EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Publisher, Editorial office
University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 669 794
E-mail: eujap@ffri.hr

Web address: https://www.eujap.uniri.hr

Printed by Impress, Opatija (Croatia)
30 copies

TABLE OF CONTENTS

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 165.172:616.89
https://doi.org/10.31820/ejap.16.2.1

RATIONALITY IN MENTAL DISORDERS: TOO LITTLE OR TOO MUCH?

Valentina Cardella
University of Messina

Original scientific article – Received: 14/4/2020 Accepted: 17/7/2020

ABSTRACT

The idea that mental illnesses are impairments in rationality is very old, and very common (Kasanin 1944; Harvey et al. 2004; Graham 2010). But is it true? In this article two severe mental disorders, schizophrenia and delusional disorder, are investigated in order to find some defects in rationality. Through the analysis of patients' performances on different tests, and the investigation of their typical reasoning styles, I will show that mental disorders can be deficits in social cognition, or common sense, but not in rationality (Sass 1992; Johnson-Laird et al. 2006; Bergamin 2018). Moreover, my claim is that psychopathological patients can also be, in some circumstances, more logical than normal controls (Kemp et al. 1997; Owen et al. 2007). From a philosophical point of view these data seem to be very relevant, because they help us to reconsider our idea of rationality, and to challenge the common way to look at sanity and mental illness.

Keywords*: Rationality; schizophrenia; delusional disorder; common sense*

1. Introduction

> What is meant by 'rational?' Whatever it is, mental disorders are shortcomings or departures from it, and only those disorders which involve the absence of it are to count as mental disorders. (Edwards 1981, 314)

In this essay, significantly called "Mental Health as Rational Autonomy", Edwards displays the common conceptualization of madness shared by

Correspondence: vcardella@unime.it

both psychology and philosophy: in people with mental disorders reason is severely affected, and, on the other side, emotions are abnormal and unrestrained (Edwards 1981). The idea that in mental disorders there is a loss of reason is very old and dates back at least to the Enlightenment. The prevailing paradigm of insanity, back then, was that people with mental disorders had to be treated like beasts (tied up with stripes and chains, beaten, constantly terrorised, forced to endure inhuman treatments like surprise baths and so on), because they *were* beast, in some sense. Actually, the conception of madness, as noted by Scull, was characterized by almost exclusive emphasis on disturbances of the reason, and this "was to imply that in losing his reason, the essence of humanity, the madman had lost his claim to be treated as a human being" (Scull 1981, 108).

For this reason, psychopathological subjects have always been considered as people who lack autonomy, rationality, responsibility, and, in some sense, discipline, because they are not able to control their reactions and emotions. As stated by Edwards:

> Mental illness means only those undesirable mental/behavioral deviations which involve primarily an extreme and prolonged inability to know and deal in a rational and autonomous way with oneself and one's social and physical environment. In other words, madness is extreme and prolonged practical irrationality and irresponsibility. (Edwards 1981, 312)

This idea is perfectly consistent with the attitude of common people who still consider patients with mental disorders as weak, unpredictable, irrational, irresponsible, and even dangerous.

The cognitive version of this ancient idea, which dates back to Beck's work (1976), is well conceptualized by Harvey et al. (2004). The authors aim to identify the cognitive impairments shared by psychological disorders, and that clearly characterize mental disorders in general, thereby creating a sharp line between normal and pathological subjects. The central idea of this paper is that psychopathologies imply various deficits in different cognitive abilities, such as reasoning (interpretative biases), memory (selective retrieval), attention (self-focus), thought processes (rumination), beliefs (confirmation bias) and behaviour (avoidance, safety conducts). For example, the authors describe the case of a schizophrenic patient, Henry, who interprets the coughing of strangers around him as a signal that his thoughts have been broadcast, and starts to avoid looking at anybody because this feeling makes him very anxious. The patient concentrates on his own thoughts in order to avoid every external signal, and ends up never leaving his home for several days. The case is described as a series of

attentional anomalies: an attentional bias (selective attention) makes Henry select certain stimuli (coughing) instead of others (conversation), while another attentional bias (avoidance) causes Henry to escape environmental cues, and this avoidance at last produces another anomaly, namely self-focused attention. In this perspective, the different mental disorders' features are conceived as the result of different impairments at the cognitive level, and therapy would thus consist in correcting the errors committed by psychopathological subjects.

Thus, people with mental disorders usually make many mistakes, which, besides impairing their different reasoning and cognitive abilities, make the way they think and reason essentially different from the 'normal' one. A large amount of recent research, though, has started to challenge this view in a significant way. On the one hand, studies on reasoning in very different disorders (anxious disorders, mood disorders, psychoses) seem to show that people who suffer from psychological disorders are not bad at reasoning, but rather they follow the same rules as healthy people (Smeets et al. 2000; Harvey et al. 2004; Mancini et al. 2007).

On the other hand, the idea that 'normal' people are usually rational has been disputed. Starting from the pioneering work of Wason (1966), a wide range of experiments and tasks show that if we consider rationality as a possible synonym of logic, we are actually largely irrational, because we usually don't follow the formal rules of logic (Kahneman et al. 1982; Johnson-Laird 1983, 2006; Evans, 1989, 2002; Johnson-Laird and Byrne 1991; Rips 1994; Baron 2000; Smeets et al. 2000). It seems that our way of thinking is heavily influenced by our beliefs, by the aims we have, by the context in which the performance occurs, and thus, we commit more logical errors than expected.

But, is rationality just a matter of formal logic? In other words, how can we define rationality in a broader sense? This will be the topic of the next section.

2. What Does it Mean to be Rational?

Rationality is an umbrella term that obviously encompasses many different aspects. I will start with another quotation by Edwards, who focuses on the common notion of rationality, that is what matters the most here, and tries to highlight its defining elements:

> […] there is widespread agreement among both philosophers and non-philosophers that rationality involves (1) being able to

> distinguish means from ends and being able to identify processes and manifest behaviors which likely will result in the realization of consciously envisioned goals; (2) thinking logically and avoiding logically contradictory beliefs; (3) having factual beliefs which are adequately supported by empirical evidence, or at least avoiding factual beliefs which are plainly falsified by experience; (4) having and being able to give reasons for one's behavior and beliefs; (5) thinking clearly and intelligibly, and avoiding confusion and nonsense; (6) having and exhibiting a capacity for impartiality or fat mindedness in judging and adopting beliefs; (7) having values which have been (or would be) adopted under conditions of freedom, enlightenment, and impartiality. (Edwards 1981, 314-15)

This seems to be what common people think of rationality, and everyone can see himself/herself in this definition. Or can s/he not? Actually, definitions like number 6 leave some of us a bit uncomfortable. Are we always impartial in judging beliefs? Or do we have the tendency to judge in a more positive way the beliefs that match the most what we already think? Even if we are not experts in cognitive psychology, we find the idea that we treat the beliefs we share and those we don't in the same way difficult to sustain. The elements contained in definition 7 are also controversial. Do we adopt our values freely or have our values been, for the most part, instilled in us by our parents either directly or indirectly? And what about number 5, according to which rationality involves thinking clearly and intelligibly? Yes, we try to avoid nonsense, and yes, we try to think clearly, but none of us could be sure to always succeed in this task. And being able to give reasons for what we believe or do, as stated in number 4, is not so simple, and often the reasons we adduce don't correspond to the truth, because we often don't know what we want, and we act under the power of unconscious desires and forces. I will omit the comments on the definition number 3 for now, because cognitive and experimental psychology have many things to say on the way in which we confirm and maintain our beliefs, as we will see in the third section.

But those objections are only philosophers' concerns, and are too specious, too contorted, as claimed by Edward himself, and can be easily sidestepped. We can make mistakes, we can be confused or biased, at times, but we are not irrational. And if we think of people suffering from mental disorders, they seem to violate all of these defining elements.

> Most people are not very rational, but most people are nevertheless sane. Extreme departures from sanity are not as

> difficult to identify in practice as some sceptical critics, especially lawyers and philosophers who have never spent any time around mentally disturbed persons, would have us to believe. Cases on the borderline of such extremities are the ones which understandably give headaches to mental health professionals, but such professionals can also cite many clear cut cases involving extreme and prolonged incompetence and self-defeating performances in selecting effective means to avowed ends, of radically inconsistent practical belief systems, items of which are plainly controverted by empirical facts, of inability to cite reasons for belief and behavior, of persisting and pervasive conceptual confusions, and of intrenched inabilities to adopt fair minded perspectives on either factual or valuational beliefs. (Edwards 1981, 315)

Philosophers seem to doubt the possibility of drawing a sharp line between madness and sanity, but philosophers doubt everything, says Edwards, it's their vocation, while in most cases recognizing mental disorder and its irrationality is rather easy (Edwards 1981). Thus, since rationality seems to be clearly affected in mental illness, irrationality should be one of the core features of mental disorders' very definition. After all, mad people are mad because they are irrational, and madness is the realm of nonsense, so irrationality is supposed to be a crucial part of the definition of madness. Let's now have a look at the way in which mental illness is defined in different professional diagnostic manuals, such as DSM-5 or ICD-10. Let me say something first; the definitions I will discuss have been conceived only for practical (medical) purposes, because it is not clear that the category of mental disorder corresponds to a natural kind (for discussion, see Beebee and Sabbarton-Leary 2010; Kendler et al. 2011; Kohne 2015; Zachar 2015). So, I will not assess the validity of such definitions here, as it lies outside the scope of this paper. I will only investigate the role of rationality in some of the descriptions of mental disorders.

I will begin with the 'Bible' of psychiatry, the Diagnostic and Statistical Manual of Mental Disorders, published by the American Psychiatric Association (APA 2013). Here, one can find the following definition of mental disorder:

> A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or

> other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behaviour (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above. (APA 2013, 20)

This definition clearly reflects the neurobiological perspective which has increasingly gained popularity within American psychiatry. The basic claim is that in mental disorders there is always some kind of dysfunction that, in a bottom-up fashion, causes the clinical symptoms. As one can see, this description refers to many aspects of human cognition and behaviour, like emotion, distress, disability in social activities, deviant conduct, while rationality is not mentioned. It is worth noting that the most important thing in defining mental disorder is, apart from the set of symptoms, the degree of distress created by them, and the amount of freedom they leave to lead an ordinary life.

This more subjective element is particularly stressed in another perspective which deals with mental illness, that of psychopathology. While psychiatry aims at treating people suffering from mental disorders, psychopathology has a more theoretical approach, in that it investigates the nature and origin of mental disorders, as well as the possibility of understanding them. In his General Psychopathology, Scharfetter (1980) claims that the core elements of mental illness are the following three: suffering, failure, and alienation. First of all, suffering from a mental disorder means suffering, indeed; the ill person "suffers from himself and from the world to an extent that is qualitatively and quantitatively beyond the average" (Scharfetter 1980, 8). The second aspect is failure: people with mental disorders, whilst often coping with not very extreme circumstances, fail to conduct themselves in life and in the world (ibid.). But the third aspect is also important, and that is alienation. A mentally disordered person is not normal, and this is often immediately recognized; as pointed out by almost every exponent of philosophical psychiatry, their deep alterity makes people with mental disorders very isolated, and prevents them to establish vital contacts with other people.

But what exactly makes the mentally disordered people so extraneous and so different? A possible answer is the lack of rationality. People with mental illnesses are alone, extraneous, failing to live in the world because they are irrational, they can't think properly, and we can't predict their actions and feelings because they have no sense, they lack an inner reason.

Man is a rational animal, and a mad man is not. Scharfetter (1980) doesn't mention rationality, yet, and generally speaking it is very hard to find a scientific definition of mental illness which refers to illogicality or irrationality.

There are two kinds of reasons behind this absence. First, man is not a rational animal, after all. Irrationality is widespread in the general population. We are far from being rational, especially if by rationality we mean logic. When we reason, we make errors, which are not rare, but rather systematic, because our rationality is bounded (we don't have enough cognitive resources to always apply correct strategies, see Simon 1991), and because we use heuristics, i.e. quick mental shortcuts which reach satisfactory, even if not rational, solutions (Kahneman et al. 1982; Girotto 1994).

But also if by rationality we mean other things, as those suggested by Edwards (1981), like having beliefs that are consistent and supported by facts, it seems that the vast majority of us, not to say all of us, is not rational. We hold beliefs that are not supported by evidence, or that break the norms of rationality: superstitious beliefs, racist beliefs, magic beliefs, paranoid beliefs. As claimed by Bortolotti (2013), although these beliefs are not supported by evidence, and fail to meet the criteria of epistemic rationality, they are not pathological. Therefore, it is not in our supposed rationality that we can find the demarcation line between sanity and madness.

But, even if we have to face the fact that we are not as rational as we thought, the idea that, however irrational we are, people with mental disorders are much more irrational than us, still lingers on. This is exactly what I will call into question here. Rationality and mental sanity do not overlap in this way either, because one can find many cases, much more than expected, of people who are perfectly rational while being affected by some kind of mental disorder. And this applies not only to mild mental disorders but also to severe psychoses, like schizophrenia and delusional disorder, as I show in the next sections.

3. Too Rational to be Sane, Part 1: Schizophrenia

The typical features of schizophrenia, such as hallucinations, delusions, disorganized speech and behaviour, make this mental disorder one of the most severe, to the point that people with schizophrenia embody the common representation of irrationality: schizophrenics say things that make no sense, believe in unbelievable things, behave in a bizarre and

often incomprehensible way. Thus, the answer to the question "are schizophrenic patients irrational?" seems to be crystal clear. And admittedly, the first studies on reasoning abilities in schizophrenia seem to confirm the presence of an impaired logic in those patients. The works by von Domarus (1944), corroborated by Arieti (1964), showed that schizophrenics break the rules of conventional logic, like the principle of non-contradiction, or the identity principle, and for this reason have many difficulties in deductive reasoning.

But more recent studies failed to validate these data (Williams 1964; Belvin 1964; Wason 1966; Maher 1992; Kemp et al. 1997), showing no significant differences in performances on syllogistic reasoning between schizophrenic patients and control subjects. It seems that when schizophrenic patients make mistakes about the judgment of a syllogism's validity, they do it because of a general weakness in cognitive performance (e.g., a lower I.Q.), rather than because of a specific impairment of schizophrenia (cf. Mirian et al. 2011; Revsbeck et al. 2015). The same applies to conditional reasoning, that is, the form 'if p, then q', which is very frequent in everyday situations. For instance, the work by Kemp and colleagues (1997) showed that, when the tasks' content on conditional reasoning was neutral, there was no relevant difference in the performance of schizophrenic patients and controls.

Furthermore, the vast majority of studies on reasoning abilities in schizophrenia focuses on probabilistic reasoning, because the tendency to 'jump to conclusions', used to explain delusional thinking in those patients, seems one of the core features of schizophrenics' reasoning style (Garety et al. 1991; Bentall 1994; Garety and Hemsley 1994; Bentall et al. 2001; Freeman et al. 200; Garety et al. 2005; Freeman 2008). Delusional patients seem to show the tendency to request minimal information in situations where information is available, and to report a high level of confidence in their decisions. But also in this kind of reasoning, the differences in performances between patients and controls usually disappear when other variables, like memory, are included in the tasks. This probabilistic bias might be due to impairments in working memory or executive functions, rather than schizophrenia (Cardella and Gangemi 2014).

Thus, recent studies are beginning to show that there are no comparable reasoning deficits in schizophrenia, and that, when these are present, they are linked to a more general cognitive deterioration. But the most relevant fact is that, in some cases, schizophrenic people seem to reason even better than control subjects, or, in other words, they seem to be more rational than healthy people. For example, with respect to deductive reasoning, Owens and colleagues (2007) tested both "pure reasoning", using valid and invalid

syllogisms, and common sense, using syllogistic content that conformed to practical knowledge or departed from it. The tasks contained syllogisms that were valid but against common sense, and invalid but commonsensical. In these experiments, the idea was to see how schizophrenic subjects behave when common sense and logic conflict. Results show that people with schizophrenia perform even better than controls in the first series of syllogisms, the non-common sense ones, since they succeed in recognizing the validity of an argument also when its conclusion goes against common sense.

Better performances have been shown by schizophrenic subjects in conditional reasoning, as well. Mellet and colleagues (2006) administered a reasoning task to schizophrenic patients and healthy participants, where the subjects had to falsify conditional rules. The results showed that control subjects exhibited a reasoning bias linked to the formulation of the conditional rule, while schizophrenics didn't get diverted by the formulation and didn't get caught in the heuristics traps that lead normal subjects to give the incorrect answer.

Another case where schizophrenics' performances are better than those of healthy people regards probabilistic reasoning. Kemp et al. (1997) administered a probabilistic task to both control and schizophrenic subjects, where, after hearing a description of four people, subjects were asked to judge the likelihood of different alternatives regarding their possible employment. Control subjects use the representativeness heuristic to make judgments (Tversky and Kahneman 1982), that is, tend to choose the alternative that most closely matches the description, but in doing so, they fall victim to the conjunction fallacy, because they tend to value the conjunction of two events as more likely than one of the events alone. Schizophrenics gave more correct responses, because they didn't value the representativeness of the descriptions and proved therefore to be less sensitive to the conjunction fallacy.

But why are schizophrenic people more logical than healthy ones in those cases that encompass many different kinds of reasoning? My claim is that what represents a severe impairment in schizophrenics' daily life, namely the loss of common sense, becomes an advantage in the experimental conditions involving the tasks I described above. The fact is that, when common sense conflicts with logic, as one can see not only in the experiment by Owen et al. (2007), but also in everyday life, we tend to privilege common sense. But common sense is exactly what is at stake in schizophrenia. The typical schizophrenic attitude towards the self, the others and the world is marked by a detachment from common sense, or, in other words, from the web of beliefs, attitudes and categories which

represents, on the one side, the grounds of the self, and, on the other side, the background of everyday life. Thus, schizophrenia can affect the inner sense of self, as showed by the following remark made by a schizophrenic patient:

> I am no longer myself (…) I feel strange, I am no longer in my body, it is someone else (…) I walk like a machine; it seems to me that it is not me who is walking, talking, or writing with this pencil. (Parnas and Handest 2003, 126-127)

Or it can involve the world, as in this example:

> I live in a sort of bubble, where the world does not matter. I lack synchrony with the people around me. (Henriksen and Nordgaard 2014, 437)

Blankenburg (1971) claimed that the central deficit of schizophrenia, that is usually hidden by the more striking positive symptoms like hallucinations or delusions, regards the loss of natural evidence. What would normally seem self-evident causes amazement and perplexity in schizophrenic subjects that find it very hard to cope with everyday practical and social activities. As declared by a schizophrenic patient "I have to do everything anew from the beginning" (Nagai 1990, 363). Anna Rau, the patient made famous by Blankenburg, says:

> It is just a matter of mere feeling, sensing what is appropriate. One has this from nature…it is such a strange feeling, when one doesn't know the simplest of thing! (Blankenburg 1971, 308)

But what I would like to stress here is that these patients, in order to *compensate* this detachment from common sense, rely precisely on rationality. In other words, the reaction to the loss of this intuitive attunement is "an intellectual, introspective, metacognitive type of hyper-reflexivity" (Pérez-Álvarez et al. 2016, 2). Some examples will clarify this crucial point.

> C. D. B. is a 27-year-old insightful man with schizophrenia. He says that nothing is obvious to him; everything can be uncanny. The world is complicated, difficult to understand: 'Why does this happen? What does that mean? How to explain these facts? Why?' Facts are not self-evident. Only explanations can give a shape to his experiences. He feels the need for a general theory

> that makes the world understandable and his actions possible. (Stanghellini 2000, 777)
>
> I study people. I am curious. I want to understand how they are inside. (Mancini et al. 2014, 431)
>
> In my case, everything is just an object of thought. (Blankenburg 1971, 79)

This reliance on rationality, this intellectual attitude toward the world, is something that the philosophical perspective on psychopathology has stressed in different manners. Minkowski (1927) called morbid rationalism "the effort to submit some or all aspects of life under schematic and often algorithmic rules, […] deviating from a common sense attitude" (Urfer-Parnas 2019, 104). Binswanger (1956) pointed out that it is the excess of logic that often leads schizophrenic patients to act in a very bizarre way. Sass stated that schizophrenia, and its typical hyper-reflexivity, is "an alienation not from reason but from the emotions, instincts and the body" (Sass 1992, 4).

Thus, we can say that the problem of schizophrenics is not that they are irrational. Schizophrenia, the most severe among the most severe mental disorders, i.e. psychoses, doesn't seem to affect rationality after all, and, on the contrary, schizophrenics can be more logical than healthy people, and often rely on their intellectual faculties to compensate the deficits typical of the disorder itself.

4. Too Rational to be Sane, Part 2: Delusional Disorder

Mrs. A, a subject with delusional disorder, was examined by a forensic psychiatrist after she had killed her infant. This is how she justified this action:

1. When we die our souls are judged,
2. They are judged on the basis of our actions and decisions,
3. My baby has neither made decisions nor performed actions.
4. Therefore she did not have a soul.
5. Therefore it did not matter that I killed her. (Gillett 1990, 319)

The author comments this example with those words: "this chain […] is itself clearly insane despite the fact that the irrationality is hard to discern", and concludes this way: "an abiding theoretical problem for psychiatry and philosophy is that, on most accounts of rationality, a severely deluded

patient may qualify as quite rational" (Ibid.). This could seem a paradox, because delusions, by definition, seem to be the typical example of irrationality. So, how could it be possible for delusional patients not to be irrational?

Let's begin with the definition of the disorder itself. The main criteria for delusional disorder are the following:

A. The presence of one (or more) delusions with a duration of 1 month or longer.
B. Criterion A for Schizophrenia has never been met [that is, apart from delusions, no other symptoms of schizophrenia, like hallucinations, disorganized speech or behaviour, are present].
C. Apart from the impact of the delusion(s) or its ramifications, functioning is not markedly impaired, and behavior is not obviously bizarre or odd. (APA 2013, 297.1)

Delusions are fixed beliefs that are not amenable to change in light of conflicting evidence. Depending on its content, delusional disorder can be of the persecutory type (the individual believes to be conspired against, cheated, spied on, followed, poisoned, maligned, harassed and so on), erotomanic type (the central theme of this delusion is that another person is in love with the individual, usually of higher status), grandiose type (the conviction of having some great, but unrecognized, talent or insight), or jealous type (the belief that the spouse or lover is unfaithful).

Delusions could also be bizarre, that is, clearly implausible, and not deriving from ordinary life experiences, like the conviction to be able to read other's thoughts, or to have one's internal or external organs replaced by persecutors. While in these cases it's easy to recognize that someone is delusional, in other cases the "distinction between a delusion and a strongly held idea is […] difficult to make and depends in part on the degree of conviction with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity" (Ibid.).

At any rate, apart from the theoretical difficulties regarding the definition of delusion, it is sure enough that, when facing someone who holds delusional beliefs, we can often discern them as suffering from delusions. How is it possible? Is it because we immediately recognize the irrationality of these kinds of beliefs?

At first glance, an unusual belief, that is poorly supported by evidence and resistant to change, seems a clear example of irrationality. Many authors (see for instance Bermudéz 2002; Gilleen and David 2005; McKay et al.

2005) interpret this feature as the proof that delusions are typical cases of epistemic irrationality. Irrationality, in other words, doesn't concern the delusion's content *per se* (in delusional disorder the content is often plausible), but the way in which the individual holds it and maintains it despite contradictory evidence. As claimed by Speechley and Ngan:

> The hallmark of delusional beliefs is their persistence in the face of overwhelming contradictory evidence. It is this feature that sets them apart from normal erroneous beliefs, and it is this feature that sets them apart from normal psychology, moving delusional belief systems into the realm of psychiatric and medical pathology. (Speechley and Ngan 2008, 1211)

In other words, delusional beliefs appear to be experienced as self-evident, and patients seem unable to even contemplate the idea that their beliefs might be incorrect. To give a clearer idea of how delusional patients behave when asked to talk about their delusions, or when the content of their delusional beliefs is challenged, I will present a few examples:

> Psychiatrist (PS): How did you realize that you were decomposing?
> Patient (PA): Because of the strong smell.
> PS: Ok, but if one is decomposing, parts of the flesh should be missing, should they not? Because when the worms enter a dead body, I don't know, in a forest, and nobody finds it there, for three months, the worms….
> PA: I put hydrochloric acid on it, the bleach kills them.
> PS: Where do you put it?
> PA: On those parts where they grow.
> PS: Did you put hydrochloric acid on your skin?
> PA: Yep.
> […]
> PS: So… do these worms eat organs too?
> PA: I think so.
> PS: How do you survive then, when these worms eat your organs?
> PA: Well, how do I know?
> […]
> PS: Your parents, what do your parents for example say?
> PA: They say that it is not true…
> PS: Okay, they say that it is not true. And the fact that they say that it is not true, does not make you think that it is possible that it is not true, that it is perhaps rather your perception that you have worms in the body?
> PA: I am decomposing.

> PS: Why did you feel at the centre of attention? What did you notice that gave you the impression that you were the centre of attention?
> PA: Well, I felt like that for very long time, and now I understand.
> PS: What made you understand it?
> PA: From the content of my thoughts.
>
> Ps: Do you think there is any slightest chance that this is something you are exaggerating? Or that you are possibly wrong?
> Pa: Nooo… I am not wrong at all. (Zangrilli et al. 2014, 3-6)

The way these patients react when questioned about the validity of their beliefs, that are obviously not supported by evidence, is always the same. They put themselves in a defensive position, don't consider any different explanations for their experience, regard their feelings and thoughts as evidence of their beliefs, and don't reject their beliefs even when they realize the others couldn't share them, like in this example:

> PS: So, they wanted to kill you and sell your organs?
> PA Yes, and... sell the meat to restaurants where cannibals go...
> PS: Are there restaurants for cannibals?
> PA: Yes, these are secrets that the police do not know
> PS: Really?
> PA: This seems to be a bit difficult to believe, honestly... . (Ibid., 5)

Thus, this patient recognizes his belief as unusual, yet keeps on holding it. This seems clearly an irrational behaviour that 'normal people' wouldn't display. But things are much more complicated than expected. Normal subjects, actually, are not in general or epistemically rational as one would think. Clinical delusions are rare enough, but the world is full of beliefs which share the same epistemic features as delusions. I will show a few examples of this kind of beliefs, that are not supported by evidence and that are impervious to new counter-evidence.

An example involves positive illusions or the tendency to over-estimate our capacities and abilities, and to adopt an optimistic vision of the future (Taylor 1989; Jefferson et al. 2017). The vast majority of us has many illusions, like thinking that our future will be rich of positive events, or that we are above average in different domains, or that we are able to control the most important events of our lives. This kind of beliefs are epistemically irrational. We often don't possess the requested evidence to hold them, we tend to ignore counter-evidence or alternative descriptions of our successes, and systematically remember positive outcomes and forget our failures (Sharot et al. 2011). In other words, when it comes to our vision of ourselves, we seem to be irrational. The curious fact is that

there is a subset of people who don't hold these optimistic illusions, and that therefore are more rational, but these very subjects are affected by another kind of severe mental disorder: depression (Alloy and Abramson 1979; Moore and Fresco 2012). Once again, pathological subjects are more rational than us, but this is not an advantage for them (Cardella and Gangemi 2018).

There are other, even more striking, examples of irrational beliefs that are widespread in the general population. For example, paranormal or superstitious beliefs are very common. In the Gallup Survey (Moore 2001) 41% of Americans showed to believe in extrasensory perception (ESP), 37% in haunted houses, about 30% in ghosts and telepathy, 25% in clairvoyance and astrology, about 20% in witches, reincarnation and mental communication with dead people. The cumulative percentage showed that more than one-fifth of all Americans, 22%, believe in five or more of the items listed above, 32% believe in at least four items, 57% believe in at least two paranormal items, and 73% believe in at least one of them. In a more recent analysis on the presence of conspiracy theories among Italian people (Mancosu et al. 2017), four conspiracy theories have been presented: moon landings never happened, vapour traits left by aircraft are chemical agents deliberately sprayed, vaccines harm the immune system and expose it to diseases, the Stamina method has been obstructed by big pharmaceutical groups. About half of the sample considered one of the theories proposed plausible, 30% of the sample does so for two or more conspiracy theories, and about 10% of the sample considers all four stories likely to be true.

Thus, we can believe in things which are not that different from those believed by delusional patients. At any rate, these beliefs are not different in the way they are endorsed and preserved in spite of counter-evidence. As claimed by Bortolotti, the most likely scenario is

> a picture of continuity between so-called *normal* and *abnormal* cognition. Irrationality is a feature of normal cognition, and as such it cannot be the criterion of demarcation between beliefs that are 'healthy' and beliefs that are 'pathological'. (Bortolotti 2018, 113; see also Lancelotta and Bortolotti this issue)

But now we can come back to the initial question. How is it possible that we can often readily discern when someone is delusional? If it's not the irrationality of the beliefs that distinguish 'normal beliefs' by 'pathological ones', what is?

An interesting way to answer this question is the one suggested by Rhodes and Gipps (2008; see also Wilkinson this issue). Following Wittgenstein's observations included in his *On Certainty* (1969), the authors claim that delusional patients question what we can call bedrock beliefs. This kind of beliefs don't need justification, because they are foundational, and we hold them with certainty. For instance, propositions like 'I have a body', 'there are other people in the world', 'physical objects can't speak', 'this is my hand', are bedrocks beliefs. But the curious thing is that, even if we are absolutely sure of these beliefs, or maybe exactly because we are absolutely sure of these beliefs, we are not able to provide evidence, or arguments, in their favour. As Wittgenstein put it:

> If someone said to me that he doubted whether he had a body I should take him to be a half-wit. But I shouldn't know what it would mean to try to convince him that he had one. And if I had said something, and that had removed his doubt, I should not know how or why. (Wittgenstein 1969, §256)

The impossibility to provide grounds for these beliefs is due to the fact that "such propositions convey our direct, pre-reflective and practical *grasp* of the world" (Rhodes and Gipps 2008, 298), and if someone expresses doubts about them, we can't appeal to other beliefs with a higher grade of certainty to justify them. In other words, it is impossible to justify what stands beyond the need for justification. So, how do I know that this hand is my hand, and not a robot hand which my persecutors implanted on me while I was sleeping? Am I able to provide some sort of justification for this kind of knowledge? What would it be to find some evidence for it? Is it logically impossible to have a hand that is identical to a human hand but is actually a robot hand? Is it logically impossible that someone has replaced my hand with a robot hand while I was sleeping? The answer is no, it's not a kind of logical impossibility, but something very different. In delusional patients, bedrock beliefs are damaged, and, as observed by Rhodes and Gipps, this results in two things:

> a lack of constraint in acquiescing in beliefs which would normally be regarded as incredible, or a willingness to entertain doubts about everyday certainties that would normally be regarded as unassailable (Rhodes and Gipps 2008, 301).

Thus, once again, even in subjects with delusional disorder, what is affected is not rationality or logic, but natural evidence, or common sense, in other words what everyday experience of the world had taught us. I showed, in the previous section, that schizophrenic people, far from being irrational, lean on rationality and logic in order to compensate their lack of

common sense. Patients with delusional disorder, on the other hand, rely on their delusions, to give a new meaning to a world that has suddenly lost its evidence. But rationality plays an important role in this disorder, too, because they use all their reasoning abilities to protect their delusional beliefs. As one can see in the first quote of this section, logic and reason are used by these patients to justify their ideas, a mechanism previously detected by Jaspers, who, already in his General Psychopathology (1963), observed that delusions are frequently accompanied by the fully preserved capacity for reasoning and formal logic. The following example will clarify this point. A paranoid patient is sure that his neighbour entered his room one night and installed a tracking device in his abdomen. He then describes the reaction of his doctors:

> The doctors latched on to this story, eager to show me the irrationality of it all. How could he have gotten in? My door and windows had been locked, and there was no sign of tampering. I answered from the Deep Meaning that had revealed it to me.
> "He atomized himself."
> "Atomized?"
> "Yes. You know—when you dismantle something into its component atoms, pass these tiny pieces through the barrier, and reassemble them again on the other side." Didn't physics have some similar concepts?
> "And the tracker in your abdomen?"
> "Atomization again. Otherwise there'd be an incision," I reasoned, rational. But the doctors concluded differently. Delusion and paranoia were their words, their explanations. (Stefanidis 2005, 422-3)

This patient, who, by the way, was a graduate student in the neuroscience program at the University of British Columbia, strives to be rational. Diving in his world of auditory hallucinations and delusions, he tries to rely on his reasoning abilities to make sense of it all. Logic is not lost, it remains the last bastion.

> I was in fact fighting to preserve my rationality in the face of the irrational. I valued my logical mind so dearly that when it began to be challenged by schizophrenic hallucinations, delusions, and disorders of the ability to ascribe meaningfulness, I used everything available to me to try and figure out what were the most rational explanations. I craved rationality, and rationality to me was taking all evidence and making conclusions. Even if they didn't conform to everyone

> else's ideas of what was rational, I was fighting to maintain, at the very least, the integrity of my own rationality. (Stefanidis 2005, 423)

5. Conclusions: The Slippery Bounds of Rationality

I started this paper with the following quote by Edwards: "What is meant by 'rational?' Whatever it is, mental disorders are shortcomings or departures from it". These words equate irrationality and abnormality as used in psychiatric discourse. The general view in classic psychiatry is that psychopathologies are marked out by their association with different kinds of reasoning mistakes and logical errors. In order to challenge this widespread opinion, I chose to focus on two cases of severe mental disorders, schizophrenia and delusional disorder. It seems uncontroversial to think that, for example, subjects with anxiety disorders for the most part do not exhibit failures in their reasoning abilities. However, other disorders, such as psychotic disorders, seem to clearly affect the ability to reason. As stated by Parnas and colleagues:

> in the common sense understanding, which precedes and helps founding psychiatric terms, psychosis is a predicate that we ascribe to a person who has seriously transgressed the intersubjective bounds of rationality. (Parnas et al 2010, 32)

But when we pay close attention to schizophrenia and delusional disorder, the picture is very different, and, in some sense, exactly the opposite of what we could think. Those disorders show that, both in experimental settings and in everyday life, rationality and logic are fundamentally preserved, and used to compensate impairments in other areas, like common sense and social knowledge. In sum, one can learn two lessons from these examples. First, rationality is an important component of our cognition, but we don't use it as much as we think, because we often rely on other strategies when we have to judge or believe or decide something. Second, we intuitively attribute irrationality to mental disorders, but this attribution lacks any evidence, since both the experimental tasks and the autobiographical reports of single patients outline a different, even opposite, picture. Thus, why do we keep considering irrationality a crucial feature of mental disorder? Maybe we have the tendency to call madness what we don't understand, and to stop considering madness what we are able to comprehend. In other words, maybe the problem is of a conceptual kind; irrationality is a crucial part of our common conception of madness, for reasons that are historical, psychological and sociological. And it's very

hard to change a notion that lies so deeply inside us. As perfectly expressed by Frith:

> Neuroscience research has had considerable success in elucidating and sometimes curing various disorders, but after each success the disorder either becomes invisible or ceases to be considered an example of madness. So it seems strangely inevitable that madness can only ever be associated with disorders that we do not understand. It is not the patients' reason that has failed, it is ours. But then reason has never been a strong point with mankind, however civilized. (Frith 2016, 639)

REFERENCES

Alloy, L. B., and L. T. Abramson. 1979. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychological General* 108(4): 441-485.

American Psychiatric Association 2013. *Diagnostic and Statistical Manual of Mental Disorders Fifth Edition*. Washington: American Psychiatric Association.

Arieti, A. 1964. *Interpretazione della Schizofrenia*. Milano: Feltrinelli.

Baron, J. 2000. *Thinking and Deciding*. Cambridge: Cambridge University Press.

Beebee, H., and N. Sabbarton-Leary. 2010. Are psychiatric kinds real? *European Journal of Analytic Philosophy* 6(1): 11-27.

Beck, A. T. 1976. *Cognitive Therapy and the Emotional Disorders*. New York: Meridian.

Belvin, W. E. 1964. Deductive reasoning in schizophrenia. *The Journal of Abnormal and Social Psychology* 69(1): 47-61.

Bentall, R. P. 1994. Cognitive biases and abnormal beliefs: Towards a model of persecutory delusions. In *The Neuropsychology of Schizophrenia*, eds. David A. and J. Cutting, Hove, UK: Lawrence Erlbaum Associates Ltd.

Bentall, R. P., R. Corcoran, R. Howard, N. Blackwood, and P. Kinderman. 2001. Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review* 21(8): 1143-1192.

Bergamin, J. A. 2018. An excess of meaning: Conceptual over-interpretation in confabulation and schizophrenia. *Topoi* 39: 163–176.

Bermúdez, J. L. 2001. Normativity and rationality in delusional psychiatric disorders. *Mind and Language* 16: 493-457.

Binswanger, L. 1956. *Drei Formen missglückten Daseins. Verstiegenheit, Verschrobenheit, Manieriertheit.* Tübingen: Niemeyer.

Blankenburg, W. 1971. *Der Verlust der natürlichen Selbstverständlichkeit*. Stuttgart: Ferdinand Enke Verlag.

Bortolotti, L. 2013. Rationality and sanity: The role of rationality judgments in understanding psychiatric disorders. In *The Oxford Handbook of Philosophy and Psychiatry*, eds. M. Davies, R. Gipps, G. Graham, J. Sadler, G. Stanghellini, and T. Thornton, Oxford: Oxford University Press.

Bortolotti, L. 2018. Delusions and three myths of irrational belief. In *Delusions in Context*, ed. L. Bortolotti, London: Palgrave McMillan.

Cardella, V., and A. Gangemi. 2018. *La Logica della Follia. Razionalità e Irrazionalità nella Psicopatologia*. Roma: Corisco editore.

Cardella, V., and A. Gangemi. 2014. Reasoning in schizophrenia. In *Aberrant Beliefs and Reasoning*, ed. N. Galbraith, London: Psychology Press.

Edwards, R. B. 1997. Mental health as rational autonomy. In *Ethics of Psychiatry: Insanity, Rational Autonomy, and Mental Health Care*, ed. R. B. Edwards, New York: Prometheus Books.

Evans, J. 1989. *Bias in Human Reasoning: Causes and Consequences*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J. 2002. Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin* 128: 978-996.

Freeman, D., P. A. Garety, E. Kuipers, D. Fowler, and P. E. Bebbington. 2002. A cognitive model of persecutory delusions. *British Journal of Clinical Psychology* 41(Pt 4): 331-347.

Freeman, D. 2008. The assessment of persecutory ideation. In *Persecutory Delusions*, eds. D. Freeman, R. Bentall, and P. Garety, Oxford: Oxford University Press.

Frith C. D. 2016. Understanding madness? *Brain* 139(2): 635-639.

Garety, P. A., D. R. Hemsley, and S. Wessely. 1991. Reasoning in deluded schizophrenic and paranoid patients: Biases in performance on a probabilistic inference task. *Journal of Nervous and Mental Disease* 179: 194-201.

Garety, P. A., and D. R. Hemsley. 1994. *Delusions: Investigations into the Psychology of Delusional Reasoning*. Oxford: Oxford University Press.

Garety, P. A., D. Freemann, S. Jolley, G. Dunn, P. E. Bebbington, D. Fowler, E. Kuipers, and R. Dudley. 2005. Reasoning, emotions and delusional conviction in psychosis. *Journal of Abnormal Psychology* 114: 373-384.

Gilleen, J., and A. S. David. 2005. The cognitive neuropsychiatry of delusions: From psychopathology to neuropsychology and back again. *Psychological Medicine* 35: 5-12.

Gillett, G. 1992. Rationality and Delusional Disorders. In *Phenomenology, Language and Schizophrenia*, eds. M. Spitzer, F. Uehlein, M. A. Schwartz, and C. Mundt, New York: Springer.

Girotto, V. 1994. *Il Ragionamento*. Bologna: Il Mulino.

Graham, G. 2010. *The Disordered Mind*. London and New York: Routledge.

Harvey, A., E. Watkins, W. Mansell, and R. Shafran. 2004. *Cognitive Behavioural Processes across Psychological Disorders: A Transdiagnostic Approach to Research and Treatment*. Oxford: Oxford University Press.

Henriksen, M. G., and J. Nordgaard. 2014. Schizophrenia as a disorder of the self. *Journal of Psychopathology* 20: 435–441.

Jaspers, K. 1963. *General Psychopathology*. Translated by J. Hoenig, and M. W. Hamilton, Chicago: University of Chicago Press.

Jefferson, A., L. Bortolotti, and B. Kuzmanovic. 2017. What is unrealistic optimism? *Consciousness and Cognition* 50: 3–11.

Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. 2006. *How We Reason*. Oxford: Oxford University Press.

Johnson-Laird, P. N., F. Mancini, and A. Gangemi. 2006. A hyper-emotion theory of psychological illnesses. *Psychological Review* 113: 822-842.

Johnson-Laird, P. N., and R. M. J. Byrne. 1991. *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Kahneman, D., A. Tversky, and P. Slovic, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kasanin, J. S., ed. 1944. *Language and Thought in Schizophrenia*. Berkeley: University of California Press.

Kemp, R., S. Chua, P. McKenna, and A. David. 1997. Reasoning and delusions. *The British Journal of Psychiatry* 170: 398-405.

Kendler K. S., P. Zachar, and C. Craver. 2011. What kinds of things are psychiatric disorders. *Psychological Medicine* 41: 1143–1150.

Kohne, A. C. J. 2015. The dubious reality of a psychiatric classification. *Tijdschrift voor psychiatrie* 57(6):433-440

Lancellotta, E. and Bortolotti, L. this issue. Delusions in the two-factor theory: Pathological or adaptive? *European Journal of Analytic Philosophy*.
https://doi.org/10.31820/ejap.16.2.2

Maher, B. A. 1992. Models and methods for the study of reasoning in delusions. *Revue Européenne de Psychologie Appliquee* 42: 97-102.

Mancini, F., A. Gangemi, and P. N. Johnson-Laird. 2007. Il ruolo del ragionamento nella psicopatologia secondo la *Hyper Emotion Theory*. *Giornale Italiano di Psicologia*: XXXIV, 4.

Mancini, M., S. Presenza, L. Di Bernardo, P. P. Lardo, S. Totaro, F. Trisolini, L. Vetrugno, and G. Stanghellini. 2014. The life-world of persons with schizophrenia. A panoramic view. *Journal of Psychopathology* 20: 423-434

Mancosu, M., S. Vassallo, C. Vezzoni. 2017. Believing in conspiracy theories: Evidence from an exploratory analysis of Italian survey data. *South European Society and Politics* 22(3): 327-344.

McKay, R., R. Langdon, and M. Coltheart. 2005. "Sleights of mind": Delusions, defences, and self-deception. *Cognitive Neuropsychiatry* 10: 305–326.

Mellet, E., O. Houdé, P. Brazo, B. Mazoyer, N. Tzourio-Mazoyer, and S. Dollfus. 2006. When a schizophrenic deficit becomes a reasoning advantage. *Schizophrenia Research* 84: 359-364.

Menon, M., E. Pomarol-Clotet, P. J. McKenna, and R. A. McCarthy. 2006. Probabilistic reasoning in schizophrenia: A comparison of the performance of deluded and nondeluded schizophrenic patients and exploration of possible cognitive underpinnings. *Cognitive Neuropsychiatry* 11(6): 521-536.

Minkowski, E. 1927. *La Schizophrénie. Psychopathologie des Schizoïdes et des Schizophrènes*. Paris: Payot.

Mirian, D. R., W. Heinrichs, and S. McDermid Vaz. 2011. Exploring logical reasoning abilities in schizophrenic patients. *Schizophrenia Research* 127: 178-180.

Moore, D. W. 2001. *Three in Four Americans Believe in Paranormal*. https://home.sandiego.edu/~baber/logic/gallup.html

Nagai, M. 1990. Bunretsubyōsha no jikoishiki niokeru 'bunretsubyōsei' [The 'schizophrenic' in the self-consciousness of schizophrenic patients]. In *Seishinbunretsubyō: Kiso to rinshō [Schizophrenia: Its Basic and Clinical Aspects],* eds. B. Kimura et al., 361–368. Tokyo: Asakura Shoten.

Owen, G. S., J. Cutting, and A. S. David. 2007. Are people with schizophrenia more logical than healthy volunteers? *British Journal of Psychiatry* 191: 453-454.

Parnas, J., J. Nordgaard, and S. Varga. 2010. The concept of psychosis: A clinical and theoretical analysis. *Clinical Neuropsychiatry* 7(2): 32-37

Parnas, J., and P. Handest. 2003. Phenomenology of anomalous experiences in early schizophrenia. *Comprehensive Psychiatry* 44(2): 121–134.

Pérez-Álvarez, M., J. M. García-Montes, O. Vallina-Fernández, and S. Perona-Garcelán. 2016. Rethinking schizophrenia in the context of the person and their circumstances: seven reasons. *Frontiers in Psychology* 7: 1650.

Revsbech, R., E. L. Mortensen, G. Owen, J. Nordgaard, L. Jansson, D. Sæbye, T. Flensborg-Madsen, and J. Parnas. 2015. Exploring rationality in schizophrenia. *British Journal of Psychiatry open* 1(1): 98–103.

Rips, L. J. 1994. *The Psychology of Proof*. Cambridge, MA: MIT Press.

Rhodes, G., and J. Gipps. 2008. Delusions, certainty, and the background. *Philosophy, Psychiatry, and Psychology* 15(4): 295-310.

Sass, L. A. 1992. *Madness and Modernism: Insanity in the Light of Modern Art, Literature, and Thought*. Oxford: Oxford University Press.

Scharfetter, C. 1980. *General Psychopathology. An Introduction*. Cambridge: Cambridge University Press.

Scull, A. 1981. Moral treatment reconsidered: Some sociological comments on an episode in the history of British psychiatry. In *Madhouses, Mad-doctors, and Madmen: The Social History of Psychiatry in the Victorian Era*, ed. A. Scull, Philadelphia: University of Pennsylvania Press.

Sharot, T., C. W. Korn, and R. J. Dolan. 2011. How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience* 14: 1475–1479.

Simon, H. 1991. Bounded rationality and organizational learning. *Organization Science* 2(1): 125-134.

Smeets, G., P. J. De Jong, and B. Mayer. 2000. If you suffer from a headache, then you have a brain tumour: Domain-specific reasoning 'bias' and hypochondriasis. *Behaviour Research and Therapy* 38(8): 763-776.

Speechley W. J., and E. T. Ngan. 2008. Dual-stream modulation failure: A novel hypothesis for the formation and maintenance of delusions in schizophrenia. *Medical Hypotheses*: 70(6), 1210-1214.

Stanghellini, G. 2000. Vulnerability to schizophrenia and lack of common sense. *Schizophrenic Bulletin*: 26(4): 775-787.

Stefanidis, E. 2006. Personal account: Being rational. *Schizophrenic Bulletin*: 32(3): 422-423.

Taylor, S. E. 1989. *Positive Illusions: Creative Self-deception and the Healthy Mind*. New York: Basic Books.

Tversky, A., and D. Kahneman. 1982. Evidential impact of base rates. In *Judgment under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky, Cambridge: Cambridge University Press.

Urfer-Parnas, A. 2019. Eugène Minkowski. In *The Oxford Handbook of Phenomenological Psychopathology*, eds. G. Stanghellini, M.

Broome, A. V. Fernandez, P. Fusar-Poli, A. Raballo, and R. Rosfort. Oxford: Oxford University Press.

Von Domarus, E. 1944. The specific laws of logic in schizophrenia. In *Language and Thought in Schizophrenia*, ed. J. S. Kasanin, 104-114. Berkeley: University of California Press.

Wason, P. C. 1966. Reasoning. In *New Horizons in Psychology*, ed. B. M. Foss, Harmondsworth: Penguin.

Williams, B. E. 1964. Deductive reasoning in schizophrenia. *The Journal of Abnormal and Social Psychology* 69(1): 47-61.

Wilkinson, S. this issue. Expressivism about delusion attribution. *European Journal of Analytic Philosophy*. https://doi.org/10.31820/ejap.16.2.3

Wittgenstein, L. 1969. *On Certainty*. Oxford: Basil Blackwell.

Zachar, P. 2015. Psychiatric disorders: Natural kinds made by the world or practical kinds made by us? *World Psychiatry*, 14(3): 288–290.

Zangrilli, A., G. Ducci, P. L. Bandinelli, J. Dooley, R. McCabe, and S. Priebe. 2014. How do psychiatrists address delusions in first meetings in acute care? A qualitative study. *BMC Psychiatry* 14: 178.

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 165:159.972
https://doi.org/10.31820/ejap.16.2.2

DELUSIONS IN THE TWO-FACTOR THEORY: PATHOLOGICAL OR ADAPTIVE?

Eugenia Lancellotta
University of Birmingham

Lisa Bortolotti
University of Birmingham

Original scientific article – Received: 15/6/2020 Accepted: 13/7/2020

ABSTRACT

In this paper we ask whether the two-factor theory of delusions is compatible with two claims, that delusions are pathological and that delusions are adaptive. We concentrate on two recent and influential models of the two-factor theory: the one proposed by Max Coltheart, Peter Menzies and John Sutton (2010) and the one developed by Ryan McKay (2012). The models converge on the nature of Factor 1 but diverge about the nature of Factor 2. The differences between the two models are reflected in different accounts of the pathological and adaptive nature of delusions. We will explore such differences, considering naturalist and normativist accounts of the pathological and focusing on judgements of adaptiveness that are informed by the shear-pin hypothesis (McKay and Dennett 2009). After reaching our conclusions about the two models, we draw more general implications for the status of delusions within two-factor theories. Are there good grounds to claim that delusions are pathological? Are delusions ever adaptive? Can delusions be at the same time pathological and adaptive?

Keywords*: Delusions; adaptiveness; pathology, two-factor theories; delusion formation*

Correspondence: L.Bortolotti@bham.ac.uk

1. Introduction

Delusions are symptoms of mental disorders. Does that mean that they inherit from disorders their *pathological* status? Or should they be seen instead as emergency responses to a critical situation and thus described as *adaptive*? Could they be simultaneously pathological *and* adaptive? In this paper we are interested in the answers that the two-factor theory of delusions provides to such questions.

We are aware that delusions come in different forms and contents and that the two-factor theory has interesting things to say about all types of delusions—and other kinds of beliefs too. However, in this paper we shall refer to monothematic delusions and in particular the Capgras delusion as our standard example. This is for two reasons: (1) the two-factor theory was initially put forward to account for monothematic delusions,[1] even though its scope has been gradually extended to account for a wider range of phenomena;[2] (2) the Capgras delusion is the standard example in the papers proposing the two models of the two-factor theory we have chosen to focus on.

1.1. Delusions: The Pathological and the Adaptive

Delusions are unusual beliefs that are considered as symptomatic of a number of mental disorders, such as schizophrenia and delusional disorder. Monothematic delusions revolve around one theme and their content is often wildly implausible: someone with Capgras delusion believes that their spouse has been replaced by an impostor who looks just like the spouse; someone with Cotard delusion believes that they are disembodied or dead; someone with mirrored-self misidentification believes that they can see a stranger—and not their own image—in the mirror. The two-factor theory of delusion formation is a very influential theory proposing that monothematic delusions are caused by at least two factors. Factor 1 is a neuropsychological deficit responsible for anomalous data that may also result in an anomalous experience. Factor 2 is a cognitive process (described as either dysfunctional or biased) explaining either the initial endorsement of the delusional belief or the prolonged maintenance of the delusional belief in the face of mounting counterevidence. Multiple versions of the two-factor theory have been put forward, where the main difference between them lies in the description of Factor 2 and its role in the process of delusion formation.

[1] Some authors suggest that the two-factor theory is best suited to account for monothematic delusions, and that has been built around the Capgras delusions (e.g., Corlett 2019).
[2] See for instance the discussion of self-deception in McKay et al. (2005).

According to the two-factor theory, are delusions pathological? Are they adaptive? Following the most popular ways to characterise what counts as a disorder in the philosophy of medicine in general and in psychiatry in particular, a belief counts as 'pathological' when it is either (1) the output of a dysfunctional process (*naturalism*); (2) harmful (*normativism*); or (3) the output of a dysfunctional process and harmful (*harmful-dysfunction account*) (Bortolotti 2020). Beliefs are sometimes regarded as pathological when they deviate from some norm to which they are expected to conform—but that use of the term 'pathological' is an extension and we shall not consider it here.

Beliefs are usually called 'adaptive' if they enhance a person's wellbeing, purpose in life, or good functioning (*psychological adaptiveness*); or if they enhance an individual's chances of survival and reproduction (*biological adaptiveness*). It has been shown that arguments for the biological adaptiveness of delusions are less common and overall less persuasive than claims about their psychological adaptiveness (McKay and Dennett 2009; Lancellotta and Bortolotti 2019) and when some delusions are presented as psychologically adaptive, their contribution to wellbeing or good functioning is often regarded as partial or temporary. We will spend more time on the psychological adaptiveness claim simply because the biological adaptiveness thesis has been defended (to our knowledge) only within the predictive-processing account of delusion formation (Fineberg and Corlett 2016) and not within the two-factor theory. To make our task more manageable, we shall confine our attention to forms of psychological adaptiveness that are explained by a shear-pin mechanism (McKay and Dennett 2009).

1.2. The Shear-pin Hypothesis

According to the "shear-pin" hypothesis (McKay and Dennett 2009), some false beliefs that prevent a cognitive system from being overwhelmed can count as adaptive (*adaptive misbeliefs*). This might happen for instance when people experience such a traumatic event that they would succumb to suicidal thoughts if their negative emotions were not managed. One example is anosognosia ("denial of illness"), where a person, who has lost the use of a limb as a result of physical trauma, denies paralysis or does not acknowledge the full extent of the ensuing impairment (Ramachandran and Blakeslee 1998; McKay et al. 2005). Someone's delusion that they can clap their hands when their right arm is paralysed would act as a motivated belief which serves to reduce the harmful impact of their new disability on their wellbeing and sense of self. McKay and Dennett (2009) suggest in their paper that, in situations of extreme stress, motivational influences are allowed to intervene in the process of belief evaluation. As a result, people

come to believe what they desire to be true ("My arm is not paralysed"; "I can clap!") and not what they have evidence for ("My arm is not moving because it is paralysed"). This is designed to permit the cognitive system to continue operating.

According to the shear-pin hypothesis, the situation in which adaptive misbeliefs emerge is already seriously compromised.

> What might count as a doxastic analogue of shear pin breakage? We envision doxastic shear pins as components of belief evaluation machinery that are "designed" to break in situations of extreme psychological stress (analogous to the mechanical overload that breaks a shear pin or the power surge that blows a fuse). Perhaps the normal function (both normatively and statistically construed) of such components would be to constrain the influence of motivational processes on belief formation. Breakage of such components, therefore, might permit the formation and maintenance of comforting misbeliefs – beliefs that would ordinarily be rejected as ungrounded, but that would facilitate the negotiation of overwhelming circumstances (perhaps by enabling the management of powerful negative emotions) and that would thus be adaptive in such extraordinary circumstances. (McKay and Dennett 2009, 501)

The person is already experiencing high levels of stress and can come to more serious harm unless their negative emotions are managed. Thus, adaptive misbeliefs prevent the situation from worsening. McKay and Dennett talk about the "extraordinary circumstances" in which motivational influences on belief are not just tolerated but desirable. Such influences intervene not by accident but by design, and this is what makes the resulting beliefs adaptive despite their falsehood.

McKay and Dennett consider the possibility that some delusions count as biologically adaptive misbeliefs but argue that in the case of delusions the extent to which desires are allowed to influence belief formation is excessive. They leave it open whether some delusions can count as psychologically adaptive.

1.3. The Two-factor Theory

According to Max Coltheart (2007), who is the founder of the two-factor theory, a satisfactory theory of delusions should be able to answer two questions about the genesis and maintenance of delusional beliefs:

1. Where does the delusion come from?
2. Why is the delusion adopted and then maintained in the face of disconfirming evidence?

Two-factor models of delusions provide an answer to these questions by advocating two factors in the generation and maintenance of a delusional belief (Coltheart 2007).

Factor 1 answers the first question and results in anomalous data/experience. Consider for example the Capgras delusion where the person comes to believe that a loved one has been replaced by an identical impostor. Factor 1 is an autonomic failure in the face recognition system, so when the person sees their spouse, the well-known face does not trigger the usual feelings of familiarity.[3] This generates an anomalous experience of a face which is recognised but does not *feel* familiar. On the model, Factor 1 explains the content of the delusion. Factor 1 varies from delusion to delusion and may even vary across individual cases of the same delusion. Two-factor theories hold that Factor 1 is necessary but not sufficient to explain the phenomenon of delusions. This is mainly due to the fact that there seem to be people who have the deficit playing the Factor 1-role but do not report delusional beliefs. To differentiate these cases from delusional ones, another factor (Factor 2) is required to explain the transition from the data resulting in an anomalous experience to the delusional belief. The move from not feeling that a well-known face is familiar to believing something like: "The person I see in front of me is not my spouse but an impostor" is due to a process of either endorsement or explanation of the content of the anomalous experience.

Whilst Factor 1 differs from one delusion (or person) to the next, Factor 2, broadly described as a problem in belief evaluation, is supposed to be constant across all delusions. However, two-factor theorists disagree on the precise nature of Factor 2. Some proposals identify Factor 2 with a lesion to the right dorsolateral prefrontal cortex (Coltheart et al. 2018) but there is disagreement about whether this locus is specific to delusions or shared with other neuropsychological conditions (see Tranel and Damasio 1994; Corlett 2019). Another open question about two-factor theories is whether Factor 2 contributes to the *adoption* or to the *maintenance* of the delusional belief.

[3] We are aware that the way of describing the conscious experience of people with Capgras when they look at their loved one is controversial, but we will not engage in questions about the nature of their experience as it is not relevant to our discussion. In this paper, we shall talk about their failing to experience a "feeling of familiarity". Also, there is a debate about how to accurately characterise the content of the Capgras delusion. In this paper, we shall talk about people believing something like the following: "The person I see in front of me is not my beloved one but an impostor".

Let us describe two competing models of the two-factor theory—the most influential and detailed—and map their differences.

1.4. The Coltheart Model

On what we shall refer to as *the Coltheart model* (Coltheart et al. 2010), Factor 1 is a neuropsychological deficit which results in anomalous data and can manifest at conscious level as an anomalous experience.

Factor 1 operates at the belief adoption stage. What happens at the belief adoption stage? The anomalous data are accounted for by a process of inference *to the best explanation* (abductive inference): given the very unusual nature of the data, the delusional explanation is the best possible explanation among a range of candidate hypotheses. Abductive inference is understood in Bayesian terms. Bayes' theorem stipulates the best way of choosing among candidate hypotheses to explain a given piece of evidence (O). A hypothesis (H) is more apt than another hypothesis (H') to explain O if its posterior probability is higher than the posterior probability of H'. The posterior probability of a hypothesis is the product of the hypothesis' prior probability (the probability of the hypothesis before O) and its likelihood (how likely it is to observe O if the hypothesis was true). On this account, given O, it is possible for H to be a better explanation than H' even if H has a low prior probability providing that the likelihood of H given O offsets its low prior probability.

Consider the Capgras delusion. In the Coltheart model, the impostor hypothesis ("That woman is not my wife but an impostor") can be a better explanation than the spouse hypothesis ("That woman is my wife") with regard to evidence O. Even if the impostor hypothesis has a lower prior probability than the spouse hypothesis, as impostors are not a frequent occurrence, its likelihood can be much greater than that of the spouse hypothesis, to the point of making its posterior probability higher than that of the spouse hypothesis. In this scenario, the impostor hypothesis is the most rational explanation for the absence of a feeling of familiarity: people have intact reasoning capacities when adopting the delusional hypothesis. Their reasoning is compromised when evidence against the delusional belief start accumulating.

Factor 2 is a cognitive deficit inhibiting the rejection of an endorsed belief even in the presence of strong counterevidence—Factor 2 makes the belief virtually impossible to revise. On this model, Factor 2 operates at the belief maintenance stage. What happens then, at the belief maintenance stage? On the Coltheart model, there is a second dysfunction responsible for the delusion (Factor 2) which amounts to a *deficit* in belief evaluation. This

allows the delusional belief to be preserved in the face of evidence to the contrary.

In the case of Capgras delusion, the person faces overwhelming evidence against the impostor belief but that is not sufficient reason for the person to abandon or revise that belief. Evidence may include the testimony from relatives and friends confirming that the person accused to be an impostor is in fact the spouse. The person who adopted the delusional belief is unable to step back from it and to consider alternative explanations even when the belief receives serious challenges.

1.5. The McKay Model

Ryan McKay puts forward several objections to the Coltheart model which are important to understand his own proposal (McKay 2012), what we shall call *the McKay model*. As the objections are also relevant to our assessment of the status of delusions, we shall consider some of them here, albeit briefly.

First, the novel contribution in the Coltheart model (Coltheart et al. 2010) is that adopting the delusional hypothesis (e.g., the impostor hypothesis in the Capgras delusion) is Bayesian-rational because the hypothesis is the best explanation for the anomalous data. But for McKay the rationality of the endorsement of the delusional hypothesis is overestimated in the Coltheart model, because the model does not take into account how incredibly unlikely the state of affairs which makes up the content of the delusion is. As McKay says, it would be akin to a miracle if an impostor were to take the place of one's spouse and be also perfectly identical to the spouse. Thus, it is not plausible to suppose that there is nothing problematic in the reasoning step that leads from the anomalous data and the resulting experience to the delusional belief.

Second, how do we account for the experiences of ventromedial frontal patients who, similarly to Capgras patients, experience an autonomic failure to familiar faces but who, differently from Capgras patients, do not adopt the impostor belief? In the Coltheart model, the assumption is that ventromedial frontal patients initially adopt the impostor belief—as the best possible explanation of the anomalous data which sometimes results in an anomalous experience—but do not maintain it. When faced with disconfirming evidence, differently from Capgras patients, they abandon the impostor belief. This can be accounted for if ventromedial frontal patients share Factor 1 with Capgras patients but not Factor 2.

McKay's objection to this proposal is that it is implausible that ventromedial frontal patients first adopt the impostor belief and then reject it. It is implausible that the spouse hypothesis is dismissed at the stage of belief adoption but then embraced once the person receives evidence against the impostor belief. The conjunction of new evidence (i.e. testimony from relatives and friends which contradicts the impostor belief) and old evidence (i.e. the absence of a feeling of familiarity which confirms the imposter belief and protestations for the alleged impostors that they are not impostors) does not favour the spouse hypothesis over the impostor belief in the circumstances. Why would the spouse hypothesis explain the total evidence any better than the impostor belief? More precisely, it is not clear why the testimony of others should radically change the distribution of likelihoods between the impostor belief and the spouse hypothesis, considering that, according to McKay, the spouse's testimony was presumably already dismissed at the stage of the adoption of the impostor belief.

A possible response in defence of the Coltheart model is that the testimony of the spouse does not count as evidence in favour of the spouse hypothesis: it is easy to see that a good impostor would still convincingly pretend to be someone's spouse even when explicitly confronted about it. The testimony of friends and family seems a more reliable source of evidence in favour of the spouse hypothesis. Hence, it might be the case that ventromedial frontal patients initially adopt the impostor belief because it is the one which best explains the evidence at hand—the absence of feelings of familiarity and the testimony of the spouse—but then correctly dismiss it in the face of the testimony of friends and family.

The third criticism of the Coltheart model is probably the most compelling. It concerns the *chronology* of Factor 1 and Factor 2. If people with Capgras delusion are unable to revise their impostor belief in the light of contradicting evidence because of Factor 2, this means that they cannot acquire Factor 2 prior or at the same time of Factor 1, otherwise they would be unlikely to abandon the spouse hypothesis and would dismiss the evidence for the impostor hypothesis (i.e., the absence of a feeling of familiarity). In other words, if people who develop the Capgras delusion are *conservative* with their existing beliefs at the maintenance stage, why should they be *revisionist* with their existing beliefs at the adoption stage? The Coltheart model seems to require that people with Capgras acquire Factor 2 *after* Factor 1, that is, after endorsing the impostor belief and before facing the testimony of family and friends which counts against it.

McKay overcomes this objection by putting forward his own model, according to which Factor 2 operates at the adoption stage, just like Factor

1: the impostor hypothesis is adopted because people suffer from a neuropsychological impairment responsible for the anomalous data and resulting in the anomalous experience (Factor 1), and because they have a bias towards *explanatory adequacy* (Factor 2) which leads them to accept hypotheses that seem to explain their experiences even when such hypotheses have low prior probability and conflict with their existing beliefs.

> An individual with a bias towards explanatory adequacy will update beliefs as if ignoring the relevant prior probabilities of the candidate hypotheses. (McKay 2012, 345)

The McKay model builds on previous work by Stone and Young (1997), Aimola Davies and Davies (2009), and McKay himself. It largely agrees with the Coltheart model about the nature of Factor 1. Factor 1 is a neuropsychological deficit and in the case of Capgras delusion it causes the absence of a feeling of familiarity towards well-known faces.

However, the model offers a different account of Factor 2. In the McKay model, Factor 2 is activated in the transition from the anomalous experience to the belief. Due to the explanatory adequacy bias, salient perceptual experience is taken at face value, causing the person to adopt a hypothesis which explains the experience in question but does not fit with the person's previous beliefs (e.g., the impostor hypothesis in Capgras). Ventromedial frontal patients who may also fail to experience feelings of familiarity towards well-known faces (Factor 1) but who do not come up with the impostor belief may just lack the explanatory adequacy bias (Factor 2). In the model, Factor 2 is thus already present when the delusional belief is adopted whereas the Coltheart model is supposed to locate Factor 2 at the belief maintenance stage.

For McKay, given the extreme low prior probability of the impostor hypothesis, it is not rational to adopt it as an explanation of the anomalous experience, so some bias needs to be involved in the acceptance of the delusional belief. The delusion is adopted due to the fact that people discount the prior probabilities of the delusional hypothesis in favour of how well the hypothesis explains ('fits') the data. So, people who develop Capgras adopt the impostor belief despite its low prior probability because it matches the absence of a feeling of familiarity towards well-known faces better than the spouse hypothesis.

Here is a way of describing the difference between the McKay model and the Coltheart model: for McKay the delusion emerges when the impostor belief is adopted, as Factor 1 and Factor 2 have contributed by then to the

person endorsing an unusual explanation for an unusual experience. For Coltheart and colleagues, the impostor belief is adopted as a result of Factor 1, but it becomes a delusion only when it grows resistant to counterevidence at the maintenance stage as a result of Factor 2.

1.6. Interim Summary and Plan

We have introduced two models of the two-factor theory, explaining how they differ (see table 1 for a summary). In section 2 we shall ask whether the models are compatible with delusions being pathological. In section 3 we shall ask whether they are compatible with delusions being adaptive.

	Factor 1	*Factor 2*
The Coltheart Model (Coltheart et al. 2010)	A *neuropsychological deficit* manifesting in an unusual experience leads the person to adopt an unusual belief.	A *cognitive deficit in belief evaluation* leads the person to preserve the unusual belief in the face of counterevidence.
	Factor 1 explains belief adoption and Factor 2 the belief maintenance.	
The McKay Model (McKay 2012)	A *neuropsychological deficit* manifesting in an unusual experience contributes to the person adopting an unusual belief.	An *explanatory adequacy bias* contributes to the person adopting a belief with low prior probability.
	Factor 1 and Factor 2 together explain the adoption of the delusional belief.	

Table 1: Differences in two influential versions of the two-factor theory of delusion formation

2. Are Delusions Pathological?

In this section we ask whether the claim that delusions are pathological beliefs is compatible with the two-factor models of delusions described in section 1, the Coltheart model and the McKay model. We structure the discussion around three ways in which we can understand what it means for delusions to be pathological, which map the notions of disorder defended in the philosophy of medicine: *naturalism* (the system is disordered if it is dysfunctional); *normativism* (the system is disordered if it causes harm); the *harmful-dysfunction* view (the system is disordered if it is dysfunctional and it causes harm).

2.1. The Naturalist View

For naturalists, the pathological nature of a delusional belief depends on whether the belief's aetiology involves a dysfunction. More precisely, the claim is that for a belief to be pathological, there must be a dysfunction in the mechanisms responsible for how the belief is adopted or maintained.

In statements about the two-factor theory of delusion formation, the words 'deficit' and 'dysfunction' are indeed used and delusions are recognised as pathological: "[W]e advocate a deficit model of delusion formation, that is, delusions arise when the normal cognitive system which people use to generate, evaluate, and then adopt beliefs is damaged" (Langdon and Coltheart 2000, 184). And again: "Essentially, we view delusion as a dysfunctional belief, a doxastic state of a particular pathological severity" (McKay et al. 2005, 315). We know by now that in the two-factor theory, the two factors are a neuropsychological deficit resulting in anomalous data/experience and, more relevant to assessing the pathology of a belief, a problem with reasoning. Factor 2 is described as a *cognitive bias* (e.g., Fine et al. 2007; Langdon et al. 2010; McKay 2012) or as a *cognitive deficit* (e.g., Coltheart 2007; Coltheart et al. 2010).[4]

In two-factor theories advocating cognitive biases, people reporting delusional beliefs are found to reason differently from people who do not, but the difference is not a disadvantage independent of the context in which the bias operates. This suggests that there is no deficit or dysfunction involved in forming the delusion given the anomalous nature of the experience. The presence of biases in the belief fixation process is not sufficient for the resulting belief to qualify as pathological, and indeed many non-pathological beliefs are the output of biased reasoning. The same bias can be beneficial in some contexts and detrimental in other contexts, and biased reasoning does not imply the presence of an underlying deficit. The McKay model is a good example of the bias approach: the problem identified in the inference from the experience to the belief (Factor 2) is an *explanatory adequacy* bias. People who have it tend to disregard a hypothesis's low prior probability if the hypothesis seems to explain well the data salient to them. The opposite tendency, often called *doxastic conservatism*, consists in resisting a hypothesis that does not fit with previous beliefs even if the hypothesis seems to explain well the data. It is a form of inertia where the person's existing model of the world is protected from change. Whether one bias or the other leads to

[4] If the only problem with the delusion was the anomalous data it explains, then one might come to the conclusion that the delusional belief itself is not pathological as there is nothing dysfunctional in the way in which belief fixation mechanisms operate.

better outcomes (the adoption and maintenance of true and rational beliefs) depends on the context. Thus, on naturalist grounds alone, delusions are not pathological in the McKay model.

In two-factor theories explicitly advocating a cognitive deficit or a doxastic dysfunction, Factor 2 is to be identified with such a deficit or dysfunction: examples would be the failure for the belief fixation system to inhibit implausible hypotheses or the failure for the belief maintenance system to abandon or revise a belief that has received disconfirmation by further evidence after its adoption. This suggests that the role of Factor 2 in the formation of delusions is sufficient for the delusion to count as pathological on naturalist grounds. The Coltheart model fits such a description: impostor beliefs may not be pathological when they are adopted, as the impostor hypothesis is the best explanation for the person's anomalous data/experience. However, the belief becomes pathological at the stage in which it is maintained in the face of powerful counterevidence, because its maintenance is due to a dysfunction affecting belief evaluation.

2.2. The Normativist and the Harmful-dysfunction View

Normativists agree that the pathological nature of a belief depends on whether the belief causes harm or otherwise leads to undesirable consequences for the agent—as judged by the agent or by society, depending on the preferred version of the view. Harms and disadvantages may include impaired functioning, loss of agency, negative emotions, failure to fulfil one's goals, and so on. It is plausible to claim that delusions (differently from many non-delusional irrational beliefs) are generally disruptive and can negatively affect a person's wellbeing causing impaired functioning, social isolation and withdrawal.

However, for a belief to be pathological, we would expect *the belief itself* to be the cause of harms or other disadvantages. It is not clear in the case of delusion whether the belief is the cause of the harm or disadvantage or is instead a response to a situation that is already critical for the person. The difficulty for normativism here is that what we know about so-called pathological beliefs does not usually enable us to determine whether the harm or disadvantage is caused by the beliefs themselves. Indeed, it may be caused by something else but ultimately explain why the beliefs are adopted or maintained; or it may just happen alongside the adoption and the maintenance of the belief.

For instance, on some accounts of delusions in schizophrenia, the delusion is seen as a response to the uncertainty in the prodromal phase of psychosis (e.g., Jaspers 1963; Mishara 2010). More relevant to monothematic

delusions, in anosognosia the adoption of the belief that one's arm is not paralysed (say) can be seen as a reaction to the physical and psychological trauma the person experienced (e.g., Turnbull et al. 2014). In such a case, the delusion seems to be a response to a critical situation as opposed to the source of the harm or disadvantage (although the maintenance of the delusion may become a source of further harm or disadvantage). In the case of monothematic delusions like Capgras, it is not clear whether the delusion causes or is a response to harm or disadvantage: psychodynamic accounts of Capgras tended to see it as a motivated delusion, but more recent cognitive-deficit accounts do not make room for the delusion to be part of a defence mechanism (McKay et al. 2005).

There are cases in which unquestionable harm or disadvantage is associated with believing the delusional content (e.g. when the content is distressing, causing guilt, fear, or anxiety). There are also cases in which the harm or disadvantage is caused by the reaction of the surrounding social environment to the person reporting the belief: individuals whose beliefs have similar surface features may experience drastically different responses, ranging from being supported by their social circle to being vulnerable to exclusion and isolation. In sum, there is a significant link between delusions and harm or disadvantage even when a person's overall functioning is not impaired by the delusion (e.g., Jackson and Fulford 1997).

Where does this leave our two models? Are delusions pathological on normativist grounds for the two-factor theory? The most plausible answer is yes. McKay is explicit about delusions causing harm—functioning is disrupted by the extent of the mismatch between the content of the delusion and the reality as experienced by those who are non-delusional (McKay et al. 2005; McKay and Dennett 2009). Factor 1 and Factor 2 are both responsible for this mismatch, the data being anomalous and the delusional hypothesis being so implausible that it would be 'miraculous' for its content to turn out true. The Coltheart model does not explicitly discuss negative psychological consequences of the delusion but that delusions cause harm or disadvantage is often implied.

On views of the pathological nature of delusions according to which both a harmfulness condition and a dysfunction condition are combined (the so-called 'harmful-dysfunction' views inspired by the work of Jerome Wakefield), delusions still result as pathological on the Coltheart model but not on the McKay model unless Factor 2 is described as a cognitive dysfunction as opposed to a cognitive bias.

2.3. Summary of Section 2

The two-factor theory aims at providing an account of the pathological nature of delusions, so it is not surprising that the claim that delusions are pathological is compatible with both the Coltheart model and the McKay model (see table 2 for a summary).

	Naturalism	*Normativism*	*Harmful Dysfunction*
The Coltheart Model (Coltheart et al. 2010)	The delusion is pathological because its maintenance is due to a cognitive dysfunction.	The delusion is pathological because its maintenance disrupts psychological functioning.	The delusion is pathological because its maintenance is due to a cognitive dysfunction and disrupts psychological functioning.
The McKay Model (McKay 2012)	The delusion is not pathological because it is due to a cognitive *bias,* not a cognitive dysfunction.	The delusion is pathological because it disrupts psychological functioning.	The delusion is not pathological because it disrupts psychological functioning but is not due to a cognitive dysfunction.

Table 2: Are delusions pathological?

3. Are Delusions Adaptive?

In this section, we ask whether the claim that delusions are adaptive is compatible with the Coltheart model and the McKay model. In the philosophical, psychological, and psychiatric literature there have been recent explorations of the idea that some delusions may be adaptive *in some sense* (Lancellotta and Bortolotti 2019), psychologically (McKay and Dennett 2009), biologically (Fineberg and Corlett 2016), even epistemically (Bortolotti 2015; 2016).

As anticipated, we shall focus on the shear-pin hypothesis as the best (most detailed) conceptualisation of adaptiveness as applied to delusional beliefs. The shear-pin metaphor illustrates one of the ways in which delusions could be considered as adaptive. By disabling some of its parts, shear pins allow a system which is about to collapse to continue operating, albeit in an imperfect manner. In shear-pin accounts, an adaptive misbelief is the

outcome of a process that is designed to prevent the collapse of the cognitive system. The misbelief is biologically adaptive if it enhances genetic fitness and psychologically adaptive if it contributes to wellbeing or good functioning. As we saw, after careful consideration, McKay and Dennett (2009) conclude that delusions are *not* biologically adaptive misbeliefs.[5] However, they do not rule out that some delusions can be psychologically adaptive.

Based on our analysis in section 2, both the Coltheart and the McKay models identify a factor responsible for anomalous data. In the Coltheart model the adoption of the belief is Bayesian-rational but its maintenance is due to a cognitive deficit; in the McKay model, the adoption of the delusion is due to a cognitive bias. Do such accounts leave room for delusions to be described as an adaptive emergency response?

3.1. The Coltheart Model and the Shear-pin Hypothesis

In the Coltheart model as applied to monothematic delusions such as Capgras, does the *adoption* of the unusual belief (1) emerge in the context of a crisis and (2) rescue the cognitive system from collapsing? As we saw, the unusual belief is an explanation—the best possible one—of the anomalous data brought about by Factor 1. When people lack feelings of familiarity towards a familiar face, the cognitive system produces a belief ("The woman in front of me is not my wife but is an impostor") which is false, but Bayesian-rational. The adoption of the unusual belief can hardly be interpreted as the response to a critical situation, and there seem to be no reason to believe that it would be rescuing the cognitive system from collapsing. This strongly suggests that the adoption of the unusual belief is not the outcome of a shear-pin mechanism.

Let's move now to the Coltheart model of belief *maintenance*. Does preserving the unusual belief in the face of counterevidence (1) emerge in the context of a crisis and (2) rescue the cognitive system from collapsing? In a delusion like Capgras and in the context of a deep tension between what one believes and what other people believe, remaining convinced that one's spouse has been replaced by an impostor could have some psychological benefits over believing that one has serious mental health

[5] Revisiting McKay and Dennett's shear-pin hypothesis in the light of their predictive-coding approach, Sarah Fineberg and Phil Corlett (2016) argue that the breakage of the shear pin and the consequent formation of the delusion allow an individual's cognitive system to keep functioning in the face of anomalous data. Such data, if left unexplained, would lead to the paralysis of the processes by which an individual engages in automated learning, significantly damaging the cognitive system. By explaining the anomalous data, the delusion allows automated learning to resume and the cognitive system to keep functioning. However, the cost is that all anomalous data are likely to be interpreted through the lens of the delusional belief which become more entrenched as the default explanation.

issues. Continuing to believe that one has veridical experiences and is the victim of a malicious third party (i.e., the impostor) would help preserve one's positive self-image, whereas acknowledging that one's experience is unreliable and gave rise to an implausible belief would not. In the light of this, Factor 2 could be interpreted as the sign that the shear pin has broken. If the goal is to salvage the cognitive system at the cost of disabling some of its parts, Factor 2 could be understood as the cost—the disabling of the capacity for belief evaluation.

However, the compatibility of the Coltheart model with the shear-pin hypothesis is compromised by the model branding Factor 2 as a cognitive dysfunction. Factor 2 emerges as a deficit in belief evaluation—an inability to revise one's existing beliefs in the face of disconfirming evidence. Due to such a deficit, the belief becomes resistant to counterevidence and is preserved. Factor 2 cannot be a shear-pin mechanism because it is characterised not as a design feature, but as a dysfunction, and thus the delusional belief cannot be regarded as adaptive.

What we can say, then, is that the shear-pin hypothesis is incompatible with belief adoption in the Coltheart model, because belief adoption does not respond to a crisis, and could be compatible with belief maintenance in the Coltheart model if the delusion were not branded as the outcome of a dysfunction. The delusion would be a design feature which prevents the system from collapsing.

3.2. The McKay Model and the Shear-pin Hypothesis

We saw that McKay sees the delusion as irrationally formed, that is, as a non-optimal explanation of the anomalous data caused by Factor 1. The main difference with the Coltheart model is that Factor 2 gets activated at the belief adoption stage rather than at the maintenance stage. Thus, we need not distinguish between belief adoption stage and belief maintenance stage in the McKay model because both Factor 1 and Factor 2 operate at the belief adoption stage and the unusual belief qualifies as a delusion then.

In the McKay model, then, do delusions (1) emerge in the context of a crisis and (2) rescue the cognitive system from collapsing? As with the Coltheart model, in the Capgras case the adoption of the delusion can hardly be interpreted as the response to a critical situation, and there seem to be no reason to believe that it would be rescuing the cognitive system from collapsing. Rather, the adoption of delusions is the outcome of a cognitive bias operating on anomalous data. When people with Capgras lack feelings of familiarity towards a familiar face, the cognitive system

produces a belief ("The woman in front of me is not my wife but is an impostor") which is false, but "fits" those feelings.

Can delusions more generally be seen as the output of a shear-pin mechanism in the McKay model? For the shear-pin hypothesis to apply, there needs to be a crisis the delusion is a response to (e.g., overwhelming negative emotions to manage) and this response prevents the cognitive system from collapsing. It is well known that unexplained anomalous experiences may generate uncertainty (Fineberg and Corlett 2016) and by providing an explanation of those experiences, delusions would contribute to relieve the ensuing anxiety. An example of a delusion that could be explained by the shear-pin hypothesis is the Reverse Othello syndrome (McKay et al. 2015). After recently becoming disabled, a man comes to believe that his previous partner is still in love with him and that they married, whereas his partner has moved on and is in another relationship. The realisation that his partner had left him on top of the many other changes caused by his new disability might have led the man to depression and even suicide, threatening the continued functioning of his cognitive system. In this case, it is easy to see how the shear-pin could intervene to avoid the collapse of the person's cognitive system. The adoption of the delusion (e.g., "My partner and I still are in a happy relationship") could be interpreted as a sign that the shear pin has broken: the man's desires have been permitted to exercise a powerful influence on his beliefs (see also Mele 2006). In the instance of Reverse Othello syndrome examined by McKay (Butler 2000), the man then gradually abandoned the conviction in the delusional belief that his former partner still loved him and had become his wife which suggests that the delusion did not have long-term negative consequences for the man's functioning. However, in an alternative hypothetical case in which the delusion persisted after the initial crisis had been managed, the delusion might have lost its adaptive role and become a serious hindrance.

Our conclusion is that the shear-pin hypothesis is compatible with the McKay model, because the adoption of the delusion is not due to a cognitive dysfunction, and the delusion can in some contexts be formed as a response to a crisis that prevents the cognitive system from collapsing. That said, the Capgras would not a be a good example of a delusion that is the outcome of a shear-pin mechanism and even for other types of delusions for which the shear-pin hypothesis is more plausible, it is not clear that the psychological benefits of adopting the delusion outweigh the potential long-term costs of maintaining the delusion.

3.3. Summary of Section 3

The two models of the two-factor theory we are discussing do not explicitly address the question whether delusions are adaptive, although Ryan McKay has considered the question elsewhere (McKay and Dennett 2009). It is an interesting issue, though, whether the two-factor theory is compatible with the claim that delusions are adaptive at least in the short-term, a claim that is not implausible for at least some delusions in some contexts.

We argued that the McKay model can make room for a shear-pin explanation of the adaptive nature of some delusions, whereas for the Coltheart model things get trickier (see table 3). We also observed that the overall plausibility of claims about delusional beliefs being adaptive cannot be generalised and depends on the content of the delusional belief and the context in which it emerges.

Delusions as adaptive outputs of a shear-pin breakage	
The Coltheart Model (Coltheart et al. 2010)	The maintenance of the delusion in the face of counterevidence could be a response to a crisis that prevents the cognitive system from collapsing so it could be due to a shear-pin breakage. However, this is not compatible with the belief being the outcome of a cognitive dysfunction.
The McKay Model (McKay 2012)	The adoption of some delusions is a response to a crisis that prevents the cognitive system from collapsing so it could be due to a shear-pin breakage. This is compatible with those delusions being the outcome of a cognitive bias.

Table 3: Are delusions adaptive?

4. Conclusions and Implications

We asked what two influential models of the two-factor theory of delusion formation have to say about the potential pathological nature and adaptiveness of delusions, with a special focus on monothematic delusions such as Capgras. Throughout, we made some observations which have implications for further investigations into the nature of delusions.

First, delusions can be pathological on a normativist reading of disorder, where delusions simply need to be harmful to count as pathological, although it is not clear that delusions are always the source of harm as

opposed to a response to an existing crisis that causes harm (Bortolotti 2015). Some delusions may enable the person to cope with adversities and preserve their self-esteem (Gunn and Bortolotti 2018). In one case, Barbara started believing that God was communicating with her by telepathic messages because she was his child and she was good: "as God was talking to me he was making sure that I knew there was nothing wrong with me. And he's always there, whether I'm right, whether I'm wr… well, he, he says I'm never wrong, God says I'm never wrong". Barbara developed the delusion after hearing voices for some time and her delusional belief may be considered as an explanation for her unusual experiences. Furthermore, Barbara's belief that she was special and that God was supporting her followed a very difficult time in her life, when her unfaithful husband had left her permanently and she was feeling both vulnerable and guilty about earlier decisions she made in her life. In the short term, the delusion might have protected Barbara from negative feelings about herself and prevented a suicidal attempt which was on her mind.

It is even more dubious that we can base the pathological nature of delusions on a naturalist or harmful-dysfunction reading of disorder, where delusions need to be the outcome of a dysfunctional process to count as pathological. That is because we cannot easily show that the cognitive process responsible for delusion formation is a dysfunctional process in itself as opposed to a cognitive process that operates in non-ideal conditions (such as a process whose input is the outcome of a dysfunction, a process affected by biases or performance errors, etc.).

Second, whether delusions are the outcome of a shear-pin breakage is also very difficult to ascertain in general terms. It is possible that a shear-pin mechanism works to protect a person's cognitive functioning by relieving that person from the anxiety which comes with anomalous experiences, helping the person manage negative emotions, or salvaging the person's positive self-image. However, whether the alleged benefits ever outweigh, even temporarily, the costs of having the delusion is by no means obvious and needs further examination. Some progress could be made with the issue whether delusions are psychologically adaptive if it were possible to compare the psychological profile of people with delusions with the psychological profile of people who have the same experiences as people with delusions but develop no delusions. If delusions are an emergency response which is devised to help in the face of a crisis, then people facing the same crisis as people with delusions but with no delusions should be psychologically worse off. This would help clarify if delusions are the problem or the imperfect solution to a problem (Lancellotta forthcoming).

Finally, one interesting upshot of our investigation is that in a version of the two-factor theory of delusions the same belief can be adaptive and pathological (though not at the same time). This marks an important difference between the Coltheart model and the McKay model. In the McKay model, some delusions can prevent the person's cognitive system from breaking down at the time of their adoption (and thus be adaptive as the outcome of a shear-pin breakage) *and* disrupt the person's psychological functioning in the long-term (and thus count as pathological on a normativist account). However, in the Coltheart model, delusions cannot be adaptive *and* pathological, because by being the outcome of a dysfunctional process and counting as pathological in a naturalist and harmful-dysfunction sense, the possibility that they are also the outcome of a shear-pin mechanism which breaks by design is ruled out.

REFERENCES

Davies, A. M., and M. Davies. 2009. Explaining pathologies of belief. In *Psychiatry as Cognitive Neuroscience*, eds. M. Broome and L. Bortolotti, ch. 15. Oxford: Oxford University Press.

Bortolotti, L. 2015. The epistemic innocence of motivated delusions. *Consciousness & Cognition* 33: 490-499.

Bortolotti, L. 2016. The epistemic benefits of elaborated and systematised delusions in schizophrenia. *British Journal for the Philosophy of Science* 67(3): 879-900.

Bortolotti, L. 2020. Doctors without disorders. *Aristotelian Society Supplementary* 94(1): 163–184.

Butler, P. V. 2000. Reverse Othello syndrome subsequent to traumatic brain injury. *Psychiatry* 63(1): 85-92.

Clutton, P., and S. Gadsby. 2018. Delusions, harmful dysfunctions, and treatable conditions. *Neuroethics* 11: 167–181.

Coltheart, M., R. Cox, P. Sowman, H. Morgan, A. Barnier, R. Langdon, E. Connaughton, L. Teichmann, N. Williams, and V. Polito. 2018. Belief, delusion, hypnosis, and the right dorsolateral prefrontal cortex: A transcranial magnetic stimulation study. *Cortex* 101: 234-248

Coltheart, M. 2007. Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology* 60: 1041–1062.

Coltheart, M., P. Menzies, and J. Sutton. 2010. Abductive inference and delusional belief. *Cognitive Neuropsychiatry* 15(1-3): 261-287.

Corlett, P. R. 2019. Factor one, familiarity and frontal cortex: A challenge to the two-factor theory of delusions. *Cognitive Neuropsychiatry* 24(3): 165–177.

Davies, M., M. Coltheart, R. Langdon, and N. Breen. 2001. Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry and Psychology* 8(2/3): 133–158.

Jackson, M., and K. W. Fulford. 1997. Spiritual experience and psychopathology. *Philosophy, Psychiatry, and Psychology* 4: 41-65.

Lancellotta, E. forthcoming. Is the biological adaptiveness of delusions doomed? The case of predictive coding. *Review of Philosophy and Psychology*.

Lancellotta, E., and L. Bortolotti. 2019. Are clinical delusions adaptive? *WIREs Cognitive Science* 10: e1502.

Jaspers, K. 1963. *General Psychopathology*. Manchester: Manchester University Press.

Gunn, R., and L. Bortolotti. 2018. Can delusions play a protective role? *Phenomenology and the Cognitive Sciences* 17, 813–833.

McKay, R. 2012. Delusional inference. *Mind & Language* 27(3): 330-355.

McKay, R., R. Langdon, and M. Coltheart. 2005. "Sleights of mind": Delusions, defences, and self-deception. *Cognitive Neuropsychiatry* 10(4): 305-326.

McKay R. T., and D. C. Dennett. 2009. The evolution of misbelief. *Behavioral and Brain Sciences* 32(6): 493-561.

Mele, A. 2006. Self-deception and delusions. *European Journal of Analytic Philosophy*, 2(1), 109-124. https://hrcak.srce.hr/91611

Miyazono, K. 2015. Delusions as harmful malfunctioning beliefs. *Consciousness and cognition* 33: 561–573.

Miyazono, K., and R. McKay. 2019. Explaining delusional beliefs: A hybrid model. *Cognitive Neuropsychiatry* 24(5): 335-346.

Mishara, A. 2010. Klaus Conrad (1905-1961): Delusional mood, psychosis, and beginning schizophrenia. *Schizophrenia Bulletin* 36: 9-13.

Ramachandran, V. S., and S. Blakeslee. 1998. *Phantoms in the Brain: Probing the Mysteries of the Human Mind.* New York: William Morrow.

Sakakibara, E. 2016. Irrationality and pathology of beliefs. *Neuroethics* 9(2): 147–157.

Turnbull, O., A. Fotopoulou, and M. Solms. 2014. Anosognosia as motivated unawareness: The 'defence' hypothesis revisited. *Cortex* 61: 18-29.

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 165.2:159.97
https://doi.org/10.31820/ejap.16.2.3

EXPRESSIVISM ABOUT DELUSION ATTRIBUTION

Sam Wilkinson
University of Exeter

Original scientific article – Received: 4/9/2020 Accepted: 22/9/2020

ABSTRACT

In this paper, I will present and advocate a view about what we are doing when we attribute delusion, namely, say that someone is delusional. It is an "expressivist" view, roughly analogous to expressivism in meta-ethics. Just as meta-ethical expressivism accounts for certain key features of moral discourse, so does this expressivism account for certain key features of delusion attribution. And just as meta-ethical expressivism undermines factualism about moral properties, so does this expressivism, if correct, show that certain attempts to objectively define delusion are misguided. I proceed as follows. I start by examining different attempts at defining delusion, separating broadly psychiatric attempts from epistemic ones. I then present a change of approach, according to which we question whether the term "delusion" is in the business of (merely) describing reality. I then support this proposal, first, by borrowing standard lines of argument from meta-ethics (including ontological reluctance, intrinsic motivation, and deep disagreement) but also, by inference to the best explanation of some the features we see when we try to theorise about delusion (namely that it is hard to define, and that our delusion attributions are elicited by a plurality of norms).

Keywords*: Delusion attribution; expressivism; non-factualism; epistemic norms; folk epistemology*

Correspondence: s.wilkinson@exeter.ac.uk

1. Defining Delusion

There is an ambiguity in the question "What is delusion?" In particular, is this question *paradigmatic* or *parametric*? By this I mean: is the question asking us to provide paradigmatic examples of delusion? In other words, is it asking us to point to or describe the sorts of things that get called delusions? Or is it asking us to provide parameters that strictly categorize any phenomena, even hypothetical phenomena, as delusional or not? The standard way to think of such parameters is in terms of necessary and sufficient conditions, free from counter-examples. These can then function as a sorting algorithm. You input the target phenomenon and it tells you whether it is a delusion or not. This is what is often meant by a definition, at least in philosophy.

The paradigmatic answer is often thought to be unsatisfying since it invites the follow-up question: "Yes, but *why* are these paradigmatic instances of delusions?" This invites attempts to define delusion, which I will broadly present now. I will present psychiatric approaches, then epistemic approaches, and then motivate a total change of tactic.

1.1. Psychiatric Approaches

The prominent psychiatrist Tony David speaks of the "impossibility of defining delusions" (1999) in a paper of that title:

> Most attempted definitions begin with "false belief", and this is swiftly amended to an unfounded belief to counter the circumstance where a person's belief turns out to be true. Then caveats accumulate concerning the person's culture and whether the beliefs are shared. Religious beliefs begin to cause problems here and religious delusions begin to create major conflicts […]. The beleaguered psychopathologist then falls back on the "quality" of the belief - the strength of the conviction in the face of contradictory evidence, the "incorrigibility", the personal commitment, etc. Here, the irrationality seen in "normal" reasoning undermines the specificity of these characteristics for delusions […] as does the variable conviction and fluctuating insight seen in patients with chronic psychoses who everyone agrees are deluded […]. Finally we have the add-ons: the distress caused by the belief, its preoccupying quality, and its maladaptiveness generally, again, sometimes equally applicable to other beliefs held by non-psychotic fanatics of one sort or another. In the end we are left with a shambles. (David 1999, 17-18)

This lament (and it surely is a lament with strong negatively laden terms like "beleaguered" and "shambles") is revealing of two interesting things. First, when theorists talk of "defining" delusion, they seem to want a clear presentation of necessary and sufficient conditions. This is seen by the fact that counter-examples are seen to be damaging to such a definition. What provide counter-examples to any definition (e.g. the DSM definition) are cases where the delusional status (and indeed pathological status) has already been recognised. That's why they are counter-examples! This means, though, that any definition isn't guiding our judgements about delusional status. We've already made these judgements intuitively.

The second thing is that David seems to be assuming that delusions must, as a matter of conceptual necessity, be pathological. This can be seen from the fact that he takes irrationality in the "normal", healthy, population to undermine a definition that might be based only on irrationality, rather than allowing that delusions in healthy people might not be a contradiction in terms.

Contrast this with, for example, what philosopher Kengo Miyazono (2015) writes in a paper explaining what it is that makes some delusions pathological: "I do not assume that all delusional beliefs are pathological […] I only discuss typical delusional beliefs that are pathological" (Miyazono 2015, 561, fn.1). Similarly, Valentina Petrolini (2017), also a philosopher, presents a fascinating account of what makes delusions pathological in terms of dysfunctional relevance detection. But, like Miyazono, there is no assumption that delusions must by definition be pathological, only that, when they are, and the canonical ones are, this explains why. And again, in a similar vein, Lancellotta and Bortolotti (this issue) examine the implications that different accounts (different versions of two-factor accounts, to be precise) of the Capgras delusion have for whether the delusion should be counted as pathological. This enterprise would only make sense on the (in my view very sensible) assumption that delusions aren't pathological *by definition*.

This contrast reveals two ways of approaching delusion. First, you can think of it as a diagnostically important psychiatric concept, a Jaspers-style "marker of madness", in light of which non-pathological delusion is indeed a contradiction. Second, and alternatively, you can think of it as a concept that has to be carefully defined in terms of epistemology, or, at least, not centrally as a medical phenomenon, but as a (more abstract, perhaps) mental state that is subject to various normative evaluations, most notably, epistemic ones.

One can easily understand why clinicians will tend to see delusions as pathological by conceptual necessity: delusions that are (deemed) pathological are the ones that are likely to come to their attention. But suppose that, instead of thinking of delusion as necessarily pathological, we thought of the relationship between mental illness and delusion as less direct. On such a view, the sorts of things that we call delusions tend to be pathological, but they aren't by conceptual necessity. Indeed, even if it were in practice impossible to have a delusion that wasn't pathological, it wouldn't be a contradiction in terms. For Miyazono (2015), for example, delusions might well be caused by, and indicative of, pathology (construed as harmful dysfunction (Wakefield 1992)), but aren't conceptually tied to this. We will return to the relationship between delusion and pathology later.

1.2. Epistemic Approaches

Whatever we take delusion to be, one thing that seems fairly obvious is that they (and the subjects who have them) are breaking norms, and, in particular, epistemic norms. As David puts it, at a minimum, they seem, most centrally, to be "unfounded belief". Echoing this, Max Coltheart (2007, 1043) writes: "couldn't a true belief be a delusion, as long as the believer had no good reason for holding the belief?" Indeed, the DSM 5 (American Psychiatric Association 2013) has picked up on this by dropping the falsehood requirement.[1]

This way of putting things introduces the notion of reasons, and epistemic rationality.[2] Unfounded beliefs are epistemically irrational, but epistemic rationality is a broader notion that encompasses, but is not exhausted by, evidentiary grounding (i.e. the "founding" of belief). Epistemic rationality is to be contrasted with practical rationality. Rationality in general can be thought of in terms of the attainment of certain *aims*. Practically rational

[1] Many thanks to Valentina Petrolini for pointing this out to me.

[2] In the context of a different debate, surrounding the question of whether delusions count as beliefs, Lisa Bortolotti (2009) distinguishes three kinds of rationality: procedural, epistemic, and agential rationality. Procedural rationality is about how a belief relates to other mental states, epistemic rationality about how it relates to evidence, and agential rationality about how it relates to action. The assessed claim is that these forms of irrationality, present in delusions, prevent delusions from being counted as *beliefs*. Bortolotti convincingly argues that these forms of irrationality are present in non-delusional beliefs too, and so if we were to deny belief-status for delusions, we'd have to do it for many other things that we count as beliefs. Note that this pertains to belief-status, rather than delusion-status. But since Bortolotti is keen to show that these forms of irrationality are present in non-delusional belief, she will agree with me that they can't function as fully definitive markers of delusion.

action is action that maximizes our chances of fulfilling our own aims ("motives", "desires"). So an irrational action is an action that does a bad job of fulfilling these. For example, a reluctant addict's behaviour (e.g. someone who wants to stop smoking but can't) is perhaps a prime example of practical irrationality. So what is an epistemically rational belief? Whether *we* can aim at anything while believing is controversial (Williams 1970), however, the idea that belief *itself* aims at truth is seen by many as highly plausible (Velleman 2000; Wedgewood 2002). There are many characterisations of epistemic rationality, but a simple one that suits our purposes is belief-formation that has maximized its chances of achieving its goal, namely, *truth*. So, what is actually involved in epistemic rationality, such that it is (as we say) truth-conducive?

A combination of things might count. As we've said, using adequate evidence in the formation of a belief, giving due weight to evidence that might cause you to revise your belief, not allowing motivational influences to derail your tracking of the truth (i.e. wishful thinking), having a certain degree of consistency among the beliefs that you hold, not compartmentalising information that is inconsistent, and so on. Here is the core question we come to: Can delusion be defined in terms of epistemic irrationality, thus construed?

1.2.1. Not Sufficient: Non-Delusional Irrationality

There are two issues that it is important to separate. One is a relatively minor issue: where do we draw the line? I call this issue "relatively minor", because it allows that irrationality could *in principle* do the job, but there's a challenge about where we place the threshold. Of course, one might think that a valuable revision to our practices is to think of delusion as being on a gradation, rather than something that is binary. On such a view, people aren't simply delusional or not, but are rather *more* or *less* delusional. The threshold at which someone tips from non-delusional irrationality to delusional levels of irrationality is arbitrary, or at best drawn on the basis of non-epistemic considerations, such as how well the person "functions", whether the delusion causes suffering, whether other symptoms are present, and so on. This is perhaps an attractive position.

The more damaging issue undermines even this more relaxed gradualism. What the gradualism is minimally committed to is some kind of *correlation* between irrationality and this gradual "delusionality". In other words, the more irrational you are, the more delusional you are. The worry is that this correlation may not hold. Stated plainly, Person A might be *more* epistemically *irrational* than Person B, but in fact turn out to be *less delusional*.

Consider, for example (from Nozick 1993, cited in Murphy 2012), a mother whose son has been convicted of murder. We can understand that she will be *highly* resistant to evidence that suggests that he is guilty. We will not, however (I would suggest) be tempted to call her delusional. (rather, this would be classified more naturally as *self-deception* (see Mele 2006)) People in these situations are believing in ways that are *epistemically* deeply irrational (they are far from being truth-conducive), but they are intuitively not delusional. Why is this? I would suggest that it is because we can recognise their motivations, and we can recognise the influences that these can have on belief-formation and maintenance. This means that we find their epistemic irrationality unsurprising and *understandable*. This is just part and parcel of our *folk models* of other human beings. We might even recognise (implicitly or explicitly) that in similar circumstances we would do similarly. We might even be repulsed by a mother who calmly and dispassionately evaluated evidence pertaining to her son's guilt accurately. We model other human beings (and ourselves) as understandably biased and emotional creatures. Of course, there is an extreme level of evidence-resistance at which a threshold could be crossed and we might be tempted to call the mother in our example delusional. But, crucially, the threshold is significantly higher as a result of our folk understanding of motivational influences on belief. This shows that degree of epistemic rationality alone cannot determine delusional status.

This, I would suggest, points towards a major change of approach. However, before moving on to this new approach, let's look at the other reason why delusion cannot be defined in terms of epistemic rationality, namely, that it might not even be necessary, let alone sufficient.

1.2.2. Not Necessary: Rational Delusion?

Might there be cases of delusion that don't involve *any* irrationality in the sense we have just sketched? There are two very different kinds of grounds one might have for claiming this. One is on the basis of already existing (and in principle empirically testable) theories about how certain cases of delusion come about. The other is a conceptual argument that can be supported with thought experiments.

With advancements in cognitive neuropsychiatry we have moved beyond the observable behaviour of delusional individuals to some understanding of what might underpin the formation of these delusions. In particular, there has been increasing support for the view that these delusions are in fact formed on the basis of some kind of anomaly at the level of experiential input. To put it in more intuitive terms, if you or I were to experience what these patients experience, then we too would form the

delusions that they form. As Brendan Maher presciently put it, at a time before neuropsychological theories of delusions were available, "The delusional belief is not being held "in the face of evidence strong enough to destroy it", but is being held because evidence is strong enough to support it" (Maher 1974, 99). The point is that we can think of (at least some) delusions as arising from correct use of very bizarre input (what Maher calls "evidence"), instead of from a misuse of normal input.

As we are about to see, most philosophers and neuropsychologists in the field agree that many paradigm cases of delusion have at least some experiential grounds. The main source of contention is whether this experiential anomaly is strong enough (carries enough epistemic weight) to explain why the delusion is maintained for so long, or whether we need to postulate a bias of some kind (Langdon and Colheart 2000, for example, think that we do). In the latter case, the delusional patient would be charged with epistemic irrationality.

However, whether or not there actually are biases at work is an empirical question, and our aim is to ascertain, regardless of whether certain real-world delusional patients are epistemically irrational or not, whether, *if* there *were* people who believed these bizarre things on the basis of fully adequate private grounds, and hence are plausibly epistemically rational (or at least *as* epistemically rational as "normal" people), we *would* still rightly consider them to be delusional. To put it another way, if Maher's theory happened to be correct (regardless of whether it actually is or not) would these patients still count as delusional?

Jennifer Radden (2010) calls these, rather aptly, "perceptual delusions". Her view is that, as "reasonable inferences from misleading perceptual experiences, "perceptual delusions" are *not epistemic lapses of the sort by which delusional states are identified*" (2010, 28, emphasis added). This amounts to us retrospectively revising our delusion attributions in light of a *stipulation* that delusions are tied to epistemic irrationality, and the *discovery* that a significant proportion of delusional states aren't after all irrational in the requisite way. In other words, it may turn out that some paradigmatic cases of delusion aren't delusional after all, since they aren't really irrational in the required way. This position is coherent, but such an overarching revision of what we deem to be a delusion needs to be thoroughly motivated, and I fail to really see such motivation. Are we really ready to say that a paradigmatic delusion like, for example, the Capgras delusion isn't really a delusion? Surely what matters is not the individual, experiential evidence that the person has, but how their beliefs

and believing (assertions and behaviours) fit with our social epistemological landscape.[3]

So, in contrast to Radden, I am tempted to say: "Yes, these patients still count as delusional". This is for reasons related to the change of approach I am about to present. Notice that this is also in line with Maher's implicit view. It certainly wasn't his intention to show that these patients, who we previously had taken to be paradigm cases, were not, after all, "really" delusional. Rather, the question he is answering is: *granting* that they are delusional, how can we explain their delusional state?

1.2.3. Murphy's Clue: An Alternative Approach

In a 2013 paper, Dominic Murphy doesn't quite go as far as making the proposal that I am about to, but offers an important clue that leads to it. He writes:

> A delusion is a false belief, just as knowledge is true belief, but, as with knowledge, philosophers do not rest there. Knowledge is true belief plus something else. So too, philosophers try to find *that extra property* of the false belief that converts it from a mere false belief into a delusion. (Murphy 2013, 115, emphasis added)

Putting aside the issue of delusions being accidentally true (since though truth is a way for a belief to be good, there are other ways in which it can be bad) this is to my mind a very important observation. It reminds me of something that Hartry Field wrote 15 years earlier in a wonderful paper presenting *Epistemological Nonfactualism*. He writes:

> Debates in epistemology […] often sound as if what is under discussion is the presence or absence of some mysterious justificatory fluid […]. Admittedly, one might reject the justificatory fluid picture and still regard epistemological debates as fully factual: one might say that the factual question is about which […] policies have such properties as reliability.

[3] In a very recent paper that is highly amenable to what I am saying here, Miyazono and Salice (2020) argue for the view that delusion should be seen through the lens of social epistemology, and, in particular, in terms of its relationship to what they call "social sources of evidence" rather than "individualistic sources of evidence". Adopting a similar tactic, Cardella (this issue) examines the fascinating hypothesis that delusions do not centrally involve irrationality (construed individualistically), but rather deficits in social cognition or common sense. On the contrary, she argues, delusional individuals are by some measures *more* rational and *better* at reasoning logically than non-delusional individuals.

> But this "naturalization move" obscures the fact that we are interested in which policies have factual properties like reliability only insofar as this bears on the practical question of which policies to employ. It is the practical question that is primary, and it is not itself a factual question. (Field 1998, 7)

This criticism, levelled at epistemology in general, I think could equally be levelled at attempts to define delusion, too. And Field's epistemological nonfactualism has inspired me to reflect on a similar position when thinking about delusion. To simplify somewhat, what Field is arguing is that, (i) you are going to struggle to find a descriptive, factual recipe that picks out all and only the things that count as epistemically good (e.g. "knowledge"), but even if you could (but you can't) it misses the fact that the whole point of the epistemically good is about the practical question about what epistemic policies to employ. And the question "What policy should I employ?" just isn't a factual question. Similarly, for delusion, the question "What beliefs and ways of believing should I avoid?" also isn't a factual question.

2. The Expressivist Proposal

Nonfactualism and expressivism are closely associated, although they do not strictly entail one another. Nonfactualism is an ontological position, a claim about reality, about the world, concerning whether there are facts corresponding to certain domains of discourse. Expressivism, in contrast, is a claim about the nature of the discourse itself. It is possible to be a nonfactualist, but not an expressivist about a given domain (e.g. an error theorist or fictionalist), and conversely to be an expressivist, but a factualist (e.g. a quasi-realist). However, in the absence of certain facts, a popular way of accounting for a particular domain of discourse is to be an expressivist about that particular domain, namely, to claim that, although it looks like the domain is in the business of describing facts, it is actually doing something else (namely, "expressing" something in a sense that I will make clear shortly).

2.1. What is Expressivism?

Expressivism about a certain kind of discourse is a position concerning the meaning of that discourse, or, which perhaps (depending on one's views of language) comes to the same thing, what we are *doing* when we are engaged in that discourse. Expressivists tend not to be expressivists about all kinds of discourse, so, expressivists about ethics are making a claim about ethical discourse, and usually distinguish that from other domains of

discourse, and, in particular, fact-stating discourse. "Grass is green" means what it does in virtue of the fact that it can be used to describe a fact, namely, the fact that grass is green. In terms of the psychological state of someone who uses that sentence, it is common to say that asserting sincerely (and without conceptual confusion) "Grass is green" is taken to express the belief that grass is green. Expressivists about ethical discourse who can agree that this picture is roughly correct, however will deny that it generalises to ethical discourse (see, e.g. Hare 1952). They will say that ethical sentences don't describe facts, and that the psychological states of those who sincerely assert ethical sentences, namely those that are expressed by their assertions, are not factual beliefs but something else with various proposals, including emotions (Ayer 1952), desire-like prescriptions (Hare 1952), attitudes of being for (or against) (Gibbard 1990; Schroeder 2008), impassioned beliefs (Ridge 2014), and so on.

An important step towards understanding expressivism is to understand this notion of "expression". What is *expressed*, in the sense relevant to understanding expressivism, is to be distinguished from what is *said* or *articulated*. Thus "Ouch!" is an expression of being in a state of pain, whereas the utterance "I am in pain" is an articulation of that state. Expressivism wants to think of moral claims as expressions in a way somewhat analogous to the way that "Ouch!" is an expression of pain. What a certain utterance expresses, in the relevant sense, is the mental state that it *reveals that you have*, not that it *describes you as having*. Note that fact-stating assertions express things too, but, unlike "Ouch!", they express in virtue of describing. "The cat is black" is an articulation that the cat is black, but, if sincerely asserted, is an expression of my *belief* that the cat is black; stipulating sincerity on my part, it *reveals* that I have that belief.

2.2. Two Kinds of Evaluation and Evaluative Discourse

When we say that people are delusional, we are *evaluating* them negatively. Everyone will agree with this. However, it is vitally important to distinguish two different kinds of evaluations. One we might call descriptive evaluations. What you do when you *descriptively evaluate* is you describe a benchmark, and say that the thing in question is attaining or failing to attain said benchmark. For example, you might be selecting a basketball team, and have the policy that only players over 6ft2 will be considered. There's a purely descriptive sense in which shorter players are deemed "inadequate". The assertion that "this player is too short" need only (indeed *will* only) express factual belief (e.g. the belief that this player is 6ft1). In philosophy these benchmarks are everywhere, and they are theoretically rich and informed. For example, theorists in philosophy of biology will provide conditions for biological proper function. Traditional

epistemology ("naturalized" or otherwise), of the kind Field derides, does the same for knowledge. Don't let the theoretical sophistication at play hide the fact that, like the basketball team selection, these evaluations are fact stating, descriptive. They describe a certain state of affairs and are true if and only if that state of affairs obtains. To put it another way, they are often taken to be unproblematically reducible to non-evaluative facts. In a sense, they aren't *really* evaluations; they are descriptions.

Some evaluations contrast with descriptive evaluations in being what we might call *deeply evaluative*. These evaluations are not about picking out a benchmark and stating that the thing in question either attains of fails to attain that benchmark. They are claims we make when we are evaluating as opposed to describing. This is most simply unpacked in terms of being in *evaluative* rather than *descriptive* mental states. Typical candidates of such deep evaluations are moral evaluations (right and wrong, good and evil). An expressivist about delusion attribution would take the attribution of delusion to be an evaluation in this deep and irreducible sense. This does not mean that calling someone delusional is negatively evaluating them *morally* (in fact, it often has quite the opposite effect). Rather, what moral discourse and delusion attribution have in common is that they are both evaluative in a way that doesn't allow them to be analysed in factual, non-evaluative terms.

Expressivism takes the sincere claim "Murder is wrong" to express something other than a straightforward factual belief.[4] The precise details of these are not what interest us now, but rather the view that there are kinds of claims, domains of discourse, that do not describe, that are not expressions of factual beliefs, but that do something else by expressing something else. This philosophical move, with regards to a certain kind of discourse, might be called the "basic expressivist move". One of the nicest general articulations of this move was made by Sellars (who was writing too early to have ever called himself an expressivist):

[4] At this point, a vital point of clarification is needed, related to this stipulation of sincerity. The "meaning" of the word, in the relevant sense, is preserved whether or not the speaker actually is in the relevant mental state. Even less is the meaning straightforwardly derived from the speaker's mental state. Language is a public and socially distributed affair. Rather, the meaning of the word, on the expressivists account, is derived from the mental state that the word has the *function* of expressing. There's nothing odd about this. It applies quite naturally to other uses of language. An assertion still does what it does, and means what it means, if I don't believe or otherwise endorse its content. But we understand what the assertion "The cat is black" means because its default function is to express the belief that the cat is black. Indeed, lying works precisely because it exploits this function.

> [O]nce the tautology 'The world is described by descriptive concepts' is freed from the idea that the business of all non-logical concepts is to describe, the way is clear to an *ungrudging* recognition that many expressions which empiricists have relegated to second-class citizenship in discourse are not *inferior*, just *different.* (Sellars 1957, 282)

My central suggestion is that "delusion" is not (or at least not primarily) in the business of describing. But it is not thereby inferior, just different.

2.3. Why be an Expressivist about Delusion Attribution?

Many of the considerations that motivate expressivism about ethics apply to delusion. These are:

1. Ontological reluctance
2. Intrinsic pragmatism
3. Deep disagreement

I'll go through these quite quickly, in turn, since I think that what is really interesting lies beyond this.

What I'm calling "ontological reluctance" is sometimes called (in Mackie's rather dated terminology) the "argument from queerness", although I take it to be broader and more general. Some theorists are generally reluctant to posit a strange ("queer", namely, "of a very strange sort, utterly different from anything else in the universe" (Mackie 1977, 38)) realm of moral properties or facts. But more generally, regardless of what we are calling these things (properties, facts etc.) there can also be a general reluctance to engage in ontology in the classical sense (e.g. social ontology being exempt) *at all*, when we can account for the phenomenon in question without any mysteries outstanding. Within the context of this ontological reluctance, expressivists about ethical discourse feel a certain calm when they reflect on the fact that social creatures like ourselves will have sought to regulate behaviour in a pro-social way by expressing (revealing) to conspecifics their disapproval, and thereby motivating the community at large to reward and punish so as to secure adherence to social norms (morality, politeness etc.). Similarly, the argument would go, there are no *sui generis* delusion-pertaining (or indeed, knowledge-pertaining) facts or properties. Social creatures like us who communicate and try to live in groups, are going to give rough-and-ready seals of approval (thumbs up) to good epistemic states and practices, and give thumbs down to poor ones. The fact that the *words* "knowledge" and "delusion" emerged in English, and became roughly regimented, is just a distraction.

This relates to the second consideration. Moral discourse is intrinsically motivating. There is a certain contradiction to sincerely claiming "Murder is wrong" while not thereby feeling motivated in certain ways, e.g. a *ceteris paribus* reluctance to murder, encourage others not to murder, etc. Similarly, delusion discourse seems intrinsically motivating: it would be inconsistent to regard someone as delusional, and yet have no inclination to refrain from taking what they are saying seriously, no inclination to not argue against them, and so on. Of course, these inclinations are multi-track and dispositional. You don't have to act in accordance with them, and they don't have to be exhaustively listed. The evaluative state is not simply the aggregate of these motivations, rather the motivations fall out of the evaluative state.

Finally, there is the consideration behind deep disagreement. In these instances, all of the facts pertaining to a particular case are agreed by two individuals, and yet there is still disagreement about where something is morally wrong. There is no further fact that can be learnt in order to bring the two disagreeing subjects in line with one another. Therefore, it is not a disagreement about facts, but about something else. Of course, on many very serious moral infringements (murder), unanimity is not hard to find, but for more contentious culturally specific "beliefs" (sex before marriage, homosexuality, abortion etc.) these deep disagreements are rife. A similar thing could be said for delusion. There are not only disagreements about what counts as good/bad, acceptable/unacceptable belief *contents*; there are also disagreements about what counts as good *methods and procedures* for forming beliefs. Murphy (2013) presents this example in a paper that very much follows the spirit, if not to the letter, of what I am saying here.

> Boyer, (2001, 69-70) reporting fieldwork done by Wendy James in the Sudan, discusses ebony trees that are believed to be a source of social information. The trees record conversations, and are privy to the plans of witches. You can learn what they know by burning an ebony twig, dipping it in water and reading the pattern of ashes in the water. A belief in cognitive interaction with ebony trees counts as culturally normal, and hence not delusional or otherwise suspect. (Murphy 2013, 22)

This final "delusional or otherwise suspect" is very much in keeping with the picture I'm presenting. To call something delusional is to express your folk-epistemic disapproval, to flag it as suspect. Aside from these theoretical considerations, there is a far more intuitive consideration one can appeal to: it *just seems right*. Just consider something you might overhear in public between two friends: "You're *delusional* if you think

that Manchester United can qualify for the Champions League!" You'll grant me that this seems like an expressive rather than descriptive use of language. But is this a particularly exotic and non-standard use of the word? Is this making expressive use of a linguistic tool that is originally purely descriptive? Or is it a hyperbolic use of a linguistic tool that is already to some extent expressive? I would be tempted to say the latter. Just because the word "delusional" is being uttered calmly by someone in a white lab-coat holding a thick book, doesn't make it any less expressive in its semantics. (Recall that the speaker does not have to be in the emotional state canonically expressed by the word, any more than I have to believe every single descriptive assertion that I utter.)

2.4. The Consequences of Expressivism about Delusion Attribution

The consequences of expressivism serve to lend further support to it. In a sense, we can adopt expressivism as an inference to the best explanation, since some of its consequences align with what we already observe.

2.4.1. Inability to Define is to be Expected

The inability to define delusion is not only to be expected, but embraced. If delusion talk expresses (reveals) our reactive folk epistemological attitudes, then we would certainly not expect these attitudes to track consistent parametric properties that can be captured by necessary and sufficient conditions. These are not going to survive the scrutiny of counterexamples. But, again, where do the counterexamples come from in the first place? I'd say, our reactive folk epistemological attitudes. The definitions don't function to tell us what's delusional: we have a sense of that already. Similar things of course can be said of our sense of right and wrong. And, again, we would expect all sorts of things to interfere with any clear, factual, theoretically informed judgment of delusion. Culture, motivation, even the way in which the case is presented, may influence the extent to which someone deems a belief (or assertion) to be bad. This relates to the second consequence.

2.4.2. Disjunctive Norm Pluralism

There are many different ways in which a belief (and related phenomena, like inquiry, reasoning, etc.) can be good or bad. Two obvious ways are the process-independent dimension of truth and falsehood, and the process-dependent dimension of rationality. Both of these contribute to the "badness" of the belief, and all that matters is that there is enough folk-detectable badness. It doesn't matter where it comes from. Take reverse Othello delusion. The belief content that the subject's wife is not cheating

on him is a perfectly plausible content taken in isolation. It is true (one hopes) of hundreds of thousands of people around the world. What makes it delusional is the subject's baffling blindness in the face of counterevidence. On the other end of the spectrum, it matters little what evidence a delusional patient might cite for the claim "I am the left foot of God." We just don't see how that could possibly be true. There are likely many other epistemic norms that we detect (e.g. cognitive flexibility, relevance detection etc.) and they may all be involved in tipping the balance toward the (folk-epistemically) bad or good. Since it doesn't matter on what types of grounds the belief is deemed to be bad, we have what we might call *disjunctive norm pluralism* when it comes to something being delusional or not.

A related point concerns "understandability". Like we saw in the case of the mother in an *understandable* level of denial about her son's guilt, a sort of pluralistic criterion of understandability tracks our delusion attributions better than something like rationality. Interestingly, Jaspers is often quoted as saying that delusions are "un-understandable", and this is often interpreted as meaning that they cannot be theoretically understood. Whether or not this is the correct interpretation, there is an interpretation of this claim according to which it approximates an accurate claim. That is, if we think of all of the different folk-epistemic norms as constituting this rough criterion of "understandability", which is basically about whether somebody adheres to our predictive models of how humans should behave (i.e. they should be resistant to evidence that casts their loves ones in a very bad light, up to a certain point). Calling something delusional is to say: "Wow, this person is flying in the face of the models I use to make sense of people!"

3. Delusion and Pathology Revisited

We examined the idea that delusion and pathology should not be conceptually tied to one another. However, suppose that we are expressivists about delusion attribution. What does that say about the relationship between delusion and pathology? There are a number of options, depending on how we think of pathology.

One way to go is to think that, whereas delusion is a folk concept that is deeply evaluative, pathology is a theoretical notion (or at least *should* be (see Boorse 1975)). Then that theoretical notion is to be thought of in objective and fact-stating terms. This would grant total conceptual independence between delusion and pathology. It might turn out that many of the things that we deem to be delusional are the results of things that,

according to this factual, theoretical notion, are also pathological. To take an imperfect, but still helpful, analogy: I might dislike the taste of tomatoes. A chemist might be able to isolate the exact compound in the tomato that arouses my dislike. The claim that tomatoes contain that compound is a factual claim, my assertion "tomatoes are yucky" is not (note, though, the claim that "SW dislikes tomatoes" is clearly factual – just like claims about whether an individual attributes delusion is factual although the attribution itself is not).

A closely related view would have a hybrid approach to pathology, e.g. Wakefield's harmful dysfunction account, where the factual component of dysfunction is necessary but not sufficient: the value-laden notion of harm is needed in addition. This means, similarly, that some of the things that we deem to be delusional are pathological. Note that not only is this consistent with Miyazono's paper, his task is to use Wakefield to tell us why the things we deem to be delusions that are pathological count as pathological. He is, in a way that is very much in the spirit of what I am saying here, not interested in delineating the realm of the delusional.

Finally, we could be non-factualists about pathology too, for similar reasons to those motivating expressivism about delusion attribution (although I'd be tempted like Boorse to distinguish disease from illness, where the latter may warrant non-factualism, but the former notion could sensibly be introduced as a factual notion). Then there is a further bifurcation. We should first establish whether the reactive dispositions that underpin our delusion attributions and those that underpin our attributions of illness (or perhaps specifically mental illness) embody norms that are the same, or similar, or completely different. In other words, we need to ask: What is the relationship between our folk-epistemology and our folk-psychiatry, as embodied in our reactive dispositions? The sorts of conditions under which someone (perhaps specifically "a Westerner") might call someone "delusional", "crazy" or "unwell", might well overlap substantially, but not entirely. In short, we might revert to a very close connection between delusion and pathology, but one that looks very different to the presumed factualist orthodoxy. Indeed, we might even revert back to a simple "definition" of delusion as "pathological belief". This is not to be unpacked as a factualist, theoretically-informed, definition, but rather as a way of flagging that our multi-track sensibilities deem belief to be pathological in the simple folk sense that it can't be "understood", is weird, alien, flies in the face of how human beings *ought* to be, and needs *correcting*.

4. Conclusion

First I'd like to clarify that my proposal is incomplete in that, although it claims that delusion attribution does not describe and does not function to express factual belief, I have not given a clear positive account of what it does instead, and what it does express. This would have to be left for another day, but I would suggest that it would be consistent with our folk-epistemic practices in general, and that these more approximately track social epistemic rather than individual epistemic norms (Miyazono and Salice 2020). As for what is expressed by delusion-talk, my hunch is that this is not going to be something individual like a mental state, but something socially distributed. For want of better terminology, what I have in mind is something like "flagging as suspect", or "enjoining to action". In short, it's about the role it plays in a community, more than the mental state that the individual is in.

Although, theoretically, my central proposal here might seem radical, in practical terms it is not requiring much revision to existing work. Indeed, much of the philosophical work on delusion glosses over strict definitions of delusion, or, at best, provides working or rough definitions, backed up by canonical examples. Then, philosophers focus on the canonical cases themselves asking questions like: "Are they really beliefs?", "What makes them pathological?", "On what grounds might the delusional judgements be made?" What I suggest here has no direct bearing on these questions, for they deal with the phenomena themselves, as already picked out. What I'm talking about here is how the picking out itself seems to work. My suggestion is that, not only is it untidy: it is not even descriptive.

There is, however, *some* impact of what I'm saying on this work (aside that some may find it interesting). It might be that there is an expectation that delusion could be cleanly defined; it just hasn't been achieved yet. And the take-home message here is that this would be a mistake, and we should rest comfortable in the understanding of the kind of term that "delusion" really is. In a related manner, a lot of philosophical work on delusion lacks clear quantification. You see questions like: "Are delusions beliefs?", "Are delusions irrational?", "What makes delusions pathological?", "Are delusions harmful malfunctioning beliefs?" A logician presented with these questions would ask: *All* delusions? And if so, is it by definition or contingently the case? My proposal makes explicit what remains largely implicit: that these questions deal with the paradigms, not the parameters, of delusion.

Finally, what I am suggesting here in no way undermines careful psychiatric taxonomy, and diagnostic clinical practice. In fact, it siphons

off the question of whether something is a delusion or not as something that is not of theoretical relevance. In a modification of G. E. Moore's Open Question Argument (Moore 1903), you can describe any condition in the greatest detail, and someone could still without inconsistency or ignorance ask, "Yes, I understand this condition, but is it *delusional*?" Establishing delusion-status is not a scientific or theoretical enterprise, and it saves time, energy and confusion to recognise this.

Acknowledgements

I would like to thank Valentina Petrolini, Emiliano Loria and Marko Jurjako for extremely helpful comments.

REFERENCES

American Psychiatric Association. 2013. *Diagnostic Statistical Manual of Mental Disorders*, Fifth edition (DSM-5).

Ayer, A. J. 1952. *Language, Truth and Logic*. New York: Dover Publications, first Dover edition.

Blackburn, S. 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.

Boorse, C. 1975. On the distinction between disease and illness. *Philosophy and Public Affairs* 5: 49–68.

Bortolotti, L. 2009. *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Boyer, P. 2001. *Religion Explained*. New York: Basic Books.

Cardella, V. this issue. Rationality in mental disorders: Too little or too much? *European Journal of Analytic Philosophy*. https://doi.org/10.31820/ejap.16.2.1

Coltheart, M. 2007. Cognitive neuropsychiatry and delusional belief (The 33rd Sir Frederick Bartlett Lecture). *The Quarterly Journal of Experimental Psychology*, 60(8): 1041–1062.

David, A. S. 1999. On the impossibility of defining delusions. *Philosophy, Psychiatry, and Psychology* 6(1):17-20.

Field, H. 1998. Epistemological nonfactualism and the aprioricity of logic. *Philosophical Studies* 92(1): 1-24.

Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

Hare, R. M., 1952. *The Language of Morals*, Oxford: Clarendon. https://doi.org/10.1007/s11229-020-02863-1

Jaspers, K, 1963. *General Psychopathology*. Translated by J. Hoenig and M. Hamilton, Manchester: Manchester University Press.

Lancellotta, E. and Bortolotti, L. this issue. Delusions in the two-factor theory: Pathological or adaptive? *European Journal of Analytic Philosophy*. https://doi.org/10.31820/ejap.16.2.2

Langdon, R., and M. Coltheart. 2000. The cognitive neuropsychology of delusions. In *Pathologies of Belief*, eds. M. Coltheart and M. Davies, 183–216. Oxford: Blackwell.

Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong.* Harmondsworth: Penguin.

Mele, A. 2006. Self-Deception and delusions. *European Journal of Analytic Philosophy* 2(1): 109-124. https://hrcak.srce.hr/91611

Miyazono, K. 2015. Delusions as harmful malfunctioning beliefs. *Consciousness and Cognition* 33: 561-573.

Miyazono, K., and A. Salice. 2020. Social epistemological conception of delusion. *Synthese*.

Moore, G. E. 1903. *Principia Ethica*, Cambridge: Cambridge University Press.

Murphy, D. 2012. The folk epistemology of delusions. *Neuroethics*, 5(1): 19–22.

Murphy, D. 2013. Delusions, modernist epistemology and irrational belief. *Mind & Language*, 28(1): 113–124.

Nozick, R. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

Petrolini, V. 2017. What makes delusions pathological? *Philosophical Psychology* 30(4):1-22.

Radden, J. 2010. *On Delusion*. Abingdon and New York: Routledge.

Ridge, M. 2014. *Impassioned Belief*. Oxford: Oxford University Press.

Schroeder, M. 2008. *Being For*. Oxford: Oxford University Press.

Sellars, W. 1957. Counterfactuals, dispositions, and the causal modalities. In *Minnesota Studies in the Philosophy of Science*, eds. H. Feigl, M. Scriven, and G. Maxwell, 225–308. Minneapolis: University of Minnesota Press.

Velleman, J. D. 2000. *The Possibility of Practical Reason*. Oxford: Clarendon.

Wakefield, J. C. 1992. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47(3): 373–88.

Wedgwood, R. 2002. The aim of belief. *Noûs* 36: 267–297.

Williams, B. 1970. Deciding to believe. *Language, Belief, and Metaphysics*, eds. Howard E. Kiefer and M. K. Munitz, 95-111. Albany: SUNY Press.

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 1:616.89-008
https://doi.org/10.31820/ejap.16.2.4

TOO MUCH OR TOO LITTLE? DISORDERS OF AGENCY ON A SPECTRUM

Valentina Petrolini
University of the Basque Country (UPV/EHU)

Original scientific article – Received: 18/7/2020 Accepted: 22/8/2020

ABSTRACT

Disorders of agency could be described as cases where people encounter difficulties in assessing their own degree of responsibility or involvement with respect to a relevant action or event. These disturbances in one's sense of agency appear to be meaningfully connected with some mental disorders and with some symptoms in particular—i.e. auditory verbal hallucinations, thought insertion, pathological guilt. A deeper understanding of these experiences may thus contribute to better identification and possibly treatment of people affected by such disorders. In this paper I explore disorders of agency to flesh out their phenomenology in more detail as well as to introduce some theoretical distinctions between them. Specifically, I argue that we may better understand disorders of agency by characterizing them as dimensional. In §1 I explore the cases of Auditory Verbal Hallucinations (AVH) and pathological guilt and I show that they lie at opposite ends of the agency spectrum (i.e. hypoagency versus hyperagency). In §2 I focus on two intermediate cases of hypo- and hyper- agency. These are situations that, despite being very similar to pathological ones, may be successfully distinguished from them in virtue of quantitative factors (e.g. duration, frequency, intensity). I first explore the phenomenon of mind wandering as an example of hypoagency, and I then discuss the phenomenon of false confessions as an example of hyperagency. While cases of hypoagency exemplify situations where people experience their own thoughts, bodies, or actions as something beyond their control, experiences of hyperagency provide an illusory sense of control over actions or events.

***Keywords**: Agency; auditory verbal hallucinations; guilt; mind wandering; false confessions*

Correspondence: valentina.petrolini@ehu.eus

Introduction

The sense of agency of an individual is normally characterized in terms of *self-attribution* and *self-ascription* and is usually connected with an appropriate assessment of one's actions. Feelings related to agency importantly include the sense of being able to do something, of being the agent of an action (Proust 2013), as well as the sense of being in control (Pacherie 2008). These capacities allow individuals to correctly determine the scope of their thoughts and actions, and also to reliably distinguish between self-generated and other-generated stimuli. When assessing one's sense of agency, it is important to distinguish between the correctness of *self-attribution* and the subjective *feeling of agency or control*.[1] With respect to the former, one may self-attribute agency concerning the things she has not done or fail to self-attribute agency concerning the things she has in fact done. In other words, self-attribution may be correct or incorrect. By contrast, the feeling of agency or control comes in degrees: one may be more or less sure or confident about having performed an action. Disorders of agency could thus be described as cases where people encounter difficulties in the two senses just described. On the one hand, they may experience issues in terms of self-attribution and thus fail to correctly determine whether they performed the relevant action. On the other hand, they may experience a diminished (or unduly strong) feeling of agency or control. As I show later in the paper, there are cases in which these two senses of agency come apart and others in which they go together.

In this paper I first characterize disorders of agency as lying on a spectrum. I then show that disturbances at both ends of this spectrum are connected to some mental disorders. On the one hand, a person may be unsure of whether she initiated an action that others attribute to her, or she might deny having done so despite evidence to the contrary. I call this kind of disturbance *hypoagency*. Extreme cases of hypoagency encompass phenomena such as auditory verbal hallucinations (AVH henceforth), thought insertion or alien hand syndrome, where people experience their thoughts or bodies as something acting beyond their control. On the other hand, a person may feel that events that are completely unrelated to her actions (or thoughts) fall under her own responsibility and therefore experience unbearable guilt as a result. This happens at times with

[1] In this paper I treat "agency" and "control" as synonymous, although I am aware that some finer-grained distinctions may be drawn between them (Pacherie 2007). For the purposes of my discussion, the feeling of agency and control seem to go hand in hand: gaining or losing control fundamentally implies augmenting or deteriorating one's sense of agency. As a consequence, the two notions cannot significantly come apart, i.e. one cannot be in control without at the same time experiencing a sense of agency.

schizophrenic individuals, who tend to blame themselves for natural disasters, terrorist attacks, or murders committed by others. In these cases, subjects attribute to themselves a greater degree of agency and control than they actually possess, thereby exhibiting *hyperagency*. As I explain in more detail below, in both cases an underlying sense of agency appears to be compromised. On the one hand, extreme cases of hypoagency exemplify a situation in which self-attribution of agency is incorrect (i.e. thoughts and bodies are not experienced as one's own) and the subject lacks a robust feeling of agency or control. On the other hand, extreme cases of hyperagency exemplify a situation in which self-attribution is also incorrect—albeit in the opposite direction (i.e. one believes to have performed actions that she has in fact not performed)—but the subject reports a strong feeling of agency.

The paper is structured as follows. In §1 I explore the cases of AVH and pathological guilt and I explain how they lie at opposite ends of the agency spectrum. In §2 I focus on two *intermediate cases* of hypo- and hyper-agency: these are situations that—despite being very similar to pathological ones—may be successfully distinguished from them in virtue of quantitative factors. As an intermediate case of hypoagency I explore the phenomenon of *mind wandering*, where intrusive thoughts, memories, and feelings tend to pop up and interfere with the completion of other tasks. As an intermediate case of hyperagency, I discuss the phenomenon of *false confessions*, where people end up pleading guilty for crimes they did not in fact commit.

1. Hypoagency and Hyperagency: Extreme Cases

1.1. Hypoagency: Auditory Verbal Hallucinations (AVH)

Disorders of hypoagency can be characterized as situations in which a person loses grip over her own thoughts or actions, thereby experiencing them as alien and beyond her control. One extreme example is the occurrence of AVH, also known as "hearing voices". Although there is evidence that these experiences are frequent in non-clinical populations (Johns et al. 2014; Allen et al. 2006), as well as in depressive disorders (Toh et al. 2015), AVH are often taken to represent one of the hallmarks of schizophrenia (Henriksen, Raballo and Parnas 2015). Many researchers have suggested that AVH would result from failures in self-monitoring mechanisms (Frith 1992; Jones and Fernyhough 2007). These views characterize self-monitoring issues as failures to correctly predict action outcomes in several domains, such as motor behavior (e.g. self-tickling), cognition (e.g. planning difficulties), or inner speech (e.g. AVH). Issues

with self-monitoring are also likely to affect cognitive control and executive functioning at various levels, from implementing basic goals to carrying out higher order plans (Petrolini, Jorba and Vicente 2020). Applied to inner speech production, these difficulties may particularly affect what has been labeled "dialogic inner speech" (Fernyhough 2004), which refers to the conversations we have with ourselves. In this respect, inner speech in people experiencing AVH exhibits peculiar characteristics. For instance, AVH subjects appear to experience more intrusions in inner speech, often in the form of *other people* being present (Alderson-Day et al. 2014). Many describe the voices as exhibiting a markedly "alien" character and as differing sharply from first-person inner speech (Nayani and David 1996). People experiencing AVH also tend to appraise their inner speech as more *negative* (Hugdahl et al. 2012), *dystonic*—i.e. failing to align with the person's self-attributed thoughts and emotions (Lopez-Silva 2016), and *fragmented*—i.e. distributed across more than one "voice" without being temporally coordinated or synchronized (Langland-Hassan 2008). Besides their relevance to inner speech, AVH showcase relevant facts about agency (Proust 2006) and ownership (Maiese 2015). Notably, they qualify as an experience in which sense of *agency* (i.e. X is caused by me, I am the author of X) and sense of *ownership* (i.e. X is mine, X is part of my experience) come apart. Indeed, in AVH subjects experience voices as alien—thereby denying authorship—but still as occurring within their bodily or mental boundaries in some significant sense—thereby preserving ownership (Proust 2013). In other terms, AVH experiences exhibit self-misattribution as well as a diminished sense of agency or control.[2]

A more detailed phenomenology of AVH may be garnered through the first-person account offered by Longden (2013). In her vivid report about the experience of "voice-hearing", Longden recalls the first appearance of this phenomenon during her early college years. She describes her younger self as struggling with severe anxiety and worries about the future, but also as exhibiting a strong tendency towards suppressing her feelings. The first voice makes its appearance one evening while Eleanor is going home after a class: she characterizes it as neutral, similar to her own voice but narrating all her actions in third person, like a running commentary—e.g. "She is leaving the room"; "she is opening the door". In the following weeks voices grow in number and intensity, becoming more persistent and menacing: in particular, *they* start threatening Eleanor and make her

[2] Another pathological case of hypoagency would be Obsessive Compulsive Disorder (OCD), where subjects experience being compelled to act in a particular way or report a sense of performing an uncontrollable action (Szalai 2019). Yet, as opposed to what happens with AVH, OCD subjects tend to self-attribute actions correctly while experiencing a diminished sense of agency/control. In this sense, OCD may qualify as a clinically relevant, but less extreme disorder of agency.

comply with a series of bizarre tasks with the promise of "getting her old life back". These tasks are experienced by Eleanor as some sort of "Labors of Hercules" over which she has absolutely no control, but that she nonetheless feels forced to carry to completion. She describes them as initially quite small (e.g. pull out a few strands of hair) but then as progressively more extreme (e.g. harm yourself) or violating social norms (e.g. pour a glass of water on the head of the instructor during a lecture). Notably, she experiences overwhelming feelings of powerlessness because she lacks the resources to exercise any form of control over the voices. Her agency appears so compromised that at one point she attempts suicide by trying to drill a hole in her head in order to get the voices out.

The second part of Longden's report is devoted to her process of recovery, which begins once she gets in touch with the UK-based *Intervoice* movement, founded in 1988 by psychiatrists Romme and Escher. The tenet of this therapeutic movement consists in claiming that voices should be treated as experiences rather than symptoms, and that the content of the voices often provides important insights into the person's life story and personality. The primary goal of this approach is not to get rid of the voices *per se*, but to accept them while learning a series of coping strategies focused on "taking the power back" from them. The turning point towards recovery consists in realizing that voices may be appropriate responses to traumatic life experiences (e.g. childhood abuse) or ways to get in touch with one's repressed emotions. For Longden this was clearly the case. During therapy she realizes that many of the voices—especially the more aggressive ones—were mirroring her hidden emotions: "Whenever I repressed anger (and that happened very often) the voice sounded frustrated" (Longden 2013). Another patient describes this phenomenon as follows:

> When the voices said: "See how awful she looks", it happened on days when I felt myself pretty awful. But they always made such exaggerated statements. By exploring this I started to realize that in a certain way the voices expressed my own thoughts. It is rather strange, but they are your own thoughts about an emotion. (Romme and Morris 2013, 263-264)

The treatment proposed by Romme and Escher appears particularly interesting for our purposes because it focuses on coping strategies to regain control over the voices (Romme and Escher 1993). Indeed, it could be seen as a way to *enhance agency* in people that experience a significant diminution in their power of controlling their mental events. Romme and Morris (2013) characterize recovery as a process of progressively gaining control over the voices by creating a dialogue with them, while at the same

time setting boundaries and avoiding being overwhelmed. Romme and Escher's approach thus appears to counter hypoagency by strengthening a sense of familiarity with the voices. The more the patient learns to incorporate the voices in her experience and to treat them as legitimate (or at least revealing) aspects of her personality, the more agency over them is restored.

1.2. Hyperagency: Pathological Guilt

Pathological guilt represents an extreme case of hyperagency which is commonly experienced by people suffering from depression, although it may also be present, albeit in a different form, in schizophrenic patients. People experiencing pathological guilt tend to feel responsible for things that they have not done or feel deeply disturbed by actions and thoughts that are regarded as innocuous by others. What these cases have in common is the subject's inability to properly assess the scope of their (moral) responsibility. Pathological guilt may manifest itself in different ways. Some people with schizophrenia attribute to themselves actions for which they are in fact not responsible—e.g. a murder that someone else committed. For example, Saks (2007) reports being filled with anxiety when reading the newspaper because she would blame herself for every violent crime reported in the area. Alternatively, some people suffering from depression assign a particularly negative valence to self-generated thoughts and events—e.g. feeling extremely guilty about finding another person annoying. Unlike AVH experiences, cases of pathological guilt combine incorrect self-attribution with an exaggerated feeling of agency or control over the relevant action or event.

One interesting example comes from one of Freud's earliest case histories, Emmy von N. (Freud and Breuer 1893, 48-105). Frau Emmy is a 40-year-old woman who suffers from recurring hallucinations and from a number of tic-like movements, in particular an idiosyncratic "clacking sound" that would come up whenever she is anxious or frightened. While analyzing her case, Freud notices that the patient tends to be overly hard on herself and to feel directly responsible "for the least signs of neglect": "If the towels for the massage are not in their usual place or if the newspaper for me to read when she is asleep is not instantly ready to hand" (Ibid., 65). One day, Freud arrives to the patient's house to continue the therapy and finds her in a state of great distress, repeating: "Am I not a worthless person? Is it not a sign of worthlessness what I did yesterday?" Freud cannot recall what happened the day before to justify such a "damning verdict" (Ibid., 70). Despite Freud's repeated admonitions not to feel guilty over small things, Emmy keeps behaving like "an ascetic medieval monk, who sees the finger of God or the temptation of the Devil in every trivial

event of his life and who is incapable of picturing the world even for a brief moment or in its smallest corner as being without reference to himself" (Ibid., 66). Notably, after a two-year long therapeutic process, Emmy is able to recover from the majority of symptoms—i.e. hallucinations, tics—but her inclination to torment herself over "indifferent things" never vanishes completely.

More recent accounts of melancholia—such as the one offered by Radden (2009)—suggest that Freud contributed to conceptualize depression as a state of mind characterized by *self-criticism*, where "dissatisfaction with the self on moral grounds" and "delusional expectation of punishment" stand out among the most typical clinical features (Freud 1917, 153). This point allows us to connect extreme forms of hyperagency with disturbances in one's sense of confidence. Indeed, diminished confidence may play a role in over-attributing guilt to oneself in the face of negative events (e.g. "It happened to me *because I am bad person*"). It is thus unsurprising that pathological guilt is often found in the context of depressive disorders, in which self-loathing tends to feature prominently (see Plath 1963; Styron 1991 for some first-person accounts).

The pervasive presence of guilt feelings in some psychiatric disorders has also been explored by authors working in the field of psychology and philosophy of emotions. Frijda (1985), for instance, connects guilt with the sense of being in control: "[Guilt feelings] may provide an explanation for one's misery, an explanation that provides an aspect of controllability, some shred of it, in the morass of helplessness; it permits acts of contrition and efforts at paying penance" (Frijda 1985, 431). In this sense, hyperagency may arise as an attempt to control and therefore justify or explain feelings of worthlessness and helplessness experienced in depression. Ratcliffe (2010) rather characterizes depressive guilt in terms of *depth*. As opposed to a circumscribed feeling of guilt about a specific action or event, depressed subjects tend to experience guilt as an "all-encompassing way of being" (Ratcliffe 2010, 609). First-person reports of depressed patients support this idea: "The reason my life is so awful at these times is *because I am a terrible, wicked, failure of a person*"; "Everything that goes wrong in my life is *directly my fault*" (reported by Ratcliffe 2015, 135. Italics mine). In these cases—such as Freud's patient Emmy—guilt shapes one's perception and appraisal of other people, objects, and events. In this sense, pathological guilt shares important similarities with delusional beliefs: one's belief of being responsible brings about an experience of reality in which environmental stimuli are overwhelmingly interpreted in light of such conviction (Bortolotti and Miyazono 2015).

In the next section I discuss some intermediate cases of hypo- and hyper-agency. Although these examples bear important similarities to the ones analyzed in §1.1. and §1.2., I show that they may be successfully distinguished from them by appealing to quantitative factors (e.g. duration, frequency, intensity). As an example of hypoagency I introduce phenomena such as distraction and daydreaming, where the sense of control over one's thoughts appears moderately diminished. As an example of hyperagency I discuss the phenomenon of false confessions, in which people over-attribute responsibility to themselves to the point of accepting punishment for crimes they did not commit.

2. Hypoagency and Hyperagency: Intermediate Cases

2.1. Hypoagency: Mind Wandering

Phenomena like distraction, daydreaming or mind wandering are extremely common in our everyday experience. We are working on an important project and we suddenly start thinking about the grocery list or our plans for the evening. We try to concentrate on a task when memories pop up and absorb us for some time before we are able to resume our previous activity. In most cases these thoughts arise automatically and are difficult to regulate. They can be seen as paradigmatic cases of *hypoagency* in which self-attribution is correct but the feeling of agency appears at least moderately diminished.

Despite their pervasiveness in our ordinary life, phenomena of mind wandering have only recently become the object of systematic scientific investigation, mostly due to the growing number of neuroimaging results about brain activity in rest conditions. This neural pattern has come to be known as the Default Mode Network (DMN henceforth) and its discovery suggests that mind wandering might constitute a psychological baseline from which people depart when engaging in demanding tasks and to which they return when their attention is not allocated elsewhere (Mason et al. 2007; Andrews-Hanna 2012). Although cases of excessive mind wandering have been at times granted pathological status (Schupak and Rosenthal 2009), this phenomenon has also been associated with an increase in creativity and problem-solving abilities. Indeed, the neural profile of brains in DMN is similar to the one exhibited by subjects engaged in conceptual processing and problem-solving tasks (Smallwood and Schooler 2006). In the past decade, researchers working in different fields—philosophy of mind, psychology and neuroscience in particular—have attempted to shed light on the nature of mind wandering while formulating hypotheses of its adaptive value. Mind wandering has been

also characterized as a form of "mental autonomy loss" because of its spontaneous, automatic and task-unrelated nature (Metzinger 2013). The notion of mental autonomy proposed by Metzinger partially overlaps with what I call agency in this paper, and comprises the ability to causally determine one's actions (self-attribution) as well as the ability to control the conscious content of one's mind (feeling of agency or control). Due to the ubiquitous interruptions caused by mind wandering, Metzinger suggests that we should regard mental autonomy as "the exception rather than the rule" (Metzinger 2013, 5).

On this view, mind wandering has several advantages, such as allowing individuals to maintain a baseline arousal activity where past, present and future mental events hang together in a (virtually) unitary whole. Similarly, mind wandering has been connected with a number of positive effects on psychological functioning, such as consolidating memories, planning future events and delaying gratification (Smallwood and Andrews-Hanna 2013). This activity thus appears to grant the mind some freedom from the "here and now" and allows agents to perform mental actions that are not simply responses to the outside world. If this is correct, it becomes easier to see how mind wandering might be connected to creative and problem-solving processes. In a recent study on the topic, Baird et al. (2012) assigned the Unusual Uses Task (UUT) to 145 participants, asking them to generate as many uses as possible for a common object (e.g. a brick) in a given amount of time. After having read the list of objects, three groups of participants were subject to an incubation period during which some subjects were administered a demanding task, others an undemanding task and still others were allowed to rest. A fourth group proceeded to solve the problem without taking a break. The results indicate that participants engaging in the non-demanding task during the incubation period performed significantly better than the ones who were assigned a demanding task, no task at all or that did not have an incubation period (Baird et al. 2012, 5). The researchers suggest that engaging in a simple task allowed participants to mind wander during the incubation period and this in turn helped them formulating more creative solutions to the UUT.

As I suggest above, mind wandering can be regarded as a paradigmatic instance of hypoagency. It typically starts out as an automatic and spontaneous mental phenomenon over which we have little control. Moreover, we often have a hard time accounting for the content and origin of thoughts generated during mind wandering (e.g. when a song is stuck in our head). Notably, an instance of mind wandering may act as detrimental or beneficial from a psychological viewpoint: in other words, mind

wandering exhibits a dual nature.[3] Let us assume that I have an important interview coming up and that I cannot concentrate on my PowerPoint preparation because my thoughts keep drifting away. Our discussion shows that this particular instance of mind wandering may acquire different valence depending on the context. On the one hand, external circumstances (e.g. how competitive the interview process is), my current emotional state (e.g. anxiety level), and broader personality traits (e.g. I may be prone to pessimistic fantasizing) may negatively affect my performance. On the other hand, as illustrated by Baird and colleagues, mind wandering while preparing for an interview might also turn out to be adaptive—e.g. if it allows me to creatively come up with original ideas or strategies. Another representation of the dual character of mind wandering comes from fiction. In Billy Wilder's movie The Seven Year Itch (1955), the protagonist Richard Sherman experiences acute and recurring episodes of daydreaming. Throughout the movie, Richard indulges in several episodes of mind wandering that mostly revolve around seducing his new neighbor (interpreted by Marilyn Monroe). In one of his raving monologues, Richard vindicates imagination as one of his most defining character traits: "It's just my imagination. Some people have flat feet. Some people have dandruff. I have this appalling imagination". These mind wandering experiences, however, produce positive as well as negative effects. On the one hand, they give Richard—who is normally quite shy and neurotic—the necessary confidence to invite her neighbor over for a drink and then out on a date. On the other hand, they fuel Richard's paranoid thoughts as he keeps fantasizing about what would happen if his wife were to find out about the (still imaginary) affair.[4]

What distinguishes the cases just described from extreme instances of hypoagency such as AVH? The two phenomena appear *prima facie* very similar in terms of duration and frequency. On the one hand, patients affected by AVH report that the experience of voice hearing becomes particularly distressing when the voices grow in number and intensity,

[3] See Lazarus and Folkman (1984) for a detailed discussion on dual factors, i.e. factors that act as risk-inducing or protective depending on the context.

[4] One might argue that even in these milder cases agency is impaired: we can't get rid of the song stuck in our head, Richard Sherman cannot control his daydreaming episodes, etc. I do grant this point, although there seem to be different *degrees of severity* at play. Although in mind wandering cases the feeling of agency is surely diminished, correct self-attribution is preserved: that is, we perceive the tune as "popping up from nowhere" but *not* as externally generated or inserted by someone else in our mind. By contrast, in extreme cases (such as AVH) the sense of agency is so disrupted that we completely lose the sense of what is self-generated and within our boundaries. Nothing in my account prevents this from happening with songs, provided that self-attribution also becomes incorrect and the song is then perceived as inserted, implanted, etc.

acting like a "running commentary" of one's life (Longden 2013). On the other hand, researchers studying mind wandering indicate that subjects "spend almost half a day engaged in the experience" (Smallwood and Andrews-Hanna 2013, 1) or even "roughly two thirds of their lifetime" (Metzinger 2013, 6). A crucial difference between the two cases seems to be the person's capacity to exercise a sufficient degree of *control* over the phenomenon. For instance, some aspects related to task-context (i.e. how demanding the activity is) might heavily influence the nature of the mind wandering episode, making it adaptive or disruptive as a result (Smallwood and Andrews-Hanna 2013). When we are engaging in a relatively non-demanding task, the experience of mind wandering is likely to be less disruptive and more conducive to positive outcomes (e.g. creative solutions) because our mental resources need not be fully absorbed in the completion of the task at hand. Conversely, when the current task requires our undivided attention an episode of mind wandering qualifies as a distressful interruption. Therefore, one's ability to regulate the context in which mind wandering episodes occur appears to play an important role: one might learn to confine mind-wandering to non-demanding situations—e.g. washing dishes—while fending it off from demanding ones (e.g. work or study). One might also learn to compartmentalize working or study time in order to devote designated unstructured spaces to mind wandering. This strategy appears to be successful as studies on creativity have consistently shown that original solutions to problems are more likely to arise when people allocate some unstructured time to mind wander (Dijksterhuis and Meurs 2006). Lots of interesting examples on how to implement these strategies are offered by the comedian John Cleese in his lecture about creativity (1991). While planning his weekly work schedule, Cleese makes sure to always leave a couple of slots open for creative thinking and treats them as serious commitments on a par with meetings, appointments, etc. He describes the rewards as extremely valuable: "If you just keep your mind resting against the subject in a friendly but persistent way, sooner or later you will get a reward from your unconscious, probably in the shower later. Or at breakfast the next morning, but suddenly you are rewarded, out of the blue a new thought mysteriously appears" (Cleese 1991).

Notably, the process of gaining control over internally generated thoughts and speech acts is similar to the one described by recovering AVH patients. For instance, Longden (2013) learns to incorporate the voices in a larger autobiographical narrative and starts regarding them as neglected parts of her own self. Similarly, one of the patients treated by Romme and Escher (2013) talks about setting boundaries and being able to push back the unwanted intrusions to a later time: "I was already able to talk back to my voices with my thoughts, but I learnt to make a specific time of day, the evening, when I would focus, and simply tell the voices 'later' if they came

at another time" (263). The ability to exercise a certain degree of control within a paradigmatically uncontrolled activity may therefore be crucial to distinguish between ordinary, or even adaptive, cases of mind wandering and their pathological counterparts.

2.2. Hyperagency: False Confessions

False confessions are usually characterized as situations in which someone confesses to a crime that he or she has not committed, or significantly overstates his or her involvement during custodial interrogation (Gudjonsson 2003). These cases qualify as instances of hyperagency because someone who falsely confesses to a crime incorrectly self-attributes an action that someone else has actually performed.

The idea of non-mentally disordered people willing to face legal charges for something they have not done appears very counterintuitive. Yet, studies in forensic psychiatry show that false confessions are relatively frequent, although their exact number is obviously difficult to determine. For example, in the early Eighties 10% of the defendants assessed in Birmingham and 24% of those in the London pleaded "not guilty" at their trial after having provided the police with a written confession (Gudjonsson 2003, 184). In his extensive work on the topic, Gudjonsson shows that false confessions are not confined to the mentally ill and that "the view that apparently normal individuals would never seriously incriminate themselves when interrogated by the police is wrong" (Ibid., 243). Forensic psychologists usually group false confessions into three categories: a) *voluntary*, where one spontaneously confesses without being interrogated, either to protect someone else or for pathological reasons—e.g. self-punishment; b) *coerced-compliant*, where one confesses as the result of an interrogation to obtain some immediate gain—e.g. escape from an intolerable situation, having one's sentence reduced; c) *coerced-internalized*, where one confesses as the result of an interrogation because he comes to believe that he has committed the crime (Gudjonsson 2003, 192-195). Obviously c) cases are the most relevant to our purposes, because they comprise a mistaken self-attribution that the subject genuinely endorses. However, the discussion of real-life examples shows that the boundary between b) and c) is not always clear-cut.

A famous case of coerced-internalized false confession is the one portrayed in Ava DuVernay's series *When They See Us* (2019) which involves the men who came to be known as the "Central Park Five" (and later as the "Exonerated Five"). The series covers the prosecution and incarceration of five males of color, following the rape and assault of a white woman in Central Park in 1989. The first episode is almost entirely devoted to the

interrogations of the five suspects and provides several insights on how their false confessions came about. Following the trial, the five teenagers received sentences ranging from five to fifteen years in prison, until the actual perpetrator confessed to the rape in 2001 and the men were finally released.

The way in which the Exonerated Five came to confess to a crime that they did not commit shows that the issue is quite complex. First, the methods used by the police during the interrogation play an important role, as well as the conditions in which the custodial confinement occurs—e.g. sleep-deprivation, under- or over-stimulation, inadequate diet and physical discomfort. Some studies suggest that interrogation techniques may be responsible for eliciting memory distrust and distortion when combined with situations of emotional shock or extreme stress (Henkel and Coffmann 2004). The case of the Exonerated Five is particularly illustrative in this respect. Kevin Richardson, who was 14 at the time, was kept in police custody and interrogated for 18 hours nonstop without any family member present. Raymond Santana spent most of the interrogation in the presence of his grandmother, who did not speak English and only received spotty translations about crucial details of the crime. Antron McCray's father was blackmailed by a police officer because of a past conviction that might have cost him his job, and ended up convincing his son to confess: "I want you to do what the police wants you to do. You need to say what they want you to say".

Second, false confessors usually exhibit a set of traits that make them particularly vulnerable to suggestion: young age, low self-confidence, exaggerated trust in authority, eagerness to help and difficulty in detecting discrepancies between what is recalled and what is suggested (Ofshe 1989). Again, this is apparent in the Exonerated Five case, where the young age of the suspects (ranging from 14 to 17), the techniques of brutal coercion employed by the police, and the racially-informed power dynamics played a crucial role. In DuVernay's series, the suspects are effectively manipulated by the detectives, who play them against one another in order to obtain partial confessions that would allow them to incriminate the group as a whole. Police officers use a variety of techniques that make it difficult to understand whether the resulting confessions would be merely compliant or also (partially) internalized. For instance, they pressure suspects by falsely claiming that others have already confessed and incriminated them ("'Ray did it', that's what they say"), they blackmail them ("The sooner you tell us, the sooner you go home"), and they ask leading questions ("Who took off her shirt? Was it Antron?"). This way five people end up confessing to a crime they neither committed nor witnessed, either by admitting partial involvement ("It was like, I came

over to where everybody was at and where the lady was at, and I was trying to stop it and help her out, and I think, no… she scratched me, that's how I got the scratch", Kevin Richardson) or by fully confessing ("This is my first rape", Korey Wise).

What makes false confessions different from the instances of pathological guilt discussed above? There are some striking similarities between the two situations: in both cases, a subject falsely, although sincerely, comes to believe that s/he has done something that falls beyond his/her control, and takes moral as well as legal responsibility for it. In this sense, both internalized false confessions and cases of pathological guilt hinge on incorrect self-attributions originating from false memories.[5] Moreover, a strong feeling of guilt features in both kinds of confessions. Many false confessors, for instance, feel guilty for not having been in control when the crime was committed (e.g. because of alcohol or drug intoxication), or for not being able to trust their memory in recalling events without confusion (Gudjonsson 2003). Despite these similarities, mentally disordered subjects appear to exhibit a pre-existing feeling of guilt that makes some of their actions particularly salient (e.g. Emmy von N), while false confessors experience guilt after having lost confidence about their ability to recollect what happened. As a consequence, the degree of internalization with respect to their confession differs; while voluntary confessions are rarely retracted, coerced-internalized confessions are usually taken back by the subject even if the timing of retraction varies from a few hours to several years (Gudjonsson 2003). In this sense, *duration* can be taken as a reliable indicator to distinguish between extreme and intermediate cases: the least pressured and the hardest to retract the confession, the higher its pathological import. This also leaves room for borderline cases: some false confessions may be characterized as transitory mental disorders from which people recover soon after the stressful situation has ended, while longer processes may indicate that the person has crossed a clinically relevant threshold.

[5] Assessing the degree of agency/control in these situations is obviously complex given that past events are involved. One option may be that false memories themselves originate from a disturbance in the sense of agency/control applied to the past. Alternatively, such a disordered sense of agency/control may apply to the subject's own thoughts in the process of recollection, which might make it more difficult to distinguish between real and imagined (or witnessed) events. In this sense, internalized false confessions would be quite similar to illusion of control cases (Wegner and Wheatley 1999; Hohwy 2004), where agency misattributions are not simultaneous with the action but rather occur at a (slightly) later time.

Pathological and non-pathological cases may also differ in terms of *urgency* and *intensity*. For instance, psychotic subjects voluntarily contact the police and appear distressed for having committed the crime in question ("I did it"; "It was me"), whereas false confessors initially proclaim their innocence and then come to confess in a tentative fashion ("I must have done it"; "I think I did"). Protective factors such as *strength* and *control* play an important role as subjects often confess after a prolonged period of physical discomfort and psychological stress. Gudjonsson describes the process as follows: "The forces pushing people towards confessing are strengthened (e.g. persuading people that it is in their own interest to confess, that there is substantial evidence to link them to the crime) whilst forces maintaining resistance are weakened (e.g. by tiredness, lack of sleep, exhaustion, emotional distress)" (2003, 189). In this sense, one important difference between pathological and non-pathological cases may lie in the degree of effort required by the subject to regain a sufficient level of control over the situation. In some cases, the state of confusion and memory distortion leading to the false confession would fade quite easily, while in others the recovery process may take longer or fail to occur at all.

3. Concluding Remarks

In this paper I discuss disorders of agency as cases in which people encounter difficulties in assessing their own degree of responsibility (self-attribution), and/or as disturbances in their sense of being in control of their actions (feeling of agency or control). I substantiate the idea that agency should be conceived in dimensional terms by discussing examples where agency may be seen as "too little" (hypoagency) or "too much" (hyperagency). Notably, *extreme cases* of hypo- and hyper-agency map onto phenomena that are usually conceived as disordered, such as AVH or pathological guilt. However, seeing agency on a spectrum also allows us to discuss *intermediate cases* in which the sense of being in control is disturbed without giving rise to clinically relevant manifestations. Although some intermediate cases may still turn out to be problematic (e.g. false confessions), I show that others exhibit an adaptive nature in many circumstances (e.g. mind wandering). Discussing these examples also contributed to a better understanding of how different aspects of agency can come apart. For instance, in AVH both self-attribution and the feeling of agency appeared to be disrupted; in other cases—such as mind wandering—self-attribution is usually correct while the subject experiences a feeling of diminished control with respect to the relevant thoughts or actions. Obviously, there are many other cases that could be assessed along these dimensions and the examples discussed here are not

meant to be exhaustive. In the synthetic table below, I provide some further suggestions as well as a summary of the examples discussed in the paper.

Self-attribution	Feeling of agency/control	Example	Classification
Incorrect	Seriously diminished	AVH	Hypoagency (extreme)
Correct	Moderately diminished	Mind wandering OCD	Hypoagency (intermediate)
Incorrect	Moderately strong	False confessions Illusions of control	Hyperagency (intermediate)
Incorrect	Unduly strong	Pathological guilt	Hyperagency (extreme)

Another core aspect of my discussion concerns the role played by quantitative factors such as duration, frequency, or intensity when it comes to distinguishing intermediate and extreme cases. These factors may play an important role in clinical practice, as they allow clinicians to improve case formulations and diagnoses of borderline or at-risk cases (Fusar-Poli et al. 2013). The focus on quantitative factors would also contribute to better track "the course of an illness" in longitudinal assessments, by monitoring how a patient's sense of agency evolves over time and in correspondence of turning points such as onset, development, relapse, and (possibly) remission (McGorry et al. 2018). The work I propose here on the sense of agency is part of a broader project that includes multiple dimensions (i.e. familiarity, confidence, salience) that may come to be altered in different circumstances, giving rise to clinically relevant conditions. In this sense, agency should be taken as only one of the relevant dimensions of functioning whose extreme disruption gives rise to mental conditions as we know them. At the same time, embracing a dimensional approach also implies acknowledging that "disordered" states are only quantitatively different from "normal" ones, and that the boundaries around normality and pathology are unlikely to be discrete and clear-cut.

Acknowledgments

The original version of this paper benefitted from discussions with Heidi Maibom, Peter Langland-Hassan, Thomas Polger, and Johannes Brandl. I would like to thank Agustín Vicente, Marta Jorba, Cecilia Fava, and Nicola Falde for offering insights on different aspects of the paper. I am also grateful to Emiliano Loria and two anonymous referees of EuJAP for their valuable feedback on earlier versions of the manuscript.

Funding

Valentina Petrolini's research is supported by Ministerio de Ciencia, Innovación y Universidades, postdoctoral fellowship FJC2018-036191-I, by Agencia Estatal de Investigación, grant number: PGC2018-093464-B-I00; by the Basque Government, grant number: IT1396-19; and by the University of the Basque Country (UPV/EHU), grant number: GIU18/221.

REFERENCES

Alderson-Day, B., S. McCarthy-Jones, S. Bedford, H. Collins, H. Dunne, C. Rooke, and C. Fernyhough. 2014. Shot through with voices: Dissociation mediates the relationship between varieties of inner speech and auditory hallucination proneness. *Consciousness and Cognition* 27: 288-296. https://doi.org/10.1016/j.concog.2014.05.010.

Allen, P., D. Freeman, L. Johns, and P. McGuire. 2006. Misattribution of self-generated speech in relation to hallucinatory proneness and delusional ideation in healthy volunteers. *Schizophrenia Research* 84(2–3): 281-288. https://doi.org/10.1016/j.schres.2006.01.021.

Andrews-Hanna, J. R. 2012. The brain's default network and its adaptive role in internal mentation. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* 18(3): 251-270. https://doi.org/10.1177/1073858411403316.

Baird, B., J. Smallwood, M. D. Mrazek, J. W. Y. Kam, M. S. Franklin, and J. W. Schooler. 2012. Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science* 23(10): 1117-1122. https://doi.org/10.1177/0956797612446024.

Bortolotti, L., and K. Miyazono. 2015. Recent work on the nature and development of delusions. *Philosophy Compass* 10(9): 636-645. https://doi.org/10.1111/phc3.12249.

Cleese, J. 1991. Lecture on creativity. *Video Arts*.

Dijksterhuis, A., and T. Meurs. 2006. Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition* 15 (1): 135-46. https://doi.org/10.1016/j.concog.2005.04.007.

DuVernay, A. 2019. *When They See Us*. Netflix.

Fernyhough, C. 2004. Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology* 22(1): 49–68. https://doi.org/10.1016/j.newideapsych.2004.09.001.

Freud, S. 1917. *Mourning and Melancholia*. Standard Edition 14. London: Hogarth.

Freud, S., and J. Breuer. 1893. *On the Psychical Mechanism of Hysterical Phenomena: Preliminary Communication*. Standard Edition 2. London: Hogarth.

Frijda, N. H. 1986. *The Emotions: Studies in Emotion and Social Interaction.* Cambridge: Cambridge University Press.

Frith, C. D. 2003. *The Cognitive Neuropsychology of Schizophrenia*. Reprinted. Essays in Cognitive Psychology Series. Hove: Psychology Press.

Fusar-Poli, P., S. Borgwardt, A. Bechdolf, J. Addington, A. Riecher-Rössler, F. Schultze-Lutter, M. Keshavan, et al. 2013. The psychosis high-risk state: A comprehensive state-of-the-art review'. *JAMA Psychiatry* 70 (1): 107-120. https://doi.org/10.1001/jamapsychiatry.2013.269.

Gudjonsson, G. H. (2003). *The Psychology of Interrogations and Confessions: A Handbook*. John Wiley & Sons.

Gram Henriksen, M., A. Raballo, and J. Parnas. 2015. The pathogenesis of auditory verbal hallucinations in schizophrenia: A clinical-phenomenological account. *Philosophy, Psychiatry, & Psychology* 22 (3): 165-181. https://doi.org/10.1353/ppp.2015.0041.

Henkel, L. A., and K. J. Coffman. 2004. Memory distortions in coerced false confessions: A source monitoring framework analysis. *Applied Cognitive Psychology* 18 (5): 567-588. https://doi.org/10.1002/acp.1026.

Hohwy, J. 2004. The experience of mental causation. *Behavior and Philosophy* 32(2): 377-400.

Hugdahl, K., E.-M. Løberg, L. E. Falkenberg, E. Johnsen, K. Kompus, R. A. Kroken, M. Nygård, R. Westerhausen, K. Alptekin, and M. Özgören. 2012. Auditory verbal hallucinations in schizophrenia as aberrant lateralized speech perception: Evidence from dichotic listening. *Schizophrenia Research* 140(1): 59-64. https://doi.org/10.1016/j.schres.2012.06.019.

Johns, L. C., K. Kompus, M. Connell, C. Humpston, T. M. Lincoln, E. Longden, A. Preti, et al. 2014. Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia Bulletin* 40 (Suppl_4): S255-264. https://doi.org/10.1093/schbul/sbu005.

Jones, S. R., and C. Fernyhough. 2007. Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition* 16 (2): 391-399. https://doi.org/10.1016/j.concog.2005.12.003.

Lazarus, R. S., and S. Folkman. 1984. *Stress, Appraisal, and Coping*. *Springer*.

López-Silva, P. 2016. Schizophrenia and the place of egodystonic states in the aetiology of thought insertion. *Review of Philosophy and Psychology* 7 (3): 577-594. https://doi.org/10.1007/s13164-015-0272-1.

Longden, E. 2013. Learning from the voices in my head. In Monterey, CA: TED Conferences.

Maiese 2015. Auditory verbal hallucinations and sense of ownership. Paper presented at the *Southern Society of Philosophy and Psychology*: New Orleans.

Mason, M. F., M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. N. Macrae. 2007. Wandering minds: The default network and stimulus-independent thought. *Science (New York, N.Y.)* 315 (5810): 393-95. https://doi.org/10.1126/science.1131295.

McGorry, P. D., J. A. Hartmann, R. Spooner, and B. Nelson. 2018. Beyond the "at risk mental state" concept: Transitioning to transdiagnostic psychiatry. *World Psychiatry* 17(2): 133-142. https://doi.org/10.1002/wps.20514.

Metzinger, T. K. 2013. The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00931.

Naranjo, J. R., and S. Schmidt. 2012. Is it me or not me? Modulation of perceptual-motor awareness and visuomotor performance by mindfulness meditation. *BMC Neuroscience* 13(1): 88. https://doi.org/10.1186/1471-2202-13-88.

Nayani, T. H., and A. S. David. 1996. The auditory hallucination: a phenomenological survey. *Psychological Medicine* 26(1): 177-189. https://doi.org/10.1017/s003329170003381x.

Ofshe, R. 1989. Coerced confessions: the logic of seemingly irrational action. *Cultic Studies Journal* 6(1): 1-15.

Pacherie, E. 2007. The sense of control and the sense of agency. *PSYCH* 13: 1-30.

Pacherie. E. 2008. The phenomenology of action: a conceptual framework. *Cognition* 107(1): 179-217. https://doi.org/10.1016/j.cognition.2007.09.003.

Petrolini, V., M. Jorba, and A. Vicente. 2020. The role of inner speech in executive functioning tasks: Schizophrenia with auditory verbal hallucinations and autistic spectrum conditions as case studies. *Frontiers in Psychology* 11. https://doi.org/10.3389/fpsyg.2020.572035.

Plath, S. 1963. *The Bell Jar.* Faber & Faber

Proust, J. 2006. Agency in schizophrenia from a control theory viewpoint. https://doi.org/10.7551/mitpress/2457.003.0006.

Radden, J. 2009. *Moody Minds Distempered: Essays on Melancholy and Depression.* Oxford: Oxford University Press.

Ratcliffe, M. 2010. Depression, guilt and emotional depth. *Inquiry* 53(6): 602-26. https://doi.org/10.1080/0020174X.2010.526324.

Ratcliffe, M. 2015. *Experiences of Depression.* Oxford: Oxford University Press.

Romme, M., and S. Escher. 1993. *Accepting Voices*. London: Mind Publications.

Romme, M., and M. Morris. 2013. The recovery process with hearing voices: accepting as well as exploring their emotional background through a supported process. *Psychosis* 5(3): 259-269. https://doi.org/10.1080/17522439.2013.830641.

Saks, E. 2007. *The Center Cannot Hold: My Journey Through Madness*. Hachette UK.

Schupak, C., and J. Rosenthal. 2009. Excessive daydreaming: A case history and discussion of mind wandering and high fantasy proneness. *Consciousness and Cognition* 18(1): 290-292. https://doi.org/10.1016/j.concog.2008.10.002.

Smallwood, J., and J. Andrews-Hanna. 2013. Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00441.

Smallwood, J., and J. W. Schooler. 2006. The restless mind. *Psychological Bulletin* 132 (6): 946-58. https://doi.org/10.1037/0033-2909.132.6.946.

Styron, W. 1991. *Darkness Visible.* Random House.

Szalai, J. 2019. The sense of agency in OCD. *Review of Philosophy and Psychology* 10(2): 363-380. https://doi.org/10.1007/s13164-017-0371-2.

Toh, W. L., N. Thomas, and S. L. Rossell. 2015. Auditory verbal hallucinations in bipolar disorder (BD) and major depressive disorder (MDD): A systematic review. *Journal of Affective Disorders* 184: 18-28. https://doi.org/10.1016/j.jad.2015.05.040.

Wegner, D. M., and T. Wheatley. 1999. Apparent mental causation. sources of the experience of will. *The American Psychologist* 54 (7): 480-492. https://doi.org/10.1037//0003-066x.54.7.480.

Wilder, B. 1955. *The Seven Year Itch*. Twentieth Century Fox.

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 165.63:616.89](450)"196"
https://doi.org/10.31820/ejap.16.2.5

RATIONALITY, IRRATIONALITY AND IRRATIONALISM IN THE ANTI-INSTITUTIONAL DEBATE IN PSYCHIATRY AROUND THE SECOND HALF OF THE 1970S IN ITALY

Matteo Fiorani
University of Rome, Tor Vergata

Review article – Received: 24/6/2020 Accepted: 21/8/2020

ABSTRACT

The movements and protests of 1968 worldwide criticized the traditional idea of normality. From the 1970s onwards, psychiatry and antipsychiatry became an ideological battleground centered on the boundaries between normality and madness. In this scenario, characterized by a deep cultural and political transformation within the Left, the traditional concept of rationality and its very connection with irrationality was called into question. As a consequence, the very ideal of reason was questioned. This paper will explore the debate on rationality, irrationality and irrationalism within the so-called anti-institutional psychiatry and its reception in the Italian New Left during the second half of the 1970s.

Keywords*: Antipsychiatry; psychiatric reforms; New Left; Italy*

1. The Boundaries Between Reason and Madness: The Italian Case

Psychiatry built, or at least set out to build, its scientific and professional foundations on the capacity to define the boundaries between reason and madness. Nevertheless, while trying to turn irrational into rational, it had to cope with some identitarian uncertainties from the beginning (Whooley 2019). Always on the boundary between human and natural sciences, psychiatry often comes back to the question of comparing quantitative and qualitative models to find the best tools to understand and explain humanity. The public debate about psychiatry has often been reduced to a rigid fight between "a romantic tradition, even irrationalist, hazily sentimental and [...] a classically rationalist and objectivist tradition"

Correspondence: matteofiorani22@gmail.com

(Jervis 2007). This simplification does not allow us to understand the various factors (e.g. technical, social, cultural, political, psychological, humanitarian) that have always contributed to defining and redefining the moving boundaries between normal and pathological.[1]

During the 1960s, questions on the boundaries between reason and madness started receiving attention from society and politics like never before. Psychiatry, through best sellers and mass media, overcame the institutional and disciplinary boundaries and started affecting mass culture and everyday life. During 1968, psychiatry was shaken by anti-authoritarian protests and by radical critics of the diagnostic systems, as well as of institutional assistance and cure. The libertarian and anti-institutional stances of some psychiatrists fascinated younger political movements, which absorbed them and elected such specialists as model intellectuals. A deep reflection, both cultural and political, was born at the time about mechanisms of social exclusion and the very idea of mental insanity. The walls of psychiatric hospitals were perceived as the concrete boundaries between reason and unreason. These had been built by society, which selected fools and expelled them through law. The Foucaultian ideas about the Big Internment (Foucault 1961) supported such a thesis, as did the arguments from the radical sociologists about deviance and total institutions (Goffman 1961; Becker 1963).

An analysis of the Italian case confirms the general dynamics described above but also presents some interesting peculiarities. In Italy, the protest movements started in 1968 extended way past the 1970s, and ended up affecting various disciplinary paths. Common ground was found not only within the workers' movement but also within politicized psychiatrists who were engaged in daily fights against the institutionalized system of asylums. The influence of feminist movements was particularly strong, with its interest in the relationship between private and public, between subjectivity and body, sharing with anti-psychiatry the search for new ethics. Finally, one of the most striking and well-known events of Italian history was the complete abolition of asylums. The Law 180/1978, which established such abolition, became the symbol (or the fetish) of an era as it was perceived as a historic turning point. Psychiatrist Franco Basaglia (along with his group)[2] was considered the hero who "freed the fools from asylums" and is still considered one of the most influential anti-psychiatrists in the West (Berlim, Fleck, and Shorter 2003). Actually, Law 180 wrongfully came to be known as the "Basaglia Law". This

[1] For an in-depth discussion of this topic, see Canguilhem (1966).

[2] For an intellectual and human biography of Basaglia, see Colucci and Di Vittorio (2001) and Pivetta (2012).

representation has deeply conditioned historiography (Foot 2014, 2015; Burns 2019; Burns and Foot 2020), despite the efforts to go beyond such a simplification (Micheli 2019).

It is interesting, then, to analyse the debate on rationality, irrationality and irrationalism in a crucial phase such as the second half of the 1970s, woven together with anti-institutional psychiatry and the New Italian Left. This label denotes the movements and political groups started between the 1960s and the 1970s outside (and often in opposition to) traditional Left-wing parties. This was a period of crisis and new ideas during which "psy" disciplines became a tool or even a shelter for people who were disoriented by the end of 1968 movements. Such people were attracted by movements characterized by extremism, individualism, depoliticization and irrationalism (Donolo 1976; Jervis 1976a; Crainz 2003). In this scenario, psychological conflicts became political conflicts as well. Mental disturbances began to be interpreted – by a minority overall, but by a large group within the Left movements – not as a disorder or suffering caused by experience, but as the result of conventions and prejudices, or even as an expression of freedom and creativity that could help to overcome the bourgeois regime. Anti-psychiatry became a word that was used, abused, mythologized, misunderstood: the focus of an ideological battle about the boundaries between normality and madness. The general scenario was one of profound cultural and political transformation within the Left. Thus, the discussion about normality and madness was related to a more general one about rationality and irrationality within the crisis of reason.

2. Other Perspectives and Definitions

Madness also became a matter of perspective, to be considered either from the outside or the inside of the asylum walls. During the 1970s, as the idea of a definite separation between normality and madness had been set aside, psychiatrists working outside the institutions made it clear that the boundary between reason and unreason was not marked by the walls of the asylums. No longer confined within the hospitals, psychiatrists and mental health workers worked in the community where new problems and issues emerged, related to work, existential sadness, and other preoccupations. Moving to the cities and suburbs, they reached local health centres and community-based outpatient clinics, and they started to attend houses, schools and factories, often with the goal of prevention. Psychiatry was led back to its social matrix, where the relationship between sanity and malady

had always been defined differently with respect to institutions.[3] Even the fact that the patient was assessed by a multi-disciplinary team composed by psychiatrists, psychologists, social workers, and nurses, as stated by Law 431/1968, was a significant change. Similarly, the virtual impossibility to reside inside a psychiatric hospital, as expressed by the absence of beds, brought about major changes within Italian psychiatry. The opposition between a rational world outside and an irrational one inside did not work, and other means of interpretation and analysis were necessary.

Giovanni Jervis began working in the community coming from Gorizia's asylum which had been repudiated as a place of treatment by Franco Basaglia from 1961.[4] Jervis was in charge of the outpatient psychiatric services in the province of Reggio Emilia from 1969 to 1977. He was called to do so by the province administration, led by the Italian communist party at the time, in order to reform and democratize the assistance to psychiatric patients. Jervis had a solid social psychiatry background and had also been part of the team of Ernesto De Martino, the ethnologist and historian of religion who had chosen him to take part in the interdisciplinary research team on tarantism in Puglia at the end of the 1950s.[5]

Jervis published his *Manuale critico di psichiatria* in 1975 [Critical handbook of psychiatry]. The volume was reprinted several times, translated abroad and considered a landmark publication for various generations of psychiatrists and mental health workers. The last chapter was dedicated to a critique of normality analysed from a Marxist viewpoint, in light of needs and desires, class struggle, everyday life, subjectivity and "rationality of revolutionary conscience" (Jervis 1975a, 194-225). This way, Jervis was trying to give back to psychiatry more than just a new complicated vocabulary, but rather a historical, cultural and political background. However, he did not give up classifying and describing, as shown by the *Piccolo dizionario ragionato di psichiatria*

[3] Psychiatrists had been working in the community for over a hundred years, although not for anti-institutional purposes, as shown by the project realized under the scientific leadership of Patrizia Guarnieri, *Fuori dal manicomio: Gli archivi della salute mentale dall'Unità d'Italia alla legge 180* (Outside the asylum: The archives of mental health from the Unity of Italy to Law 180), https://siusa.archivi.beniculturali.it/cgi-bin/pagina.pl?RicProgetto=preg-tos-fuoman. Accessed May 7, 2020.

[4] *L'istituzione negata* (The denied institution) is the title of a book, edited by Basaglia and ideated by Jervis, that described the experience within Gorizia's psychiatric hospital. It was published in 1968 by Einaudi and became a cult book for the students' movement that regarded the abolished psychiatric institution as a realized utopia (see Foot 2014, 131-152).

[5] For more details on Jervis and his work, see Marraffa (2014) and Fiorani (2016).

[Little thoughtful dictionary of psychiatry], which was inserted as an appendix to his 1975 manual. The effort to redefine the boundaries between normality and madness during the 1970s was tightly connected to such a context: it concerned the institutional and community practices of a discipline (i.e. psychiatry) that sought a new identity. It was the object of a political and cultural discussion that was held not on specialized journals but mostly on the New Left venues, as the Left movements were beginning to reason about their crisis.

3. The Crisis of Normality

A special effort to treat such questions was made by the *Quaderni Piacentini [Piacenza's Notebooks]*, a journal of dialogues and ideas from the New Left, open to discussing a variety of topics related to psychiatry and psychoanalysis. In the October of 1976 a double issue of the *Quaderni* was devoted to reflections on mythization and dogmatization. Sociologist Carlo Donolo opened the issue with a paper lucidly describing the transition taking place within the New Left.[6] In his opinion, it was necessary to move beyond 1968 and its movements, with more objective and non-ideological analyses, in order to restore the ideal and intellectual heritage of the movement and to avoid the risk of being wiped out by a tide of de-politicization, individualism and irrationalism (Donolo 1976). Other contributions of the issue included the philosopher Franco Rella on the mythization of Freud; the Jungian analyst Silvia Montefoschi on the myth of feminism; and the militant feminist and psychoanalyst Manuela Fraire on the women's movement. A psychiatric point of view was needed in order to understand the cultural change within the Left and this was provided by Jervis with the essay *Il mito dell'antipsichiatria* [The myth of antipsychiatry], also part of the special issue.

The essay by Jervis (later translated also into French and German) joined a fierce discussion that included both scientific and political matters. Both the dominant image of normality and the role of psychiatrists were undergoing a crisis. In such a scenario, Jervis wrote that "anti-psychiatry" had become an abused expression, a source of lies and illusions. Within the more extreme Leftist groups, mental malady was discussed through a seductive jargon, thereby generating confusing and imprecise discourse. At the same time, references to Deleuze and Lacan, as well as to Foucauldian anti-authoritarianism, had become expressions of cultural fashion. Laing and Cooper were often cited, albeit wrongfully: in fact, the

[6] Donolo's essay has often been considered a perfect example of such a phase (see Crainz 2003, 542).

former never acknowledged the term "anti-psychiatry" and the latter overtly stated that defining his position as anti-psychiatric was a misunderstanding (Jervis 1975a, 1976a, 1977a). Many ideas defined as anti-psychiatric were not novel: the quantitative relationship between normal and abnormal, sane and insane, had been already established by Freud, dynamic psychiatry, and interpersonal theories as opposed to being discovered by recent anti-psychiatric movements (Jervis 1976a).

This misunderstanding had practical as well as cultural effects, especially within psychiatric services, where some young militant clinicians were convinced that anti-authoritarianism on its own would solve the problem of mental health. On the other hand, people within the movement had adopted the "inconsistent and unusable" idea of madness as freedom, mistaking anti-authoritarian struggle for "typically bourgeois" permissiveness (Jervis 1976a, 47, 60). Such an ideological position, according to Jervis, puts the necessary affirmation of a different conception of normality and madness at risk. His view, rooted in Marxism and social psychiatry, does not necessarily regard biological damage as the origin of mental illness: rather, social and class contradictions would also feature as important causal factors.[7] Extremism and the mythologization of anti-psychiatric battles were, on the contrary, compatible with the most traditional positions of bourgeois rationality that they wanted to reject.[8] In other words: "on the one hand, antipsychiatric tendencies and theories demonstrated the politic potential of a crisis; on the other hand, they managed to turn this very crisis into a bourgeois intellectual theory. Indeed, such tendencies and theories psychologized the disease instead of historicizing it; they labeled it through a formula and a series of deceptions. In the end they pretended to solve it through conservative, and even reactionary, methods" (Jervis 1976a, 40). Jervis had been warning against idealization, dogmatism and sectarianism in psychiatry for years (Jervis 1972, 36-37). Although he held a minority position, he was never alone (Guarnieri 2012). Others negatively considered the regressive simplification (from a cultural, operational, political point of view) originating at the beginning of the 1970s (Ajmone 1976; Gleiss 1976) as an attitude that was common within the protest movements but was also shared by humanist intellectuals and, even if not overtly, by some militant psychiatrists.[9]

[7] To historically define psychiatry, Jervis cited—not by chance—authors such as Dörner (1969) and Ellenberger (1970).

[8] Cooper (1978) briefly responded by insisting on the relationship between madness and the need for autonomy.

[9] Most Italian psychiatrists, including Basaglia, rejected the notion of anti-psychiatry, and only used it as a synonym of anti-specialism (see Colucci and Di Vittorio 2001, 78). Jervis

The crisis within the New Left was then tied to the very crisis of the concept of normality. It concerned the uncertainties of anti-institutional psychiatry and emerged within the organization called *Psichiatria democratica* [Democratic psychiatry], founded in 1973 by Basaglia, which tried to unify, without success, the various alternative psychiatric movements in Italy.

In this sense, it is worth noting that the attitudes that Jervis and others criticized were not representative of the whole complex anti-institutional environment in Italy. There were different places where outpatient psychiatry was successfully practiced, and it was also possible to discuss rationality and the crisis of reason with international experts and in scientific journals.[10] Nevertheless, these experiences were tangential with respect to the debate within the New Left. Even people with similar ideas did not always manage to find a common ground for dialogue.

4. Subjectivity, Normality, Madness

In this phase, within the New Italian Left, and especially through feminist movements, special attention was dedicated to the reflection on personal needs and desires in connection with the public environment. More specifically, the relationship between subjectivity and collective movements, and between body and sexuality, was explored. Several themes were discussed, such as couple relationships, free and conscious reproductive rights, and sexuality. The ensuing battles also led to major transformations from a legal point of view, such as the Italian laws legalizing divorce and abortion.

In his *Manuale*, Jervis stated that "even everyday life is a political problem, because what is personal is political" (Jervis 1975a, 15). Criticizing the concept of normality, he regarded feminist groups as the only ones capable of developing and advancing a battle for a new conscience, one able to keep together new political urges: "a battle on the various fronts of everyday life" (Jervis 1975a, 213). Nevertheless, already in 1976, the theoretical references began to change and the efforts of "keeping together

came back to reflect on this period in a dialogue with medical historian Gilberto Corbellini in a book with the significant title *La razionalità negata: psichiatria e antipsichiatria in Italia* (Denied rationality: Psychiatry and Antipsychiatry in Italy; Corbellini and Jervis 2008).

[10] The experience of Perugia was considered a positive example of outpatient psychiatric assistance and de-institutionalization (see Guarnieri 1997). Another successful example was Grosseto, where mental patients were assisted, like elsewhere in Italy, without having a mental hospital in the vicinity (see Fiorani 2012).

Marxism and the individual through the theory of needs of Agnes Heller" remained separate (Jervis 1976c; see also Crainz 2012, 62). In this time, the interest in subjectivity and the political dimension of the personal became a tool to reflect on the loss of political sense and on the irrationalist drive. Also some Italian feminists started asking themselves:

> if the feminist practice (meaning taking consciousness, to be clear, not the public manifestations) can start from subjectivity, without becoming subjectivism, or if we are flooding in a sea of irrationalism and intimism, whose ties with struggle, radical transformation, taking power are becoming weak. (Ravera and Usai 1976, 35)

In the newspaper *la Repubblica* (located in the reformist Left), journalist and writer Enzo Forcella wrote about the book *Porci con le ali* [Pigs with wings]—a bestseller that was reprinted many times and translated in many countries—as a sign of the crisis within the revolutionary Left (Forcella 1976). This book was written by Marco Lombardo Radice and Lidia Ravera and described the stories, both intimate and political, of Rocco and Antonia, two young high school activists. It was published by Savelli (a publisher close to the radical Left) and was the first of a series significantly called *Il pane e le rose* (Bread and roses), edited by the same Lombardo Radice and Ravera, together with Giaime Pintor and Annalisa Usai. The title referred, purposefully, to the English idiom "pigs have wings", quoted by David Cooper in *The Death of the Family* (Cooper 1971). A quote from Cooper's book appeared on the back cover of *Porci con le ali*. The authors, who were both active in the debate about subjectivity, love and sex, shared the ideas of a movement called *Lotta Continua*, which was broken up during the same year because of internal conflicts that feminism encouraged. This was interpreted as an emerging gender conflict, which was going to become more radical than the one between wage labour and capital.

While *Porci con le ali* was Ravera's first book which launched her career as a professional writer, Lombardo Radice was a 27-year-old physician working as the chair of Psychophysiology at the University of Rome. He participated in the 1968 protests and, as a specialist, wrote with some colleagues in the newspaper *l'Unità* (the official voice of the communist party) to support the necessity of evaluating both social and biological aspects when assessing mental illness (Lombardo Radice, Venturini, and

Ruggieri 1974).[11] Afterwards, within the debate about social roles, he tried to introduce the idea of evaluating both the biological and psychological factors of behaviour (of men, women, heterosexuals, homosexuals, etc.): on his view, these factors went beyond the results of "social conditioning and learning". He expressed these ideas in an essay published in the activist journal *Ombre Rosse* (Lombardo Radice 1976b, 54). In the same year of *Porci con le ali*, Lombardo Radice also edited (together with Riccardo Venturini) the Italian edition of *Le motivazioni biologiche* (Biological motivations), written by the soviet physiologist Konstantin V. Sudakov. In the introduction he proposed a psychophysiological view of mental health within the debate about needs (Lombardo Radice and Venturini 1976). In the end, surely, Lombardo Radice could not be accused of naive wishful thinking.[12] *Porci con le ali* was harshly criticized—albeit in a "friendly" way—in the same journal *Ombre Rosse*, as an individualistic and irrational product (Manconi and Sarno 1976).

5. Rationality, Irrationality and Alternate Normality

Following a path already started by feminist movements (Lonzi 1970), the New Left called into question the traditional meaning of rationality and the very connection between rationality and irrationality (Crainz 2012, 63). Irrationalism, started as an eminently cultural movement, became a political and ideological category. In this debate, the so-called drug ideology (which mostly concerned young people) played a central role and attracted militant psychiatrists and psychologists, especially with regard to the possibility of an alternative concept of normality.

In an essay from 1976 called *Giovani senza rivoluzione* [Young people without revolution], Lombardo Radice wrote about the connection between irrationality and irrationalism. He distinguished between the veterans of 1968 and the young people of the 1970s and, referring to the drug ideology, affirmed that the irrationalist ideas of a "deeply politicized and radically anti-bourgeois youth" could be important to "reappropriate a bond with nature, creativity, madness in a structurally and violently anti-bourgeois and revolutionary way" (Lombardo Radice 1976a, 21). Irrationalism, on the other hand, should be rejected because it denied the

[11] The letter opened a discussion within the Marxist environment about the origin of mental illness (social or biological), which was gathered in a volume edited by two leading figures of the communist party: Berlinguer and Scarpa (1975).

[12] Lombardo Radice (1977) later wrote about institutionalized psychiatric violence within *Lotta Continua*, insisting on the paradox of the preposterous precision of diagnostic definitions and the sadism of psychiatrists. For a biography of Lombardo Radice, see Fiorani (2019).

use of reason to interpret and change reality. Thus, the battle against the "irrationalist ideology" of psychotropic drugs was warranted in those cases where the 'trip' was perceived as the only means of knowledge (which was believed by very few people). However, the same battle was unwarranted if it wanted to restrict the search for an alternative rationality, which was not necessarily unreasonable. This way, Lombardo Radice was trying to save juvenile irrationalism, which could also be evaluated from a psychophysiological point of view, and which also represented a

> drive to recover an important part of humanity, meaning a 'visceral brain', which, for an animal so highly corticalized as man, is somehow naturally submitted to the 'rational brain', but is today crushed and annihilated, for complex reasons, related to the evolution of history and society. (Lombardo Radice 1976a, 21; see his 1975)

Macondo, a former factory in Milan that had been transformed into a community space in the late 1970s, with restaurants, discos, and rooms dedicated to music, reading, art exhibitions, and markets, was shut down in 1978 because of pervasive drug use. Mauro Rostagno, a leader of the Italian 1968 movements who had just left *Lotta Continua*, was among the founders of the space,[13] which had been an effort to answer, in an alternative and creative way, to the crisis and disorientation of Leftist young people. In a way, it was a concrete experiment of new forms of normality. Indeed, changing the concept of normality appeared to be an ever more complicated enterprise, as it clearly emerges from the book devoted to the experience of Macondo (Rostagno and Castellacci 1978). Fiorello—member of the youth collective of Stadera, district of Milan—published a critical review of the book in *Ombre Rosse* and highlighted that "gay people, feminists, freaks, addicts and so on" stayed away from gurus (such as David Cooper and André Glucksmann) who praised their liberation and participated in the same events. In this sense, the efforts to make them the "active subjects of transformation", to define their misery as "non-integration" had not worked out (Fiorello 1979, 165).

Jervis, in an article dedicated to drug ideology, described two essays by Rostagno and Romano Madera (another militant of the 1968 movement, later turned philosopher and Jungian analyst) as both mystifying and dangerous. These articles appeared in November 1975 in one of the most

[13] Rostagno's intellectual development was significant. After the closure of Macondo, he went to India, where he joined an Orange community in Poona. In 1981 he moved back to Italy and founded the Saman Community in Sicily, devoted to the recovery of drug addicts, based on a "pact among free men" and specific techniques inspired by Indian meditation. See Bigaran (2017) for more details.

popular "counterculture" journals, *Re nudo* [Naked king, i.e. Emperor's New Clothes]. The praise of mental illness, the appreciation of psychotropic drugs such as LSD as a revolutionary path, and the appeal to a "non-rational liberation of the brain" that would replace political action, were all issues that, according to Jervis, the authors should be held responsible for. But the same responsibility applied to people who did not understand that "in the hand of the new post-1968 decadence, spontaneity had become improvisation, subjectivity had become subjectivism". In the end, Jervis wrote that it was "our fault too" if these questions did not receive appropriate answers (Jervis 1976b, 32-33).

On Jervis' view, mystical and esoteric temptations, hippie theories and fashions (which had arrived in Italy ten years later with respect to the United States), escape and self-destruction through drugs (ever more often heroine), a revaluation of madness as liberation and of marginality and deviance as revolutionary, irrationalist and regressive ideologies, should not be considered in terms of fighting the rational (class fight and organization) through the irrational (youth counterculture). This attitude, underscored also by Lombardo Radice (cited by Jervis, 1976b, 5, 29), did not meet the needs of young people: while adults were playing with their late 1968 subjectivism (thereby betraying the position that "the personal is political"), teenagers did not have a proper culture to refer to (Ibid., 33).

Despite the difference in age and experience, Jervis (born in 1933) and Lombardo Radice (together with others in the teams behind *Ombre Rosse* and *Quaderni Piacentini*) highlighted the contradictions of the movements but at the same time their potentialities, which should not be forgotten. This way, they both tried to address the needs of those young people, often students and mental health clinicians and workers, who were sincerely interested in discussing and understanding the crisis of normality and reason in all its cultural, scientific, and political aspects.

6. Irrationalism and the Crisis of Reason

The explosion of the so-called 1977 movements (which lasted almost a year and had Rome, Milan, and Bologna as their epicenters), along with their irony and ferocity towards their 1968 predecessors, contributed to the search for new behaviours and politics in order to radicalize the fight between reason and un-reason, as well as between rationalism and irrationalism.

Together with the classical texts of the earlier culture—*The Death of Family* by Cooper (1971); *The Divided Self* and *The Politics of Experience*

by Laing (1960, 1967)—it is a shared opinion that 1977 brought to the fore new reference books. One of them was the *Anti-Oedipus* by Deleuze and Guattari. By contrast, Foucault was not only regarded as the theoretician of the big institutional internment of madness, but also as the one who first highlighted the molecular power of human relationships.[14]

A critical piece about the popularity of the so-called *nouveaux philosophes* was published in *Ombre Rosse* by Luigi Manconi, Gad Lerner and Marino Sinibaldi, who were both customary collaborators of the journal and well-known members of *Lotta Continua*. Their essay rejects the standard representation, usually endorsed by the "bourgeois press", of the movement as culturally homogeneous and uniformly seeking meaning through irrationalistic categories (Lerner, Manconi, and Sinibaldi 1977). "Microphysics of power", "desiring flow" and "desiring machines" were concepts that were not very well-known in Italy at the time—who had actually read the *Anti-Oedipus*? (Deleuze and Guattari 1972) Manconi, Lerner and Sinibaldi wondered—and as a consequence not very clear to most people. On the other hand, reason and unreason, rationality and irrationality, kept being concepts to be reasoned and discussed about within the New Left. A 1977 photograph by Tano D'Amico portrayed a girl lying down, with a book on her chest entitled *Donne, povere matte: inchiesta nell'Ospedale psichiatrico di Roma* [Women, poor fools: Inquiry into the psychiatric hospital of Rome],[15] exemplifies another crucial issue, namely the relationship between feminism and antipsychiatry.

Both criticizing authoritarianism and the political culture of the Left, antipsychiatry and feminism shared the goal of tracking subjective and intersubjective paths, which were alternative to the process of homologation through normality. As Marthe Van De Meulebroeke wrote in 1976 in the feminist journal *Effe*: "What Laing accepted to call anti-psychiatry could become the whole psychiatry; it could come out of psychiatry and change our everyday relationships with others", (Van De Meulebroeke 1976). More cautious were the comments to the symposium *Donna e follia* [Woman and madness] held in Florence on November 12, 1977 and published in the same journal the following year. In this issue they highlighted the contradiction between the mythization of the fool, who joined the way of madness as a "conscient political choice against the

[14] *Microfisica del potere* (Microphysics of Power) was published in Italy in 1977 (Foucault 1977).

[15] For the book, see Harrison (1976). The photograph has been published in a special issue of *Robinson*, a journal periodically attached to the newspaper *la Repubblica*, whose title was "Settantasette. Parole e immagini" [Seventyseven. Words and images], issued February 12, 2017,
https://www.repubblica.it/static/robinson/numero-11/settantasette/, accessed May 1, 2020.

imposed norm", and the negative definition of fools, applied to women who broke the rules of a male chauvinist society (see Tagliaferri 1977; Vitas 1977; Vitas et al. 1977).

In 1977 the philosophy journal *Aut Aut*—which had been discussing the theme of needs during the 1970s—decided to dedicate an issue on *Irrazionalismo e nuove forme di razionalità* [Irrationalism and new forms of rationality], given the "political hardening" and the ideological struggle around these questions. Intellectuals from different backgrounds and experiences were called up to express their opinion. Among them was Giovanni Jervis, who considered the conflict about rational and irrational to be in the heads and behaviour of people.

From as far back as Freud and Jung, who had well demonstrated the links between rational and irrational, Jervis stated that in the "human psychic structure" reason and unreason have always lived together. Psychology, psychoanalysis, and Marxism had already shown that "personality and the human psyche cannot be divided into a socio-rational and a natural-instinctual part" (Jervis 1977b, 41). Jervis later referred to Herbert Marcuse's work (1955), and in particular to the illusion, based on a "metahistorical naturalism", of considering certain behaviours (e.g. imagination, eversion, spontaneous, madness) as liberatory. By contrast, in his opinion they were functional to the system, because

> romantic models, literary avantgardes, decadent forms of irrationalism, ideologies of instincts, violence, immediacy, irreflexivity, partying and totalization [...] might deceive pleasure and desire and end up leading them to authoritarian grounds, where the abuse becomes law. (Jervis 1977b, 41)

Irrationalism was then not only unhelpful to overcome capitalistic rationality, but it actually ended up reinforcing the stability of the capitalist system. Indeed, if irrationality

> is just another face of bourgeois reason, irrationalism also shares its structure: it favors the same interest of the system toward stability, or even more authoritarian forms of social control. (Jervis 1977b, 43)

The essay ended quite bitterly, with a sort of personal and collective assessment: The New Left had been unable to clarify the themes of rationality and irrationality. In opposition to the "rational bourgeois" positions of the communist party, the only answers had been desperately irrational attitudes.

In the same issue, feminist Lea Melandri harshly criticized Jervis for his position on needs (which she considered as old-fashioned Marxist) and his criticism of the mythization of anti-psychiatry and drugs. The main problem was in the conception of the connections between normality, madness and irrationalism. The cautious attitude of Jervis, Melandri wrote, more than the effort of saving

> normality and madness, feast and necessity, individual and history, leaks out censorship, dogmatic rigidity, Manichaean moralism, which distinguishes between, on the one hand, the sane reason [...], on the other, 'irrationalism', 'drift', 'regression', 'disorder', 'instinctivism', 'bad faith'. (Melandri 1977)

The needs of life that have been extensively discussed had transformed classical rationality into something brittle and precarious. "The ways of thinking could be several and reason was in crisis" wrote philosopher Aldo Gargani when introducing the 1979 volume *La crisi della ragione* (The crisis of reason) – a collection of interdisciplinary essays (Gargani 1979). By criticizing classical notions of reason, Gargani tried to deepen the debate on rationalism and irrationalism in order to escape from ideological positions. His answer gave relief to militants, younger and older, who were uncomfortable with the denial of social roles and professional competence and with anti-psychiatric generalizations. It worked well as a manifesto for those who were not searching for expedients but rather for new ideals of rationality and political paths.[16] Against extreme irrationalist positions, these people appreciated discussions of classical reasons, such as those collected by Gargani and also internationally by Kuhn and Feyerabend. In these works, science was criticized without being devalued – as had always been done by Italian neo-idealistic thought. Similarly, new models of rationality, less univocal and more complex than the ones offered by the Galilean tradition, were embraced.[17] In satisfying this need for knowledge, an important role was played by Left-leaning publishers (such as Feltrinelli, Einaudi, and others), who translated works especially from the USA.

[16] Historian Patrizia Guarnieri wrote about the reasons why, as a "student or little more", it was comforting for her to read Jervis, with respect to the simplifications and equivocations of certain anti-psychiatry during the second half of the 1970s (Guarnieri 2012, 69). It is interesting also to highlight the unexpected re-reading of Nietzsche by young people, stated by historian Guido Crainz, in order to overcome the "old categories of rationality" and the search for "new behaviors and new politics" (Crainz 2012, 63).

[17] This is what Guarnieri notes in a testimony about these years (personal communication). I thank Patrizia Guarnieri for the testimony and for her observations on this contribution.

7. A Cul de Sac?

Despite the anxious quest for answers, at the end of the 1970s Italian activists appeared to be in a *cul de sac*. Even Marxism (in all its variations) seemed unable to satisfy the collective and individual need for knowledge. Despite extensive discussions, even the ideas about normality and madness were still not well-defined, as it clearly emerges from the journals of the movement during these years. Here, harsh judgements prevailed over the efforts to recover the positive aspects of the battle against mental institutions and the critics of normality.

The "systematic praise of madness", as Alfonso Belardinelli and Giovanni La Guardia noted in *Quaderni Piacentini*, was the symbol of how the New Left was getting old. Such a psychiatric reversal was not even useful in integrating the dropouts, as noted by Roberto Polce of the Gay Collectives from Milan (formerly *Nostra signora dei fiori*) in the journal *Ombre Rosse.* In this essay, Polce compares the contradictions of the path towards the liberations of gay people and mad people who came out from the institutions just to be marginalized again from society (Polce 1978). First embraced as a "conceptual subversive model of schizophrenia and mental illness", through abstract representations of suffering, the conception of pathology as "areas which were independent from the normalization power" was not useful to understand the real bites of illness, as Mirella Serri wrote also in *Ombre Rosse* the same year (Serri 1978). Reacting against the inhumanity of psychiatric institutions had been rightful, noted Alberto Mellucci in the *Quaderni Piacentini*, but it failed to provide answers to the basic questions: what is madness, how to interpret it within society along the new praxis and how to use one's body and identity (Mellucci 1978).

At the end of the 1970s, difficulties and anxiety seem to characterize Italian anti-institutional psychiatry in an even more evident way. This climate can be perceived by looking at the third meeting of the *Réseau internazionale di alternativa alla psichiatria* [International Network of Alternatives to Psychiatry], entitled *Il circuito del controllo* [The circuit of control], which was held in September 1977 at the psychiatric hospital of Trieste. This venue was highly symbolic because it had been directed since 1971 by Franco Basaglia, who was soon going to announce its closure before the Law 180/1978 would close all Italian mental institutions. There were 4,000 participants (among them Cooper and Guatari): not only psychiatrists and mental health clinicians and workers, but also young people and collectives (both Italian and French). There was a crescendo of tension and anti-psychiatric slogans. During one turmoils, Basaglia broke a rib. Nonetheless, he declared that he wanted dialogue at all costs because

"beyond sanity, the life of oppressed people matters" (Giliberto 1977). What one of the collectives declared to the press seemed significant to understand the ongoing confusion among social roles:

> Yesterday we shouted that we wanted Basaglia as principal of the Asinara prison. No one understood that it was not an insult but a hope. Many comrades that are closed in that concentration camp would like to have a person like Basaglia in charge: as a first thing, he would keep away the bars from the windows and open the doors, as he did in the mental hospital of Trieste. (Giliberto 1977)

Another article written by Enzo Siciliano, a writer and literary critic, entitled *La psichiatria democratica in cerca di senso* [Democratic psychiatry in search for meaning], in 1978 describes the situation as follows. Between Laing and Cooper, Guattari and Foucault, Basaglia's "romanticism" and Jervis' "rationalism", illness and norm, institution and community, the direction seemed uncertain. Despite the results achieved, Siciliano noted that many radical questions posed during the previous ten years were still unanswered. Can anti-psychiatry help the insane? Can drugs? Can psychoanalysis? Is an individual good even if s/he exhibits deviance? Article was written just before the reform that would have closed mental institutions at the end of the so-called season of movements. Reflecting on a decade of militant psychiatry, and on the connections between madness and politics, this seemed to be ever more urgent and necessary.

Jervis followed up on this issue by giving voice to his own subjectivity, through an individual and collective biography from 1951 to 1976 that prefaced a selection of published and unpublished works (Jervis 1977c). Basaglia himself tried to sum up his experience in the volume *La nave che affonda* [The Sinking ship] from 1978, which consisted in the transcripts of heated discussions held in his Venice house with his wife Franca Ongaro, Agostino Pirella (from the Gorizia team and director of the Arezzo psychiatric hospital at the time), and journalist Salvatore Taverna. The sinking ship was symbolizing the mental institution: the focus of discussions and the core issue of future psychiatry. On the same wavelength, in 1979, Basaglia and Ongaro Basaglia tried to define, *in extremis*, the entries "madness" and "delusion" for the *Einaudi Encyclopedia* (Basaglia and Ongaro Basaglia 1979).

Two years after the approval of Law 180/1978, the sociologist Ota De Leonardis, while reflecting on the concept of deviance, affirmed that the battle against mental institutions did not achieve the goal of re-defining the

relationship between normal and pathological, as various problems emerged in the process of de-institutionalizing and re-socializing mentally ill people. The very same pretense of giving voice to madmen and, through them, to the revolutionary conscience had been re-absorbed within the administrative management of mental illness after the law became effective. Marginalized deviant people, bringers of the un-reason, had been reinstated into the subordinate role of the consumer (De Leonardis 1980). Despite the closed hospitals and the variety of alternative outpatient psychiatric experiences, there was still a felt necessity to reckon with a recent but cumbersome past as well as with the efforts of giving madness the role of a new form of rationality, in opposition to a society that was considered a mass institution itself.[18] In such a scenario of political and social crisis, the individual had become an abstraction, only capable of saving the innocence of un-reason in an intimist and irrationalist regression.

The psychiatric questions—such as the definitions of normal and pathological—following the de-institutionalization process, should be (by law) brought into the community to find a new theoretical and practical sense. The dissatisfaction of many, both from the New Left and anti-institutional psychiatric movements in Italy,[19] was well-expressed by Carlo Manuali, head of the outpatient psychiatric services of Perugia. Manuali was concerned with the incapacity to think about psychiatry outside the institution, given the everyday difficulties brought up in the context of where people lived. Mental illness was, as he wrote in 1980, "a biographic event, and biographic needs have a more rapid rhythm than the social development on the whole" (Manuali 1980).[20] There was still a long road ahead, also beyond political movements, given the "return of naturalism" bursting out around the early 1980s.

Acknowledgments

I thank Marco Innamorati for the help.

[18] See Basaglia and Ongaro Basaglia (1971, 1975).
[19] Such a dishomogeneity clearly emerges in the effort of De Salvia and Crepet (1982) to track "a critical epidemiology of the reformation".
[20] See also Manuali et al. (1978) on how psychiatry was no longer only struggling with institutional marginalization, but also with social factors of illness.

REFERENCES

Ajmone, T. 1976. Partecipazione e politica del territorio. *Inchiesta* 22: 3-14.

Basaglia, F., and F. Ongaro Basaglia. 1971. *La Maggioranza Deviante. L'Ideologia del Controllo Sociale Totale*. Turin: Einaudi.

Basaglia, F., and F. Ongaro Basaglia. 1975. *Crimini di Pace. Ricerche sugli Intellettuali e sui Tecnici come Addetti all'Oppressione*. Turin: Einaudi.

Basaglia, F., and F. Ongaro Basaglia. 1979. Follia/delirio. In *Enciclopedia Einaudi*, vol. VI, 262-287. Turin: Einaudi.

Becker, H. S. 1963. *Outsiders. Studies in the Sociology of Deviance*. New York: Free Press of Glencoe.

Belardinelli, A., and G. La Guardia. 1978. Restaurazione e liberazione. Osservazioni sull'invecchiamento della nuova sinistra. *Quaderni Piacentini* 69: 11-17.

Berlim, M. T., M. P. Fleck, and E. Shorter. 2003. Notes on antipsychiatry. *European Archives of Psychiatry and Clinical Neuroscience* 253: 61-67.

Berlinguer, G., and S. Scarpa, eds. 1975. *Psichiatria e Società*. Rome: Editori riuniti.

Bigaran, M. P. 2017. Rostagno Mauro. In *Dizionario biografico degli Italiani*, vol. 88. Roma: Istituto della Enciclopedia Italiana. http://www.treccani.it/enciclopedia/mauro-rostagno_%28Dizionario-Biografico%29/. Accessed April 26, 2020.

Burns, T. 2019. Franco Basaglia: A revolutionary reformer ignored in Anglophone psychiatry. *Lancet Psychiatry* 6(1):19-21.

Burns, T., and J. Foot, eds. 2020. *Basaglia's International Legacy: From Asylum to Community*. Oxford: Oxford University Press.

Canguilhem, G. 1966. *Le Normal et le Pathologique*. Paris: Presses Universitaires de France; Italian translation, Turin: Einaudi, 1998.

Colucci, M., and Di Vittorio, P. 2001. *Franco Basaglia*. Milan: B. Mondadori.

Cooper, D. 1971. *The Death of the Family*. London: Penguin; Italian translation, Turin: Einaudi, 1972.

Cooper, D. 1978. L'antipsichiatria demitologizzata. Risposta a Giovanni Jervis. *Aut Aut* 165/166: 200-2004.

Corbellini, G., and G. Jervis. 2008. *La Razionalità Negata. Psichiatria e Antipsichiatria in Italia*. Turin: Bollati Boringhieri.

Crainz, G. 2003. *Il Paese Mancato. Dal Miracolo Economico agli Anni Ottanta*. Rome: Donzelli.

Crainz, G. 2012. *Il Paese Reale. Dall'Assassinio di Moro all'Italia di Oggi*. Rome: Donzelli.

De Leonardis, O. 1980. Devianza: crisi e critica. *Quaderni Piacentini* 74: 71-83.

De Salvia, D., and P. Crepet, eds. 1982. *Psichiatria senza Manicomio*. Milan: Feltrinelli.

Deleuze, G., and F. Guattari. 1972. *L'Anti-Oedipe. Capitalisme et Schizophrénie*. Paris: Deninuit; Italian translation, Turin: Einaudi, 1975.

Donolo, C. 1976. Oltre il '68. La società italiana tra mutamento e transizione. *Quaderni Piacentini* 60/61: 3-38.

Dörner, K. 1969. *Bürger und Irre. Zur Sozialgeschichte und Wissenschaftssoziologie der Psychiatrie*. Frankfurt: Europäische Verlagsanstalt; Italian translation, Rome-Bari: Laterza, 1975.

Ellenberger, H. F. 1970. *The Discovery of the Unconscious. The History and Evolution of Dynamic Psychiatry*. New York: Basic Books; Italian translation, Turin: Bollati Boringhieri, 1972.

Fiorello, 1979. Macondo. *Ombre Rosse* 27/28: 164-165.

Fiorani, M. 2012. *Follia senza Manicomio. Assistenza e Cura ai Malati di Mente nell'Italia del Secondo Novecento*. Naples: Edizioni Scientifiche Italiane.

Fiorani, M. 2016. Jervis Giovanni. In *Dizionario biografico degli italiani*. Roma: Istituto della Enciclopedia Italiana, http://www.treccani.it/enciclopedia/giovanni-jervis_%28Dizionario-Biografico%29/. Accessed May 15, 2020.

Fiorani, M. 2019. Lombardo Radice Marco. In *Dizionario biografico degli italiani*. Roma: Istituto della Enciclopedia Italiana, http://www.treccani.it/enciclopedia/marco-lombardo-radice_%28Dizionario-Biografico%29/. Accessed May 20, 2020.

Foot, J. 2014. *La "Repubblica di matti". Franco Basaglia e la Psichiatria Radicale in Italia, 1961-1978*. Milan: Feltrinelli.

Foot, J. 2015. *The Man who Closed the Asylums. Franco Basaglia and the Revolution in Mental Health Care*. London: Verso.

Forcella, E. 1976. "Porci con le ali": perché il sequestro? *la Repubblica*, December 8.

Foucault, M. 1961. *Histoire de la Folie à l'Âge Classique*. Paris: Plon; Italian translation, Milan: Bompiani, 1963.

Foucault, M. 1977. *Microfisica del Potere*. Turin: Einaudi.

Gargani, A. ed. 1979. *La Crisi della Ragione*. Turin: Einaudi.

Giliberto, F. 1977. Trieste: un convegno "modello di follia"? *La Stampa*, September 16.

Gleiss, I. 1976. Il contenuto conservatore dell'antipsichiatria. *Psicoterapia e Scienze Umane* 3: 6-15.

Goffman, E. 1961. *Asylums. Essays on the Social Situation of Mental Patients and Other Inmates*. Garden City, NY: Doubleday; Italian translation, Turin: Einaudi, 1968; introduction by F. Basaglia and F. Ongaro Basaglia.

Guarnieri, P. 1998. Per una storia della psichiatria anti-istituzionale. L'esperienza del rinnovamento psichiatrico in Umbria 1965-1995. *Annali di Neurologia e Psichiatria* 2 (Suppl.): 5-63.

Guarnieri, P. 2012. Presente e passato. L'interesse di Jervis per la storia. *Medicina nei Secoli* 1: 55-78.

Harrison, L. 1976. *Donne, Povere Matte. Inchiesta nell'Ospedale Psichiatrico di Roma*. Rome: Edizioni delle Donne.

Jervis, G. 1972, Manicomi eccetera. *Inchiesta* 5: 33-37.

Jervis, G. 1975. *Manuale Critico di Psichiatria*. Milan: Feltrinelli.

Jervis, G. 1976a. Il mito dell'antipsichiatria. *Quaderni Piacentini* 60/61: 39-60.

Jervis, G. 1976b. L'ideologia della droga. *Quaderni Piacentini* 58/59: 3-32.

Jervis, G. 1976c. Quali bisogni. Alcune note. *Ombre Rosse* 17: 5-11.

Jervis, G. 1977a. Non sparate al terapeuta. *Quaderni Piacentini* 62/63: 192-193.

Jervis, G. 1977b. Un parere su razionalità, irrazionalità e irrazionalismo. *Aut Aut* 161: 39-44.

Jervis, G. 1977c. Il buon rieducatore. In Id. *Il buon rieducatore. Scritti sugli usi della psichiatria e della psicoanalisi*, 9-42. Milan: Feltrinelli.

Jervis, G. 2007. Complessità e ricerca in cinquant'anni di psichiatria in Italia. *Epidemiologia e Psichiatria Sociale* 2: 139-143.

Laing, R. D. 1960. *The Divided Self. A Study of Sanity and Madness*. London: Tavistock Publications. Italian translation, Turin: Einaudi, 1969; preface by L. Jervis Comba.

Laing, R. D. 1967. *The Politics of Experience*. New York: Pantheon books. Italian translation, Milan: Feltrinelli, 1968.

Lerner, G., L. Manconi, and M. Sinibaldi. 1977. Le altre stagioni del movimento di primavera. *Ombre Rosse* 22/23: 3-39.

Lombardo Radice, M. 1975. I giovani e la droga. *Ombre Rosse* 9/10: 56-66.

Lombardo Radice, M. 1976a. Giovani senza rivoluzione. *Ombre Rosse* 15/16: 9-23.

Lombardo Radice, M. 1976b. Tre domande nel dibattito sulla famiglia. *Ombre Rosse* 14: 53-56.

Lombardo Radice, M. 1977. L'ultimo paziente dello psichiatra Coda. *Lotta Continua*, December 6.

Lombardo Radice, M., R. Venturini, and V. Ruggieri. 1974. Psichiatria ed emarginazione sociale. *l'Unità*, July 18.

Lombardo Radice, M. and R. Venturini. 1976. Preface. In K. V. Sudakov, *Le Motivazioni Biologiche*, V-IX. Florence: Giunti-Barbera.

Lonzi, C. 1970. *Sputiamo su Hegel*. Rome: Editoriale Grafica.

Manconi, L. and M. Sarno. 1976. Porci (?) con le ali (?). *Ombre Rosse* 17: 58-62.

Manuali, C. 1980. Without title. In *Dove Va la Psichiatria? Pareri a Confronto su Salute Mentale e Manicomi in Italia dopo la Nuova Legge*, eds. L. Onnis and G. Lo Russo, 98-150. Milan: Feltrinelli.

Manuali, C. et al. 1978. Ipotesi sulla malattia mentale e criteri di operatività sul territorio. *Fogli di informazione* 50: 336-340.

Marcuse, H. 1955. *Eros and Civilization. A Philosophical Inquiry into Freud*. Boston: Beacon Press; Italian translation, Turin: Einaudi, 1964; introduction by G. Jervis.

Marraffa, M. 2014. Giovanni Jervis: la ricerca della concretezza. In G. Jervis, *Contro il sentito dire. Psicoanalisi, psichiatria e politica*, ed. M. Marraffa, XIII-XCIV. Turin: Bollati Boringhieri.

Melandri, L. 1977. Ascetismo rosso. *Aut Aut* 161: 58-66.

Mellucci, A. 1978. Dieci ipotesi per l'analisi dei nuovi movimenti. *Quaderni Piacentini* 65/66: 3-19.

Micheli, G. A. 2019. Not just a one-man revolution: The multifaceted anti-asylum watershed in Italy. *History of Psychiatry* 30(2): 133-149.

Pivetta, O. 2012. *Franco Basaglia, il Dottore dei Matti. La Biografia*. Milan: Dalai.

Polce, R. 1978. Il vizio dello struzzo (ovvero: c'è uomo e uomo). *Ombre Rosse* 24: 86-94.

Ravera, L., and A. Usai. 1976. I tempi delle donne sono i tempi che le donne si danno. *Ombre Rosse* 15/16: 30-38.

Rostagno, M., and C. Castellacci. 1978. *Macondo*. Milan: SugarCo.

Serri, M. 1978. Fratelli. *Ombre Rosse* 26: 102-103.

Tagliaferri, M. 1977. Dalla follia alla liberazione. *Effe* 12: http://efferivistafemminista.it/2014/11/dalla-follia-alla-liberazione/, Accessed, May 20, 2020.

Van De Meulebroeke, M. 1976. La religione e la donna. *Effe* 2: 12-15.

Villa, A. 1979. Tra il cane e il lupo. *Ombre Rosse* 27/28: 110-118.

Vitas, E. 1977. Le così dette pazze. *Effe* 10: http://efferivistafemminista.it/2014/11/le-cosi-dette-pazze/, Accessed May 20, 2020.

Vitas, E. et al. 1977. Testimonianze. *Effe* 12: http://efferivistafemminista.it/2014/11/testimonianze-3/, Accessed May 20, 2020.

Whooley, O. 2019. *On the Heels of Ignorance. Psychiatry and the Politics of Not Knowing*. Chicago: University of Chicago Press.

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 17:616-08ECT-053.2
https://doi.org/10.31820/ejap.16.2.6

A DESIRABLE CONVULSIVE THRESHOLD. SOME REFLECTIONS ABOUT ELECTROCONVULSIVE THERAPY (ECT)

Emiliano Loria
Sapienza - University of Rome

Review article – Received: 17/8/2020 Accepted: 5/10/2020

ABSTRACT

Long-standing psychiatric practice confirms the pervasive use of pharmacological therapies for treating severe mental disorders. In many circumstances, drugs constitute the best allies of psychotherapeutic interventions. A robust scientific literature is oriented on finding the best strategies to improve therapeutic efficacy through different modes and timing of combined interventions. Nevertheless, we are far from triumphal therapeutic success. Despite the advances made by neuropsychiatry, this medical discipline remains lacking in terms of diagnostic and prognostic capabilities when compared to other branches of medicine. An ethical principle remains as the guidance of therapeutic interventions: improving the quality of life for patients. Unfortunately, psychotropic drugs and psychotherapies do not always result in an efficient remission of symptoms. In this paper I corroborate the idea that therapists should provide drug-resistant patients with every effective and available treatment, even if some of such interventions could be invasive, like Electroconvulsive Therapy (ECT). ECT carries upon its shoulders a long and dramatic history that should be better investigated to provide new insights. In fact, ECT has attracted renewed interest in recent years. This is due to the fact that antidepressant drugs in younger patients show often scarce effectiveness and unpleasant side-effects. Moreover, I show that, thanks to modern advances, ECT may work as a successful form of treatment for specific and rare cases, such as severe depression (with suicide attempts) and catatonia.

***Keywords**: ECT; neuroendocrinology; psychopharmacology; history of child psychiatry*

Correspondence: emilianoloria@gmail.com

> "The story of psychiatry is a story of tensions that have not yet played out. It is a history where what we decide is the truth regarding our past has immediate and profound implications for how we view ourselves and our futures and how we treat others when they are at their most vulnerable. It is a history that affects all of us."
> (Healy 2002, 8)

1. The History of a Euphoria

Electroconvulsive therapy has an almost centennial history that began in Rome (Italy) in 1938, at the Clinic of Nervous and Mental Diseases, run at the time by (psychiatrist) Ugo Cerletti. ECT still represents one of the most important and controversial therapeutic discoveries in the field of psychiatry. It is based on the passage of short electrical impulses that cause convulsions. Unlike insulin, and still more than the drug Metrazol used by von Meduna to provoke convulsions but with severe side effects (Shorter and Healy 2007), electricity could be administered to patients in subtly graded doses. This way, the convulsions, which were considered by psychiatrists to be the essential therapeutic ingredient, would be less dangerous and easier to manage (Passione 2004, 89).

After the first official presentation of electroshock (E.S., as ECT was called at that time) to a restricted audience, Cerletti waited two years before the publication of a detailed account of his research in the *Rivista sperimentale di freniatria* (Cerletti 1940). Cerletti was probably aware of the potential risks involved in the dissemination of ECT, and he wanted to avoid hasty imitators applying E.S. with too little experience. He also believed that publicity about his discovery would have sounded inappropriate before a thorough analysis of "the best modes of application of E.S." was completed (Cerletti 1940). The long and detailed essays by Bini (1940) "La Tecnica e le manifestazioni dell'Elettroshock", testify that the inventors of this technique, namely Cerletti and Bini himself, were extremely careful and cautious to carry out electroconvulsive treatment (Passione 2004, 90-91).

However, their advice was not followed. The rapid dissemination of ECT in Western countries was due to Lothar Kalinowsky, a Jewish student of Cerletti emigrated from Italy to England because of the racial laws promulgated in 1938 during the Fascist regime. The application of the therapy occurred in every conceivable way (Berrios 1997; Shorter 1997). From England to France, to Germany and to the United States, where another collaborator of Cerletti, Renato Almansi, taught the E.S. technique to David Impastato, who treated the first American patient, a 29-year-old schizophrenic woman in 1940 (Passione 2004, 93).

At the XXIII Congress of the Italian Society of Psychiatry held in 1946, Cerletti denounced the non-scientific use of the recent shock therapeutic discoveries. In the same congress, psychiatrists Baldi and Reale raised the same question about the indiscriminate abuse of shock therapies, in particular ECT, for all types of mental illness. They posited the need to "arrive at a clearer definition of its field of application and restrict its use accordingly" (Baldi and Reale 1947; Passione 2004, 96). Max Fink, one of the most famous American psychiatrists who practiced and taught ECT throughout his long career, has well described these initial phases in an interview with Shorter and Healy in 2002. He claimed: "The problem at the beginning was what is necessary for an effect. It was [easier] to get seizures. But they didn't know what was important, so they tried sub convulsive, they tried convulsive, they tried multiple seizures, they tried different electrode pairs, with drugs, without drugs, daily treatments, twice daily" (Shorter and Healy 2007, 140).

As patient and clinical reports attest, it was not uncommon for E.S. to be administered without a real therapeutic need and in combination with insulin therapy and psychotropic drugs.[1] In some cases, E.S. was unfortunately used as a deterrent or even as a threat. These repeated occurrences contributed to the creation of stigma surrounding ECT and to its wrongful assimilation to a barbaric practice.

Every catatonic or schizophrenic patient was the ideal candidate, together with adolescents and children. The first minor patients treated with ECT were believed to have been in France starting in 1941 by George Heuyer and colleagues, and in the United States starting in 1942 by Lauretta Bender (see Shorter 2013).[2] However, from the examination of the Archives of Pediatric Neuropsychiatry in Rome—a section of the Roman Clinic specifically dedicated to minors[3]—it emerges that the first child ever administered with ECT (September 18, 1940) was a 7-year-old boy diagnosed with "dementia praecocissima".[4] After 23 applications of the technique, the child was discharged and returned to the father, who had assisted him during the preparation of the first session. Another child (a

[1] For the Italian case, see Nemec (2015) who discusses several clinical reports and treatments relative to the post-WWII era in Trieste and Gorizia.
[2] See also Walter et al. (2010).
[3] The hospital section devoted to minors was established by the Director Sante de Sanctis in 1930 (Coccanari de' Fornari et al. 2017).
[4] Historical Archives of Child Neuropsychiatry – Department of Human Neuroscience (Sapienza University of Rome), volume n. 18, record 1682. The ordering of the medical records which are in the Historical Archives follows an arrangement by years. For a review about childhood dementias according to the diagnostic criteria of that time, see Campailla (1945, 264-265).

six-year-old boy) with the same diagnosis was treated with E.S. in 1943. Before and after them, a few other children and adolescents were treated with insulin shock therapies. The first catatonic young patient treated with E.S. was a 14-year-old girl in 1944. After 11 applications, the girl had a remission of symptoms and was discharged from the clinic and returned to her family. She relapsed years later and was readmitted to the clinic in December 1950. Accornero and Anderson (1948) reported a sample of E.S. treatments which regarded infant and young adolescent patients of the Roman clinic. Among these cases is also the first one I mentioned.[5] Compared to the original medical record, the article by Accornero and Anderson does not provide many other data. In general, the medical records of those years often report a detailed anamnesis. However, the outcome of the treatments is hard to obtain. Accornero and Anderson's essay has the merit of deepening this aspect for some of the 15 cases of prepubertal schizophrenia they examined. This way, we come to discover that Gianluigi, this is the name of the first child, is discharged from the Clinic in the same conditions in which he entered. Furthermore, the frequency of treatments is well specified in the paper (the first 12 applications twice a week and the other ones three times a week). The authors also specify the standard voltage: 110 or 120 Volt in 1/10 per second (Accornero and Anderson 1948, 233-234). Accornero and Anderson's conclusive analysis is pessimistic about the efficacy of ECT in childhood psychosis. The authors were aware of the small sample of patients, and mightily complained of the difficulty in providing a more detailed diagnosis. The vagueness of the diagnosis increased the empiricity of the therapeutic method which was already empirical.

The archives of the Roman Clinic of Nervous and Mental Diseases also bring to light the existence of an informed consent form that families (or caretakers) were asked to sign, albeit in a primitive form. In the E.S. Collection, established by the Director Cerletti, there are the primitive informed consent forms that the parents, relatives or tutors had to fill in. Initially, in 1937, the informed consent only concerned insulin therapy and it had later been extended to ECT. It recited as follows:

> [On letterhead *R.[oyal] University of Rome. Clinic of Nervous and Mental Diseases*] I have been made aware of the dangers that shock therapies pose in some cases. Given that these treatments are today those that give some hope for improvement, I adhere to the proposal to practice it in my son

[5] It is worth noting that Felice Accornero helped Cerletti and Bini in making the ECT device (Sirgiovanni and Aruta 2020, 321). Thus, two remarkable researchers had already reported the very first case of E.S. in children, although they did not underline this peculiarity.

> [or other family member] [name and surname of the patient], releasing the R[oyal] Clinic from any responsibility. [Original signature].[6]

Nowadays, patients are routinely told why a certain treatment is prescribed, and the administration procedures are explained in detail. At the time, of course, each patient had to sign informed consent forms where all this information was illustrated in writing. The family members also had to agree.

2. From Euphoria to Barbarities

From its beginning in 1938 throughout the 1950s, ECT gained wide fame for the treatment of schizophrenia. Those who have the opportunity to consult old medical records dating back to those years may realize the series of therapeutic success, partial symptomatic remissions, cyclic relapses, and consequent readmissions, as well as complete failures of ECT. Although this scenario cannot be considered a full-fledged victory, the therapeutic success attained was already an exceptional improvement, which could illuminate the darkness of therapeutic nihilism in which psychiatry as a science had been struggling for centuries.

The ECT practice also revealed its efficacy for catatonia and affective disorders like severe depression. In a very few years, from the era of medical frustration where there was no other remedy than to sedate, purge and contain patients, and "the treatment of mental illness represented a vast wasteland of hopelessness" (Shorter and Healy 2007, 4), ECT appeared to provide fast alleviation of psychotic symptoms, above all in the acutely ill, without too many significant risks (see below for a discussion on ECT's side effects).

Throughout the 1960s and 1970s, the wide availability and dissemination of antipsychotic drugs, together with increasing opposition to the use of ECT, that went into a gradual decline (Fink 2001). During the years of contestations, i.e. from 1968 onward, a destructive as well as misleading connection was generated between the movements of psychiatry renovation and antipsychiatry. The former worked on changing the methods of interventions and limiting the restrictive measures adopted in asylums. The latter was against both the asylums and psychiatric interventions as a whole, including shock therapies and drugs. Drugs, which had not much earlier been acclaimed "as awakening the chronically

[6] My translation is from original documents, which are kept in the E.S. Collection.

psychotic from irretrievable madness, were now castigated as chemical straitjackets" (Healy 2002, 5). All of psychiatry's past was interpreted as a series of barbarities culminated in E.S. and psychosurgery. Similarly, for many philosophers at that time, the methods of psychiatric discipline represented a clear and potent symbol of the irrationality of modern society, generated by the union between science and capitalism.[7]

This political and cultural short circuit led to two dramatic changes, first of all in Italy, which acted as a pioneering country in this field. The first was the closure of psychiatric asylums, which in many cases had become institutions of arbitrary and unmotivated restraints (see Babini 2009 on the Italian case). The second was the abandonment of those therapeutic practices that were considered too invasive. This way, psychiatric treatments of proven effectiveness such as E.S. were banned, little by little, in many (Western) countries (Healy 2002, 5). A resurgence of interest in ECT occurred in the 1980s, following the finding that a large portion of patients do not tolerate or respond to drugs, or appear to be drug resistant. This has prompted the investigation of combining drugs with ECT. Furthermore, many psychiatrists suggest ECT as the treatment of choice for catatonic schizophrenia (Fear 2005, 30).

3. Barefaced and Raw ECT

Given these historical premises, we cannot disagree with Edward Shorter when he claims that ECT provides an exciting insight into the role of culture against science in clinical decision-making (Shorter 2013). Why should we reconsider ECT after a massive and historically-laden disengagement from this therapy? The answer is simple: ECT is an effective and safe treatment, with potentially low side-effects risks for some kinds of severe disorders. Quoting the UK National Institute for Clinical Excellence (NICE): "Electroconvulsive therapy is a proven effective treatment for depression. It is a safe form of treatment even in the medically ill, the elderly and in pregnancy. There are benefits in using it in emergencies. It should not be relegated to a treatment of last resort" (Lamprecht et al. 2005, 19). In fact, ECT could be the treatment of choice when severe forms of depression are associated with attempts of suicide (or with strong suicidal plans) and urgent intervention is needed, or when patients present significant psychomotor retardation, stupor, depressive

[7] I suggest reading Giovanni Jervis on this point (2014). He was a lucid witness of the cultural climate in Europe and also contributed to a profound renewal of psychiatric interventions in Italy. For more details see also Fiorani's article (this issue).

delusions or hallucinations (Lamprecht et al. 2005, 13; see also UK ECT Review Group 2003).

Contrary to their original hopes (and promises), some kinds of psychotropic drugs like selective serotonin reuptake inhibitors (SSRIs) have been accompanied by several distressing side effects. Among them, decreased libido and even impotence, delayed orgasm, anorgasmia, anhedonia. Impotence has proven particularly troublesome, as Clayton and colleagues showed (2002; but see also Jacobsen et al. 2020). Furthermore, twenty years ago in the UK, two-thirds of SSRI patients had discontinued treatment by the end of the third month (Shorter 2009, 201).[8]

When pharmacotherapy fails to improve depressive symptoms, then, "response rates of about 50–60% can be achieved by ECT" (Eser et al. 2007, 2). For this reason, particularly in depressed patients at high risk of suicide, ECT should be recommended earlier than its conventional "last resort" position. In fact, the risks of suicide have been shown to relieve quickly through ECT, when administered in continuity with previous treatments, that are essential to sustain its benefits. Look, for example, at Kellner et al.'s (2005, 977) study, in which expressed suicidal intent in 131 depressed patients was rapidly relieved by ECT, already after one week (i.e., three ECT sessions) in 38.2% of the patients, and, at the end of the course of treatment, in 80.9% of the patients.

Furthermore, in all stages of pregnancy, ECT may represent the best solution in case of severe mood or thought disorders, when antidepressants and antipsychotics cannot be administered (Abrams 2002; Fink 2009). In this sense, ECT may represent a safer treatment for these more difficult cases.

In addition to drug-resistant major depression, which remains the most frequent indication for ECT, other psychiatric illnesses like delusional mania, catatonia, malignant neuroleptic syndrome and delusional mania should be considered "urgent first-line indications for ECT treatment" (Eser et al. 2007, 2). Patient choice is obviously crucial in most cases. In fact, if some sufferers have already experienced ineffective or intolerable medical treatment, they may choose ECT; alternatively, if they had previous successful experiences of recovery with ECT, they can express the will to be administered a further cycle of treatment.[9] Although ECT has

[8] For a review of these issues, see Holtzheimer and Mayberg (2011).

[9] Several handbooks and essays about psychiatry insert positive biographical reports about ECT treatments (see references below). In this regard, see the recent collection edited by

been recognized as "an effective treatment for mood and psychotic disorders, it is among patients with catatonia that the most remarkable efficacy is observed" (Petrides et al. 2004, 151). Although for uncomplicated cases of catatonia the first-line treatment should be benzodiazepines (Rosebush and Mazurek 2004), "for malignant catatonia, or in circumstances where use of benzodiazepines is contraindicated, ECT should be the first-line treatment" (Fear 2005, 40). On the contrary, ECT is not helpful for those people affected by a lifelong history of emotional dysfunctions, or suffering of personality disorders, addictions, sociopathy (Fink 2009).

4. The Procedure

The image of a fast, shift-like psychiatry is far from the truth: "ECT is not a surgical excision that removes the defect", claims Fink. "It is more like the treatment of diabetes in which repeated administration of insulin and dietary control are needed to maintain healthy serum glucose levels" (Fink 2013, 21).

Thanks to the evolution of ECT instrumentation, which often includes EEG[10] and EKG (for a review see Weiss 2018, 16-18), the current procedure involves the use of brief-pulse stimulation techniques which induct a series of generalized epileptic seizures under anesthesia and muscle relaxation. This procedure is "one of the best tolerated and safest biological treatment strategies with low risk for severe complications" (Eser et al. 2007, 2; see also Abrams 2002). A typical course of ECT treatment consists of two or three inductions a week for two to seven weeks (Fink 2009, 4). Fink explains the procedure as follows:

> The patient is asked to empty the bladder and is then taken to the treatment room, where she lies down on a stretcher. A nurse or physician inserts a needle into a vein in the arm or foot, attaches a bottle of fluid (usually sugar in water), and sets the fluid flowing at a slow rate. […] This intravenous line allows the easy and painless administration of medications during the treatment. Adhesive monitoring electrodes - flat, disposable pads or reusable discs to which electrical connections can be

Kirov (2020). In the Bini archive, there are letters from patients containing thanks to Lucio Bini and Ugo Cerletti for the improvement of the conditions after the E.S. treatments. (Passione 2007, 89)

[10] "The EEG facilitates discrimination between electrode positions and different stimulus doses. […] EEG may indicate threshold changes highlighting the need to increase the stimulus dose or reduce the anesthetic induction agent" (Weiss 2018, 17).

> made - are applied to the skin, a painless procedure. Three electrodes are put in place for the electroencephalogram (EEG); two stimulating electrodes for the electrical stimulus; three for the electrocardiogram (ECG) and heart rate; and two to measure motor movements during treatment. A recording electrode placed on the patient's finger or toe measures the blood oxygen saturation. A blood pressure cuff on the arm measures the blood pressure, and a second one may be placed as a tourniquet on a leg to allow the psychiatrist to record the duration of the muscular signs of the seizure. (Fink 2009, 14)

During the seizure, the medical team needs to monitor heart rate and rhythm, blood pressure and the brain's electrical activity. Some dental conditions require a personalized plastic braces for each treatment.

Electrodes placement affects the procedure: treatments based on bilateral electrode placement usually exhibit shorter duration. The successful use of bilateral ECT is reported, for example, for severe forms of catatonia. I provide more details on this below. Another important issue is the combination of ECT therapy with antipsychotics. Since ECT is not considered as therapy of first choice in catatonia, patients often come to ECT after benzodiazepines administration or while they are receiving this medication. In these cases, a drastic withdrawal is not recommended, because it may cause worsening or relapse of the most severe symptoms. Indeed, a synergic effect of ECT and benzodiazepines has been reported by Petrides and colleagues, who also claim that "benzodiazepines are often useful and can be continued for many months after recovery" (Petrides et al. 2004, 156-157; Weiss 2018, 70-71).

5. The Underlying Mechanism of ECT: A Neuroendocrinological Perspective

An objective and scientific-based reason that can explain why ECT, by giving patients epileptic-like seizures, should be beneficial to some psychiatric illness remains quite obscure. What we know is that adrenocorticotropic hormone (ACTH), thyrotropin releasing hormone (TRH), and vasopressin are well-studied hormones that are known to be released by seizures. The amount of hormones released varies with electrode placement, stimulation dosage, number and frequency of treatments. Prolactin, for example, is quickly released within the serum with a peak at 20 minutes. According to Fink, "it is the massive outpouring of these hormones that characterize the seizures in effective ECT" (Fink 2013, 21).

As it is known, the hypothalamus plays a crucial role in sustaining life because it regulates the autonomic nervous and neuroendocrine systems. We also know that hormonal functions in mental illness are disordered. For instance, in severe depressive states, the axis linking hypothalamus, pituitary gland, and adrenal and thyroid glands appears to be dysfunctional, involving increases of cortisol and other hormones. ECT with bitemporal electrode placement allows the passage of electricity from one temple to the other, and as a result the hypothalamus is directly stimulated. The first effects of such stimulation are transitory, but by the fourth or fifth stimulus—according to Max Fink—"the normal feedback actions of the hormones of the hypothalamic-pituitary-adrenal axis are again in place. Feeding and sleep become normal, and improvements in motor activity, mood, memory, and thought to follow quickly" (Fink 2013, 24). It is no coincidence that ECT is most effective in patients who present neuroendocrine dysfunctions (e.g. "abnormal dexamethasone suppression test or diminished thyroid-stimulating hormone response to thyrotropin-releasing hormone"), which can disappear with effective treatment (Petrides et al. 2004, 158).

However, the neurotransmitter pathway has also been explored to better understand the ECT effects. For what concerns depression, several studies show that ECT can attenuate serotonergic and noradrenergic neurotransmission (for a review see Eser et al. 2007, 3). Two phenomena have been observed in this respect. On the one hand, ECT would increase the availability of the serotonin precursor thus contributing to the therapeutic effects of ECT (Palmio et al. 2005). On the other hand, in line with the GABA deficit hypothesis of depression, ECT would exert a compensatory increase in γ-aminobutyric acid (GABA) neurotransmission. For instance, a study based on proton magnetic resonance spectroscopy showed that "occipital cortex GABA concentrations are increased in depressed patients treated with ECT" (Eser et al. 2007, 3; see also Sanacora et al. 2003).

The behavioral effects of seizures require repetition until the pathological symptoms have remitted, and new patterns of behavior have stabilized. The prescription of antidepressants after a successful ECT administration reduces the risk of early relapse (Weiss 2018). As I mention above, antidepressants should not be suddenly discontinued before ECT, especially those with a short half-life or SSRIs. Those patients previously medicated with SSRIs should be initially administered "a low electrical dose at the first treatment" (Scott 2005, 111).

6. Side-effects

Currently, the mortality rate of ECT is estimated at two deaths per 100,000 treatments. The causes of death referring to adults during ECT are from cardiovascular disorders, most often pre-existing conditions, such as cardiac arrhythmia or hypertension, when these fail to be taken into account by clinical practitioners and anaesthesiologists. According to proton magnetic resonance spectroscopy studies, it seems that neuronal damage or cell death are not induced by ECT, because it does not provoke a significant decrease in the N-acetylaspartate signal, which constitutes a sign of cell atrophy (Ende et al. 2000; Eser et al. 2007, 4).

No deaths have been ever reported in an adolescent or a child directly due to ECT. Mortality associated with ECT has markedly decreased with technological improvement and medical monitoring during the applications. For this reason, nowadays, ECT can be considered a low-risk procedure, "even among older patients with cardiac disorders" (Consoli et al. 2013, 141). Anaesthesia may obviously provoke other side effects, including disturbance in cardiac rhythm, variations in blood pressure, respiratory incidents, and allergic reactions. Less serious side effects of ECT include headache and muscle pain, nausea (with or without vomiting), temporary confusion immediately after the sessions, and understandable fear about ECT. "Up to 45% of patients report headache after ECT, which can be treated using analgesics such as acetylsalicylic acid or paracetamol and, if severe, by changing the induction medications" (Eser et al. 2007, 10). Even if rarely, nausea may occur after intravenous anesthesia; in this case metoclopramide can be used successfully. Since ECT is not a first-line treatment, when psychiatrists suggest or prescribe ECT, patients often present a very compromised medical status. When ECT is early administered in the course of catatonia and in severe forms of depression, then more favourable results and reduced morbidity are obtained. Furthermore, such severe psychiatric disorders often entail cognitive impairments *per se*. Thus, it is challenging to evaluate eventual cognitive damages due to ECT or to illness.

In the last twenty years, an increasing number of studies have focused on ECT effects on memory. The meta-analysis conducted by Fraser and colleagues (Fraser et al. 2008) suggests that ECT may cause autobiographical memory impairment. Still, such memory loss is relatively short-term, i.e., about six months post-treatment. The potential cognitive deficits refer to autobiographical episodes and events that occurred in the period close to the beginning of treatment. Furthermore, implicit memory, procedural memory and semantic memory are not modified by ECT (Consoli et al. 2013). Therefore, it is reasonable to claim that memory loss

due to ECT is temporary and circumscribed to some particular events close to the sessions. The degree of these temporary memory damages depends on many factors such as the number of sessions, the features of the electrical current (e.g. brief pulsed or sinusoidal), the placement of the electrodes, and the presence of persistent depressive symptoms before ECT (Sackeim et al. 1991). In brief, variations in the administration methodology may affect cognitive function after ECT. From randomized studies, it seems that "sinusoidal waveform causes more memory impairment than brief-pulse ECT" (Lamprecht et al. 2005, 18), and, for this reason, brief-pulse stimulation techniques represent the most common stimulation method nowadays.

This brief overview shows how 80 years after the invention and introduction of ECT, the methodology of its application has improved both to avoid collateral physical damage and to enhance the effectiveness of treatment according to the type and severity of the disease.[11]

Electrode placement

Given the importance of lateralization and, more generally, of the allocation of the electrodes for what concerns efficacy and risks of ECT, some further considerations are necessary. First, it is established that "stimulus intensity depends on electrode placement" (Eser et al. 2007, 5). In this sense, bilateral ECT seems to be more effective than unilateral (UL) ECT, which requires a higher stimulus dosage to gain the same results. Many practitioners doubt the effectiveness of UL ECT in treating patients who have severe or life-threatening psychopathologies (Weiss 2018). Consequently, the final choice of the more effective positioning of the electrodes requires a balance of costs and benefits for the suffering patients. The electrode placement, in fact, constitutes a crucial factor for cognitive side effects; in this regard, it seems that bifrontal electrode placement is as efficacious as bitemporal placement in inducing less severe cognitive impairment (Bailine et al. 2000).[12] The UK ECT Review Group recommends that bilateral arrangement is preferred when rapid and complete recovery has priority, whereas UL is preferred when minimizing cognitive adverse effects has priority.

> When treatment is not urgent, an initial trial of unilateral ECT will significantly shift the cost–benefit balance because of the

[11] It is worth noting that the World Health Organization promoted the ban on unmodified ECT (Leiknes et al. 2012).

[12] Notably, in schizophrenic patients the influence of electrode placement (or stimulus intensity) is less obvious than in depressive patients (Eser et al. 2007).

> substantial reduction in the risk of severe or persistent retrograde amnesia. […] Clinical monitoring of symptoms and possible cognitive adverse effects is necessary in any case throughout treatment, and a lack of satisfactory improvement may lead to an increase in the electrical dose or a switch to bilateral electrode placement if there had been no clinical improvement. (Scott 2005, 135-136)

What is the optimal frequency of treatments? According to Scott (2005, 141), in the case of bilateral treatment, the optimal frequency is twice per week, which may be reduced if cognitive adverse effects emerge. A three times per week administration could be justified only in severe, life-threating depression, since it leads to a faster reduction of depressive symptoms, but "at the cost of more pronounced cognitive adverse effects" (Scott 2005, 140). For what concerns UL ECT, the optimal frequency seems to be twice per week.

7. ECT in Pediatric Care

The stigma of ETC is even more evident when talking about its applications in children and adolescents. Any public and professional discourse over ECT in minors is likely "to be emotive, rhetorical, and unbalanced" (Robertson et al. 2013, 59). A (young) person receiving ECT is liable to experience stigma because of his or her age, illness, and treatment. It is a fact that negative attitudes toward ECT contribute to this stigma, which entails negative consequences such as concealment of illness and social exclusion (McDonald and Walter 2013, 51). However, when people who hold these prejudices encounter individuals who have received ECT, or gain accurate information about the treatment, they view the treatment in a more positive light. As Fink claims, "patients undergoing ECT have proved to be its best advocates" (2009, 11). It is the case of young patients and their parents as the studies of Walter et al. (1999) and Flamerique et al. (2017) show. In both studies, the researchers used a self-administered questionnaire to assess the experience, knowledge, and attitudes of parents of adolescents who had been treated with ECT about the treatment. In the former, 28 parents were interviewed and expressed favorable opinions about ECT: 17 of 28 parents claimed that ECT had been helpful. The latter investigated parents of adolescents (under the age of 18) diagnosed with schizophrenia and treated with ECT. They were compared with a randomly selected group of parents of adolescents treated with drugs.

> Most parents in the ECT group claimed that they had received adequate information about the ECT procedure (94.7%), most of them thought it had been helpful for their children (73.7%) and none thought that it had made things worse. The large majority of parents in the ECT group (80%) thought that the illness had been worse than ECT or medication, and none thought that ECT was the worst. (Flamerique et al. 2017, 1)

In adolescence,[13] treatment-resistant mood disorder constitutes a severe debilitating illness which may be successfully treated with ECT when treatment resistance is well identified (Ghazziudin 2013). Depression and catatonia in children and adolescents constitute a severe and debilitating disorder that can be life-threatening, and even in less severe cases, it significantly affects the quality of life of patients and their families. Furthermore, these kinds of disorders negatively affect normal growth and development. The treatment of prepubescent children still occurs in rare cases, while for adolescents the scientific literature confirms that prescriptions, safety and efficacy are the same as for adults (Fink 2009, Abrams 2002; Fink and Taylor 2003; Cohen et al. 2000; Consoli et al. 2012).

Nevertheless, there is a gap in the guidelines of many national institutes when this unpleasant failure of treatment occurs. For example, the American Academy of Child and Adolescent Psychiatry Practice Parameters for the assessment and treatment of children and adolescents with depression do not include ECT as a treatment option. The first-line treatment of moderate to severe adolescent depression is usually SSRI combined with evidence-based psychotherapy (CBT or IPT). Failure to respond to one or two SSRIs is often followed by treatment with a noradrenergic antidepressant. Ghazziudin estimates that "approximately 71% of adolescents may respond to initial combination treatment (fluoxetine and CBT) by the end of a 12-week period, although remission rates may be higher when treatment is administered over a longer duration" (Ghazziudin 2013, 185). It is worth noting that antipsychotic agents often have substantial side effects, such as weight gain and obesity associated with an elevated risk for diabetes mellitus (Holt 2019). Additionally, an increased risk of suicide rates has been associated with Antiepileptic drugs (AEDs), that can function as mood stabilizers and are often used for augmenting antidepressants. Still, they result in a "black box" warning by the US Food and Drug Administration (FDA). Ghazziudin found that "one in every four or five adolescents diagnosed with MDD may not respond to two sequentially used antidepressant agents administered inadequate dose

[13] ECT is an extremely rare procedure in children below 12 years of age.

and duration" (2013, 177). Some of these young patients might reasonably benefit from ECT.

ECT side effects in adolescents

ECT is usually well-tolerated by many adolescents. Some studies reported in Consoli et al. (2008, 155) indicate that "most former adolescent recipients report a positive experience and attitude toward ECT", despite initial and understandable apprehension. The possible presence of cognitive impairments among adolescents treated by ECT has been analyzed by Ghazziudin and colleagues (2000):

> Comparison of pre-ECT and the first post-ECT testing administered during the first 10 days of the treatment yielded significant impairments of concentration and attention, verbal- and visual-delayed recall, and verbal fluency. *A complete recovery of these functions* was noted at the second post-ECT testing. There was no deficit in the ability to problem solve during the initial or the subsequent testing. (Ghazziudin et al. 2000, 269, italics added)

Although caution is in order while we wait for further confirmation from larger samples, these results attest the evidence already emerged among adults: there is no long-term injury in terms of concentration, attention, visual memory, and verbal fluency. Side effects of ECT in young persons are generally transient, as in the adult population. As David Cohen, director of the child and adolescent psychiatry service at the Pitié-Salpêtrière hospital group in Paris, claimed: "there is no ethical reason to ban ECT use in adolescents". On the contrary, "unrealistic fears regarding ECT" lead to untreated minors, even in cases of dramatic conditions (Cohen et al. 2000, 1).

8. Legal and Ethical Concerns

In the USA, despite the updated recommendations of the American Psychiatric Association (APA) for ECT practice, the legislative framework concerning ECT differs among the member States. For example, in California and Texas there is a stricter legislation than the norms recommended by the APA. The EU boasts the creation of a European Forum for Electroconvulsive Therapy (EFFECT – founded in 2005)[14] and

[14] "EFFECT brought together clinicians and researchers to improve the practice of ECT, and to reduce its stigma" (https://www.theeffect.eu/).

a Task Force on ECT within the World Federation for Societies of Biological Psychiatry (WFSBP). Nevertheless, there is no unified European recommendation on ECT yet (Robertson et al. 2013). In Slovenia and Luxembourg, ECT is not available to treat people of any age. In Romania people are sometimes treated without anaesthesia because anaesthesiologists are not always available, while in Spain, Austria, Slovakia, Greece, Switzerland, and Germany, ECT is available and separate consent for anaesthesia is required (Gazdag et al. 2012, 6). In Italy, Ireland, and Latvia a written informative consent is needed for each ECT session, while in Portugal and the UK, one informative consent is valid for a definite number of sessions (12 or 15) (Gazdag et al. 2012).

ECT is so stigmatized that its use is severely limited, and its merits are neglected or even denied. The numerous attacks that ECT has received are often supported by ideological reasons not based on science or clinical experience. Nevertheless, these attacks have obtained political attention and often led to legislative restrictions (Ottosson and Fink 2004, 19). Misconceptions regarding ECT are associated with negative attitudes toward it, but prejudices and folk beliefs should not be involved in the scientific, ethical, and political discourse. Psychiatry – as Shorter and Fink note - "has swung wildly from fashion to fashion from asylum care to psychoanalysis to lobotomy to psychopharmacology without having an underlying scientific rationale for doing so. More than any other medical field, psychiatry has been guided by cultural preferences and political persuasions" (Shorter and Fink 2010, vi). It is time to "regroup" and adopt a "multifaceted approach" to change this bleak picture (McDonald and Walter 2013, 51). We must surrender to the idea that there is a group of psychiatric patients who are drug-resistant, or they cannot assume drugs for several reasons. In a nutshell, the ethical puzzle that I raise is the following. Is it possible to administer a therapy to help these suffering patients (be them adults or minors)? If the answer is yes, at least for some types of severe diseases, why should we refrain from administering that therapy? Why should we ignore, for ideological reasons, robust therapeutic possibilities? Why should we prolong the severity of the symptoms by making the life of patients and their family members unbearable? One therapy for the improvement of some severe psychotic symptoms exists and is practicable. ECT is such a therapy. In this sense, ECT does not constitute an alternative model of treatment, but an additional therapeutic tool that does not replace, but rather integrates pharmacotherapy and psychotherapy.

In conclusion, it is worth noting that the last decades have seen the promising development of several invasive treatments for brain stimulation aimed to relieve severe neuropsychological impairments in neurological

and psychiatric patients, like for example Transcranial Magnetic Stimulation (TMS), and Deep Brain Stimulation (DBS) (see Pycroft et al. 2018 for a review). If ECT is a well-established treatment for depression and catatonia, DBS is applied in patients who have severe Parkinson's disease, as well as depression (Marcolin and Padberg 2007, vii). Many of these brain stimulation techniques "converge in terms of underlying mechanisms of action based on fundamental principles of brain function" and differentiate each other in virtue of specific characteristics regarding the levels of invasiveness and the duration of intervention (Ibid., viii). For example, DBS, is, *de facto*, a permanent stimulator into the brain and it represents "a long-term treatment, particularly suitable for chronic or frequently relapsing disorders" (Ibid.). These novel methods find great attention without triggering reprobation, while ECT is still condemned.

9. Conclusion

Electroconvulsive Therapy represents an appropriate treatment for severely ill psychiatric patients who need hospital care. In those Western countries where ECT is permitted and practiced, it is usually recommended after long, unsuccessful courses of psychotherapies and psychotropic drugs. To summarize in a nutshell the current practice, I follow Yuval Bloch's words, according to whom the major reason for child and adolescents referrals to ECT is represented by the severity of symptoms, particularly catatonia and suicidal behaviour, while the major reason for referral of adults is the long and persistent failure of response to pharmacotherapy (Bloch et al. 2008). In this paper I set out to show that a further step is possible to promptly provide appropriate and effective care in those difficult cases. However, a cultural change is necessary to get started with this process.

Acknowledgments

I would like to thank Valentina Petrolini, Marko Jurjako, and Matteo Fiorani, excellent *travel* companions.

REFERENCES

Abrams, R. 2002. *Electroconvulsive Therapy*. New York: Oxford University Press.

Accornero, F., and M. Anderson. 1948. L'elettroshock nella psichiatria infantile. *Il Lavoro Neuropsichiatrico* 3(2): 231-250.

American Psychiatric Association. 2001. *Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging* Washington: APA.

Babini, V. 2009. *Liberi Tutti. Manicomi e Psichiatri in Italia: Una Storia del Novecento*. Bologna: Il Mulino.

Bailine, S. H., A. Rifkin, E. Kayne, J. A. Selzer, J. Vital-Herne, M. Blieka, and S. Pollack. 2000. Comparison of bifrontal and bitemporal ECT for major depression. *American Journal of Psychiatry* 157, 121–123.

Baldi, F., and G. Reale. 1947. L'elettroshock nella terapia neuropsichiatrica. *Rivista Sperimentale di Freniatria* 72: 239–42.

Berrios, G. E. 1997. The scientific origins of electroconvulsive therapy: A conceptual history. *History of Psychiatry* 8: 105–119.

Bini, L. 1940. La tecnica e le manifestazioni dell'elettroshock. *Rivista Sperimentale di Freniatria* 18: 361–458.

Bloch Y., D. Sobol, and Y. Levkovitz. 2008. Reasons for referral for electroconvulsive therapy: A comparison between adolescents and adults. *Australasian Psychiatry* 16: 191–194.

Campailla, G. 1945. Le demenze infanto-prepuberali. *Archivio di Psicologia Neurologia Psichiatria e Psicoterapia* 4(4): 262-338.

Cerletti, U. 1940. L'elettroshock. *Rivista Sperimentale di Freniatria* 18: 209–310.

Clayton, A. H., J. F. Pradko, H. A. Croft. et al. 2002. Prevalence of sexual dysfunction among newer antidepressants. *Journal of Clinical Psychiatry* 63(4): 357–366.

Coccanari de' Fornari, M. A., A. Iannitelli, and M. Biondi. 2017. Storia della clinica psichiatrica della Sapienza Università di Roma nel policlinico Umberto I. *Rivista di Psichiatria* 52(1): 1-8.

Cohen, J., M. Flament, O. Taieb, C. Thompson, and M. Basquin. 2000. Electroconvulsive therapy in adolescence. *European Child & Adolescent Psychiatry* 9: 1-6.

Consoli, A., J. Cohen, N. Bodeau, V. Guinchat, and L. Wachtel. Electroconvulsive therapy in adolescents with intellectual disability and severe self-injurious behavior and aggression: A retrospective study. *European Child & Adolescent Psychiatry* 22(1): 55-62.

Consoli, A., W. De Carvalho, and D. Cohen. 2013. Side effects of ECT. In *Electroconvulsive Therapy in Children and Adolescents*, eds. N. Ghazziudin and G. Walter, pp. 140-160. New York: Oxford University Press.

Ende, G., D. F. Braus, S. Walter, W. Weber-Fahr, and F. A. Henn. 2000. The hippocampus in patients treated with electroconvulsive therapy: A proton magnetic resonance spectroscopic imaging study. *Archives of General Psychiatry* 57, 937–943.

Eser, D., C. Schüle, R. Rupprecht, and T. C. Baghai. 2007. Electroconvulsive therapy: Update and new research. In *Transcranial Brain Stimulation for Treatment of Psychiatric Disorders*, eds. M. A. Marcolin and F. Padberg. *Advances in Biological Psychiatry* 23: 1–17.

Fear, C. F. 2005. The use of ECT in the treatment of schizophrenia and catatonia. In *The ECT Handbook*, ed. A. Scott, pp. 30-46. London: The Royal College of Psychiatrists

Fink, M., and M. A. Taylor. 2003. *Catatonia: A Clinician's Guide to Diagnosis and Treatment*. Cambridge: Cambridge University Press.

Fink, M. 2011. Transcranial magnetic stimulation is not a replacement for electroconvulsive therapy in depressive mood disorders. *Journal of ECT* 27: 3–4.

Fink, M. 2013. The mechanism of action of ECT. In *Electroconvulsive Therapy in Children and Adolescents*, eds. N. Ghazziudin, G. Walter, pp. 18-28. New York: Oxford University Press.

Fiorani, M. this issue. Rationality, irrationality and irrationalism in the anti-institutional debate in psychiatry around the second half of the 1970s in Italy. European Journal of Analytic Philosophy. https://doi.org/10.31820/ejap.16.2.4

Flamarique, I., I. Baeza, E. de la Serna, A. Pons, M. Bernardo, and J. Castro-Forniele. 2017. Thinking about electroconvulsive therapy: The opinions of parents of adolescents with schizophrenia spectrum disorders. *Journal of Child and Adolescent Psychopharmacology* 27(1): 75-82.

Fraser, L. M., R. W. O'Carroll, and K. P. Ebmeier. 2008. The effect of electroconvulsive therapy on autobiographical memory: A systematic review. *The Journal of ECT* 24(1): 10-17.

Ghazziudin, N. 2013. ECT for Mood Disorder. In *Electroconvulsive Therapy in Children and Adolescents*, eds. N. Ghanzi and G. Walter, pp. 161-190. Oxford University Press.

Ghazziudin, N., D. Laughrin, and B. Giordani. 2000. Cognitive side effects of electroconvulsive therapy in adolescents. *Journal of Child and Adolescent Psychopharmachology* 10(4): 269-276.

Gazdag, G., R. Takács, G. Ungvari, and P. Sienaert. 2012. The practice of consenting to electroconvulsive therapy in the European Union. *Journal of ECT* 28(1): 4–6.

Healy, D. 2002. *The Creation of Psychopharmacology*. Cambridge: Harvard University Press.

Heuyer G., and F. Bour. 1942. Electro-choc chez les adolescents. *Annales Médico-Psychologiques* 2: 75–84.

Hoekstra, R., W. W. van den Broek, D. Fekkes, J. A. Bruijn, P. G. Mulder, and L. Pepplinkhuizen. 2001. Effect of electroconvulsive therapy

on biopterin and large neutral amino acids in severe, medication resistant depression. *Psychiatry Research* 103, 115–123.

Holt, R. I. G. 2019. Association between antipsychotic medication use and diabetes. *Current Diabetes Reports* 19(96) doi: 10.1007/s11892-019-1220-8.

Holtzheimer, P. E., and H. S. Mayberg. 2011. Stuck in a rut: Rethinking depression and its treatment. *Trends in Neuroscience* 34(1): 1-9.

Jacobsen, P. L., E. M. Thorley, and C. Curran. 2020. Real-world patient experience with sexual dysfunction and antidepressant use in patients with self-reported depression: Across-sectional survey study. *Neurology, Psychiatry and Brain Research* 36: 57-64.

Jervis, G. 2014. *Contro il Sentito Dire. Psicoanalisi, psichiatria e politica*. Torino: Bollati Boringhieri.

Kellner, C. H., M. Fink, R. Knapp, G. Petrides, T. Rummens, M. Mueller, H. Bernstein, K. Rasmussen, K. O'Connor, G. Smith, A. J. Rush, M. Biggs, S. McClintock, S. Bailine, and C. Malur. 2005. Relief of expressed suicidal intent by ECT: A consortium for research in ECT study. *American Journal of Psychiatry* 162, 977–982.

Kirov, G. 2020. *Shocked: Insider stories about electroconvulsive therapy*. Amazon Kindle Direct Publishing.

Lamprecht, H. C., I. N. Ferrier, and A. G. Swann. 2005. The use of ECT in depressive illness. In *The ECT Handbook,* ed. Scott, A. I. p. 9-24. London: The Royal College of Psychiatrists.

Leiknes, K. A., L. J.-von Schweder, B. Høie. 2012. Contemporary use and practice of electroconvulsive therapy worldwide. *Brain and Behavior* 2(3): 283-344.

Marcolin M. A., and F. Padberg. 2007. Transcranial brain stimulation for treatment of psychiatric disorders. *Advances in Biological Psychiatry* 23.

Maayan, R., Y. Yagorowski, D. Gruppe, M. Weiss, B. Shtai, M. A. Kant, and M. A. Weizman. 2000. Basal plasma dehydroepiandrosterone sulfate level: A possible predictor for response to electroconvulsive therapy in depressed psychotic inpatients. *Biological Psychiatry* 48, 693–701.

McDonald, A., and G. Walter. 2013. ECT and Stigma. In *Electroconvulsive Therapy in Children and Adolescents*, eds. N. Ghanzi and G. Walter, pp. 29-55. USA: Oxford University Press.

Nemec, G. 2015. *Dopo Venuti a Trieste. Storie di Euli Giuliano-dalmati attraverso un manicomio di confine 1945-1970*. Trieste: Alphabeta verlag.

Ottoson, J. O., and M. Fink. 2004. *Ethics in Electroconvulsive Therapy*. New York: Brunner-Routledge (Taylor & Francis).

Passione, R. 2004. Italian psychiatry in an international context: Ugo Cerletti and the case of electroshock. *History of Psychiatry* 15(1): 83–104.

Passione, R. 2007. *Ugo Cerletti. Il romanzo dell'elettroshock.* Reggio Emilia: Aliberti.

Petrides, G., C. Malur, and M. Fink. 2004. Convulsive therapy. In *Catatonia. From Psychopathology to Neurobiology*, eds. S. N. Caroff, S. C. Mann, A. Francis, and G. L. Fricchione, pp. 151-160. Washington: American Psychiatry Publishing.

Palmio, J., M. Huuhka, P. Saransaari, S. S. Oja, J. Peltola, E. Leinonen, J. Suhonen, and T. Keranen. 2005. Changes in plasma amino acids after electroconvulsive therapy of depressed patients. *Psychiatry Research* 137, 183–190.

Pycroft, L., J. Stein, and A. Tipu. 2018. Deep brain stimulation: An overview of history, methods, and future developments. *Brain and Neuroscience Advances* 2: 1-6.

Robertson, M., J. M. Rey, and W. G. Walter. 2013. Ethical and Consent Aspects. In *Electroconvulsive Therapy in Children and Adolescents*, eds N. Ghazziudin and G. Walter, pp. 56-75. New York: Oxford University Press.

Rosebush, M. F., and P. I. Mazurek. 2004. Pharmacotherapy. In *Catatonia. From Psychopathology to Neurobiology*, eds. Caroff, S. N., Mann, S. C., A. Francis, and G. L. Fricchione, pp. 141-150. Washington: American Psychiatry Publishing.

Sachs M., and V. Madaan. 2012. *Electroconvulsive Therapy in Children and Adolescents: Brief Overview and Ethical Issues*. Charlottesville: University of Virginia Health System.

Sackeim, H. A., Devanand, D. P., and J. Prudic. 1991. Stimulus intensity, seizure threshold, and seizure duration: Impact on the efficacy and safety of electroconvulsive therapy. *Psychiatric Clinics of North America* 14: 803–843.

Sanacora, G., G. F. Mason, D. L. Rothman, F. Hyder, J. J. Ciarcia, R. B. Ostroff, R. M. Berman, and J. H. Krystal. 2003. Increased cortical GABA concentrations in depressed patients receiving ECT. *American Journal of Psychiatry* 160, 577–579.

Sirgiovanni, E., and A. Aruta. 2020. The electroshock triangle: Disputes about the ECT apparatus prototype and its display in the 1960s. *History of Psychiatry* 31(3): 311-324.

Scott, A. I. F. ed. 2005. *The ECT Handbook. The Third Report of the Royal College of Psychiatrists' Special Committee on ECT*. London: The Royal College of Psychiatrists.

Shorter, E. 1997. *A History of Psychiatry. From the Era of the Asylum to the age of Prozac*. New York: John Wiley & Sons.

Shorter, E. 2009. *Before Prozac. The Troubled History of Mood Disorders in Psychiatry*. Oxford: Oxford University Press.

Shorter, E. 2013. The history of pediatric ECT. In *Electroconvulsive Therapy in Children and Adolescents,* N. Ghazziudin and G. Walter pp. 1-17. New York: Oxford University Press.

Shorter, E., and M. Fink. 2010. *Endocrine Psychiatry: Solving the Riddle of Melancholia*. New York: Oxford University Press.

Shorter, E., and D. Heale. 2007. *Shock Therapy. A History of Electroconvulsive Treatment in Mental Illness*. New Brunswick: Rutgers University Press.

UK ECT Review Group. 2003. Efficacy and safety of electroconvulsive therapy in depressive disorders: A systematic review and meta-analysis. *Lancet* 361, 799–808.

Walter, G., K. Koster, and J. Rey. 1999. Views about treatment among parents of adolescents who received electroconvulsive therapy. *Psychiatric Service* 50: 701–702.

Walter, G., M. Robertson, J. M. Rey, N. Soh, and G. S. Malhi. 2010. Electroconvulsive therapy in young people and the pioneering spirit of Lauretta Bender. *Acta Neuropsychiatrica* 22(5): 253-254.

Weiss, A. 2018. *The Electroconvulsive Therapy Workbook. Clinical Applications*. London-New York: Routledge.

Regular article

EuJAP | Vol. 16 | No. 2 | 2020
UDC: 1:[004+504](062)
https://doi.org/10.31820/ejap.16.2.7

THE ONLINE ALTERNATIVE: SUSTAINABILITY, JUSTICE, AND CONFERENCING IN PHILOSOPHY

Rose Trappes
Bielefeld University

Daniel Cohnitz
Utrecht University

Viorel Pâslaru
University of Dayton

T. J. Perkins
University of Utah

Ali Teymoori
Helmut Schmidt University

Original scientific article – Received: 7/11/2020 Accepted: 11/12/2020

ABSTRACT

The recent global pandemic has led to a shift to online conferences in philosophy. In this paper we argue that online conferences, more than a temporary replacement, should be considered a sustainable alternative to in-person conferences well into the future. We present three arguments for more online conferences, including their reduced impact on the environment, their enhanced accessibility for groups that are minorities in philosophy, and their lower financial burdens, especially important given likely future reductions in university budgets. We also present results from two surveys of participants who attended one large and three small online philosophy conferences this year. We show that participants were in general very satisfied with presentations and discussions at the conferences, and that they reported greater accessibility. This indicates that online conferences can serve as a good alternative to in-person conferences. We also find that networking was less satisfactory in online conferences, indicating a point for improvement and further research. In general, we conclude that philosophers should continue to organize online conferences after the pandemic. We also provide some advice for those wishing to organize online conferences.

Correspondence: rtrappes@uni-bielefeld.de

Keywords: *Online conferences; accessibility; carbon footprint; carbon offsetting; inclusivity; minorities in philosophy*

1. A Natural Experiment

The pandemic has caused a collective re-think in the ways that many facets of academia traditionally proceed. The emergence of COVID-19 in early 2020 made it such that in-person conferencing, a regular part of most academics' yearly routine, was deemed too high risk and—given rapid closures of borders and universities—was soon practically impossible. Many conferences were cancelled in the early months of the pandemic; some were postponed (Philosophy of Science Association Biennial Meeting postponed one year until 2021, for example). Yet there were some conference organizers who shifted the meetings to an online format. And so arose a global groundswell of virtual conferences in philosophy.

Pre-corona, holding a conference online was not part of the mainstream. Whether large or small, local or international, conferences were just supposed to involve hotels, handshakes, and those little biscuits that inevitably come with filter coffee from an urn. Until recently, only a handful of philosophers had bucked the trend, convened online, and argued in favor of the virtual format.[1] The sudden increase in online conferencing in 2020 therefore represents a sort of natural experiment. Despite the confounding factors that a pandemic brings, we can start to look at whether online conferences in philosophy are an acceptable alternative to in-person conferences.

Assessing the suitability of online formats for conferences is especially pressing in light of a number of existing arguments in their favor. There have for some years been calls for academics to reduce their carbon footprints through limiting travel, for instance through online conference attendance. In addition, in-person conferences are often extremely financially demanding, and they present specific challenges for researchers outside North America and Europe, researchers with disabilities, and primary caregivers. The idea that these diverse challenges can be overcome

[1] *Consciousness Online Conference* organized by Richard Brown from 2009 to 2013 was one of the first online philosophy conferences (https://consciousnessonline.wordpress.com/program/).
Buckner, Byrd and Schwenkler (2015) offered a model of online conferences and argued in their favor. Byrd (2020) significantly updated the argument with new data and reasons, and some useful advice can be found in St. Croix (2020) and Calzavarini and Viola (2020). We thank an anonymous reviewer for alerting us to some of these sources.

by taking a number of conferences online are often answered with a hand-wavy "but we just *have* to meet in person".

If the pandemic has taught us anything, it's that this answer is no longer good enough. In this paper we consider four philosophy conferences that were held online between April and August 2020, presenting empirical results showing that they are in many respects a suitable alternative to in-person conferences. In particular, presentations and discussions are experienced by participants and speakers as very satisfactory in an online format. Networking does suffer to a certain extent, though we suggest that this can be partly remedied through planning to provide networking opportunities with special attention to diverse needs of the audience.

We therefore argue that online conferences, rather than just a necessary measure during acute crises like a pandemic, are a sustainable and functional alternative to—but not wholesale replacement of—in-person conferences for the future of philosophy. Taking more conferences online is crucial to reduce the carbon footprint of philosophy, to address existing systematic inequalities in conference accessibility, and to cope with likely post-pandemic economic shortfalls and the consequent restrictions on university funding. Although some of these arguments for online conferences could also be addressed with hybrid conferences permitting online attendance, we focus on online-only conferences. Hybrid conferences have their own challenges and specificities that demand a separate treatment.

We begin by looking at three arguments for holding more conferences online: the environmental impact of traditional conferencing, the accessibility problems of many in-person conferences, and the likely increasing financial restrictions of scholars and universities to attend and host in-person conferences. We then introduce the four online conferences we organized and present the results of two post-conference surveys. Based on these results, we show that online conferences are a suitable alternative to both large and small in-person conferences and that pre-recorded and live lectures are both accepted formats, and we provide some suggestions for how to schedule and structure a successful online conference. Given the three arguments for and the suitability of online conferences, we conclude by suggesting that even after the pandemic online conferences should be the new default for academic meetings along with measures to decarbonize academic conventions and offset carbon emissions from both online and in-person meetings that cannot be avoided. In-person meetings should be rare and well justified departures from the default due to the inability of the online format to offer to academic practice what the in-person format affords. Networking opportunities are

the major shortcoming of online conventions and further work is required to design them so that scholars, especially early-career ones, benefit from this important experience.

2. Three Reasons for Online Conferences

2.1. Environmental Issues

We believe that it is roughly accurate that most philosophers are committed to social justice, inclusivity and have accepted the findings and recommendations of the IPCC (Intergovernmental Panel on Climate Change). A result of this commitment is that some philosophers have addressed the moral implications of greenhouse gas pollution and the responsibility of governments and of individuals to act toward preventing, reducing and eliminating this pollution that causes widespread harm. An example of such an argument for environmental action can be found in John Broome (2016, 161): "Justice requires you not to harm other people, at least not for your own benefit. Since emissions of greenhouse gas do harm, you should not make them". Arguments like this speak in favor of online conferences.

The two models of online conferences that we present here are ways to effectively realize the moral argument for greenhouse gas reduction. In the absence of estimates for philosophy conferences, we can use those for science conventions to gauge their environmental impact. Burtscher et al. (2020) estimate the total carbon footprint of the virtual meeting of the European Astronomical Society to be 582 kg, roughly 3,000 times smaller than the carbon footprint of the 2019 in-person meeting in Lyon. Klöwer et al. (2020) estimate that travel to the 2019 meeting in San Francisco of the American Geophysical Union resulted in 80,000 tons of carbon emissions, whereas choosing a venue with the explicit goal to minimize transport emissions, increasing virtual attendance and meeting biannually in person instead of annually would have reduced about 90% of travel-related carbon emissions.

Despite the existence of philosophical arguments for reducing greenhouse gas emissions, professional organizations of philosophers have not implemented measures to effectively reduce and offset greenhouse gas emissions that result from their activities. And thirty years after the first

IPCC report, philosophers appear to do mostly business as usual.[2] This is despite the environmental impact of philosophers' research activities, and while forcefully objecting to politicians and businesspeople who advocate business as usual to ensure economic growth. Critics of philosophers would be right to label philosophers' talk not supported with substantial measures as a hypocritical and glaring departure from professed moral principles. The two models of online conferences described here allow professional organizations of philosophers to close the wide gap between their public defense of environmental causes and actual actions.

In addition to professional ethics, there is also an argument based on inclusivity towards personal preferences. Public statements of philosophical organizations show various efforts to be inclusive not just towards needs like childcare or accessibility, but also towards the preferences of those who have made principled decisions to be vegetarians and vegans. For several years, a number of philosophers have joined a

[2] Here is a sample for illustration. The 2020 edition of the Good Practices Guide of the American Philosophical Association (Railton et al. 2020, 95-102) includes for the first time a section, the last one, on sustainability, containing comprehensive advice on preventing and reducing the environmental footprint of philosophical events. Funding and/or encouraging the use of carbon offsets and incorporating to various degrees digital conferencing are among the recommendations. Yet what forced APA to move its main meetings online was the pandemic, not the recommendations from its guide of good practices. APA plans to examine the issue of carbon offsets for travel to its meetings (Amy Ferrer, personal communication with VP on October 23, 2020).
The biannual meeting sites of the International Society for the History, Philosophy and Social Studies of Biology (ISHPSSB) oscillate between North America and Europe. The Site Selection Committee reports Milwaukee, WI, as the site for the 2021 meeting and that it has received an inquiry about hosting the 2023 meeting in Australia (sic!). Of the four points the Committee makes about future meetings, not one concerns the carbon footprint. Milwaukee site organizers are silent on this topic as well. The society plans to discuss the carbon footprint of its meetings at the upcoming Milwaukee conference, and the last meeting held in Oslo implemented measures to promote reusables and reduce single-use materials, such as plastic cups and bottles (Newsletter of ISHPSSB Fall 2019). The society has hosted a number of talks on problems related to sustainability.
The Philosophy of Science Association holds its meetings biannually in the USA and Canada, and its members have been examining philosophical problems of climate science. In the summer of 2020, it established a Sustainability & Climate Task Force. One of its goals is to reconceive "the format, frequency, and location of PSA meetings given the heavy carbon footprint of the existing conference model" (Sustainability & Climate Task Force), yet the decision to hold the regular meeting in Baltimore, MD, in 2021 so as to avoid a hefty hotel cancellation fee did not come with a request to association members to offset their carbon emissions nor information on how to do it (https://psa2020.philsci.org/81-psa2020-2021-faqs). Offsetting carbon emissions were not requested at the previous meetings either.
The sites of the European Philosophy of Science Association, the British Philosophical Association, the British Society for the Philosophy of Science, and the German Society for Philosophy of Science do not contain information about their efforts to address their carbon emissions.

growing number of scientists who object to flying to conferences. Some of them have self-reported on https://noflyclimatesci.org/. These academics are conscientious climate change objectors. The traditional model of in-person conferences is not inclusive toward them. To be inclusive toward these academics, and given the moral and justice principles to which philosophers are committed as well as the aforementioned precedents, virtual models of conference participation should be implemented.

2.2. Accessibility

In-person conferences are not as accessible to researchers outside the European Union and North America, to researchers with disabilities, and to primary caregivers (often women), all of whom are underrepresented groups in philosophy (Schwitzgebel and Jennings 2017; Humanities Indicators 2019a, 2019b). We think online conferences address many in-person accessibility issues and may thereby redress systematic limitations on conference attendance.

The first group benefited by online conferences are researchers outside the European Union and North America. For many of these researchers, in-person conferences require cumbersome, time-consuming and expensive visa application procedures, and visas are often denied or not granted in time (Khalid, Ardila-Gómez, and Scott 2016; Minai 2018; Albayrak-Aydemir 2020). Increased travel time, planning and expense present additional obstacles. Online conferences obviate the need to travel and obtain visas and thereby facilitate attendance from such countries.

Online conferences also offer advantages for researchers who have disabilities. Despite steps to improve accessibility of in-person conferences, many hurdles remain (Felappi, Gregory, and Beebee 2018; Fleming 2019; Railton et al. 2020, 70-76). In online conferences, participants can utilize their own systems, such as technological and physical aids. An online format might also help some participants with networking. For instance, using breakout rooms and written chats places less burden on individuals to approach strangers and reduces sensory input in comparison to crowded conference halls.

Finally, online conferences can benefit primary caregivers. Attending from home simplifies bottle- or breastfeeding, often challenging at in-person conferences (Calisi, Working Group Mothers in Science 2018; Felappi, Gregory, and Beebee 2018; Railton et al. 2020, 70-76). Muting or turning off the video also enables parents to remain in talks rather than having to leave the room when their child is crying. Depending on the conference schedule, parents can also often utilize their usual childcare arrangements.

Online conferences are however not without accessibility issues. First, not all researchers have access to adequate technology or internet connection, in particular working- or lower-class scholars (Minai 2018). Recording talks and using written chats might enable some participation, but these are likely sub-optimal for networking. Second, time zone differences reduce the wins for researchers outside traditional conferencing regions, though not entirely (researchers in South America can usually attend meetings in North America, and similarly for researchers in Africa and the Middle East for European meetings). Third, conference schedules should respect parents and people with disabilities. Breaks are especially important for these groups; a shorter day is often also necessary (Botterill 2020). Finally, conference-provided childcare funds remain important to ensure extra childcare can be arranged at no cost to participants.

If these steps are taken, we think online conferences are likely to enhance participation from minorities in philosophy. Increasing the availability of online conferences is not only fairer, it might also contribute to reducing inequalities in career outcomes for members of minority groups, especially given the importance of attending conferences for early-career researchers (Calisi, Working Group Mothers in Science 2018; Felappi, Gregory, and Beebee 2018; Railton et al. 2020).

2.3. Financial Issues

One of the many impacts of the COVID-19 pandemic has been economic. The measures recommended to prevent transmission of infection—social distancing, reduced numbers in groups, wearing masks, isolating when exposed, restricted international travel, etc.—spell disaster for the normal maintenance of many businesses. This has had ramifications for the economy as a whole as many are not working (whether laid off or on leave) and businesses have closed, many not to open again. For the purposes of this paper, the impact that networks around universities have faced is most salient. This includes students, staff, and the institutions themselves. The online conference format, we argue, may provide some relief to the monetary strains placed on universities given the economic impact of the pandemic, and will also be worth considering even in times of relative normalcy. Further, even in times of non-acute crisis, online conferences provide those without the fiscal means to travel an opportunity to attend and be involved.

Many universities are reporting large budget shortfalls due to the recommended COVID-related changes in student activities. Just as one example, according to Lee Gardener (2020), “The University of Wisconsin system […] has estimated it will lose $170 million in the spring semester

alone from refunding room, dining, and parking fees to students, and other unexpected expenses". While each university will be impacted differently, there is no doubt that many universities will be impacted in this or similar ways.

The impacts of these budget shortfalls are trickling down into the budgets of the individual departments, often resulting in the suspension of admissions to graduate programs. As of September 28, more than 50 humanities and social sciences departments in the US have suspended PhD admissions (Zahneis 2020). Largely, the justification to cut admissions has been to allocate what little resources remain to their existing students. While this paper is not about the larger effects of COVID-19 on universities, the point is that philosophy departments are likely going to be feeling a fiscal crunch for some time. This will potentially impact travel budgets: money allocated for both sending students out and bringing guest faculty in.

Even small conferences often require many thousands of dollars for flying and housing speakers, booking conference spaces, catering, software, staffing, and social events like day-trips or city tours depending on the location (De Cruz 2015). Online conferences offer a way to alleviate a lot of the spending, and associated risks, that accompanies an in-person conference. Nevertheless, some costs will remain, including for staff, technical support, software, and potentially also reimbursements for speakers.

To offset these kinds of costs, conference organizers will often require a registration fee which can be hundreds of dollars for larger conferences. For example, the 2019 Pacific American Philosophical Association meeting registration fee ranges from $90 to $290 depending on career status (https://www.apaonline.org/page/2019P_RegInfo). Registration and travel costs are especially difficult for graduate students and early career researchers. Large proportions of doctoral students report feeling stressed about money on a regular basis (Kasia 2016). Even a domestic flight can be quite a burden for a graduate student, not to mention the costs of international travel, visas, hotels, and dining out. Attendance at conferences is thought to be a necessary component of career-advancement for early career scholars, so not attending has implications for career prospects later.

Virtual conferences alleviate much of the financial burden, enabling attendance by those affected by financial worries. In addition to students, the reduction in attendance cost promises to be especially beneficial for researchers from low-funded universities or countries with little public

funding for research, a condition that may increase in many post-pandemic economies, especially given the facts about university budgets discussed above.

Virtual formats promise to reduce many of the costs associated with organizing and attending conferences, and thereby reduce the required registration fee. Indeed, lower costs and avoiding travel were overwhelmingly recorded as positives in the survey responses (see below), suggesting that this mattered greatly to the attendees of online conferences.

3. Conference Models

In this section we introduce the online conferences that we organized. The European Congress for Analytic Philosophy (ECAP), and the colloquia *Doing Science in a Pluralistic Society* (DSPS), *Eco-Evo Mechanisms* (EEM), and *Philosophy of Biology at the Mountains* (POBAM) were planned as in-person. Because of lockdowns, organizers of all meetings, after having consulted conference participants, decided shortly before the planned events to switch to the virtual model. Here we describe some of the common aspects of and variations to their organization.

3.1. Large Scale Event: ECAP10

Every three years the European Society for Analytic Philosophy (ESAP) organizes the ECAP. With about 800 participants at the 2017 congress, ECAP is one of the biggest philosophy conferences in Europe. Plans to hold the 2020 congress in Utrecht (Netherlands) at the end of August were scrapped in mid-March, and the move was made to switch the conference online.

Background

The conference was supposed to run for a whole week with several parallel sessions (ECAP had reservations for 13 rooms that could be used in parallel), keynote lectures, invited longer talks, and symposia. For each of the 450 contributed papers, 20 minutes were allocated in the programme plus about 5 minutes of discussion. The participation fee was set at 200 € (300 € after April 1) or 70 € for students (100 € after April 1).

Conference Organization

It took until June 5 to come up with a detailed plan to have the conference online. It was clear that one couldn't have 13 parallel live video group chats

for 8hrs a day (as we would have had for the in-person conference). It would be impossible to do the troubleshooting and tech support, and nobody would be able to follow so many talks online. It was also assumed that many people would no longer be able to attend during the (whole) week for which the conference was originally planned.

The ECAP organization thus opened a registration for the reduced fee of 30 € (to cover the costs for two student assistants and the EasyChair license that was used for the review process). Talks were to be pre-recorded, with two options for Q&A: offline/asynchronous or online live Q&A. The conference would be hosted in MS Teams (see Supplementary Material Section 1 for details on the technical setup).

Registration and Participation

Of the 450 accepted speakers for the physical conference, about 300 decided to participate in the online version. Eventually, over 400 people participated in the conference. The top 7 countries where participants came from were Germany, Italy, UK, the Netherlands, Poland, Spain and the Czech Republic, but the conference was also attended by philosophers outside Europe, such as the USA and India. In total, participants came from 32 different countries.

Contributed papers

All talks (except for invited talks and keynotes) were accessible online from August 17, one week before the official start of the conference. This way conference participants had a week to watch the talks that they wanted to see. Participants could also comment directly on the pre-recorded videos.

Roughly 50% of speakers opted for a live Q&A, while the rest preferred a purely offline Q&A. On the basis of the registration, a program was made for the live Q&A talks, the invited talks and the keynote lectures. At most 4 parallel live Q&A sessions were scheduled, each with 6 papers for 1h. That way the conference programme was not too demanding on each day (see Supplementary Material section 2).

Each live Q&A session had a chair and participants were asked to watch all videos for a live Q&A meeting beforehand. At the live Q&A presenters were given a 2-minute spot to quickly remind everyone of the main thesis of their papers. For each paper there were roughly 10 minutes of discussion time allocated.

Invited talks and keynotes

Invited talks and keynotes were streamed within MS Teams at specific times and then followed by a longer live Q&A. Invited talks were 40 minutes, followed by 20 minutes live discussion, keynotes were 60 minutes, with 60 minutes discussion. The keynote lectures were also simultaneously uploaded to YouTube for a wider audience. After the live event was over, the videos of the keynotes and invited talks were also available within MS Teams for participants who weren't able to attend the live event.

Networking

In between events, participants were encouraged to use a dedicated environment within MS Teams for discussion and chats. On one evening during the conference, ECAP organized a pub quiz as a social event.

Local Team

The local organization was a team of 6 colleagues, plus two student assistants for July and August. For a physical congress of that size, a much bigger team would have been necessary. In addition to the two student assistants, the only extra cost was for the EasyChair license (Utrecht University has a license for MS Teams). That way the conference could be organized for a fraction of the costs of a physical conference.

Since MS Teams is used at Utrecht University for teaching, the local organization team was already familiar with the software and Utrecht University could provide tech support. The tech support that was requested from participants, for entering MS Teams, creating and uploading videos, navigating the conference environment, etc. was minimal. The local organization team experienced the organization of the online conference as less stressful and demanding than the organization of comparable physical conferences.

3.2. Small to Medium Scale Conferences

The other three online conferences we are comparing, all in philosophy of science, were comparatively smaller events. DSPS, EEM, and POBAM used the conferencing software Zoom (see Supplementary Material Section 1 for details on the technical set-up).

Conference Organization

DSPS, EEM and POBAM were organized to be as close as possible to in-person events. All conferences involved live presentations and Q&A sessions, with each session (including Q&A) ranging from 40 minutes to 2 hours and 50 minutes, which included a 10-minute break. Talks of speakers who gave permission were recorded and access to recordings and slides was enabled for all participants after the conference was over.

The conference schedules varied from consecutive days to spread out over the course of a week. DSPS events took place on consecutive Fridays, POBAM activities occurred on Tuesday, Thursday and Monday, while EEM meetings took place on two consecutive days, Thursday and Friday. All meetings started at 9:00 local times of organizers and continued until 17:00, but 15:40 for POBAM.

For some of the conferences, their programs were entry points to conference activities, with zoom meeting rooms linked to the names and talk titles displayed in the program. This allowed participants to attend the talks of their choice just like in an in-person conference. Programs are attached in the supplementary material for illustration (Supplementary Material Section 2).

Registration and Participation

Registration was free, yet required to prevent "zoombombing", when an uninvited person joins a virtual meeting, often with the intention of being disruptive (Gunnel 2020). As a safety measure, a password to access the meetings was sent to registrants after they had registered. We adopted this measure because of media reports about occasional zoombombing, although we are yet to have experienced such incidents ourselves. Zoom also allows moderators to block possible disruptors, another way to deal with zoombombing without the hurdle of required registration. It is also possible to set up a zoom meeting so that participants have to request to be unmuted, which provides a certain level of security against unwanted disruptions.

Registration for the conferences was as follows: for DSPS, 127 persons from 16 countries (except for USA, India and South Korea, all countries were from Europe). Similarly, POBAM's 136 registrants were mainly from the USA, Canada and Europe, but also from Mexico, Brazil, Egypt, India, Australia, and New Zealand. EMM had 100 registrations from 21 countries, primarily from Europe, but also from countries in North and South America, South-East Asia and the Middle East.

Conference participation was lower than registration. For DSPS, POBAM, and EEM participation varied between 30 and 70 participants.

Talks and Q&A Sessions

At DSPS, POBAM, and EEM, talks were delivered live and were followed by live Q&A sessions. Keynote talks were between 40 and 50 minutes long, and regular talks were between 20 and 30 minutes, followed by 10 to 20 minutes of Q&A. Attendees could raise their hand (digitally) and ask their question *via* audio/video, or they could write a question in the chat, to be read out by the moderator.

Networking and other sessions

All three conferences created opportunities for informal virtual social interaction. Separate Zoom meeting rooms were created for those events. DSPS had morning cafes, prior to the morning talks, and lunches. EEM included networking coffee breaks, a group-work session, and a happy hour at the end of the first day. EEM participants were asked for an additional registration for these sessions; registered participants and speakers were assigned to breakout rooms at random by a student assistant, though they also had the option to request to speak with a particular person which a few people did take up. POBAM's social rooms were largely unstructured. On the first day of the conference, only speakers and organizers were provided the passwords for lunchtime and post-talks happy hour social rooms. In the subsequent days the rooms were opened up to attendees to allow for more interaction. For all of the three colloquia, these virtual social events and group-work sessions were positively received; attendance fluctuated around 7-30 participants.

Poster session

POBAM was the only colloquium to hold a poster session. The session occupied a normal session spot in which seven 5-minute back-to-back presentations were given over the 40 minutes with no Q&A time allotted. Each presentation was accompanied by a single poster slide that the speaker would reference if they wanted. The presenters were given the opportunity to share their poster on the POBAM website ahead of the conference to generate discussion beforehand (say, on Twitter or some other medium). The break-out rooms were partly designed with follow-up discussion in mind between presenters and attendees for that day.

4. Survey Design

Aims, Research Questions

We conducted a survey to find out more about how participants and speakers experienced these four online conferences. Our aim was to answer the following general research questions:

1. Is the virtual format an acceptable *temporary replacement* for in person conferences?
2. Is the virtual format an acceptable *alternative* to in person conferences?
3. How do the two formats (live and pre-recorded) compare?
4. How should online conferences be organized?

As well as general attitudes towards and experience of the online conferences, we looked at evaluation of and preferences concerning the following elements of the conferences:

A. content delivery
B. feedback acquisition
C. networking
D. accessibility

Survey design and administration

We use data from two surveys. One survey was developed for the ECAP10, here Survey A. The other survey was developed for the three smaller colloquia, here Survey B. Survey A was administered using Qualtrics at the beginning of September 2020 (right after the conference). Survey B was administered using SoSci Survey (www.soscisurvey.de) in September 2020, between two and five months after the conferences. Both surveys consisted of multiple choice and open response questions. For simplicity, we focus on the quantitative results only.

Response rate and representativeness of sample

All people who registered for the conferences were invited to participate in the survey. There was a total of 124 participants for Survey A, around a third of all participants to the ECAP. For Survey B, the total number of participants was 99; 33 had registered for EEM, 27 for POBAM, and 38 for DSPS. This represents roughly one quarter to one third of the total registered participants for each conference. Amongst the registered

participants were presenters: 12 presenters from DSPS, 11 from POBAM, and 7 from EEM.

In terms of demographics, the samples for both Survey A and Survey B were fairly equally spread across career stage (Survey A: 30 graduate students, 50 junior faculty members, 31 senior faculty members, and 13 missing responses; Survey B: 30 graduate students, 36 junior faculty members, 21 senior faculty members, and 12 missing responses). Survey B included some additional demographic questions, including gender, location, and disability status. The spread of genders was fairly even, and respondents were primarily located in Europe and North America. Few participants reported that they had a disability. As we note below, due to low sample sizes we cannot address questions about how online conferencing affects researchers with disabilities and researchers outside traditional conferencing countries (see Supplementary Material Section 3 for full demographics).

Instrument and data availability

The items regarding participants' experiences in the workshop had a seven-point (survey A) or five-point (survey B) Likert scale response format and the items regarding the accessibility had a logical yes or no response format. Unless stated otherwise, the respondents are allowed to choose only one option for each item. The survey instruments and data from both surveys can be downloaded at https://doi.org/10.17605/OSF.IO/D7QEZ. A description of the results can be found in Supplementary Material Section 3, Tables S3.1 and S3.2.

5. Survey Results

General satisfaction

We found that the conferences were evaluated positively overall. In particular, in Survey B, general satisfaction with the conferences was on average high to very high (Survey B: Mean [M] = 4.32, Standard Deviation [SD] = 0.82, Min-Max responses = 1 – 5), as was willingness to attend future online conferences (M = 4.28, SD = 1.03, Min-Max responses = 1 – 5). Survey B found no significant differences between presenters and regular participants in terms of their satisfaction with the conference ($t(65.43) = -0.61$, $p = 0.54$), nor between participants at different career stages ($F(2, 81) = 1.90$, $p = 0.16$).

In addition, the online format was not only seen as a temporary replacement during acute crises like pandemics but as a legitimate alternative to in-person conferences (see Figure 1).

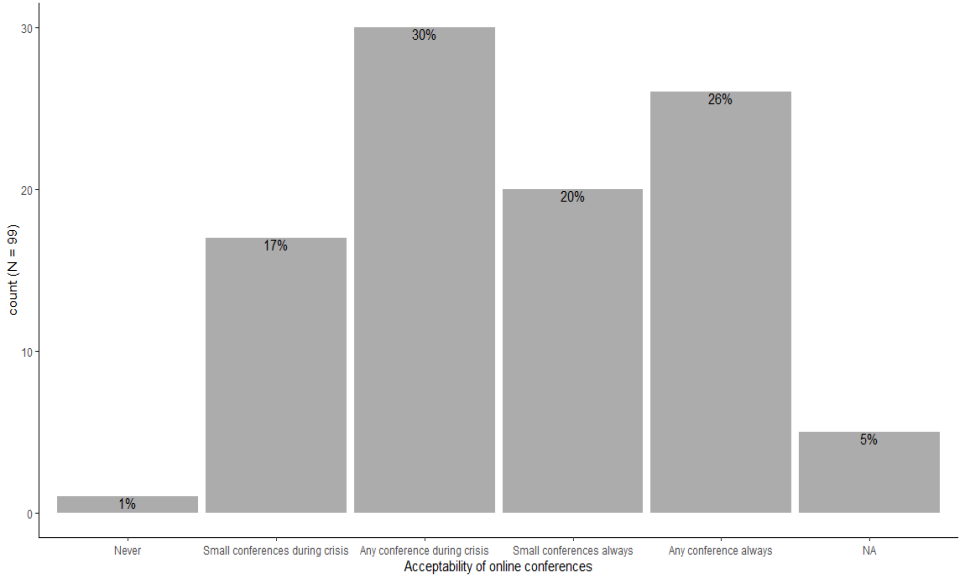

Figure 1. The frequency of participants in Survey B who agree with online conferences as an alternative format.

Experience of different aspects of the conferences

In addition to general satisfaction, data show how respondents evaluated different aspects of the conferences. Respondents reported equal levels of very high satisfaction in both surveys (pre-recorded presentations in Survey A and live presentations in Survey B) with regards to presentations ($M_A = 6.02$, $SD_A = 0.99$, $Min_A – Max_A = 1\text{-}7$; $M_B = 4.38$, $SD_B = 0.71$, $Min_B – Max_B = 1\text{-}5$) and discussions ($M_A = 5.75$, $SD_A = 1.29$, $Min_A – Max_A = 1\text{-}7$; $M_B = 3.97$, $SD_B = 0.98$, $Min_B – Max_B = 1\text{-}5$). Given the difference in the response format, we rescaled the items from 0-10 to make sure they have comparable lower and upper scores and the result of a t-test showed that the satisfaction in both surveys were equally high with regards to presentation ($t[df = 173] = -0.35$, $p = 0.73$) and discussion ($t[df = 177] = -1.44$, $p = 0.15$).

Unsurprisingly, networking suffers in online conferences. Participants responded in Survey A that opportunities to network and chat with colleagues were worse or much worse than in physical conferences

($M_{Networking}$ = 1.83, $SD_{Networking}$ = 1.23; M_{Chat} = 1.59, SD_{Chat} = 0.94, Min – Max = 1 (*much worse*) – 7 (*much better*)). This indicates that the written channels and the single pub quiz event at ECAP were not seen as sufficient for networking. The smaller conferences in Survey B included more targeted networking using break-out rooms, and it is positive to see that this seems to have improved participants' satisfaction with networking. Respondents to Survey B were not very dissatisfied, but they were still on average neither particularly satisfied nor particularly dissatisfied with the networking in the conferences (M = 2.75, SD = 1.32, Min-Max = 1-5).[3] Noteworthy to mention that there were no differences between participants in the three workshops in terms of their networking experience in the Survey B ($F(2, 90)=0.04$, $p = 0.96$).

In addition to the experience of participants, it is important to look at how presenters evaluated the experience of presenting online. In Survey A, presenters responded that communicating their work to others and getting feedback was about the same as it is in physical conferences ($M_{communicating}$ = 4.5, $SD_{communicating}$ = 1.16 $M_{feedback}$ = 4.5, $SD_{feedback}$ = 1.24, Min-Max = 1 (*much worse*) – 7 (*much better*)). In Survey B we found that presenters were on average satisfied with how their presentations went when they were live (M = 4.2, SD = 0.79, Min-Max = 1-5) but less so when their presentations were pre-recorded (M=2.67, SD = 0.58). However, this latter result is perhaps not indicative because only 3 presenters pre-recorded their presentations for the small conferences, and in general presenters were satisfied with their pre-recorded presentations at the ECAP, as seen in Survey A. Presenters were also fairly happy with the feedback they were able to get, especially for spoken feedback during Q&A sessions and in breakout rooms (M_{spoken} = 3.9, SD_{spoken} = 1.03, $M_{breakout}$ = 3.7, $SD_{breakout}$ = 1.11, Min-Max = 1-5), and to a lesser extent, but still at the mid-point of the scale, with the written feedback through the chat function ($M_{written}$ = 3, $SD_{written}$ = 1.00).

Accessibility

Another aspect to online conferences is their potential for enhanced accessibility (see subsection Accessibility). In Survey A, participants responded that the online conference accessibility was better than an in-person conference (M = 5.62, SD = 1.20, Min-Max = 1(much worse) – 7(much better)). Similarly, in Survey B, we found that 87% of participants agreed or strongly agreed that making the conference online made it easier

[3] Note that the questions about networking asked in the two surveys are not directly comparable, since Survey A asks about online networking in comparison to in-person networking, whereas Survey B asks only about satisfaction with online networking.

to attend. This positive result needs to be interpreted with caution because it was asked during a pandemic, meaning attending conferences in person would have been difficult if not impossible. Nevertheless, it is a positive indication that online conferences are accessible.

In Survey B we also asked about what factors impacted positively or negatively on participants' ability to attend the conference. Figure 2 shows the full list. Lower cost, reduced travel, and being able to attend from home were positive factors for many, whereas other work commitments, day length, and time zone hampered participation for many. It is also important to note the factors relevant to accessibility for minorities in philosophy, such as the positive impact of not having to worry about venue accessibility and the persisting negative impact of caring responsibilities. No significant differences were found between genders in terms of accessibility. As mentioned earlier, sample sizes were too low for people with disabilities and people outside North America and Europe to see if the conference being online had a positive impact for these groups.

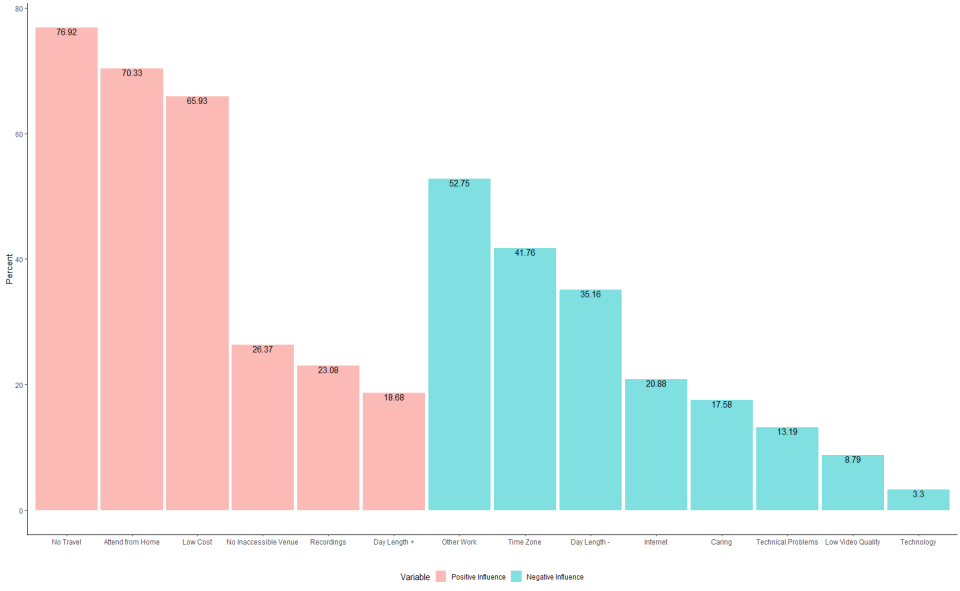

Figure 2. The percentage of participants that agree that the positive (left, pink bars) and negative (right, blue bars) factors affected their ability to participate in the conference.

Session format and scheduling

Finally, online conferences offer a range of options for formats and schedules that aren't available for in-person conferences. In Survey B we found out about participants' preferences for these different options.

For Q&A, we found that 70% of participants prefer to have the option to ask questions either through the text-based chat function or spoken out loud; text-based only or spoken only formats were far less popular (5% and 20% respectively). For networking, the most favored format is digital coffee breaks using break-out rooms. This was the format employed in the colloquia, and it is a good sign that 70% of participants were happy with it after trying it out. Group work was also a fairly popular option (50% of respondents). In addition, the text-based chat function, speed dating, and participant organized events were moderately popular options (25-30% of respondents). Perhaps most importantly, only 10% of people preferred no networking in an online conference.

Finally, there is the question of length and scheduling of sessions. The general message from our results is that online conferences need shorter sessions and shorter days. Most people prefer a keynote of 40 minutes or less (90% of all participants), and a regular presentation of under 25 minutes (55% of all participants)—not including time dedicated for Q&A. For day length, participants preferred that the schedule runs for either 2-4 hours (47%) or 4-6 hours (42%), not including any scheduled breaks. Importantly, this is much shorter than a usual in-person conference day. In addition, whereas around half the participants prefer to have the conference held over consecutive days, the other half thinks that days spread out over a week or over multiple weeks is the better format.

6. Discussion

Addressing Q1. Is the virtual format an acceptable temporary replacement for in person conferences?

Our results indicate that online conferences are in general very satisfactory and that they are accepted as a temporary replacement for in person conferences.

Addressing Q2. Is the virtual format an acceptable alternative to in person conferences?

The next step is to decide whether online conferences are acceptable not just as a temporary solution to the contact and travel restrictions during a pandemic, but as an alternative to in person conferences regardless of such acute crises. Answering this question is central to our argument that online conferences should be adopted in academic philosophy's post-corona future.

We found that online conferences seem to be a suitable alternative to in-person conferences when it comes to presentations, discussions and getting feedback, and accessibility. However, online conferences are less effective when it comes to networking, at least given current levels of familiarity with online networking using the formats we already have available. Given the importance of networking (e.g., in Survey B, networking was rated just as important as presentations and discussion), online conferences cannot be expected to totally replace all aspects of in-person conferencing.

Addressing Q3. How do the two formats (live and pre-recorded) compare?

No significant differences were found between the two data sets in terms of satisfaction with the presentations and discussions. This is perhaps a positive indication that both pre-recorded and live talks are suitable for online conferences.

Survey B did find that respondents overwhelmingly preferred live talks in comparison to pre-recorded talks (95% vs 18%; participants could choose more than one option). The preferences of participants should be taken into consideration when planning a conference. Nevertheless, it should be noted that the respondents to Survey B may not have seen any pre-recorded talks and may therefore be expressing an opinion not informed by experiences with the relevant medium.

Addressing Q4. How should online conferences be organized?

Our results indicate that online conferences should aim to have live Q&A sessions that permit both text-based and spoken contributions. Another clear indication is that days should be shorter with more frequent and longer breaks. This measure will potentially enhance accessibility, for instance for primary caregivers. In addition, shorter days and more breaks can help to combat a phenomenon known as "zoom fatigue," attributable to the particularly draining effect of social interaction in video calls due to factors such as lack of eye contact, micro-delays in audio, and even 2D representation (Lee 2020; Nadler 2020).

Aside from shorter days, the key message seems to be that online conferences can be quite flexible in their scheduling. Consecutive days, multiple days in a week or over several weeks all seem to be accepted formats, allowing conference organizers greater freedom in how they choose to schedule sessions.

A number of options are available to organizers planning networking events. Digital coffee breaks and group work, when feasible, can be recommended. But other formats such as speed dating and making use of text-based chat functions may also work. Nevertheless, networking remains a sticking point for many online conferences. There may be other formats or ways of organizing networking that make it particularly effective, which is a point for further research.

6.1. Directions for Further Work

Along with many disruptions, the pandemic created an overdue incentive for academics to experiment with online conferences. Yet they should not let this experience recede into history along with the pandemic. Because of their commitment to moral and justice principles and acceptance of the recommendations from IPCC, philosophers should lead the way to establish interdisciplinary teams with other academics from various sciences to improve the current models of online and in-person conferences with the goal of making them carbon-neutral, while preserving the desired features of conferences as mediums for exchanging and testing ideas, and forging relationships.

Regarding the result of our analysis, two cautionary remarks are warranted. First, out of those 450 participants initially registered for ECAP's physical conference, only 300 registered for the online conference. It may therefore be that those who registered for the online version of the conference and subsequently completed survey A already had a more positive attitude toward online conferencing in comparison to those who chose not to participate. Future studies might aim to conduct a more controlled survey of participants to ensure the result can be generalized to a wider circle of academics. Second, the sample size in our study was rather small. To improve explanatory power, future studies might aim for a larger sample size.

We should also note that the implicit assumption that online conferences are temporary because of the pandemic might have influenced their positive reception and the tolerance of their shortcomings. Making online conferences a permanent feature of academic life as well as incorporating online talks into traditional conferences might raise the bar for accepting

them. Interdisciplinary teams consisting of representatives of academic fields organizing conferences, psychologists, education, social and communication scientists, as well as other scientists as needed could adopt as a research project examination of extant practices of online conferences, articulation of improved models, their subsequent testing, and dissemination of best practices, which should facilitate widespread adoption of online conferences. In cooperation with those research teams, professional societies could establish platforms for communication of best practices. Furthermore, the interdisciplinary research teams and the boards of professional associations could work with specialized non-governmental and governmental organizations to identify effective ways to offset carbon emissions and to develop new ones if necessary.

For in-person conferences, professional organizations could build upon and further develop the suggestions proposed by Philosophers for Sustainability and contained in the Good Practices Guide of the American Philosophical Association, as well as those made by scientists to decrease the carbon footprint and increase the accessibility of conferences and of research (Bousema et al. 2020; Burtscher et al. 2020; Klöwer et al. 2020; Stevens et al. 2020). Several common proposals emerge from those suggestions: (1) replace in-person conferences with the online format; (2) alternate in-person with online conferences; (3) incorporate online talks in in-person conferences; (4) choose conference venues that are accessible and that would result in the lowest amount of carbon emissions, especially due to transportation; (5) mandatorily offset carbon emissions that cannot be avoided, including for online conferences. Mandatory offsetting could be achieved by incorporating it into conference registration fees.

7. Conclusion

Online conferences are a worthy alternative to in-person conferences not only in times of acute crisis, but generally. That is not to say that online conferences should be adopted exclusively as a replacement. However, we do believe online conferences should become the new normal, with in-person conferences as an alternative that must be well-justified and responsibly carried out.

The COVID 19 pandemic created the conditions for a natural experiment, shifting conferences fully online. Our survey found that the online conferences that we organized offered participants an overwhelmingly positive experience to share and engage with research. Attendees reported that the online format was more than sufficient for presentations,

discussions, and feedback, with increased accessibility and affordability, allowing scholars from institutions less financially endowed to participate. We also acknowledge some shortfalls of online conferences. Our survey highlighted that online conferences fell short in the ability of the participants to network in a fully satisfactory way. This could be because the online format is new, and networking will develop as more conferences are run online or new technologies are developed. Alternatively, it may be that the online medium is not an adequate environment for networking.

When academics opt for an in-person format, they ought to be mindful of its environmental and financial costs and its implications for accessibility. They should resort to it only when the online version is not feasible for conference goals, while taking all the possible measures to decrease the environmental cost as well as to ensure accessibility. Given that most academics accept moral and justice principles and the recommendations of the IPCC, they should end the practice of externalizing the environmental costs of conferences and adopt mandatory carbon offset measures both for in-person and for online conferences; the latter are not entirely carbon free. In addition, we should continue to strive for accessible in-person conferences through measures like accessible venues and facilities, family-friendly scheduling and visa-friendly timing of decisions. Philosophers and other academics should take the natural experiment that the pandemic brought about as an opportunity to build interdisciplinary work groups to study and establish best practices for online conferences, environmentally friendly and accessible in-person conferences, and adequate ways to offset carbon emissions.

Administrators of universities and research institutions might take the shift to online conferences as simply a justification for reducing travel funding. However, they ought to view it as a motivation to reduce environmental costs of teaching and research done at their institutions, in-person conference participation being only a part of it. Teaching and research institutions are committed to a greater common good. Engaging in activities that pursue the common good while producing pollution that threatens the wellbeing of all and especially of the vulnerable creates an inconsistency between the deeds of institutions and their stated principles. It is incumbent on the administrators of teaching and research institutions to eliminate that inconsistency.

In-person conferences, externalization of environmental costs due to professional conventions and other aspects of research and teaching, as well as minimal and rare voluntary offsets of emissions have been the default of academic practice. The three reasons for online conferences we outlined, the models of conferences of different sizes we have organized

successfully, as well as the wider recognition among academics of the environmental footprint of their research activities suggest changing the default of academic practice to online meetings, denying the externalization of environmental costs, and ensuring mandatory offsetting of unavoidable carbon emissions. In-person conferences should become rare and well justified departures from the default of the online format.

Author Contribution Statement

DC, VP, TJP, RT: Conception and design of the work, Data collection, Drafting the article, Critical revision of the article
AT, RT: Data analysis and interpretation
All: Critical revision of the article and final approval of the version to be published

Acknowledgments

Viorel Pâslaru acknowledges that the online format of the DSPS colloquium was entirely due to Zachary Piso.
Daniel Cohnitz would like to thank the local organization team of ECAP10, Brandt van der Gaast, Bjørn Jespersen, Johannes Korbmacher, Janneke van Lith, and Erik Stei, the ESAP steering committee, the ICT and Media support of Utrecht University and the Educate-It service of Utrecht University.
Rose Trappes acknowledges Marie I. Kaiser for securing funding and starting the ball rolling with organizing the EEM workshop. Many thanks also to David Lambert, Tim Nissel and Clemens Litfin for their excellent work helping to realize the online event.
T. J. Perkins thanks all involved with POBAM 2020, especially Matt Haber for his trust.
All authors would like to thank two anonymous reviewers for their comments that helped improve the article.

Ethical Approval

The ethics boards at Utrecht University, Bielefeld University and the University of Dayton deemed that the surveys did not require ethical approval.

REFERENCES

Albayrak-Aydemir, N. 2020. The hidden costs of being a scholar from the global south. *Higher Education Across Borders (LSE Blog)*. Accessed 18th September 2020. https://blogs.lse.ac.uk/highereducation/2020/02/20/the-hidden-costs-of-being-a-scholar-from-the-global-south/

Botterill, S. 2020. How to host an accessible online meeting. *Ability Net*. Accessed 18th September 2020. https://abilitynet.org.uk/news-blogs/how-host-accessible-online-meeting

Bousema, T, P. Selvaraj, A. A. Djimde, D. Yakar, B. Hagedorn, A. Pratt, D. Barret, K. Whitfield, and J. M. Cohen. 2020. Reducing the carbon footprint of academic conferences: the example of the american society of tropical medicine and hygiene. *The American Journal of Tropical Medicine and Hygiene*, 1758-1761. doi: 10.4269/ajtmh.20-1013

Broome, J. 2016, A reply to my critics. *Midwest Studies in Philosophy*, 40: 158-171. doi:10.1111/misp.12053

Buckner, C., N. Byrd, and J. Schwenkler. 2015. *The Future of Online Conferences in Philosophy*. Accessed Dec. 3, 2020. https://dailynous.com/2015/10/15/the-future-of-online-conferences-in-philosophy/

Burtscher, L., D. Barret, A. P. Borkar, V. Grinberg, K. Jahnke, S. Kendrew, G. Maffey, and M. J. McCaughrean. 2020. The carbon footprint of large astronomy meetings. *Nature Astronomy* 4 (9): 823-825.

Byrd, N. 2020. Online conferences: Some history, methods, and benefits. forthcoming in *Right Research: Modelling Sustainable Research Practices in the Anthropocene*, eds. G. Rockwell, C. Miya and O. Rossier. Open Book Publishers.

Calisi, R. M. and a Working Group Mothers in Science. 2018. Opinion: How to tackle the childcare–conference conundrum. *PNAS* 115 (12) 2845-2849. https://doi.org/10.1073/pnas.1803153115

Calzavarini, F., and M. Viola. 2020. Socially distanced, yet virtually convened: neural mechanism online's model of online conferencing. Accessed 28th August 2020. https://dailynous.com/2020/03/20/socially-distanced-yet-virtually-convened-model-online-conferencing-guest-post/

De Cruz, H. 2015. How to organize a philosophy conference. *Philosophers Cocoon*. Accessed 9th October 2020. https://philosopherscocoon.typepad.com/blog/2015/08/how-to-organize-a-philosophy-conference-or-workshop.htl

Felappi, G., A. Gregory, and H. Beebee. 2018. Guidelines for accessible conferences: A guide by the British Philosophical Association and the Society for Women in Philosophy UK. https://bpa.ac.uk/diversity/

Fleming, N. 2019. How to organize a conference that's open to everyone. *Nature* 571, S46-S47 doi: 10.1038/d41586-019-02253-9

Gardner, L. 2020. How college leaders are planning for the fall. *The Chronicle of Higher Education*. Accessed 9th October 2020. https://www.chronicle.com/article/how-college-leaders-are-planning-for-the-fall

Gunnel, M. 2020. What is Zoombombing, and how can you stop it? *How-to Geek*. Accessed 27th October 2020. https://www.howtogeek.com/667183/what-is-zoombombing-and-how-can-you-stop-it/

Humanities Indicators. 2019a. Racial/Ethnic Distribution of Degrees in Philosophy. *American Academy of Arts & Sciences.* Accessed 5th October 2020. https://www.amacad.org/humanities-indicators/higher-education/racialethnic-distribution-degrees-philosophy

Humanities Indicators. 2019b. Gender distribution of degrees in philosophy. *American Academy of Arts & Sciences.* Accessed 5th October 2020. https://www.amacad.org/humanities-indicators/higher-education/gender-distribution-degrees-philosophy

Lee, J. 2020. A neuropsychological exploration of zoom fatigue. *Psychiatric Times*. Accessed 4th December 2020. https://www.psychiatrictimes.com/view/psychological-exploration-zoom-fatigue

Khalid, F., S. Ardila-Gómez, and K. Scott. 2016. Visa headaches and inequalities in attending global health conferences. *The BMJ Opinion.* Accessed 5th October 2020. https://blogs.bmj.com/bmj/2016/11/03/visa-headaches-and-inequalities-in-attending-global-health-conferences/

Klöwer, M., D. Hopkins, M. Allen, and J. Higham. 2020. An analysis of ways to decarbonize conference travel after COVID-19. *Nature*. 583: 356-359.

Kovacs, K. 2016. Grad students' financial worries. *Inside Higher Ed*. Accessed 9th October 2020. https://www.insidehighered.com/news/2016/11/17/report-finds-many-graduate-students-are-stressed-about-finances

Minai, N. Z. 2018. Guest post: Challenges for academics in the global south—resource constraints, institutional issues, and infrastructural problems. *The Scholarly Kitchen*. Accessed 25 September 2020.

https://scholarlykitchen.sspnet.org/2018/08/16/guest-post-challenges-academics-global-south-resource-constraints-institutional-issues-infrastructural-problems/

Nadler, R. 2020. Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. *Computers and Composition*, 58, 102613.

Newsletter of ISHPSSB. 2019. Volume 30, No. 2. Issue 57. Accessed 22 October 2020. https://www.ishpssb.org/news/newsletters/196-newsletter-volume-30-no-2-issue-57-fall-2019

Sustainability and Climate Task Force. Accessed 22 October 2020. https://www.philsci.org/inclusion-outreach/sustainability-climate-task-force.html

Railton, P., M.-K. Lee, D. Michelfelder, and R. Zheng. 2020. *APA Good Practices Guide*, February 2020 Edition. https://www.apaonline.org/page/goodpracticesguide

Schwitzgebel, E., and C. D. Jennings. 2017. Women in philosophy: quantitative analyses of specialization, prevalence, visibility, and generational change. *Public Affairs Quarterly*, 31, 83-105.

St. Croix, C. 2020. The online-first model: On hosting an awesome online academic conference. Accessed 28th August 2020. https://dailynous.com/2020/03/16/online-first-model-hosting-awesome-online-academic-conference-guest-post-catharine-st-croix/

Stevens, A. R. H., S. Bellstedt, P. J. Elahi, and M. T. Murphy. 2020. The imperative to reduce carbon emissions in astronomy. *Nature Astronomy* 4(9): 843-851.

Zahneis, M. 2020. More doctoral programs suspend admissions. That could have lasting effects on graduate education. *The Chronicle of Higher Education*. Accessed 9th October 2020. https://www.chronicle.com/article/more-doctoral-programs-suspend-admissions-that-could-have-lasting-effects-on-graduate-education

Zwart, S. P. 2020. The ecological impact of high-performance computing in astrophysics. *Nature Astronomy* 4(9): 819-822.

Supplementary material can be accessed at this link: https://eujap.uniri.hr/the-online-alternative_supp/

ABSTRACTS (SAŽECI)

RATIONALITY IN MENTAL DISORDERS: TOO LITTLE OR TOO MUCH?

Valentina Cardella
University of Messina

ABSTRACT

The idea that mental illnesses are impairments in rationality is very old, and very common (Kasanin 1944; Harvey et al. 2004; Graham 2010). But is it true? In this article two severe mental disorders, schizophrenia and delusional disorder, are investigated in order to find some defects in rationality. Through the analysis of patients' performances on different tests, and the investigation of their typical reasoning styles, I will show that mental disorders can be deficits in social cognition, or common sense, but not in rationality (Sass 1992; Johnson-Laird et al. 2006; Bergamin 2018). Moreover, my claim is that psychopathological patients can also be, in some circumstances, more logical than normal controls (Kemp et al. 1997; Owen et al. 2007). From a philosophical point of view these data seem to be very relevant, because they help us to reconsider our idea of rationality, and to challenge the common way to look at sanity and mental illness.

Keywords: Rationality; schizophrenia; delusional disorder; common sense

RACIONALNOST U MENTALNIM POREMEĆAJIMA: PREVIŠE ILI PREMALO?

Valentina Cardella
University of Messina

SAŽETAK

Ideja da su mentalne bolesti poremećaji racionalnosti vrlo je stara i uobičajena (Kasanin 1944; Harvey et al. 2004; Graham 2010). No je li istinita? Ovaj članak razmatra dvije ozbiljne mentalne bolesti, shizofreniju i poremećaj deluzije, s ciljem utvrđivanja grešaka u racionalnosti. Analizom uspjeha pacijenata na različitim testovima i razmatranjem njihovog tipičnog načina zaključivanja, pokazat ću da mentalne bolesti možemo smatrati manjkavostima socijalnog spoznavanja ili zdravog razuma, ali ne i racionalnosti (Sass 1992; Johnson-Laird et al. 2006; Bergamin 2018). Nadalje, tvrdim da pacijenti u psihopatološkim stanjima u određenim okolnostima pokazuju viši stupanj logičkog zaključivanja od

kontrolne skupine (Kemp et al. 1997; Owen et al. 2007). Filozofski gledano, ovi relevantni podaci mogu nam pomoći u preispitivanju same ideje racionalnosti i uobičajenog shvaćanja duševnog zdravlja i mentalnih bolesti.

Ključne riječi: Racionalnost; shizofrenija; poremećaj deluzije; zdrav razum

DELUSIONS IN THE TWO-FACTOR THEORY: PATHOLOGICAL OR ADAPTIVE?

Eugenia Lancellotta
University of Birmingham

Lisa Bortolotti
University of Birmingham

ABSTRACT

In this paper we ask whether the two-factor theory of delusions is compatible with two claims, that delusions are pathological and that delusions are adaptive. We concentrate on two recent and influential models of the two-factor theory: the one proposed by Max Coltheart, Peter Menzies and John Sutton (2010) and the one developed by Ryan McKay (2012). The models converge on the nature of Factor 1 but diverge about the nature of Factor 2. The differences between the two models are reflected in different accounts of the pathological and adaptive nature of delusions. We will explore such differences, considering naturalist and normativist accounts of the pathological and focusing on judgements of adaptiveness that are informed by the shear-pin hypothesis (McKay and Dennett 2009). After reaching our conclusions about the two models, we draw more general implications for the status of delusions within two-factor theories. Are there good grounds to claim that delusions are pathological? Are delusions ever adaptive? Can delusions be at the same time pathological and adaptive?

Keywords: Delusions; adaptiveness; pathology, two-factor theories; delusion formation

DELUZIJE U DVOFAKTORSKOJ TEORIJI: PATOLOŠKE ILI ADAPTIVNE?

Eugenia Lancellotta
University of Birmingham

Lisa Bortolotti
University of Birmingham

SAŽETAK

U ovom članku razmatramo kompatibilnost dvofaktorske teorije s dvjema tvrdnjama: da su deluzije patološke i da su deluzije adaptivne. U središte razmatranja stavljamo dva nedavna i utjecajna modela dvofaktorske teorije: model Maxa Colthearta, Petera Menziesa i Johna Suttona (2010) i model Ryana McKayja (2012). Modeli se podudaraju u shvaćanju naravi Faktora 1, ali razilaze se u razumijevanju naravi Faktora 2. Razlike između modela odražavaju se u različitim objašnjenjima patološke i adaptivne naravi deluzija. Istražit ćemo navedene razlike s obzirom na naturalistička i normativistička objašnjenja patoloških svojstava i fokusirajući se na sudove o adaptivnosti utemeljene na „shear-pin" hipotezi (McKay i Dennett 2009). Nakon donošenja zaključaka o dvama modelima, dolazimo do općenitijih implikacija o položaju deluzija u dvofaktorskim teorijama. Postoje li prihvatljiva uporišta za tvrdnju da su deluzije patološke? Jesu li deluzije ikada adaptivne? Mogu li deluzije istovremeno biti i patološke i adaptivne?

Ključne riječi: Deluzije; adaptivnost; patologija; dvofaktorske teorije; formiranje deluzija

EXPRESSIVISM ABOUT DELUSION ATTRIBUTION

Sam Wilkinson
University of Exeter

ABSTRACT

In this paper, I will present and advocate a view about what we are doing when we attribute delusion, namely, say that someone is delusional. It is an "expressivist" view, roughly analogous to expressivism in meta-ethics. Just as meta-ethical expressivism accounts for certain key features of moral discourse, so does this expressivism account for certain key features of delusion attribution. And just as meta-ethical expressivism undermines factualism about moral properties, so does this expressivism, if correct,

show that certain attempts to objectively define delusion are misguided. I proceed as follows. I start by examining different attempts at defining delusion, separating broadly psychiatric attempts from epistemic ones. I then present a change of approach, according to which we question whether the term "delusion" is in the business of (merely) describing reality. I then support this proposal, first, by borrowing standard lines of argument from meta-ethics (including ontological reluctance, intrinsic motivation, and deep disagreement) but also, by inference to the best explanation of some the features we see when we try to theorise about delusion (namely that it is hard to define, and that our delusion attributions are elicited by a plurality of norms).

Keywords: Delusion attribution; expressivism; non-factualism; epistemic norms; folk epistemology

EKSPRESIVIZAM U POGLEDU PRIPISIVANJA DELUZIJA

Sam Wilkinson
University of Exeter

SAŽETAK

U ovom ću članku predstaviti i braniti gledište o tome što činimo kad pripisujemo deluzije, tj. kada kažemo da netko pati od deluzija. Radi se o "ekspresivističkom" gledištu koje ugrubo odgovara ekspresivizmu u metaetici. Kao što metaetički ekspresivizam objašnjava neke od središnjih značajki moralnog diskursa, tako i ovaj ekspresivizam objašnjava neke od središnjih značajki pripisivanja deluzija. I kao što metaetički ekspresivizam dovodi u pitanje činjenično shvaćanje moralnih svojstava, tako i ovaj ekspresivizam, ako je točan, pokazuje da su neki pokušaji da se deluzije definiraju objektivno na krivome tragu. Članak je strukturiran na sljedeći način. Počinjem od proučavanja različitih pokušaja definiranja deluzija, pri čemu psihijatrijske pokušaje razdvajam od epistemičkih. Zatim predstavljam drugačiji pristup, prema kojem propitujemo nastoji li se pojmom "deluzije" (samo) opisati realnost. Dajem podršku ovom prijedlogu, najprije pozivajući se na standardno argumentiranje iz metaetike (koje uključuje ontološko opiranje, intrinzičnu motivaciju i dubinsko neslaganje), ali i zaključivanjem na najbolje objašnjenje o nekim značajkama koje možemo primijetiti kada teorijski obrađujemo deluzije (da ih je teško definirati i da je pripisivanje deluzija vođeno različitim normama).

Ključne riječi: pripisivanje deluzije; ekspresivizam; ne-faktualizam; epistemičke norme; pučka epistemologija

TOO MUCH OR TOO LITTLE? DISORDERS OF AGENCY ON A SPECTRUM

Valentina Petrolini
University of the Basque Country (UPV/EHU)

ABSTRACT

Disorders of agency could be described as cases where people encounter difficulties in assessing their own degree of responsibility or involvement with respect to a relevant action or event. These disturbances in one's sense of agency appear to be meaningfully connected with some mental disorders and with some symptoms in particular—i.e. auditory verbal hallucinations, thought insertion, pathological guilt. A deeper understanding of these experiences may thus contribute to better identification and possibly treatment of people affected by such disorders. In this paper I explore disorders of agency to flesh out their phenomenology in more detail as well as to introduce some theoretical distinctions between them. Specifically, I argue that we may better understand disorders of agency by characterizing them as dimensional. In §1 I explore the cases of Auditory Verbal Hallucinations (AVH) and pathological guilt and I show that they lie at opposite ends of the agency spectrum (i.e. hypoagency versus hyperagency). In §2 I focus on two intermediate cases of hypo- and hyper-agency. These are situations that, despite being very similar to pathological ones, may be successfully distinguished from them in virtue of quantitative factors (e.g. duration, frequency, intensity). I first explore the phenomenon of mind wandering as an example of hypoagency, and I then discuss the phenomenon of false confessions as an example of hyperagency. While cases of hypoagency exemplify situations where people experience their own thoughts, bodies, or actions as something beyond their control, experiences of hyperagency provide an illusory sense of control over actions or events.

Keywords: Agency; auditory verbal hallucinations; guilt; mind wandering; false confessions

PREVIŠE ILI PREMALO? POREMEĆAJI DJELOVANJA NA SPEKTRU

Valentina Petrolini
University of the Basque Country (UPV/EHU)

SAŽETAK

Poremećaji djelovanja mogu se opisati kao slučajevi u kojima ljudi nailaze na teškoće u procjeni svojeg stupnja odgovornosti ili sudjelovanja u nekom djelovanju ili događaju. Ove se smetnje u nečijem osjećaju djelovanja čine na značajan način povezane s mentalnim poremećajima, a posebno s nekim simptomima – auditornim halucinacijama verbalnog tipa, umetanjem misli i patološkom krivnjom. Bolje shvaćanje ovih iskustava moglo bi poboljšati prepoznavanje, a možda i tretman ljudi koji pate od takvih poremećaja. U ovom članku proučavam poremećaje djelovanja kako bih detaljnije pojasnila njihovu fenomenologiju te između njih uvela neka teorijska razlikovanja. Konkretnije, tvrdit ću da ćemo poremećaje djelovanja možda moći bolje razumjeti ako ih okarakteriziramo kao dimenzionalne. U prvom dijelu proučavam slučajeve auditornih halucinacija verbalnog tipa (AVH) i patološke krivnje te pokazujem da leže na suprotnim krajevima spektra djelovanja (hipoagencija naspram hiperagencije). U drugom dijelu bavim se dvama međuslučajevima hipo i hiperagencije. Iako su vrlo slične patološkima, ove se situacije mogu uspješno razlučiti od njih na temelju kvantitativnih faktora (poput trajanja, frekvencije i intenziteta). Prvo istražujem fenomen lutanja misli kao primjer hipoagencije, a zatim razmatram fenomen lažnih ispovijesti kao primjer hiperagencije. Dok se slučajevi hipoagencije odnose na situacije u kojima ljudi osjećaju da su njihove misli, tijela i djelovanja izvan njihove kontrole, iskustva hiperagencije pružaju varljiv osjećaj kontrole nad djelovanjem i događajima.

Ključne riječi: djelatništvo; auditorne halucinacije verbalnog tipa; krivnja; lutanje misli; lažne ispovijesti

RATIONALITY, IRRATIONALITY AND IRRATIONALISM IN THE ANTI-INSTITUTIONAL DEBATE IN PSYCHIATRY AROUND THE SECOND HALF OF THE 1970S IN ITALY

Matteo Fiorani
University of Rome, Tor Vergata

ABSTRACT

The movements and protests of 1968 worldwide criticized the traditional idea of normality. From the 1970s onwards, psychiatry and antipsychiatry became an ideological battleground centered on the boundaries between normality and madness. In this scenario, characterized by a deep cultural and political transformation within the Left, the traditional concept of rationality and its very connection with irrationality was called into question. As a consequence, the very ideal of reason was questioned. This paper will explore the debate on rationality, irrationality and irrationalism within the so-called anti-institutional psychiatry and its reception in the Italian New Left during the second half of the 1970s.

Keywords: Antipsychiatry; psychiatric reforms; New Left; Italy

RACIONALNOST, IRACIONALNOST I IRACIONALIZAM U ANTI-INSTITUCIONALNOJ RASPRAVI U PSIHIJATRIJI OKO DRUGE POLOVINE 1970-ih U ITALIJI

Matteo Fiorani
University of Rome, Tor Vergata

SAŽETAK

Pokreti i prosvjedi koji su se 1968. odvijali širom svijeta, kritizirali su tradicionalnu ideju normalnosti. Od 1970-ih nadalje, psihijatrija i antipsihijatrija postali su ideološko bojište usredotočeno na granice između normalnosti i ludila. U ovom scenariju, koji karakterizira duboka kulturna i politička transformacija na Ljevici, tradicionalni pojam racionalnosti i njegova povezanost s iracionalnošću dovedeni su u pitanje. Kao posljedica toga, doveden je u pitanje sam ideal razuma. Ovaj će rad istražiti raspravu o racionalnosti, iracionalnosti i iracionalizmu unutar takozvane antiinstitucionalne psihijatrije i njezinu recepciju na talijanskoj Novoj ljevici tijekom druge polovice 1970-ih.

Ključne riječi: Antipsihijatrija; psihijatrijske reforme; Nova ljevica; Italija

A DESIRABLE CONVULSIVE THRESHOLD.SOME REFLECTIONS ABOUT ELECTROCONVULSIVE THERAPY (ECT)

Emiliano Loria
Sapienza – University of Rome

ABSTRACT

Long-standing psychiatric practice confirms the pervasive use of pharmacological therapies for treating severe mental disorders. In many circumstances, drugs constitute the best allies of psychotherapeutic interventions. A robust scientific literature is oriented on finding the best strategies to improve therapeutic efficacy through different modes and timing of combined interventions. Nevertheless, we are far from triumphal therapeutic success. Despite the advances made by neuropsychiatry, this medical discipline remains lacking in terms of diagnostic and prognostic capabilities when compared to other branches of medicine. An ethical principle remains as the guidance of therapeutic interventions: improving the quality of life for patients. Unfortunately, psychotropic drugs and psychotherapies do not always result in an efficient remission of symptoms. In this paper I corroborate the idea that therapists should provide drug-resistant patients with every effective and available treatment, even if some of such interventions could be invasive, like Electroconvulsive Therapy (ECT). ECT carries upon its shoulders a long and dramatic history that should be better investigated to provide new insights. In fact, ECT has attracted renewed interest in recent years. This is due to the fact that antidepressant drugs in younger patients show often scarce effectiveness and unpleasant side-effects. Moreover, I show that, thanks to modern advances, ECT may work as a successful form of treatment for specific and rare cases, such as severe depression (with suicide attempts) and catatonia.

Keywords: ECT; neuroendocrinology; psychopharmacology; history of child psychiatry

POŽELJAN KONVULZIVNI PRAG. NEKA RAZMIŠLJANJA O ELEKTROKONVULZIVNOJ TERAPIJI (EKT)

Emiliano Loria
Sapienza – University of Rome

SAŽETAK

Dugogodišnja psihijatrijska praksa potvrđuje sveprisutnu uporabu farmakoloških terapija za liječenje teških mentalnih poremećaja. U mnogim okolnostima, lijekovi predstavljaju najbolje saveznike psihoterapijskih intervencija. Mnogo znanstvene literature usmjereno je na pronalaženje najboljih strategija za poboljšanje terapijske učinkovitosti kroz različite načine i vrijeme kombiniranih intervencija. Ipak, daleko smo od trijumfalnog terapijskog uspjeha. Unatoč napretku koji je postigla neuropsihijatrija, ovoj medicinskoj disciplini i dalje nedostaju dijagnostičke i prognostičke sposobnosti u usporedbi s drugim granama medicine. I dalje ostaje važeće etičko načelo za vođenje terapijskih intervencija prema kojemu je cilj poboljšanje kvalitete života pacijenata. Nažalost, psihotropni lijekovi i psihoterapije nisu uvijek uspješni za ublažavanje simptoma. U ovom radu potkrepljujem ideju da bi terapeuti trebali pružiti pacijentima otpornima na lijekove svaki učinkovit i dostupan tretman, čak i ako bi neke od takvih intervencija mogle biti invazivne, poput elektrokonvulzivne terapije (ECT). ECT na svojim plećima nosi dugu i dramatičnu povijest koju bi trebalo bolje istražiti kako bi se dobili novi uvidi. Zapravo, ECT posljednjih godina privlači sve veći interes zbog činjenice da antidepresivi kod mlađih bolesnika nisu učinkoviti i često imaju neugodne nuspojave. Štoviše, pokazujem da, zahvaljujući modernom napretku, ECT može djelovati kao uspješan oblik liječenja za specifične i rijetke slučajeve, poput teške depresije (s pokušajima samoubojstva) i katatonije.

Ključne riječi: ECT; neuroendokrinologija; psihofarmakologija; povijest dječje psihijatrije

THE ONLINE ALTERNATIVE: SUSTAINABILITY, JUSTICE, AND CONFERENCING IN PHILOSOPHY

Rose Trappes
Bielefeld University

Daniel Cohnitz
Utrecht University

Viorel Pâslaru
University of Dayton

T. J. Perkins
University of Utah

Ali Teymoori
Helmut Schmidt University

ABSTRACT

The recent global pandemic has led to a shift to online conferences in philosophy. In this paper we argue that online conferences, more than a temporary replacement, should be considered a sustainable alternative to in-person conferences well into the future. We present three arguments for more online conferences, including their reduced impact on the environment, their enhanced accessibility for groups that are minorities in philosophy, and their lower financial burdens, especially important given likely future reductions in university budgets. We also present results from two surveys of participants who attended one large and three small online philosophy conferences this year. We show that participants were in general very satisfied with presentations and discussions at the conferences, and that they reported greater accessibility. This indicates that online conferences can serve as a good alternative to in-person conferences. We also find that networking was less satisfactory in online conferences, indicating a point for improvement and further research. In general, we conclude that philosophers should continue to organize online conferences after the pandemic. We also provide some advice for those wishing to organize online conferences.

Keywords: Online conferences; accessibility; carbon footprint; carbon offsetting; inclusivity; minorities in philosophy

ONLINE ALTERNATIVA: ODRŽIVOST, PRAVEDNOST I ODLAZAK NA FILOZOFSKE KONFERENCIJE

Rose Trappes
Bielefeld University

Daniel Cohnitz
Utrecht University

Viorel Pâslaru
University of Dayton

T. J. Perkins
University of Utah

Ali Teymoori
Helmut Schmidt University

SAŽETAK

Nedavna globalna pandemija dovela je do prelaska na online konferencije u filozofiji. U ovom radu tvrdimo da bi se online konferencije, više od privremene zamjene, trebalo smatrati održivom alternativom konferencijama uživo i u budućnosti. Predstavljamo tri argumenta za održavanje više internetskih konferencija, među njima navodimo njihov smanjeni utjecaj na okoliš, poboljšanu dostupnost manjinskim skupinama u filozofiji te njihovu smanjenu financijsku opterećenost, posebno važnu s obzirom na vjerojatna buduća smanjenja proračuna kojima upravljaju sveučilišta. Također predstavljamo rezultate dviju anketa sudionika koji su ove godine prisustvovali na jednoj velikoj i tri male online konferencije iz filozofije. Pokazujemo da su sudionici općenito bili vrlo zadovoljni prezentacijama i raspravama na konferencijama te da su izvijestili o većoj dostupnosti. To ukazuje da online konferencije mogu poslužiti kao dobra alternativa konferencijama uživo. Također smo otkrili da je umrežavanje bilo manje zadovoljavajuće na online konferencijama, što ukazuje na potrebu za poboljšanjem i daljnjim istraživanjima. Općenito zaključujemo da bi filozofi trebali nastaviti

organizirati online konferencije nakon pandemije. Također pružamo nekoliko savjeta onima koji žele organizirati online konferencije.

Ključne riječi: Online konferencije; dostupnost; ugljični trag; ugljično nadoknađivanje; inkluzivnost; manjine u filozofiji

Translated by Lovro Savić (Oxford), Viktor Ivanković (Zagreb), and Marko Jurjako (Rijeka). Proofread by Martina Blečić (Rijeka)

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- be sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).
- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, R. 2006. The Epistemic Role of Testimony: Internalist and Externalist Perspectives. In *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa, 77-92. Oxford: Oxford University Press.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, J. and E. Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. Normative requirements. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, J. 2010. Sentimentalism: Its Scope and Limits. *Ethical Theory and Moral Practice* 13: 125-136.
http://www.jstor.org/stable/40602550

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013.
http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 30

institutions € 40

Croatia:

individuals 100 kn

institutions 150 kn

Bank: Zagrebačka banka d.d. Zagreb

SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455

Only for subscribers from Croatia,

please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed and abstracted in The Philosopher's Index, European Reference Index for the Humanities (ERIH PLUS), Directory of Open Access Journals (DOAJ), PhilPapers, and Portal of Scientific Journals of Croatia (HRČAK)