiatry to defend itself against the investigations of certain superior intellects whose faculties of divination would be troublesome (quoted after Kotowicz 1997: 62).

In his most famous book *The Myth of Mental Illness* from 1960 Thomas Szasz raised a claim that there are no grounds for the belief that mental illness is scientific category, i.e., that it is not, unlike any genuine illness (being a scientific category), a physical lesion, a deviation from the anatomical structure (I shall devote some attention to t the value of this argument at the end of this paper); rather than being a scientific category mental illness is a myth constructed by the psychiatric authorities to support the existing social norms. Szasz's and many other anti-psychiatrists' view of schizophrenia – often called 'the conspiratorial model of madness' – has been aptly characterized in the following way:

Schizophrenia is a *label* which some people pin on other people, under certain social circumstances. It is not an illness, like pneumonia. It is a form of alienation which is out of step with the prevailing state of alienation. It is a social fact and political event (Siegler, Osmond, Mann 1969: 950)".

Thomasz Szasz claimed that that those who are called mentally ill can at best be said to suffer from 'problems in living', i.e., from not being capable of adapting themselves to the norms existing in a society. They should be allowed to choose by themselves the way they want to be helped (if they at all want to be helped): 'custodial psychiatry', as Sasz called it, should replace 'institutional psychiatry'. Szasz, as a right-wing libertarian, claimed that the state should not play any role in solving the problems of the mentally suffering; there should

be no place for involuntary, psychiatric hospitalization. As we can see, anti-psychiatrists, usually, did not maintain that schizophrenics do not need any help; but they believed it should not be the kind of help offered to them in traditional mental hospitals.

The third claim is that the mentally ill are entirely or to a large extent responsible for their illness.[12] This claim was not, strictly speaking, defended explicitly by any one of the group of anti-psychiatrists (they were not concerned with the problem of responsibility for mental illness), but it seems to be implied by what some of them said. Note the differences and similarities between this claim and the previous (second) one. This claim says that mental illness is a reality, not a myth, and the mentally ill are responsible for their illness. The previous (second) claim does not exclude that the mentally 'ill' may be responsible for their 'illness' but this (possible) ascription of responsibility loses moral weight insofar as mental illness is assumed to be a myth – no illness at all. Both claims imply that mental illness cannot be an exculpatory circumstance in criminal law.[13] In discussing the third claim at greater length, I shall focus on Ronald. D. Laing's most famous book *The Divided Self* (from 1960).

Arguably, Laing's account of mental illness can be interpreted as implying that mental illness is indeed a defect, but the one which has been caused by the mentally ill himself, so that one can speak about his responsibility for his illness. This interpretation can be justified by two strands in his argumentation. *Firs*t, Laing maintains that there is no unbridgeable abyss between sanity and madness (especially schizophrenia): in his view, schizophrenia is in fact a development of certain traits already present in the (sane) personality of a schizoid, especially of the division between the 'dis-embodied self' on the one hand, and one's body (and the 'false-self systems' associated with it) and the world on the other. The transition, in Laing's view, proceeds more or less as follows: the division between 'dis-embodied', 'inner' self and the 'rest' (body/false-self systems associated with the body/ world) – the division which is already present in a schizoid personality and which results from "ontological security" (an anxiety in the face of the world) – keeps on becoming deeper: as the false-self systems become more and more extensive and autonomous, the inner self becomes more and more volatilized, empty, unreal, impoverished, and finally dead, fragmented. *Secondly*,

[12] Patricia E. Erickson and Steven K. Erickson (2008: 14-17) call such a conception of mental illness (which views it simply a failure of personal responsibility) 'moral/punitive'.

[13] Though it should be stressed that those who defended the second claim were often also highly critical towards the institution of the 'capitalist state' and its manifold 'means of oppression', including criminal law.

in his account of mental illness, Laing assumed a humanistic perspective, i.e., he looked at the illness from an existential point of view, as a specific way of life, without invoking any biological factors in his description (his account of the etiology of mental illness is therefore psychogenic). Given this perspective, one could plausibly argue that this description implies that this way of life has been chosen or at least was not resolutely opposed by the patient. Laing did not say expressly that the schizophrenic patient is responsible for his illness but there seems to be nothing in his account that would exclude such an interpretation. It is true that he maintained that, at some stage of the movement of the schizoid towards psychosis, there may be no way back to 'normality'. It does not have to mean, however, that he believed that the schizoid could not have done it earlier, when the cleft between his 'inner' self and his 'false-self' system was not so deep. In summary, Laing's view of mental illness developed in *The Divided Self* may be interpreted as implying that the mentally ill is to a large extent responsible for his illness; though that it must be admitted that is by no means clear whether Laing himself would endorse this interpretation based mainly on the humanistic/existentialist mode of his analysis. For even though he assumed this perspective he did not deny that an illness may have also organic aspects or causes, and that an important part of the etiology of mental illness is played by the parents of the ill.[14]

By way of digression, let me remark that it is also not entirely clear whether and to what extent Laing endorsed the second claim in its stronger form. What is beyond doubt is that his views on mental illness underwent a certain evolution: in his later works he got closer to the second claim, viz. that mental illness may be a privileged form of cognition; though, interestingly, he seems to have severed this belief form the belief that it is no illness at all. He wrote already in *The Divided Self* that "our 'normal' 'adjusted' state is too often the abdication of ecstasy, the betrayal of our true potentialities, that many of us are only too successful in acquiring a false self to adapt to false realities (Laing 1969 [1960]: 12)" He developed this line of thought in *The Politics of Experience*, suggesting that the madman may have an insight into other dimension – may become "the hierophant of the sacred (Laing 1967: 109)", though he did not claim that madness *guarantees* this kind of insight or that this is the only way of achieving it. Furthermore, as already mentioned, he did not maintain

[14] Laing developed the famous 'good-bad-mad' scenario of the development of psychosis, which was to demonstrate that the suppression of the needs of a child may trigger off the process of splitting the self into the 'inner' self and 'false-self' system

that mental illness is not an illness; as he wrote, "madness need not be all breakdown. It may also be breakthrough. It is potentially liberation and renewal as well as enslavement and existential death (Laing 1967: 110)". Laing's point was therefore that a psychotic breakdown is not only an onset of the disintegration of personality: it may also (though is not bound to) be a way of coming into contact with some other – mystical – reality. Thus, as Zenon Kotowcz notices, there is no reason to maintain that Laing was "romanticizing madness" or "encouraging people to go mad as this would enrich their lives (Kotowicz 1997: 66)". As Kotowicz puts it:

> At no point did Laing lose sight of the fact that a breakdown is also 'enslavement and existential death'. So why should these views provoke such a strong reaction? In part this was because Laing's views were conflated with those of David Cooper, his collaborator in *Reason and Violence*. All of Cooper's subsequent writing are far more unrestrained than Laing's. (…) Cooper really saw the mentally ill as part of the vanguard in the fight against oppression, a position that Laing never adopted (Kotowicz 1997: 67).

To sum up, three different claims have been made within 'anti-psychiatry': (1) that the psychiatric institutions developed within the paradigm of the 'orthodox psychiatry' deprive the patients' of their dignity, violate their rights; (2) that mental illness is no illness at all (a weaker form); and, besides being no illness at all, it is a privileged form of cognition (a stronger form); (3) that mental illness is a real illness, but the one for which the patient himself/herself is responsible. As we shall see in the next section, all the three claims exerted a strong – both beneficial and negative – influence on the psychiatric and penal systems. Especially, claims (2) and (3), though different, may lead (and led to) to similar conclusions as far as criminal responsibility of the mentally ill is concerned, viz., to the equalization of the responsibility of the mentally ill and the mentally healthy: the claim (2) implies that the responsibility should be the same, because there is no such thing as mental illness (though, since on this claim there is also no such thing as 'normality', it may be also invoked to justify the abolition of all disciplinary institutions, including criminal law); the claim (3) implies that the responsibility should be the same because the mentally ill are themselves responsible for their (real) illness.

4. Models of criminal responsibility of the mentally incompetent

In this section I shall distinguish three models of criminal responsibility of the mentally incompetent. The distinction is made on the basis of two criteria: the mental state

(mentally competent and mentally incompetent), and criminal responsibility (admitted, excluded). We have therefore four possible models:

(M1)Mentally competent – Responsible, Mentally incompetent – Not-responsible

(M2)Mentally competent – Not-Responsible, Mentally incompetent – Responsible

(M3)Mentally competent – Not-Responsible, Mentally incompetent – Not-Responsible

(M4) Mentally competent – Responsible, Mentally incompetent – Responsible

I shall call (M1) 'the Standard Model', (M3) 'the Negative Equalization Model', and (M4) 'the Positive Equalization Model'. I shall examine their possible justifications, drawing on the conceptions of criminal responsibility and mental illness discussed in previous sections. I omit a rather idiosyncratic ('perverse') model (2), which was not implemented in any legal system, and which can hardly be justified.

As was already mentioned in the Introduction, The *Standard Model* assumes the distinction between the mentally ill and the mentally healthy is sound, and that the mentally competent can be held criminally responsible, whereas the mentally incompetent cannot be held criminally responsible.[15] This model is accepted by most contemporary legal systems. For instance, in the Polish Penal Code the model is worded out by the Article 31:

§ 1. Whoever, at the time of the commission of a prohibited act, was incapable of recognizing its significance or controlling his conduct because of a mental disease, mental deficiency or other mental disturbance, shall not commit an offence.
§ 2. If at the time of the commission of an offence the ability to recognize the significance of the act or to control one's conduct was diminished to a significant extent, the court may apply an extraordinary mitigation of the penalty.
§ 3. The provisions of § 1 and 2 shall not be applied when the perpetrator has brought himself to a state of insobriety or intoxication, causing the exclusion or reduction of accountability which he has or could have foreseen.

It is clear that this model implies the affirmative account of mental illness. It is less clear which of the conceptions of criminal responsibility it is based on. It is compatible with *Conceptions 2* and *5*. But, arguably, this is *Conceptions 5* which has been assumed by the authors of the code (I shall return to this issue in the last section).

[15] Let me recall that the concepts of mental illness/disorder and mental competence are connected with each other on the 'double' (mixed), and 'single' (purely psychiatric) models of mental competence, so that the rejection of the concept of mental illness implies the rejection of the concept of mental incompetence (with the exception of its variety which is caused by mental deficiency/retardation).

In *the Negative Equalization Model* neither mentally competent nor mentally incompetent can be held criminally responsible. This model must be accepted by those who assume a conception which deems free will to be a condition of criminal responsibility (*Conception 1* or *2*) and deny its (free will's) existence. This model was, for instance, accepted by the representatives of the so-called Sociological School in criminology; the most famous of them, the Italian thinker from the turn of 19th and 20th century Enrico Ferri, asserted that the notion of penalty cannot be legitimately used given the non-existence of free will; the measures taken by the state against the perpetrators of criminal acts have only a pragmatic justification: they serve '*difesa sociale*' – a self-defense of the society against potentially dangerous individuals. This model is also implied by the conception which treats being *causa sui* as a condition of criminal responsibility and denies the existence of *causa sui.*

The *Positive Equalization Model* assumes that both mentally competent and mentally incompetent can be held criminally responsible. This model can be based on various grounds. It is implied by *Conception 4* (objective/absolute responsibility), *Conception 6*, *Conception 1* (if it is assumed that free will exists and that the mentally incompetent exhibit it even though their actions are not voluntary), *Conception 7*, *Conception 2* in conjunction with the negationist account of mental illness and with the acknowledgment of the existence of free will, or *Conception 5* in conjunction with the negationist account of mental illness.[16] This model was not implemented in any legal system but its two weaker variants (which may be called 'partial-positive-equalization' models) were implemented either in a legal system and legal practice (Sweden) or at least in legal practice (USA). According to the first variant both the mentally ill and the mentally healthy can be held criminally responsible but only the mentally healthy can be punished; according to the second variant, only the mentally healthy can be held criminally responsible, but, in practice, the mentally ill are often punished. Let me discuss them in more detail.

4.1. Variant 1 of the partial-positive-equalization model (Sweden)

Sweden adopted in 1962 a rather odd solution, according to which the criminal court is obliged to decide, if the accused has committed a criminal act in the state of mental

[16] It should be noticed, though, that the negationist account undermines the very concept of internal coercion; in consequence, Conception 5 requires an appropriate modification.

incompetence, that he is *guilty/criminally responsible but mentally incompetent* (rather than, as in most other countries, those which assumed the Standard Model, that he is not guilty/not criminally responsible by reason of insanity), and subsequently, may punish him more leniently or not punish him at all but instead place him in mental hospital. As we can read in the Swedish Penal Code:

Ch. 29. Section (3) In assessing penal value, the following mitigating circumstances shall be given special consideration in addition to what is prescribed elsewhere, if, in a particular case:
1. the crime was occasioned by the grossly offensive behavior of some other person,
2. the accused, in consequence of a mental disturbance or emotional excitement, or for some other cause, had a markedly diminished capacity to control his actions,
3. the actions of the accused were connected with his manifestly deficient development, experience or capacity for judgment,
4. the crime was occasioned by strong human compassion or
5. the act, without being free from criminal responsibility, was such as is covered by Chapter 24.
The sentence imposed may be less severe than that prescribed for the crime in question if this is called for having regard to the penal value of the crime.

Ch. 30. Section (6) A person who commits a crime under the influence of a serious mental disturbance may not be sentenced to imprisonment. If, in such a case the court also considers that no other sanction should be imposed, the accused shall go free from sanction.

Ch. 31. Section (3) If a person who has committed a crime for which the sanction cannot be limited to a fine, suffers from a serious mental disturbance, the court may commit him for forensic psychiatric care if, having regard to his mental condition and personal circumstances, admission to an institution for psychiatric care combined with deprivation of liberty and other coercive measures, is called for.
If the crime has been committed under the influence of a serious mental disturbance, the court may decide that a special release inquiry under the Act on Forensic Psychiatric Care (1991:1129) shall be conducted during the time in care if there is risk for relapse into serious criminality of a serious kind by reason of the mental disturbance.
The court may, in conjunction with a committal to forensic psychiatric care impose another sanction, but not imprisonment or committal for other special care, if this is called for having regard to the previous criminality of the accused or for other special reasons.

Thus, the Swedish solution separates criminal responsibility from punishability, and abolishes the insanity defence. The mentally ill offenders are not acquitted: if they committed a crime they are declared guilty/culpable/criminally responsible; their illness is taken into account only at the stage of determining sanctions (though, arguably, forensic mental care is also regarded as a punishment, so the separability does not have a conceptual character). This also means that what is decisive for legal reaction is the mental state of the accused at the time of trial, not at the time of committing an offence.

According to the Polish sociologist and journalist Maciej Zaremba (2008) the Swedish regulation of the criminal responsibility of the mentally ill stops half-way (it does not lead to the full equalization of the criminal responsibility of the mentally ill and the mentally healthy) because it is an incoherent fusion of two different conceptions of criminal

responsibility. On the one hand, it was inspired by the Sociological School of criminology (especially by Enrico Ferri), which, as already mentioned postulated the abolition of the concept of criminal responsibility on the grounds that since determinism is true, no agent can be deemed guilty, and therefore no punishment is justified (the state can defend against perpetrators of crime but the measures taken by the state cannot be called punishment because they are not the realization of justice but '*difesa sociale*'). Ferri's conception influenced the thought of the Swedish psychiatrist Olof Kinberg who postulated that the concept of free will be eliminated (as fictitious) from legal considerations, and, consequently, also the concepts of guilt and responsibility. The state, assisted by the psychiatrists, was to isolate the most dangerous individuals, thereby realizing the utilitarian goal of maximizing social utility rather than the goal of retributive justice. This view was underlying the project of a penal code (called 'The Law of Protection') proposed by the Swedish minister of justice Karl Schlyter in 1956: the project made no avail of the concept of punishment, but that of 'consequences'; the consequences could be imposed also for an indefinite length of time; everyone (the mentally ill and the mentally healthy) could be put to criminal proceedings but no one could be declared guilty. But this project in this radical (but consistent) form was finally rejected.[17] On the other hand, the Swedish solution was influenced by the common-sense, standard view according to which the mentally ill should be treated differently from the mentally healthy: only the latter deserve to be punished. The Swedish – compromise – solution, which (unsuccessfully) tried to reconcile these two different conceptions assumes that both the mentally ill and the mentally healthy criminals are responsible for their acts but the distinction between them can be introduced at the stage of determining consequences of their actions. This is indeed paradoxical: the theory of Enrico Ferri implied that no one can be criminally responsible, while the Swedish code, even though it was influenced by this theory, implied that all (including the mentally ill) can be criminally responsible.[18] One might think that this solution, as far as its practical

[17] The deterministic view of human beings, as well as the blurring of the border between the competences of judges and psychiatrists, which were to underlay this project, were criticized, e.g., by the social-democratic politician Georg Branting in the article from 1943 'Legal corruption, or the fall of the state of law in Sweden', and by the professor of law Ivar Strahl in 1955, who was strongly against the rejection of the concept of guilt (cf. Zaremba 2008).

[18] It is worth noticing that a certain role in adopting this conception could have played the idea, advocated by some legal philosophers from the so called Scandinavian legal realism (e.g., Alf Ross), that 'responsibility' is a technical, empty, or functional concept, not carrying any moral weight, and that the notions of guilt',

consequences are concerned, does not differ from the Standard Model. But the reality is different: the Swedish solution has had several non-humanitarian consequences (cf. Zaremba 2008). Firstly, the accused who is mentally ill has to take part in the criminal proceedings. Secondly, if an agent committed the crime in a state of mental incompetence, but during the criminal process is healthy, then he will be punished (just like the criminal who committed the crime in a state of mental competence). Thirdly, a large number of the mentally ill convicts are punished rather than sent to mental hospitals.

If one considers the weaknesses and the very oddity of the Swedish solution, it seems somewhat surprising that a similar solution was proposed (independently of the authors of the Swedish criminal code) by one of the greatest 20th century legal philosophers Herbert L. A. Hart (2008, 1968), who opted for transferring the problem of mental competence of the accused (or, rather, the convict) to the post-trial stage; at the trial stage the judge was to resolve only the issues of *mens rea* (intent, negligence, or recklessness) and *actus reus*. Hart seems to have had, however, a different reason for adopting this solution than the authors of the Swedish criminal code. The reason was his specific conception of criminal responsibility (*Conception 7* on my list) according to which *mens rea* and *actus reus* are sufficient for holding an agent responsible even if his/her (criminal) act was not voluntary (in the sense of freedom from internal coercion).

4.2. Variant 2 of the partial-positive-equalization mode (USA)

The criminologists Patricia E. Erickson and Steven K. Erickson have advocated an interesting two-part thesis: (1) that in the USA there prevails

> a negative public sentiment and policy toward the mentally ill [caused by the fact that mental illness came to be regarded – W.Z.] as a failure of individual responsibility rather than an illness requiring a humane orientation (…) reframing mental illness as a failure of individual responsibility, along with a continuation of the traditional criminal law model, created a shift to a punitive stance toward the mentally ill – hence the criminalization of mental illness (Erickson, Erickson 2008: 8-9);

(2) that one of the important causes of this "reframing" of mental illness was the 'anti-psychiatric' movement.

As already mentioned, it may seem paradoxical (even though it is by no means illogical) that the 'humanistic', as one may call it, view of mental illness, which postulates

'culpability' or 'punishment' are metaphysical fictions and as such should be dispensed with. It should be stressed, though, that Ross himself was in favour of the Standard Model based on *Conception 5* (cf. Ross 1975).

that illness be regarded as a specific mode of experience to be analyzed from the 'inside' perspective of the ill, and which (on some interpretations) implies individual responsibility for mental illness and posits the existence of free will, could have had non-humanitarian consequences, i.e., could have left to the harsh treatment of the mentally ill (on the grounds that they are themselves to blame for their predicament, and that thereby mental illness is to be regarded as a moral failure), whereas the 'biological', 'reductionist', non-humanistic' account of mental illness is likely to lead to the humanitarian treatment of the mentally ill (on the grounds that they are not to be blamed for their illness). If, as it has happened in the USA, the 'humanistic' account is combined with the retributive theory of punishment, the treatment of the mentally ill is likely to be especially harsh. Thus, even though anti-psychiatry were to a large degree a reaction towards the non-humanitarian treatment of the mentally ill, it could itself have led to (as an unintended side-effect) – and arguably, did lead to – the decreasing of empathy towards the mentally ill (since if they are not mentally ill at all or are responsible for their mental illness, then the important grounds for empathy disappear).

Of course, the process which led to the domination of the "moral/punitive model" of treating the mentally ill was caused by many factors, including a genuine concern for the well-being of the mentally ill, not only by the view of mental illness as a failure of individual responsibility. The process started in the 1950s (and gained momentum in the 1970s) as a result of the critique of 'institutions' which, as was already mentioned in section 3.2., were often indeed cruel and dehumanizing (as for instance, described suggestively in the book by Ken Kesey *One Flew over the Cuckoo Nest*), e.g., patients in mental hospitals were applied cold baths, lobotomy, electroshocks: mental hospitals fulfilled not only a therapeutic function but also, as stressed by Erickson and Erickson (2008: 15), a "containment function". Erickson and Erickson, however, aptly noticed that "while the conditions in state mental hospitals rightfully came under attack, the conditions themselves do not explain deinstitutionalization and the movement to community care (2008: 30)". The previous cause (a moral indignation at the conditions prevailing in mental hospitals) was strengthened by the cultural revolution of the 1960s, which led to a growing emphasis on the importance of individual rights and personal autonomy and the concomitant protest against all – or at least all arbitrary – forms of state coercion. Furthermore, many medical experts cherished a strong conviction about the efficiency of the pharmacological treatment of mental illness,

which could be conducted outside mental hospitals. It was also believed that long-term stays in hospitals "produce institutional behavior and a tendency to chronic illness (Erickson, Erikcson 2008: 26)", and that thereby deinstitutionalization serves autonomy, liberty, and dignity of the mentally ill persons. The economic motive, connected with the fact that hospitals for the mentally ill were expensive, also played an important role. As a result of these variegated causes there took place the 'deinstitutionalization of mental illness' – "the process of moving mentally ill people out of large state mental hospitals (Erickson, Erickson 2008: 25)". As Edwin F. Torrey wrote in his book *Out of the Shadows. Confronting America's Mental Illness*, this was one of the greatest social experiments in the history of the USA: while in 1955 there were about 560 000 patients in mental hospitals (at the population of 164 million), in 1994 there were only 72 000 (at the population of 260 millions); the difference, as noticed by Erickson and Erickson, will prove to be even higher if we take into account the change in population (if the population in 1955 were such as it was in 1994, the number of patients in mental hospitals would rise to 885 000); it means that "approximately 92% of the people who would have been living in public psychiatric hospitals in 1955 were not living there in 1994 (Erickson and Erickson 2008: 27-28)". It bears also emphasizing that also the most seriously mentally ill (especially schizophrenics, about 50-60% of them) were discharged from the hospitals. What were the consequences of this 'social experiment'? It was believed that the mentally ill will take care of themselves and will be taken care of by the local communities. But this expectation was unduly optimistic. The mentally ill who were discharged from hospitals often started to take drugs and abuse alcohol, they frequently refused to participate in the pharmacological treatment (because the mentally ill rarely regard themselves as ill), and, finally, many of them turned to crime. As a result, there took place what Torrey called a 'transcarceration' of the mentally ill from asylums to prisons (which proved to be a disastrous alternative to them as compared with mental hospitals, which is attested for instance by a high suicide rate among the incarcerated mentally ill). Erickson and Erickson aptly noticed that "the proponents of deinstitutionalization (…) created a climate that was less tolerant of the mentally ill and more willing to use prisons and jails as their places of confinement (Erickson and Erickson 2008: 25); and in similar vein: "many of the older commitment laws were predicated on recognizing this difference [between the mentally ill and the mentally healthy –WZ.], while much of the 1960-70

"liberation" from psychiatry was guided by seeing the mentally ill as the same as everyone else (*ibidem*, 194)".

To sum up (somewhat crudely) the above description of changes in the situation of the mentally ill: since 'deinstitutionalization of mental illness' was, in large part, based on unduly optimistic premises (e.g., that the mentally ill can take proper care of themselves), it entangled the mentally ill in criminal behavior, for which they were held responsible, just like the mentally healthy (which was a result of the reinterpretation of mental illness as a failure of individual responsibility); in consequence, there took place 'transcarceration'.

It is clear that one should not interpret the above analyses one-sidedly – as justifying an overall negative evaluation of the changes in the approach to the mentally ill that took place in the USA in the 1960-70s, and as postulating a return to the state from before the 'anti-psychiatry' revolution. The truth is, of course, much more complex. One should not underestimate the positive effects of 'anti-psychiatry'.[19] One should also remember that 'ant-psychiatry' did not engender negative effects by itself but jointly with a number of other causes (e.g., the belief in the efficiency of the pharmacological treatment of the mentally ill), which may have played an even more important role.

5. Evaluation of the models

Even though the main goal of this paper was to distinguish various models of the criminal responsibility of the mentally incompetent and to identify their philosophical underpinnings, one cannot stop here and refrain from at least attempting at an evaluation. In point of fact, my presentation of various models, especially of the models of criminal responsibility, was critical, not purely descriptive. But the evaluation made so far must be supplemented by some additional remarks.

Let me first focus on the first dimension of the models – various conceptions of criminal responsibility. I have already argued that the *causa sui* requirement seems to be too strong, and the 'consequentialist' requirement seems to be too weak. Therefore the real choice, as far as the conditions of responsibility are concerned, is between, one the one

[19] To give one of many examples in Italy, the movement *Psychiatria Democratica*, founded by the psychiatrist Franco Basaglia, led to the adoption of the law in 1978 (the so called Basaglia's law) which improved the situation of the mentally ill, protected their civil rights.

hand freedom of will *cum* freedom from coercion on the hand, and freedom from coercion on the other. Now, while analyzing the requirements of criminal responsibility explicitly stated in most contemporary penal codes, which include, e.g., certain perceptual, conative, evaluative and volitional capacities, one will notice that none of them implies the metaphysical capacity to act otherwise, i.e., free will. It seems therefore that *penal codes do not treat free will as the condition of criminal responsibility*; freedom from coercion is presupposed by them to be a sufficiently strong concept of freedom. But one could, of course, disagree with this presupposition, claiming that an agent cannot be held responsible if he could not have acted otherwise, i.e., if he is not endowed with free will. This argument seems apt in so far as the justification provided by the free will is undoubtedly stronger than the one provided only by freedom from coercion. The question is whether the latter is *sufficient*. In my view, even though it may not be sufficient for moral responsibility (understood as responsibility in some deeper sense) it is entirely sufficient for criminal (legal) responsibility. Interestingly, the question of the responsibility of the mentally incompetent may be a test case here. We could probably say that, from a (deeper) moral point of view, neither the mentally competent nor the mentally incompetent are responsible (e.g., if free will – treated as a condition of moral responsibility – does not exist, or if we treat being c*ausa sui* as a condition of moral responsibility and deny that a human being can ever be *causa sui*). But it can hardly be denied that there is a normatively-relevant difference between the two categories of perpetrators, and that this difference must be taken into account at some level of responsibility (the level of *legal* responsibility being a natural candidate here).

As for the second dimension, the negationist view is only partly justified, and, in addition to that, it is justified in its least controversial part. It is undoubtedly true that certain types of behavior, e.g., masturbation, trans-sexuality, homosexuality, or drapetomania ("the 'disorder' that afflicted slaves who ran away from their masters (Wakefield 1993: 373)") were regarded as mental illnesses only because they were inconsistent with the prevailing social norms. But, needless to say, many other mental illnesses (e.g., schizophrenia) cannot be interpreted in this way. In other words, the argument from the lack of evidence for the physical/material basis of, say, schizophrenia, turned out to be the weakest part of anti-psychiatrists' (especially Szasz's) argumentation for the thesis that mental illness is a myth (*while being at the same its most important part*). Szasz may have been right in the sense

that at the time in which he proposed his theory, the physical bases of mental illnesses were not known, so he could be *to some extent* justified in saying that they do not have such bases But he should not have excluded the possibility that they may be discovered in the future (as it eventually happened: contemporarily our knowledge about the physical bases of many mental illnesses, though far from being perfect, is large enough to justify a peremptory rejection of Szasz's claim that mental illnesses have no physical bases). A more specific criticism of his theory, provided, among others by, Wakefield (1993), Kendell (2005), and Shorter (2011), concerns his lesion or morphological abnormality account of disorder; as Wakefield noticed:

> The account consists of two theses: (a) that a lesion (or abnormal bodily structure) is simply a statistical deviation from a typical anatomical structure and (b) that a physical disorder is simply a lesion. First, the idea that a lesion can be directly recognized by its deviant anatomical structure is incorrect. Bodily structures normally vary from person to person, and many normal variations are unusual as any lesion. Moreover, some lesions are statistically nondeviant in a culture, such as artherosclerosis, minor lung irritation, and gum recession in American culture and hookworm and malaria in other cultures. Therefore, recognition of a lesion is not simply a matter of observing anatomical deviance. Second, and more importantly, it is not the existence of a lesion that defines disorder. (…) A lesion can be a harmless abnormality that is not a disorder, such as when the heart is positioned on the right side of the body but retains functional integrity (Wakefield 1992: 375).

But this criticism, though important, is secondary: Szasz's main point was that mental illness does not have a *material* basis, so even if he worked out more carefully the concept of the 'material basis' of illness to counter the objection of Wakefield, it would still be untenable in the light of contemporary knowledge about mental illnesses.

In summary: after this long 'tour' through various models of criminal responsibility of the mentally incompetent, we return to the Standard Model as the most cogent one. This model seems to gain additional support form Peter Strawson's thesis that the distinction between the responsibility of the sane and insane cannot be abolished because it is 'built' into our reactive attitudes (or more precisely: in the way in which our impersonal reactive attitudes function) and cannot be easily (or, rather, cannot be at all, at least in the long run) dispensed with. But Strawson is less convincing when he says disparagingly about "the obscure and panicky metaphysics of libertarianism" based on the concept of "contra-causal freedom", i.e. freedom of will (cf. Strawson 1962: 23-25). *Contra* Strawson, I have argued that the justification of responsibility ascriptions may be graded: moral responsibility (in some deeper sense) may require a stronger justification (arguably based on the concept of free will) than criminal (legal) responsibility.

Bibliography

Basaglia, F. (ed.) (1973), *L'Istituzione Negata*, Torino: Einaudi.

Baumeister, R.F. (2001), *Evil. Inside Human Violence and Cruelty*, New York: Henry Holt and Company.

Cooper, D. (1967), *Psychiatry and Anti-Psychiatry*, London: Tavistock.

Dryla, O. (2018), *Pojęcie doskonalenia w sporach o cele medycyny*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Erickson, P.E., Erickson, S.K. (2008), *Crime, Punishment, and Mental Illness*, New Brunswick, New Jersey: Rutgers University Press.

Eshleman, A. (2016), "Moral Responsibility", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/moral-responsibility/>.

Foucault, M. (1971, 1961), *Madness and Civilization*, London: Tavistock.

Goffman, E. (1961), *Asylums. Essays on the Social Situation of Mental Patients and Other Inmates*, New York: Random House.

Hart, H.L.A. (2008, 1968), *Punishment and Responsibility*, Oxford: Oxford University Press.

Kendell, R.E. (2005), 'The Myth of Mental Illness'. In: J. Schaler, *Szasz Under Fire: The Psychiatric Abolitionist Faces His Critics*, Illinois: Open Court, pp. 29-48.

Kotowicz, Z. (1997), *R.D. Laing and the Paths of Anti-Psychiatry*, London and New York: Routledge.

Laing, R. D. (1967), *The Politics of Experience and The Bird of Paradise*, Harmondsworth: Penguin Books.

Laing, R.D. (1969, 1960), *The Divided Self. An Existential Study in Sanity and Madness*, Middlesex, England: Penguin Books.

Laing, R.D. (1985), *Wisdom, Madness and Folly. The Making of a Psychiatrist*, London: Macmillan.

Ross, A. (1975), *On Guilt, Responsibility, and Punishment*, University of California Press, Berkeley and Los Angeles.

Shorter, E. (2011), 'Still Tilting at Windmills: Commentary on *The Myth of Mental Illness*', *The Psychiatrist* 35(5), pp. 183-184.

Siegler, M., Osmond, H., Mann, H. (1969), 'Laing's Models of Madness', *The Britisch Journal of Psychiatry*, 115 (525), pp. 947-958

Strawson, G. (1994), 'The Impossibility of Moral Responsibility', *Philosophical Studies* 75 (1-2), pp. 5-24.

Strawson, P. (1962), 'Freedom and Resentment'. In: G. Watson (ed.), *Proceedings of the British Academy*, Volume 48, Oxford: Oxford University Press, pp. 1-25.

Szasz, T. (1974, 1960), *The Myth of Mental Illness: Foundations of a Theory of Personal Conduct*, New York: Harper&Row Publishers.

Szasz, T. (1976), 'Anti-Psychiatry: The Paradigm of a Plundered Mind', *New Review, 3(29).*

Torrey, E. T. (1997), *Out of the Shadows. Confronting America's Mental Illness*, London: Wiley.

Twardowski, K. (1983), 'Etyka i prawo karne wobec zagadnienia wolności woli', *Etyka* 20, pp. 123-157.

Wakefield, J. C. (1992), 'The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values', *American Psychologist* 47(3), pp. 373–88.

Wootton, B. (1963), *Crime and the Criminal Law*, London: Steven&Sons.

Zaremba, M. (2008), 'Sceny łowieckie z domu wariatów', transl. W. Chudoba. In: M. Zaremba, *Polski hydraulik i inne opowieści ze Szwecji*, wydawnictwo czarne, Wołowiec, pp. 175-229.