

EUROPEAN JOURNAL OF  
ANALYTIC PHILOSOPHY

UDC 101

ISSN (Print) 1845-8475

ISSN (Online) 1849-0514

<https://doi.org/10.31820/ejap>

Open access



### *Editors*

Luca Malatesti

University of Rijeka, lmalatesti@ffri.hr

Majda Trobok

University of Rijeka, trobok@ffri.hr

### *Editor-in-Chief-Elect*

Marko Jurjako

University of Rijeka, mjurjako@uniri.hr

### *Assistant editors*

Viktor Ivanković

Institute of Philosophy, Zagreb

Lovro Savić

University of Oxford

### *Editorial administrator*

Mia Biturajac

University of Rijeka

### *Editorial board*

Lisa Bortolotti (University of Birmingham), Anneli Jefferson (Cardiff University), James W. Lenman (The University of Sheffield), Luca Malatesti (University of Rijeka), Alfred Mele (Florida State University), Carlo Penco (University of Genoa), Katrina Sifferd (Elmhurst College), Majda Trobok (University of Rijeka), Rafał Urbaniak (University of Gdansk)

### *Advisory board*

Miloš Arsenijević (University of Belgrade), Elvio Baccarini (University of Rijeka), Carla Bagnoli (University of Modena and Reggio Emilia), Boran Berčić (University of Rijeka), Clotilde Calabi (University of Milan), Mario De Caro (University of Rome), Raphael Cohen-Almagor (University of Hull, UK), Jonathan Dancy (University of Reading/University of Texas, Austin), Mylan Engel (University of Northern Illinois), Katalin Farkas (Central European University), Luca Ferrero (University of California, Riverside), Paul Horwich (City University New York), Pierre Jacob (Institut Jean Nicod, Paris), Kerry McKenzie (University of California, San Diego), Kevin Mulligan (University of Geneva), Snježana Prijić-Samaržija (University of Rijeka), Michael Ridge (University of Edinburgh), Sally Sedgwick (Boston University), Mark Timmons (University of Arizona, Tucson), Nicola Vassallo (University of Genoa), Bruno Verbeek (University Leiden), Alberto Voltolini (University of Turin), Joan Weiner (Indiana University Bloomington), Timothy Williamson (University of Oxford), Jonathan Wolff (University College London)

### *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: [ejlap@ffri.hr](mailto:ejlap@ffri.hr)

*Web address: <https://www.eujap.uniri.hr>*

Printed by Impress, Opatija (Croatia)

30 copies

## TABLE OF CONTENTS

DOUBLE VISION, PHOSPHENES AND AFTERIMAGES: NON-ENDORSED REPRESENTATIONS RATHER THAN NON-REPRESENTATIONAL QUALIA Işık Sarihan.....	5
KNOWLEDGE AND ASSERTION: A CRITIQUE OF LACKEY Joshua Anderson.....	33
RIGHT TO BE PUNISHED? Adriana Placani and Stearns Broadhead.....	53
CYBERBULLYING, MORAL RESPONSIBILITY, AND SOCIAL NETWORKING: LESSONS FROM THE MEGAN MEIER TRAGEDY Raphael Cohen-Almagor.....	75
INFERENCE TO THE BEST EXPLANATION: THE CASE OF POTENTIAL ENERGY Peter J. Riggs.....	99
BOOK REVIEW: John McMillan, THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS Iva Martinić.....	117
BOOK REVIEW: Dražen Pehar, PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST- DAYTON Višeslav Raos.....	123
<b>ABSTRACTS (SAŽECI) .....</b>	<b>127</b>
<b>AUTHOR GUIDELINES AND MALPRACTICE STATEMENT.....</b>	<b>135</b>



## DOUBLE VISION, PHOSPHENES AND AFTERIMAGES: NON-ENDORSED REPRESENTATIONS RATHER THAN NON-REPRESENTATIONAL QUALIA

Işık Sarıhan  
Central European University

Original scientific article – Received: 13/08/2019 Accepted: 24/1/2020

### ABSTRACT

*Pure representationalism or intentionalism for phenomenal experience is the theory that all introspectible qualitative aspects of a conscious experience can be analyzed as qualities that the experience non-conceptually represents the world to have. Some philosophers have argued that experiences such as afterimages, phosphenes and double vision are counterexamples to the representationalist theory, claiming that they are non-representational states or have non-representational aspects, and they are better explained in a qualia-theoretical framework. I argue that these states are fully representational states of a certain kind, which I call “automatically non-endorsed representations”, experiential states the veridicality of which we are almost never committed to, and which do not trigger explicit belief or disbelief in the mind of the subject. By investigating descriptive accounts of afterimages by two qualia theorists, I speculate that the mistaken claims of some anti-representationalists might be rooted in confusing two senses of the term “seeming”.*

**Keywords:** Perception; representationalism; qualia; non-conceptual content; afterimages; double vision

## 1. Introduction

This paper deals with the question of the nature of experiential phenomena such as afterimages, double vision and phosphenes. When you close your eyes and apply pressure on them, you can experience a swirling array of colored dots or patches. Is this a representational mental state? Is it possible to say that there is a way that the world visually appears to you while having this experience, that is, does the experience consist in a visual appearance that can be judged to be truthful or misleading about the visible environment? Does the experience, as it were, tell you anything true or false about reality, does it have a truth-evaluable content? When you introspect, do you become aware of this experience by being aware of how the world qualitatively appears to be, a way which it may or may not be, or do you become aware of it by becoming aware of something else, say, some actual qualities of the mental state itself that cannot be spelled out as the qualities that the world appears to have? Or consider the visual phenomenon of double vision: Is the difference between “single” and double vision a difference about the representational content of these experiences, that is, a difference about how things look? If not, what is the difference?

One might think that the answer is straightforward: The difference between double and ordinary vision is a difference between how things visually appear to be. In double vision, objects visually look doubled in an unusual way, even though we do not take them to be doubled in the great majority of cases. In the case where I rub my eyes and undergo a phosphene experience, my vision suggests me a two-dimensional ghostly world of chaotic colors that exist slightly behind my eyes, even though I do not believe that there are such colors somewhere in there or out there. Some reasonable philosophers disagree with this account. They claim that such experiences, or some aspects of these experiences, cannot be spelled out in terms of appearance or representation.<sup>1</sup>

---

<sup>1</sup> In philosophical literature, the term “appears” and its cognates have also taken on a separate sense which is distinct from the sense expressed by the term’s use in the context of an appearance-reality distinction. Such uses of the term do not necessarily imply the experiential state’s having truth-evaluable content, rather, the term is used to capture “phenomenal character”. Throughout this paper, I use “appear” and similar terms in the sense that implies a truth-evaluable state, as in “that box appears white to Agnes” (“X appears to have the property Y to subject Z”). Perhaps there is a coherent understanding of phenomenal character distinct from content, even though I doubt that there is, but in any case, my preference regarding the usage of the term confined to its narrow sense shall be taken as a mere terminological preference rather than a theoretical attitude towards how the term should be used. But for an argument that phenomenal character logically implies content, see Siegel (2010).

Representationalism or intentionalism for phenomenal experience, in its stronger variety, is the theory that all introspectible qualitative aspects of a conscious experience are qualities that the experience non-conceptually represents the world to have.<sup>2</sup> Some philosophers have argued, on the basis of introspective examination, that there are certain experiential states which are not fully representational, with afterimage experiences being the paradigmatic example of such states (Boghossian and Velleman 1989; Block 1996; Kind 2008). According to the anti-representationalist line of analysis, such states, or some aspects of these states, do not represent the world to be in a certain way. They do not have truth-evaluable content or intentional objects. It cannot be said that they are “about something” or that they are true or false. And as we can be aware of these mental states by introspection, the properties in virtue of which we are aware of them should be non-representational or non-intentional properties, and then the point about these special experiences are generalized to all experiences by some argumentative move or other. Therefore, the argument goes, representationalism about experience is false: There is more to experience than representation, and this can be revealed by an introspective analysis of the phenomenology of certain experiences. Introspection allegedly reveals “qualia” besides the representational facts about these experiences.<sup>3</sup> I will assume that the reader has some acquaintance with the concept of a quale and its place in contemporary philosophy of mind. I will just leave it at mentioning that here we are dealing with qualia in the restricted sense of the term that applies to mental qualities<sup>4</sup> that are concretely instantiated while undergoing an experience, and not with the

---

<sup>2</sup> For seminal examples of representationalism, see Byrne (2001), Thau (2002), and Tye (2000). Representationalists of the “impure” sort (see Chalmers 2004 for the distinction) hold that the qualitateness of a phenomenal experience is not only determined by the content of a representation but also by its “mode” or “manner”. The final verdict on pure-versus-impure representationalism has no bearing on the debate dealt within this paper.

<sup>3</sup> In the times before there was a debate between qualia theory and representationalism, these revealed properties would be classified as “sensations” or “properties of a sensory field” (Peacocke 1983) or “sense-data” (Moore 1939; Jackson 1977).

<sup>4</sup> It has been suggested by an anonymous reviewer that “mental qualities that are concretely instantiated while undergoing an experience” is too weak to be rejected by representationalists. This would be the case if this characterization mentioned the instantiation of “properties” rather than “qualities”. According to representationalism, experience instantiates properties for sure, such as the property of representing something. The term “quality” is chosen here particularly to narrow down the range of properties instantiated in an experience that are acceptable by the representationalist, as non-representationalist theories rely on the actual instantiations of a quality to explain the qualitative character of experiences, such as qualia, or qualities of sense-data, or qualities of the objects in the actual environment. According to representationalism, the mere appearance of such qualities, without any actual quality concretely existing, explains the qualitateness of experience.

term “qualia” that is sometimes used as a placeholder for whatever accounts for the qualitative aspect of experience.<sup>5</sup>

In this paper, I argue that the experiential states that are brought forward as examples in such arguments for qualia are not really non-representational states, but they are a sub-class of fully representational states. These are representational states the contents of which we do not endorse at the cognitive level, and their qualitative phenomenology is exhausted by the facts about *how things appear*, that is, by facts about what qualities show up in the intentional or representational content of the experience. By examining the claims of two defenders of the qualia view, Ned Block (1996) and Amy Kind (2008), I will try to show that the invalid argument against representationalism rests on confusing non-representationality with what I want to call “non-endorsement of mental representations”. That is, these authors themselves describe such states as if they were non-endorsed representations, before making the logically illegitimate move that they are non-representational. I will also speculate about the causes of the confusion. I will suggest that the content of the experiential states deployed in such arguments differ from other non-endorsed representations in being either obviously non-veridical or being impossible, and so they lose their belief-inducing function, becoming automatically non-endorsed states. This makes it harder to see how they could be true or false in the first place, as we almost never have environment-directed beliefs, or even explicit disbeliefs, triggered by these states. When this is coupled with an ambiguity in our appearance-talk, some of us are lead to the confused conclusion that such states are non-representational. When we are helped out of the confusion, we see that there is no need to postulate qualia to analyze such states. There are, for sure, various arguments in defense of qualia, and the appeal to afterimages and the like is only one such argument. Qualia theorists, in general, can be, and often are, fine with the idea that there are representational aspects of experience and there can be a close relation between qualia and the representational content. But arguments that appeal to afterimages and the like rest on a specific phenomenological analysis of these experiences which aim to show that these experiences are (partly) non-representational. So these particular arguments for qualia shall be rendered unmotivated with the help of a full representationalist analysis of these experiences.

---

<sup>5</sup> According to this second meaning of “qualia”, qualia exist according to all theories of experience such as representationalism, direct realism and sense-data theory, though in such cases the term might constitute a case of terminological inflation, as talking of “experienced qualities” would most probably suffice.

Before we move on, it should be noted that to demonstrate that a mental state is representational, one doesn't need to demonstrate that the represented qualities are represented as environmental, material or mind-independent, or external to the body or to the subject. It is possible that these states represent something as immaterial, mind-dependent or mental (whatever "being represented as mental" may amount to), or as internal to the body or to the subject. They can represent objects as internal to the head, like right behind the eyes or on the eyes, in the ear, or somewhere in the head (as in the experience of internal speech and inserted thoughts). Representation is not only a matter of representing things in the external physical environment. As Byrne puts is:

The subject [of an afterimage experiment] attends to the world as it appears to her, just as she did in the initial experiment [involving colored chips.] ... According to some, it appears to the subject that a red filmy thing is some indeterminate distance from her eyes. According to others, it doesn't appear to the subject that the object is in her physical environment at all: Instead it appears that the object is in some inner realm. However, this does not affect the point that the subject can only discover the phenomenal character of her experience by attending to the world (either external or internal) as her experience represents it. (Byrne 2001, 13)

Also, it needn't be the case that these representational states represent objects in the ordinary sense of the term "object". It can be that they represent properties like color and shape without these properties being bound together into what some would call proper objects, so the properties end up being represented as properties of "areas" or "patches".<sup>6</sup> These states would still have intentional objects understood in the sense of "object" expressed by the term as it is used in a representational context, while these intentional objects may not be proper "objects" in the other sense of the term<sup>7</sup> – they can belong to ontological categories such as

---

<sup>6</sup> Also see the related discussion in Schroer (2004).

<sup>7</sup> These intentional objects might even be qualities instantiated in an experience. So even if qualia theorists were right in analyzing experiences as states where we *seem to find* qualities instantiated in experience, there would still be some theoretical space to say that experiences are representational states, where *our experience represent an experience as having a certain quality* (see Thau 2002 for a similar argument). And it would require a separate argument for the qualia theorist to demonstrate that we should reject the representationalist analysis of such states in favor of an analysis that relies on concretely instantiated mental qualities that constitute such experiences.

properties and events.<sup>8</sup> In any case, I will be arguing for the position that, whatever the other details about these representational states are, the properties being represented in visual phenomena like phosphene and afterimage experiences are those like color, shape and location which we also find in more ordinary visual states.

One last terminological point I should note is that the term “representation” that I rely on throughout this paper does not necessarily rest on an externalist or relational understanding of the term found in accounts that analyze experiential representation as requiring some causal-informational relation to a concrete entity that is the represented object or property.

The discussions in this paper concern a dispute between representationalists and qualia theorists, without touching the disagreement between representationalist and anti-representationalists like direct realists and sense-data theorists. Some of my assumptions will not be argued for and will be just taken for granted, and that is because it is an uncontroversial point as far as the debate between the representationalists and the qualia theorists is concerned, even though it might be found controversial by the direct realists and friends of sense-data. Nothing said here is aimed at persuading them, though perhaps the representational analysis of afterimage experiences and the like can provide them examples to get a grip on how experience in general can coherently be analyzed as a representational phenomenon.

In the next section, I elucidate the concept of a non-endorsed representation, and in section three, I introduce the concept of an automatically non-endorsed representation. In section four I offer an analysis of double vision as an automatically non-endorsed representation, and in section five I do the same for afterimages and phosphenes. In section six I borrow an argument from David Bourget (2015) to help demonstrate further the representational aspect of the aforementioned visual experience types. In section seven, I analyze passages from Block (1996) and Kind (2008), and try to show that their anti-representational arguments fail and that the failure is due to a confusion. Section eight investigates the psychological and linguistic roots of this confusion, pointing to an ambiguity in our appearance-talk and to special features of afterimages and the like that make them different from other illusory experiences that are not deployed for anti-representationalist arguments. The ninth section concludes the

---

<sup>8</sup> See Phillips for a list of features of afterimages which, according to anti-representationalists, makes afterimages “incompatible with [...] being apparent presentations of public objects” (2013, 418).

discussion, briefly hinting at similar strategies to resolve the debate about other allegedly non-representational states like pain and orgasm experiences.

## 2. Non-Endorsed Representations

Hereby I introduce the term “non-endorsed representational content”: A mental state has a non-endorsed representational content if it has truth-evaluable content but the subject doesn’t take the content as true on a higher, cognitive level.<sup>9</sup> This doesn’t mean that the subject necessarily takes the content to be false. As we will see, there are cases of non-endorsed representations which are neither taken as true nor taken as false, states that have completely lost their automatic belief (or disbelief) inducing role and also probably have lost most of their other causal roles that produce intentional behavior. Maybe it would be better to coin two different terms for these two different phenomena, to capture the difference between, on the one hand, a subject explicitly disbelieving the content of the experience, and, on the other hand, neither explicitly believing nor disbelieving it, but my terminological aims will be less ambitious. And perhaps there can be a non-conceptual feature of phenomenal experience that is something like a pre-cognitive analogue of non-endorsing, but for the purpose of this paper, we need not enter that territory.

Let me make clearer what I mean by “endorsement”. There is a type of endorsement which is directed at the representation itself, which requires the subject to understand that she is in a representational state, and to hold the belief that this state represents the world veridically.<sup>10</sup> This kind of endorsement is presumably limited to adult human beings and perhaps some animals that have the concept of representation or the concept of an appearance-reality distinction. However, there is a more basic type of endorsement, which can also be applied to non-human animals and human infants. This kind of endorsement can be simply defined by a way that the world looks to a subject, and given no disturbance from an appearance-reality distinction at the cognitive level, what appears to exist is the same as what is taken to exist.

---

<sup>9</sup> The concept of a non-endorsed representation has some parallels with the concept of an alief (Gendler 2008), but aims to capture a phenomenon at the experiential level.

<sup>10</sup> Of course, a subject who is not a philosopher or a cognitive scientist by profession does not think about her experiences *via* technical terms like “representation”, “content”, or “veridicality”, but can nevertheless think of appearances and think about whether things are as they appear or not.

I will leave it open whether creatures without higher-order beliefs directed at their own mental representations can have non-endorsed mental states or not, that is, whether there can be a purely non-conceptual state of non-endorsement, but as I remarked above, nothing I will say in this paper hinges on this. Also, I will not be concerned with the question of whether all mental states need a higher-order state to become endorsed (which would imply that some mental states will always end up non-endorsed), and I will also not be concerned with the question of whether some states, like belief, carry a “built-in” feature of endorsement and are automatically endorsed unless challenged by a higher-order doubt. In the following discussion, I am mainly concerned with non-cognitive states such as vision that lose their function of inducing beliefs about things that they ordinarily do induce beliefs about, and that will be what I am primarily trying to capture with the term “non-endorsement”.

The Müller-Lyer illusion is an often presented example for what I have called a non-endorsed representation: Two lines seem to be of different length due to the different direction of the arrowheads at the ends of the lines, but they are indeed equal. As soon as you learn that this is an illusion, you stop believing what is represented. But the phenomenal state still keeps representing the world in the way it did, it retains its visual representational content. This applies to all illusions the illusoriness of which we are aware of while undergoing them, and to pseudo-hallucinations like those had by subjects with Charles Bonnet syndrome. Lucid dreams and some hallucinogen-induced experiences are among other examples. In all these cases the world is represented to be in a certain way, it phenomenally seems to a subject that things are some way or other: That one line is longer than the other, that the clouds are getting twisted, or that there is a purple monkey in the room. However, such representations are not endorsed by the subject. We can ordinarily talk of such states in a way of saying “that looks thus and so to my eyes, but I know that it isn’t thus and so”. In these states of non-endorsement there is a clash between what is phenomenally represented and what is believed, or to put it idiomatically, one’s mind or senses are playing tricks on one while one is aware of the tricks.

There are also some states, like imagination, which, by their nature, do not make us committed to endorsing their content. When you imagine a yellow lemon, you do not believe that there is a yellow lemon, nor you disbelieve that there is a yellow lemon. Imagination does not induce true or false beliefs, at least not the kind of belief that is targeted at something in the perceptible environment that should match the content of the act of imagination, just like a painting of a yellow lemon is not in the business of making you believe that there exists or doesn’t exist a yellow lemon. Still, like the painting, the state has an intentional object: A yellow lemon. It

represents a yellow lemon. By visually imagining, one forms in one's mind a contentful visual state, representing objects and their properties as viewed from a certain perspective.<sup>11</sup> One can find other examples, from standard experiences or clinical data, where the appearance content stays the same, but the feeling of presence, or the feeling of reality, and therefore the strength of endorsement, differs.

The notion of a non-endorsed phenomenal representation can be elucidated further with similar states that are the non-sensory, cognitive counterparts of these representations: Sometimes we can entertain a thought, or have a feeling that something is the case, without believing that the content of these cognitive states are true (or false). We can understand someone else's thoughts through her utterances without believing (or disbelieving) their content. By understanding the proposition expressed, and before judging it, we represent the world in a certain way that we don't necessarily take to be true (or false).

So far, I have not said anything that contradicts the qualia theory. Many defenders of qualia will readily acknowledge the existence of what I have called non-endorsed phenomenal states, as there doesn't seem to be a reason to deny them. And it is important to note that arguments for qualia do not rely on states like the Müller-Lyer illusion, at least not in a way that depends on an alleged non-representationality.<sup>12</sup> But beginning with these phenomena will help us clarify things as we move on to more philosophically problematic phenomena, as I will later investigate why some philosophers have problems with the representational analysis of phenomena like double vision or afterimages, while they would grant that other illusions can be non-endorsed representations. Before doing that, I shall demonstrate that afterimages, phosphenes and double vision are indeed a type of non-endorsed representation, the non-endorsement of which results automatically rather than with theoretical reflection unlike in the case of the Müller-Lyer illusion.

---

<sup>11</sup> Imagination, of course, can have some element of endorsement, when, for instance, a subject imagines something to remember how it looks, and a painting, similarly, can be made for the purposes of inducing beliefs about how something looks, but these are not the necessary features of imaginings or paintings.

<sup>12</sup> The argument known as "the argument from illusion", a species of "argument from error", can rely on the Müller-Lyer illusion, or any other illusion or hallucination, where the property encountered in experience doesn't exist in the environment, and is therefore allegedly instantiated in the mind. That debate is outside the scope of this paper, and it is often dealt with a separate counter-argument which relies on considerations related to the intentionality of mental states.

### 3. Automatically Non-Endorsed Phenomenal Representations

In the previous section, I described non-endorsed representational states. These are representational states that represent the world in certain ways, but the veridicality of which the subject is not committed to. Now I want to turn to a special class of non-endorsement, a non-endorsement of an automatic sort which does not require conscious deliberation on the side of the subject regarding the illusoriness of the experience in question. My primary aim is to analyze the representational features of these states, and my remarks about *how they become* automatically non-endorsed should be taken as speculative psychology, serving here the function of helping elucidate the analyzed phenomena. The question of what causal mechanisms make these states non-endorsed is an empirical question, not a philosophical one. My suggestions regarding these mechanisms might be, and most probably are, wrong or incomplete, but these suggestions should serve as illustrative possibilities that shall help the reader understand in what way the analyzed experiential states are representational states, and what, in general, it means for a representational state to become non-endorsed.

Afterimages, phosphenes, tinnitus, and psychoactive-induced experiences all represent the world as being in weird ways, a very fundamental weirdness that contradicts the basic facts we know about the physical world and how it interacts with our sense organs.<sup>13</sup> Now, the content of other non-endorsed representations can also be weird, but I am trying to point to a more fundamental weirdness that is found in automatically non-endorsed representations. I am talking about experiences the contents of which are so weird that we don't even take these experiences as information-carrying states, and we do not even attempt to develop a proper vocabulary to talk about how the world looks while we are undergoing these states, unless a philosophical disagreement about them prompts us to.

In the previous section, I mentioned some non-endorsed representations like the Müller-Lyer illusion, which the defenders of qualia do *not* deploy for their arguments. Those states differ from afterimages and the like in that they represent the world in rather ordinary ways. The content of such states can be “extraordinary” in some sense, like when a subject with Charles Bonnet syndrome hallucinates a bright yellow monkey in the room, but still, the content is ordinary when it comes to the general

---

<sup>13</sup> Here, I use “weird” not as a technical term that is supposed to capture a clearly defined common feature of these experiences, but merely as a descriptive prelude to the examples that follow.

properties represented: There appears to be a stable middle-sized object with a certain color, shape and location in the visible environment. The extraordinariness is due to some background contextual knowledge: We do not live in a world where bright yellow monkeys pop up in one's room out of the blue. If a subject would have a strong background belief that monkeys with crazy colors were the kind of things that popped up in one's room every now and then, then there wouldn't be anything extraordinary about the content of such experiences.

Similarly, when we look at cases of non-endorsement as in the Müller-Lyer illusion, imagination, or hallucinogen-induced experiences, we see that what makes these states non-endorsed is not the basic physical features represented by the experience. Rather, the reason for the non-endorsement is due either to other acquired beliefs (e.g. we know by our own or someone else's measurement that the Müller-Lyer lines are actually equal) or the way the experience is caused (e.g. we know that we are merely imagining, or that we have taken hallucinogenic substances, etc.). But when it comes to cases like afterimages and double vision, the extraordinariness is due to some very fundamental physical properties things are represented to have, which clashes with our very basic knowledge of the perceptible environment and how we relate to it through our sensory organs. As we will shortly see, a lot of the extraordinariness seems to have something to do with the represented spatial features.

Among the representational phenomena I will describe, I leave it to the reader to decide whether what is found in their content is merely weird or impossible. Perhaps what makes these representations automatically non-endorsed is not our beliefs regarding the basic physical structure of the environment. But an incoherency within the visual content of these experiences that makes it the case that things just can't be the way they look, that is, these representations would be non-endorsed representations regardless of what we believe about the environment. Whether phenomenal states can represent impossibilities is a controversial matter, but in any case, nothing much hinges on this for the purposes of this paper, as the extraordinariness of the content, whether impossible or not, suffices to elucidate the nature of the experiences to be discussed. However, the possibility that these states might be representing impossibilities should be kept in mind as an option.<sup>14</sup>

---

<sup>14</sup> For an early discussion of experiential phenomena that might represent impossibilities, see Crane (1988a; 1988b) and Mellor (1988).

#### 4. Double Vision as an Automatically Non-Endorsed Representation

Double vision represents things as double. That is, presumably, why it is called “double vision”. We understand what is meant when people ask each other “how many fingers do you see?” to find out how drunk they are. This representational aspect of double vision, however, has been disputed, and it was claimed that double vision phenomenologically differs from ordinary vision without differing in representational content. Boghossian and Velleman, for instance, have claimed that:

If you press the side of one eyeball, you can see this line of type twice without seeing the page as bearing two identical lines of type. Indeed, you cannot even force the resulting experience into representing the existence of two lines, even if you try. Similarly, you can see nearby objects double by focusing on distant objects behind them, and yet you cannot get yourself to see the number of nearby objects as doubling. (Boghossian and Velleman 1989, 94)

What I will aim to do in this section is to convince the reader that the phenomenology described by Boghossian and Velleman above is better analyzed as a case of non-endorsed representation of doubleness.<sup>15</sup>

We can begin to demonstrate the representational aspect of double vision with the help of anecdotal examples where the illusion actually works: Once I was standing by the side of a dark road, and saw the headlights of a vehicle approaching. There were, apparently, two headlights, and they were the only visible parts of the vehicle. Given that vehicles with two headlights are cars or trucks, I thought it was a car or truck approaching. However, I was subject to alcohol-induced double vision. There was in reality only one headlight, but it visually seemed to me that there were two. As the vehicle drove closer, the two apparent headlights collapsed into a single headlight, and I realized that it was not a car or a truck, but a motorbike.

This is a rare example where the representation of there being two things is endorsed by the subject, as the overall visual state in this example doesn't have some of the peculiarities of most double vision experiences. The only represented objects in the whole visual field are two headlights surrounded by darkness, while in most double vision experiences in better light we are presented with a wholly doubled world, with things spatially overlapping with each other in a strange way, while we can identify the

---

<sup>15</sup> Also see Tye (2003) for a counter-argument to Boghossian and Velleman.

two apparent instances of the doubled thing as the same thing. In such cases, it is easy to see one reason why the representation is never endorsed, as there is something logically awkward about *one* thing being *two*: Two things are conceptually identified as the same thing in many occurrences of double vision, possibly because we know that there is only one thing, and unlike in the headlights case above, we see the “two things” as somehow superimposed. When we see something double, we see one thing (on the cognitive or conceptual level of representation) as double (on the visual level of representation.) There are also spatial peculiarities about how the two things are represented: Their locations can be identified in relation to each other, say, one being on the left, but still, it is not like the two things are taking up space in the same three-dimensional spatial field. They are usually partly superimposed in an unusual, ghostly way. That is quite different from how we ordinarily experience the world to be and how we think it actually is.

Perhaps the causal history of how we get these experiences also contributes to the automatic non-endorsement of these representations. Most of the time, we know that causes of double vision come from inside and not from the world being really that way; it is either due to consumption of alcohol or some other substance, or due to a failure to focus. However, our knowledge regarding the causes of the experience might not be a necessary factor for the non-endorsement; we can speculate that the spatial way things are represented is most probably sufficient. Because of the peculiarities of the content we find it difficult to see how double vision *could* be endorsed in the first place, how to even exactly describe the spatial properties objects should have for this appearance to be truthful. The disregard is so automatic that the relevant feature of the representation loses its belief-triggering function at the cognitive, non-experiential level (except for occasionally triggering beliefs about the contents of one’s visual experiences). Still, what separates double vision from ordinary vision is a representational fact, no matter how weird that represented world may be, and how hard it might be to describe the truth conditions of this representation.<sup>16</sup>

---

<sup>16</sup> Bourget (2015) suggests that the difficulty we might have with very clearly articulating the contents of such experiences might be due to the limitations of our concepts relating to their phenomenology, and why we have these limitations might be due to the fact that we are not interested in communicating them, as the representational contents that make these experiences the type of distorted experiences they are do not correspond to anything in the actual world.

## 5. Phosphenes and Afterimages as Automatically Non-Endorsed Representations

Phosphenes and various other closed-eye visuals are examples of visual phenomena which are sometimes described in English via phrases like “the play of shadow and light”, or “booming-buzzing confusion” for more anarchic cases. A world of unstable, two-dimensional chaotic colored patches is much different than the world we are accustomed to seeing, which is a world of stable, middle-sized, three-dimensional objects that behave in broadly predictable ways. Some might be tempted to say that these chaotic visual experiences are not “world-suggestive”, but this would only amount to saying that they do not suggest a world that we are familiar with, and not that these experiences do not suggest a world at all. There could be such worlds, after all. Some artists try to depict them, and some photographers find little worlds in nature that look like psychedelic closed-eye visuals and draw our attention to the similarity, like natural patterns on a stone, or the northern lights. Lycan similarly suggests that “[...] given any visual experience [...] there is *some* technological means of producing a veridical qualitative equivalent—e.g. a psychedelic movie shown to a subject in a small theater” (Lycan 1987, 90). However, as Schroer (2004, 536) also notes, we don’t need to put things this way in order to point to the representational nature of these experiences. It might be the case that there is *no* technological means to produce a veridical equivalent of these experiences, just like there is no technological means to produce the veridical equivalent of a belief in a magical wand, a round square, or the world of double vision. What is represented might be a nomological or logical impossibility.<sup>17</sup>

As in the case of double vision, there might be factors other than the content that might be playing a role in the automatic non-endorsement of this kind of experiences. We have experiences of phosphenes and afterimages from childhood on. We are not surprised by such experiences and we do not ever wonder if they are truthful. We know how to get them, and how we get them is not by orienting ourselves towards some perceptible external object to be explored by sight, but by playing around with our visual apparatus, say, by closing our eyes and rubbing them. Moreover, the spatial properties of these experienced objects make it impossible for them to be objects that can be explored by ordinary visual means – at least for phosphenes, which are not represented to be in front

---

<sup>17</sup> Indeed, Lycan is probably wrong in suggesting that a veridical equivalent of these experiences can be produced in a movie theater, because that perceived scene would be in front of the subject and experienced as such, while some of these experiences represent their objects as being somewhat located *inside* the eyes of the subject, while at the same time representing properties we see *with* our eyes.

of the eyes, but rather on something like, as it were, an internal screen which has some of the characteristics of the ordinary visible environment, like color, shape and location.<sup>18</sup>

Afterimages require a more complicated analysis when it comes to their experienced location, as they are usually experienced with eyes open, but still linger when you close your eyes. They can switch between being “out there” and being on an “internal screen”. Afterimages also seem to occlude other things in the visual field (which, by the way, is another fact that should make it clear that having an afterimage is a modification of how the perceptible environment is represented.) These represented spatial properties probably constitute the most important factor for these experience’s being automatically non-endorsed.

In order to point out that these experiences are representational or intentional, some philosophers have pointed at the fact that we can talk about them by using the vocabulary we use to describe ordinary phenomena perceptible by vision, like color, size and location. Tye claims that a red afterimage backed by a yellow background is “similar perhaps to that of viewing (in dim lighting) a bloodstain on a transparent sheet of glass suspended between oneself and a yellow background surface” (Tye 2000, 85), while Schroer provides an anecdotal example of an endorsed afterimage experience he had while fixing a lightbulb, where he

immediately took the object to be a red beanbag. [...] It did not occur to me that there really was no beanbag and that I was merely having an afterimage experience until I looked away from the table and the object in question suddenly moved and changed in several very unexpected ways. (Schroer 2004, 543)

Phillips, meanwhile, proposes that afterimage experiences are best described as “illusory presentations of light phenomena often apparently projected from the subject’s point of view” (2013, 433).

The above examples help demonstrate that these experiences are representational, but we should keep in mind that these examples are not strictly necessary: It is a possibility that the content of a representational

---

<sup>18</sup> An auditory analogue of these visual experiences might be tinnitus, or “ringing-in-the-ear”, though it is a somewhat weaker case of automatically non-endorsed representation, as they are more commonly endorsed, especially if the tinnitus experience is a novel experience for the subject and she genuinely believes that there is a source of sound somewhere inside her ear. Some tinnitus experiences, after all, are veridical experiences of a sound produced in the ear.

experience may not be similar to anything that can be veridically represented, or for some experiences it can be the case that there are absolutely no cases where the experience is taken to be veridical. Also, if our aim is to provide anecdotal examples to make a case for the contentfulness of these experiences, we don't need to search for cases where the experience is taken to be veridical, as it will do equally good to find cases where it is taken to be *illusory*, since the claim of the anti-representationalist is that these states are *neither veridical nor illusory*, as these are, allegedly, not contentful states at all.

Another way to demonstrate the representational nature of these experiences is to think of situations where an automatically non-endorsed representation regains its belief-inducing role. Floater experiences provide us with an example to use as a nice springboard for this purpose. Floaters are specks in one's eyes which are normally transparent, but may become visible due to degeneration in the eyes, which makes subjects experience these specks as floating in front of, or inside, one's eye. Floaters are especially vivid against a white or monochromatic background, and they move along with one's gaze. Some of the spatial properties of floaters make them similar to other visual noise like afterimages. Indeed, when I first learned about them, I was surprised to hear that experiencing a floater is a type of veridical perception, as I have always thought that they were a species of visual illusion created by the condition of my eye. When I acquired this information, floater experiences regained their object-targeted belief-inducing function (replacing the occasionally induced belief that there is something wrong with my visual system because I am experiencing illusory patches).

Floater experiences initially lose their belief-inducing function most probably because floaters are spatially similar to the patches experienced in other visual illusions like afterimages (having a strange location, moving along with the eye, etc.), and because one cannot behaviorally interact with a floater any more than one can interact with an afterimage, besides moving them along with one's gaze. One cannot touch or hold a floater, or visually investigate it by further usual means.

Floaters are not the kind of phenomena that are brought up in defense of the qualia theory, so why did I bring up this example? Because floaters constitute a real-world example to think of how other non-endorsed representations could regain their belief-inducing function. Think of a scenario where visual scientists discover that, like floaters, phosphenes are actual colored patches that temporarily exist in your eye, or think of a naïve person getting fooled into believing that in phosphene experiences, she

experiences colored things in a parallel reality inaccessible to others,<sup>19</sup> or rare real-life cases like retinal detachment (discussed in Gow 2019 in the context of the representationalism-qualia dispute) where subjects undergo phosphene experiences with eyes open and initially take the phosphenes to be actual flashes of color and light. The fact that we can coherently imagine these scenarios should help us understand that these experiences have visual representational content of a non-endorsed type.<sup>20</sup>

## 6. Bourget's Argument from the Loss of Visual Information

Before moving on, I will briefly look at an argument by Bourget (2015) where he similarly argues for representationalism for cases which he classifies as cases of visual distortion, including double vision and blur. Bourget invites us to consider cases of ordinary vision, such as seeing a

---

<sup>19</sup> Several philosophers have recently suggested that some phosphene experiences are instances of veridical perception. Based on empirical work by Salari et al. (2017), among others, Ben-Yami (unpublished) suggests that in some cases “the phosphenes perceived are due to ultraweak bioluminescent photon emission of cells in the visual system”, and Ali (2018) makes a similar claim that some phosphenes are generated internally, basing this claim on Davis et al. (1976). Perhaps these findings could also be interpreted as the electromagnetism produced by the eye causing illusory experiences rather than us veridically perceiving electromagnetism produced by the eye, but I will not discuss this here, for the examples discussed by the above authors involve only some types of phosphenes (such as those experienced in a dark room), and if it is indeed true that we veridically experience something in these experiences, this can be accommodated within a representationalist framework: They are veridical representations. The qualia theorist, on the other hand, should provide a strong argument to the effect that the veridical experience claim is confused: That the scientific evidence cannot be interpreted in such a way because, according to the qualia theory, we cannot make sense of phosphene experiences being veridical, as phenomenological reflection allegedly shows that these experiences lack intentional objects or representational contents.

<sup>20</sup> An anonymous reviewer has suggested that even if the endorsement/non-endorsement framework is on the right track, there is still the possibility that the difference between endorsement and non-endorsement is due to a change in phenomenal character without a change in content. This, however, is not possible, given that the notions of endorsement and non-endorsement are, by stipulation, cognitive notions related to the formation of beliefs or belief-like states. Perhaps my analyses in this paper are wrong and there is nothing in the mental realm that can be captured by the notions of endorsement and non-endorsement, but if there are, then by definition endorsement and non-endorsement refer to content-related differences. At best, there might be a phenomenal difference between endorsement and non-endorsement, alongside representational and functional differences, if there is a phenomenology specific to various cognitive states. The question of whether that phenomenological difference can be captured by representational terms or not would be a general question regarding the cognitive phenomenology of all experiences, because all experiences are either endorsed or non-endorsed, and the question would be out of the scope of this paper, as this paper takes up issue with the claim that particularly afterimage experiences and the like are resistant to a representationalist analysis for reasons specific to these type of experiences.

square, and then think of a series of cases where the visual effect (blur, double vision, etc.) is gradually amplified, and draws our attention to the fact that in cases of higher distortion, it is obvious that there is an impoverishment in the representational content of the experience (e.g. we cannot see a square as a square anymore as the blur or double vision gets extremely amplified.) The cases of extreme distortion are not everyday cases of perceptual distortion, but Bourget asks: If, the phenomenal character is independent from the content in mild cases of distortion as anti-representationalists argue, then at what point does the content get intertwined with the phenomenal character of a distorted experience as the distortion gets amplified? Bourget concludes that all possible answers to this question are implausible and/or ad hoc.

I do think that this presents an interesting case against anti-representationalism for double vision and blur. The primary reason I bring up Bourget's argument, however, is that it provides an interesting phenomenological exercise for the qualia theorist that could help her understand in what way the allegedly non-representational aspects of experience are actually representational, and help her have a better grasp of what it means for a state to be automatically non-endorsed. Imagining a gradual transition from ordinary to extreme cases can help one see that ordinary cases also involve a loss of visual information, even though the information that is lost might be little. The exercise should also help us see why we do not prefer to have these distorted experiences, that is, why we avoid them. If what makes an experience an experience of double vision or an afterimage had nothing to do with the representational content, then our preference regarding whether to have these experiences or not would merely be a cosmetic preference, it would be a preference regarding which "mental paint" we would like our experiences to be painted with: Whether we would strive to be or not to be in such visual states would depend, perhaps, on which phenomenology we enjoy more, and not on whether we prefer to see the world as it is or see it in a distorted manner. One can, of course, from time to time, prefer to have these experiences, mainly for phenomenological curiosity or merely to enjoy a certain experience. Some children, philosophers, cognitive scientists, and recreational users of psychedelic substances do occasionally strive to have experiences of visual distortion. But for the sake of seeing the visible environment clearly and as it is (rather than blurrily or doubly), these experiences are experiences that we prefer to avoid, because their representational content is distorted. We, of course, rarely complain about ordinary experiences of afterimages or double vision, because they are temporary and mild. But imagining extreme cases of these visual distortions should make it more obvious why

it is not preferable to have these experiences for epistemic reasons, and therefore make it obvious that they have representational content.<sup>21</sup>

## 7. Confusing Non-Representationality with Non-Endorsement

I hope I have so far established that double vision, phosphenes and afterimages can be analyzed as non-endorsed representational states, and this non-endorsement is an automatic one in great majority of cases. Now I wish to demonstrate that some defenders of qualia confuse non-endorsement with non-representationality when talking about these experiences.

In his *Mental Paint and Mental Latex*, Ned Block suggests that “[Afterimages] don’t look as if they are really objects or as if they are really red. They look ... illusory” (Block 1996, 32; ellipsis in original). He uses this observation in support of non-representationality of such experiences, and therefore the existence of qualia: In such experiences, it cannot be the representational content that is responsible for the qualitative phenomenology we are introspectively aware of, so it should be a non-representational property such as a quale (or “mental paint”) which constitutes that phenomenology, such as qualities instantiated in experiences.<sup>22</sup> Interestingly, Block argues for the anti-representationalist conclusion via mentioning illusoriness, which is a feature of representations.

---

<sup>21</sup> The phenomenon of blurry vision is outside the scope of this paper, but I will just mention in passing that the information-degrading nature of experiences of blur should be pretty obvious even without considering extreme cases. It is part of the daily visual experience of many people with eye problems, and a rather undesirable one.

<sup>22</sup> This definition of qualia or mental paint as non-representational might seem to go against the definitions of some qualia theorists that define these properties as representational, in the sense that these properties are vehicles of representation, as when Block says mental paint is the “mental properties of the experience that represent the redness of a tomato” (Block 1996, 29). Mental paint can be a “representational” property in the sense that it can serve representational functions, though it can also be instantiated without representing anything. When I define these properties as “non-representational”, what I mean is that these properties are neither represented properties, nor the property-of-representing-something. If there is no mental paint, introspection can only reveal which property is represented, along with the fact that the subject is in a representational state (and perhaps, if impure representationalism is true, introspection can reveal some other properties like representational “modes”), but it does not reveal properties that do the representation in the way that one looks at a painting and becomes aware of the paint. Block claims that what is found by introspection in afterimage experiences is neither the represented property, nor the property of something being represented, but only mental paint, and in this sense mental paint is non-representational. Contrary to Block, I claim that analyzing an afterimage experience gives us only representational properties: Introspection reveals the fact that one is undergoing a representational state along with the content of this representation, but no mental paint.

I take it that the ordinary understanding of visual illusion is the following: Something visually seems to be a certain way, but it is not really that way. But what would it mean to say that afterimages “look illusory”, or that “afterimages don’t look as if they are really red”? It is easy to understand what it means to say that something does not look red, but what about something’s “not looking really red”? Whatever Block means by that, he seems to talk about something that looks or does not look some way or other, which means that the experience has an intentional object that is represented to be some way or other. The statement about an object’s “not looking really red” is a statement about how it looks, how it appears to be. In more technical terms, that is how an object is represented by the experience.

And what about “looking illusory”? Does Block mean that there is a brute represented feature of “unrealness” or “illusoriness” attached to these represented objects or properties at a pre-cognitive level?<sup>23</sup> If we could make sense of this, and if it were indeed the case, the state would still be a representational one: Redness is represented, and it is represented as “unreal”. Or does Block mean that something looks illusory in the sense that the way that thing is represented makes it unlikely that it is real, because of the weird or contradictory properties it is represented to have, so it “looks illusory” at a judgmental level? If we follow this interpretation, we should take Block as implying that the object of the experience is represented in such a way that it cannot be real, or we automatically know that it is not real. But neither interpretation can help one reach an anti-representationalist conclusion about afterimage experiences, because in both interpretations “illusoriness” is understood as a representational phenomenon.

Perhaps, one might think, in order to interpret Block properly, we should turn to what an illusion is supposed to be according to qualia theory. Qualia can be thought of as vehicles of representation that do the representing. A certain quale’s instantiation makes the mind represent a certain property. How this representational feature is to be understood depends on the details of the theory. According to externalist theories, a quale represents something in virtue of the instantiation of that quale’s being in some causal-informational relation with instances of the represented property. A “red quale” represents redness (or some other property of red-looking objects, if one is an anti-realist about colors) in virtue of being reliably

---

<sup>23</sup> An anonymous reviewer has suggested that this is possibly the correct interpretation, as Block describes afterimages (more particularly, an afterimage’s changing size) as “somehow unreal or unobjective” in his more recent work (2010, 54). This suggestion might be on the right track, but for the purposes of this paper, it is useful to explore all the possible interpretations of Block’s original remark, and follow where they theoretically go.

triggered by redness (and by fulfilling some other conditions depending on the details of the theory). According to internalist theories of representation that rely on qualia, a “red quale” represents redness either in a projectivist manner, where subjects mistakenly take these experiential qualities to be the qualities of external objects (and form a useful but erroneous representation of the world with the help of qualia and the act of projection), or somehow end up representing red-looking objects if the experiences, somehow, acquire a representational aspect which signals that these experiences are caused by external things. In all of these theories, illusion is a state where a quale is instantiated, but the thing it is supposed to represent is not there. If the qualia theorist wants to claim that an afterimage experience is the bare awareness of a quale-complex, without any representation going on, “illusoriness” is not the concept to be alluded to in order to demonstrate that. Such a state would be neither veridical, nor illusory. A qualia-theoretical analysis of an afterimage experience can be given without appealing to illusoriness, by saying that “when we introspect an afterimage experience, we find a mentally-instantiated quality, which does not represent anything”. This formulation would not face the problems that the illusion formulation faces, but it faces other problems: It has no appeal for people who, introspecting an afterimage experience, find no qualities other than those like colors that objects appear to have also in uncontroversially representational experiences, which makes it natural to think of afterimages as a special type of misrepresentation.

Is there perhaps some reason, that introspection can provide us, which would urge us to conceptualize afterimages differently, in a way that involves qualia? Let’s turn to Amy Kind, who, in her article *How to Believe in Qualia*, gives us instructions to find qualia in afterimage experiences:

In general, afterimages occur subsequent to the removal of some original (usually intense) stimulus. When a camera flash goes off, you might experience an afterimage in front of the photographer’s face. If you stare intently at a bright light for a little while and then close your eyes, there will be a lingering glow in the darkness. And if you stare at a green dot for half a minute and then shift your attention to a bright white piece of paper, you will visually experience a red dot similar in size and location to the green dot you had been staring at. But in none of these cases does it seem as if the afterimage represents something that is really there. When you close your eyes after looking at the bright light, for example, you don’t take the lingering glow to be on the inner surface of your eyelids. When you see the red afterimage against the white page, you

don't take the redness to suggest the existence of a red dot on the page. (Kind 2008, 289)

Kind puts the bottom line with the aforementioned quote by Block, where he mentions the illusoriness of afterimages. Note that Kind talks about these allegedly non-representational experiences by mentioning the same kind of experienced phenomena found also in ordinary experiences: Both before and after looking at the bright light and then turning away, we experience brightness, referred to as a "glow" by Kind; after staring at the green dot, we see a "red dot similar in size and location to the green dot" we have been staring at. Kind holds that these experiences are not representational: Unlike the experience we have when we are looking at a green dot, which represents a green dot, the corresponding afterimage does not represent a red dot. But isn't this claim of non-representationality contradictory to what Kind says, that we *experience a red dot*? Isn't the red dot the intentional object of the experience? Can we coherently assert that we "experience a red dot" while it is the case that *a red dot is not represented*, that it *doesn't visually seem to us* that there is a dot which is red? I think not. If we experience a red dot, it follows that it seems to us that there is a red dot. "Seeming" is of course used here in the pre-cognitive (or "non-conceptual") sense, in a way that allows us to say "it seems to me [in the first sense of "seeming": it seems to my eyes] that there is a red dot, but it doesn't seem to me [in the second sense of "seeming": I don't judge] that it is real".

I think what is going on here is a confusion, taking non-endorsed representations as non-representational. Saying that a non-representational state is illusory, as Block does, or saying that we experience something without that something being the object of experience, as Kind does, is a contradictory way of talking. In some experiences, of course it can appear to us that the experienced thing doesn't exist, if things can "look illusory" in some sense, but that would also create no motivation to postulate the existence of mental qualities to account for the experience, as those experienced qualities *don't seem to exist*. Merely intentional qualities of merely intentional objects that figure in the representational content would do the job, or to put it another way, the fact that things *seem* a certain way, that there seems to be an "illusory redness", can fully account for the qualitative phenomenology.

## 8. Possible Causes of the Confusion

In the previous section, I made an attempt to show that Block and Kind confuse two different phenomena, an experience's being non-representational

and its being a non-endorsed representation. Now I want to investigate the psychological and linguistic causes of this confusion. What makes some of us prone to making this logically invalid move? I speculate that one determining factor here might be a verbal ambiguity in some expressions related to mental representation, expressions in the form of “it seems to me that X” or “it appears to me that X”. As I have mentioned earlier, one can talk of seeming in a cognitive way, as opposed to a visual/experiential way, as in phrases like “it seems to me that...” followed by a clause that expresses a proposition associated with a belief or belief-like state. So when you are in a visual state where the world in front of you seems to you to have such and such objects and properties, it is still possible that it doesn’t seem to you, in the cognitive sense, that there are such objects and properties: Your mind is in two conflicting representational states at two different levels, and you endorse the higher, cognitive one. Even though at the end of the day the world doesn’t include a red patch according to *you*, the cognizing subject, there is a red patch in front of you *according to your visual system*.<sup>24</sup> The logical fallacy lies in the following move: Starting with the premise that the world doesn’t seem to be like this or that to you at the cognitive level, and moving to the conclusion that it doesn’t seem to be like this or that to you *at all*.

But why don’t we come across a similar fallacy when it comes to other non-endorsed visual states, like the Müller-Lyer illusion? I suggest that the answer can be found in the difference between how the visual state at hand becomes non-endorsed. If the illusoriness of the visual state is not so phenomenally obvious, and when we reflect on these experiences, we form higher-order beliefs that take as data the visual state *already conceptualized as a representational state*, and when we come to know about the illusoriness, we are in a state of explicit disbelief. To put it another way, when we are looking at the Müller-Lyer lines, understanding that we are in an illusory visual state involves, first of all, understanding that we are in a representational visual state, a state with a truth-evaluable content, and understanding that the content is false. So, no one would argue for qualia for a state like the Müller-Lyer illusion on the grounds of non-representationality, as this state causes explicit belief or disbelief in us, and having explicit belief or disbelief that takes visual appearances as data means that we are aware of the visual representational state as a representational state. (It is called an “illusion”, after all.) But we rarely have explicit belief or disbelief in cases of afterimages or phosphenes. The

---

<sup>24</sup> Here I don’t mean to say that your visual system is an independently conscious homunculus. It is just that in cases of non-endorsement you distance yourself from how things look to you as represented by your visual system, and endorse your cognitive, judgmental representations about the environment.

factor that would prevent a similar fallacy in the Müller-Lyer illusion case is missing from the cases of automatic non-endorsement.

One could argue that I am not being charitable here to Kind and Block, and say that even though they might be mistaken, they are not confused: Maybe I shouldn't take their descriptions of these states at face value. Let's follow this suggestion and try to read Block and Kind in a way that wouldn't make them contradict themselves.

As for Block's way of putting things, it is hard to come up with a charitable reading. How can a talk of "illusoriness" be translated into non-representational talk? I do not know. However, there can be prospects for charity for some aspects of Kind's description of afterimages. When talking about qualia, it is common that theorists use terms that are normally used to talk about represented properties, like when some philosophers say things like "a red-quale". By this, they don't mean that a certain quale is itself red, or that it necessarily represents redness. What is meant here is that the quale is the sort of quale we find in experiences that ordinarily represent redness. Also, it could be that when these philosophers talk of experience in a way where the grammar suggests an act-object structure, we shouldn't accuse them of failing to recognize that if a state such as an afterimage experience is supposed to be non-representational then it cannot have an act-object structure. Maybe when Kind says that "you might experience an afterimage in front of the photographer's face", she uses "experience" in the same way when we say that we "experience joy", where joy is not (or not so obviously is) the intentional object of the experience.

If this charitable reading is correct, then all of what Kind says should in principle be translatable to non-intentional qualia talk. According to this translation, in the above sentence, "afterimage" doesn't refer to the intentional object of the experience. Rather, the whole phrase "experiencing an afterimage" refers to a non-intentional mental state, as in "experiencing joy" (conceptualized for the sake of the argument as a non-intentional mental state), even though the "an" in "an afterimage" sets off some alarms: We can experience two afterimages at the same time, but can we experience "two joys" at the same time? Perhaps, and perhaps not, but in any case the parallelism already seems suspicious.

In any case, even if a non-intentional reading of "experiencing an afterimage" is possible, what to do with "an afterimage in front of the photographer's face"? My mental state is not in front of the photographer's face. Well, if "an afterimage" refers to an experience, and mental state tokens are identical to neural state tokens, and I am in front of the

photographer, then perhaps my experience is in front of the photographer's face in a sense, but this is obviously not what is meant by the phrase. Maybe we should understand it in the same way we are supposed to understand red-talk about qualia: Just like there can be red-afterimages in the sense that they instantiate a quale that we also find in red-representing states, there can be in-front-of-the-photographer's-face-afterimages, where we find the type of qualia, presumably shape and location related qualia, that our minds are supposed to instantiate when looking at faces or face-shaped objects. This translation doesn't seem impossible, but nevertheless, as Kind is describing an environmental situation that causes us to have afterimages, and our experience of the photographer's face, unlike the afterimage, is uncontroversially representational, it is much more plausible to read "in front of the photographer's face" as referring to nothing other than a spatial property related to the photographer's face as it is represented by the experience. So, after all, there visually appears to be something in front of the photographer's face.

Kind can insist that this natural reading is not correct, and the qualia-theoretical analysis is the correct description of the situation. But we should remember that Kind's article is titled "How to Believe in Qualia", and it is aimed at the reader who doesn't believe in qualia yet, or may not even have the concept of a quale, but who will be made to believe in qualia after following some instructions. So the initial steps of the description of the mental state and the description of the environmental factors that trigger it should not have references to qualia until the reader stumbles upon the relevant property by following the introspective steps. It would be surprising if Kind was expecting the uninitiated reader to translate this in-front-of-the-photographer's-face-talk into qualia-talk. More plausibly, what she is doing is just describing a non-endorsed representational state, with ordinary language that refers to represented objects and properties, such as the photographer's face and an illusory glow in front of it.

Perhaps there is a way that Kind could provide us with a complete qualia-theoretical interpretation of the way she describes the afterimage experience that could get around the above objections, so maybe I am mistaken in attributing her a confusion. But that still wouldn't make the qualia theory more appealing. There is, after all, no apparent reason why we should interpret the ordinary talk of afterimages in such a complicated fashion and prefer the qualia view for a philosophical analysis of the mind. We have a simpler analysis which treats these states as illusions of an automatically non-endorsed sort, an analysis that applies to ordinary talk of afterimages, a talk which sounds linguistically the same as Kind's description of afterimages, but still makes sense without being translated into something very complicated that allegedly lies beneath the surface

linguistic form and that relies on a property called qualia that we have independent reasons to be suspicious of.

## 9. Concluding Remarks

There are several other kinds of experiential phenomena, experiences with affective aspects like pain and orgasm, which are also among the armaments of the anti-representationalist. I want to briefly point to the similarity of these phenomena to the cases discussed above. These states might also be among the states which sometimes lose their belief-inducing function regarding the veridicality of their contents, not because they represent impossibilities or oddities, but because we care more about the very existence of these experiential states than their veridicality, as feeling pain on a real arm is as undesirable as feeling pain on a phantom arm.<sup>25</sup> I will not provide a defense of representationalism about pain and orgasm along these lines here as it is beyond this paper's scope, and leave it at pointing at hints for this strategy found in the literature (see the discussion in Aydede 2009, sections 3.2 and 4.2). But if this initial line of thought is on the right track, in the future it can provide us with an explanation of the confusion of the anti-representationalist in the case of affective states as well.

In any case, I hope what I have argued above makes a case for representationalism for afterimage-like experiences. These experiences are automatically non-endorsed representations confusedly taken to be non-representational states by some defenders of qualia, and when we examine the way they define these states, we see that they are actually talking about these experiences as if these states were a type of illusion, that is, a type of mental representation with truth-evaluable appearance content. And their representational nature makes the postulation of qualia unnecessary. They represent the world in certain ways, and if the represented qualities are not in the world, then they need not be instantiated in the mind or anywhere else. They simply do not exist. Afterimages and the like do not trigger explicit belief or disbelief at the cognitive level, but the lack of a representation at the cognitive level is not an obstacle to a representationalist analysis at the non-conceptual visual level. Compare with visual imagination: Imagination is also a representational state, and visually imagining a red patch does not make you believe or disbelieve that there is

---

<sup>25</sup> By "caring about the very existence of these experiential states rather than their veridicality" I do not mean that we care about some non-representational or non-intentional aspects of these states. We care about pain experiences because of the object of experience, the felt pain quality.

a red patch, while still being a representational state which has a phenomenal character that is exhausted by the qualities of what it is an imagination of. The same applies to experiences of red afterimages.

## Acknowledgments

Thanks to Katalin Farkas, Howard Robinson, Alex Byrne, Amy Kind, Gianfranco Soldati, Martine Nida-Rümelin, and an anonymous reviewer for helpful comments on earlier versions of this paper.

## REFERENCES

- Ali, R. 2018. Does hallucinating involve perceiving? *Philosophical Studies* 175: 601-627.
- Aydede, M. 2009. Pain. In *Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), ed. E. N. Zalta, <http://plato.stanford.edu/entries/pain/> Accessed 16 May 2020.
- Ben-Yami, H. Unpublished. Afterimages and related phenomena.
- Block, N. J. 1996. Mental paint and mental latex. In *Philosophical Issues*, vol. 7: Perception, ed. E. Villanueva, 18-49. Atascadero: Calif.: Ridgeview.
- Block, N. J. 2010. Attention and mental paint. In *Philosophical Issues, 20: Philosophy of Mind*, eds. E. Sosa and E. Villanueva, 23-65. Oxford: Wiley-Blackwell.
- Boghossian, P., and D. Velleman. 1989. Color as secondary quality. *Mind* 98: 81-103.
- Bourget, D. 2015. Representationalism, perceptual distortion and the limits of phenomenal concepts. *Canadian Journal of Philosophy* 45: 16-36.
- Byrne, A. 2001. Intentionalism defended. *Philosophical Review* 110: 199-240.
- Chalmers, D. J. 2004. The representational character of experience. In *The Future for Philosophy*, ed. B. Leiter, 153-181. Oxford: Oxford University Press.
- Crane, T. 1988a. The waterfall illusion. *Analysis* 48: 142-147.
- Crane, T. 1988b. Concepts in perception. *Analysis* 48: 150-153.
- Davis, F. A., D. Bergen, C. Schauf, I. McDonald, and W. Deutsch. 1976. Movement phosphenes in optic neuritis: a new clinical sign. *Neurology*, 26: 1100-1104.
- Gendler, T. 2008. Alief and belief. *Journal of Philosophy* 105: 634-663.
- Gow, L. 2019. Everything is clear: All perceptual experiences are transparent. *European Journal of Philosophy* 27: 412-425.

- Jackson, F. 1977. *Perception: A Representative Theory*. Cambridge: Cambridge University Press.
- Kind, A. 2008. How to believe in qualia. In *The Case for Qualia*, ed. E. Wright, 285-298. Cambridge MA: The MIT Press.
- Lycan, W. 1987. *Consciousness*. Cambridge MA: The MIT Press.
- Mellor, D. H. 1988. Crane's waterfall illusion. *Analysis* 48: 147-150.
- Moore, G. E. 1939. Proof of an external world. *Proceedings of the British Academy* 25: 273-300.
- Peacocke, C. 1983. *Sense and Content*. Oxford: Oxford University Press.
- Phillips, I. 2013. Afterimages and sensation. *Philosophy and Phenomenological Research*. 87: 417-453.
- Salari, V., F. Scholkmann, R. L. Pandey Vimal, N. Császár, M. Aslani, and I. Bókkon. 2017. Phosphenes, retinal discrete dark noise, negative afterimages and retinogeniculate projections: A new explanatory framework based on endogenous ocular luminescence. *Progress in Retinal and Eye Research* 60: 101-119.
- Siegel, S. 2010. Do visual experiences have contents?. In *Perceiving the World*, ed. B. Nanay, 333-638. Oxford: Oxford University Press.
- Schroer, R. 2004. Environmental representationalists on afterimages and phosphenes: Putting our best foot forward. *Southern Journal of Philosophy* 42: 531-546.
- Thau, M. 2002. *Consciousness and Cognition*. Oxford: Oxford University Press.
- Tye, M. 2000. *Consciousness, Color, and Content*. Cambridge MA: The MIT Press.
- Tye, M. 2003. Blurry images, double vision, and other oddities: New problems for representationalism?. In *Consciousness: New Philosophical Perspectives*, eds. Q. Smith and A. Jokic, 7-32. Oxford: Oxford University Press.

## KNOWLEDGE AND ASSERTION: A CRITIQUE OF LACKEY

Joshua Anderson  
Virginia State University

Original scientific article – Received: 05/01/2020 Accepted: 7/5/2020

### **ABSTRACT**

*In the literature on assertion, there is a common assumption that having the knowledge that  $p$  is a sufficient condition for having the epistemic right to assert that  $p$ —call this the Knowledge is Sufficient for Assertion Principle, or KSA. Jennifer Lackey has challenged KSA based on several counterexamples that all, roughly, involve isolated secondhand knowledge. In this article, I argue that Lackey's counterexamples fail to be convincing because her intuition that the agent in her counterexamples both has knowledge and do not have the epistemic right to assert is wrong. The article will progress as follows: In section 2, I present Lackey's argument. In section 3, I suggest some more general reasons for doubting that the agent in her counterexamples actually has knowledge. I then show that from a virtue theoretic and Edward Craig's practical explication of knowledge perspectives the agent in Lackey's counterexamples does not know. Since the agent in Lackey's counterexamples does not have knowledge, she has failed to convincingly prove that KSA is false. In section 4, I conclude by suggesting that, at most, what Lackey's counterexamples demonstrate is a problem with a simplistic evidentialist and/or process reliabilist epistemology.*

**Keywords:** *Assertion; Jennifer Lackey; secondhand knowledge; virtue epistemology*

## 1. Introduction

In the literature on assertion, there is a common assumption that having the knowledge that  $p$  is a sufficient condition for having the epistemic right to assert that  $p$ —call this the Knowledge is Sufficient for Assertion Principle, or KSA.<sup>1</sup> Recently, Jennifer Lackey has challenged KSA based on several counterexamples that all, roughly, involve isolated secondhand knowledge. In this article, I argue that Lackey’s counterexamples fail to be convincing because her intuition that the agent in her counterexamples both has knowledge and does not have the epistemic right to assert is wrong.

Lackey is correct that the agent in her counterexamples does not have the epistemic right to assert, but part of the reason that they do not have that right is because the agent does not, in fact, have knowledge. In other words, the reason that Lackey’s agent does not have the epistemic right to assert is in part because KSA, or something like it, holds. However, it is not just that I do not share Lackey’s intuitions—which I do not. It will be shown that for a variety of reasons, and according to different epistemological theories, the agent does not have knowledge.

This article will progress in the following way. In section 2, I present Lackey’s argument. In section 3, I suggest some more general reasons for doubting that the agent in her counterexamples actually has knowledge. I then show that from a virtue theoretic and Edward Craig’s practical explication of knowledge perspectives the agent in Lackey’s counterexamples does not know. Since the agent in Lackey’s counterexamples does not have knowledge, she has failed to convincingly prove that KSA is false. In section 4, I conclude by suggesting that, at most, what Lackey’s counterexamples demonstrate is a problem with a simplistic evidentialist and/or process reliabilist epistemology.

## 2. Lackey’s Argument

According to what has been identified in this article as KSA: If a subject  $S$  knows that  $p$ , then  $S$  has the epistemic right to assert that  $p$  (Lackey 2011, 252).<sup>2</sup> To be clear, KSA in its most general form simply maintains that

---

<sup>1</sup> For example, DeRose (2002), Hawthorne (2004), Simion (2016), amongst others, can all be seen as either advocating or sympathetic to something like KSA.

<sup>2</sup> What I have identified as KSA, is Lackey’s second formulation of the sufficient condition of the Knowledge Norm of Assertion (KNA-S\*). “KNA-S\*: One is properly epistemically

knowledge is sufficient for the *epistemic* right to assert. KSA does not require one to assert, and there may be additional norms governing assertion. For example, a psychiatrist, over the course of many therapeutic sessions, comes to know that her client, Tom, is an alcoholic; the norms governing doctor-patient privilege prevent the psychiatrist from asserting what the psychiatrist knows to certain people. However, the psychiatrist would not be *epistemically* blameworthy for asserting that Tom is an alcoholic.

Even though Lackey admits that there is some intuitive plausibility and, potentially, some theoretical power behind KSA, she argues that KSA is actually false. She suggests “that there are various kinds of cases in which a speaker asserts that *p*, clearly knows that *p*, and yet does not have the proper epistemic authority or credentials [i.e. the epistemic right] to make such an assertion, thereby showing that knowledge is not always sufficient for epistemically proper assertion” (Lackey 2011, 253). Lackey makes her case through a series of counterexamples that all involve isolated secondhand knowledge.<sup>3</sup> For simplicity, I will focus, predominately, on one of Lackey’s counterexamples that she calls DOCTOR.

DOCTOR: Matilda is an oncologist at a teaching hospital who has been diagnosing and treating various kinds of cancers for the past fifteen years. One of her patients, Derek, was recently referred to her office because he has been experiencing intense abdominal pain for a couple of weeks. After requesting an ultrasound and MRI, the results of the tests arrived on Matilda’s day off, consequently, all of the relevant data were reviewed by Nancy, a competent medical student in oncology training at her hospital. Being able to confer for only a very brief period of time prior to Derek’s appointment today, Nancy communicated to Matilda simply that her diagnosis is pancreatic cancer, without offering any of the details of the test results or the reasons underlying her conclusion. Shortly thereafter, Matilda had her appointment with Derek, where she truly asserts to him purely on the basis

---

positioned to assert that *p* if one knows that *p*. According to the KNA-S\*, then, knowledge is sufficient for possessing the epistemic authority for assertion” (Lackey 2011, 252). KSA simply cashes out “possessing the epistemic authority for assertion” as the epistemic right to assert. Similarly, Simion (2016, 3043) refers to it as the sufficiency claim for a Knowledge Norm of Assertion (KNA-Suff) and words it as follows: “One’s assertion that *p* is epistemically proper if one knows that *p*”. KSA is just clarifying “epistemically proper”.

<sup>3</sup> What Lackey means by “isolated secondhand knowledge” will become clear as I explicate her view, below.

of Nancy's reliable testimony, "I am very sorry to tell you this, but you have pancreatic cancer" (Lackey 2011, 253).

The first thing Lackey draws one's attention to is that in DOCTOR Matilda's knowledge is isolated and secondhand. It is isolated because the only thing that Matilda "knows" is simply the fact that Derek has cancer—Matilda's general background knowledge regarding cancer, generally, and the little information she has from her previous meeting with Derek is not significant enough to un-isolate her knowledge. Matilda's knowledge is secondhand because the only reason she "knows" is based exclusively on another person's—viz. Nancy's—testimony. There is nothing inherently wrong with isolated secondhand knowledge;<sup>4</sup> it is just that in DOCTOR it leads to problems.

Lackey maintains "that Matilda clearly knows that Derek has pancreatic cancer—it is true, she believes it, she has good reason to trust the testimony of her medical student, and Nancy is in fact a reliable source" (Lackey 2011, 254). Further, according to Lackey, it is clear that Matilda does not have the epistemic right to simply assert to Derek that he has pancreatic cancer. There are several reasons that Lackey gives to justify her claim that Matilda lacks the epistemic right to assert. Matilda is an expert oncologist and this expertise carries with it certain epistemic duties and responsibilities to fulfill before asserting a diagnosis. These duties and "responsibilities may include having reviewed the test results firsthand, possessing reasons for choosing one condition over another, knowing details about the size and nature of the cancer, and so on" (Lackey 2011, 254). Further, as an expert it is expected that Matilda should be able to justify and explain her diagnoses, in general, and Derek's cancer diagnosis, in particular. Such justification and explanation is impossible for Matilda given the isolated secondhand knowledge involved.

Based on the considerations regarding DOCTOR, Lackey concludes DOCTOR is "a case where a speaker knows that  $p$  without thereby being epistemically positioned to assert that  $p$ , thereby falsifying [what Lackey refers to as] KNA-S\*" (Lackey 2011, 258). In the terminology being used in this article: DOCTOR demonstrates that one can know that  $p$ , but lack the epistemic right to assert that  $p$ , thereby falsifying KSA. More explicitly, Lackey's argument amounts to this:

- 1) If KSA is true, then if one knows that  $p$ , then one has the epistemic right to assert that  $p$ .

---

<sup>4</sup> An example of unproblematic isolated secondhand knowledge is Lackey's own Chicago visitor case. See Lackey (2007).

- 2) In DOCTOR: One knows that  $p$ , but does not have the epistemic right to assert that  $p$ .
- 3) Therefore, KSA is not true. (by 1, 2 and Modus Tollens)

### 3. Problems with Lackey's Argument

Lackey's argument is simple and straightforward. If she is right, then she has identified a huge problem with much of the literature on assertion.<sup>5</sup> In this section of the article I argue that premise 2 of Lackey's argument does not hold, and thus she fails to demonstrate that KSA is false. I will show that in DOCTOR<sup>6</sup> it is not the case that an agent both knows and lacks the epistemic right to assert.

Although Lackey claims that it is "clear" that Matilda knows Derek has cancer, it certainly does not seem clear that Matilda knows. Further, although Lackey gives some reasons for thinking that Matilda does know Derek has cancer—viz. "it is true, she believes it, she has good reason to trust the testimony of her medical student, and Nancy is in fact a reliable source"—she is really just trading in intuitions (Lackey 2011, 254). It is obvious that she is merely appealing to intuition when she considers some modified versions of her counterexample and states that one can "compar[e] the intuitions elicited from such modified cases with those from the original" (Lackey 2011, 256).

Considering DOCTOR from the perspective of a variety of epistemological theories—e.g., Craig's practical explication of knowledge, virtue epistemology, inferentialism—it will be shown that Matilda does not actually know Derek has cancer. If Matilda does not know that Derek has cancer, then Lackey's counterexample fails, and thereby her argument that KSA is false does not go through. Moreover, though it was not gone into above, Lackey gives reasons to justify her claim that Matilda lacks the epistemic right to assert. What is interesting is her reasons for claiming that Matilda lacks the epistemic right to assert—such as the increased epistemic duties tied to Matilda's status as an expert oncologist and her inability to provide a justification for the diagnosis—are actually

---

<sup>5</sup> See for example, DeRose (2002), and if one takes assertion to be a kind of action, see Hawthorne and Stanley (2008).

<sup>6</sup> Again, I will only be primarily arguing against Lackey's counterexample DOCTOR. Although she gives several counterexamples that she thinks refute KSA, they are all roughly the same. They all involve what she calls isolated secondhand knowledge and are cases where, purportedly, the agent involved has knowledge but lacks the epistemic right to assert. The reasoning I will use to refute DOCTOR, I think can be extended to all her cases. Thus, for brevity I have only focused on the one counterexample.

reasons to think Matilda does not know. Thus, Lackey is implicitly appealing, in part, to KSA, or something like it, in here counterexample.

The advantage of demonstrating that Matilda fails to have knowledge for a variety of reasons is that it will block a couple of responses that Lackey could make. Such responses would amount to something like this: On the one hand, “I disagree that your general reasons hold”, and on the other hand “Well sure according to theory X Matilda does not have knowledge, but I have independent reasons for thinking theory X is false, therefore it has not been shown that Matilda does not have knowledge, really”. I will be considering DOCTOR in a variety of ways. First, I will give some more general reasons for doubting that Matilda has knowledge. Then I will show that from the perspectives of virtue theoretic account of knowledge and Craig’s practical explication of knowledge Matilda does not know that Derek has cancer. Even if Lackey has independent reasons for rejecting all the considerations put forward, it seems hard to believe that if all of these reasons and theories point to the fact that Matilda lacks knowledge in DOCTOR that Matilda would have knowledge. At the very least the onus would be on Lackey to do more work to prove that Matilda does have knowledge—i.e. more than the few off hand reasons she does give, and her intuitions.

### 3.1. Why Lackey’s Agent Fails to Have Knowledge

Before turning to Craig and virtue epistemology, in this sub-section I will suggest some reasons for thinking that the agent in Lackey’s counterexample does not actually have the purported knowledge that Lackey thinks the agent does. Importantly, part of my reasoning for claiming that the agent does not know is similar to Lackey’s justification for claiming that the agent does not have the epistemic right to assert.

So first, in DOCTOR, consider the content of what is to be known—viz. that Derek has pancreatic cancer—especially since Matilda is an expert oncologist it is not a straightforward proposition like “grass is green”, or “a square has four sides”. Like cancer itself, the content of the belief is complex and multifaceted,<sup>7</sup> in fact that is why multiple tests are run before a diagnosis can be made. Saying that an oncologist knows someone has cancer is not to say that the oncologist knows the truth value of a simple proposition. Thus, at least *prima facie*, Matilda does not actually know that Derek has cancer. The reason she does not know is because all that the

---

<sup>7</sup> This same type of issue holds for Lackey’s other counterexamples as well. For example, when one is speaking to the quality of a restaurant, or a student’s writing ability, these are things that are multifaceted and complex, not simple propositions Lackey (2011).

reliable testimony of Nancy conveys to Matilda is, at best, the vague understanding that there is something like pancreatic cancer present in the patient Derek, and, at worst, simply the truth-value of a statement.

Cancers are unique. They come in stages, involve different types of tumors, can be more or less deadly, sometimes they run the danger of spreading, and sometimes not. In order to really know, Matilda needs to know what Derek's cancer is. Nancy's testimony does not speak to that. Therefore, based simply on Nancy's testimony Matilda simply cannot know Derek has cancer because she does not know what Derek's cancer is. Lackey, inadvertently, draws attention to the issue under consideration here when, in justifying why Matilda lacks the epistemic right to assert, she states that "Matilda should be able to (at least partially) explain or justify the diagnosis of pancreatic cancer that she is offering to her patient" (Lackey 2011, 254). One way to understand the explication or justification of the diagnosis is to see it as explaining what the diagnosis is. In other words, to explain a diagnosis is to fill out the content of the proposition "Derek has pancreatic cancer".

Another way to think about what Lackey is saying when she is denying that Matilda lacks the epistemic right to assert is that she is, implicitly, appealing to a principle that is even stronger than KSA. The reason that Matilda lacks the epistemic right to assert is because she does not know what she is asserting. Matilda cannot explain or justify the assertion that Derek has pancreatic cancer because she does not know that he has cancer. That is to say, one has the epistemic right to assert only if one knows—i.e. knowing that  $p$  is necessary in order to have the epistemic right to assert that  $p$ . The fact that Matilda cannot explain or justify the diagnosis underscores the fact that she does not know what she is purportedly diagnosing, which, in turn, explains the impropriety of asserting the diagnosis. If that is right, then KSA has not been refuted, because Matilda does not know.<sup>8</sup>

A further issue is that it seems that the evidence needed to ground knowing that Derek has cancer needs to be fairly significant, especially considering Matilda is an expert oncologist. Again, Lackey draws attention to this increased demand for evidence. Being an oncologist carries with it "certain epistemic duties. [...] [T]hese responsibilities may include having reviewed the test results firsthand, possessing reasons for choosing one

---

<sup>8</sup> My thanks to an anonymous reviewer for drawing attention to the fact that I might not have been clear on this point. What I am suggesting is that because KNA-N fails in the case means that Matilda does not have knowledge, which means that DOCTOR is not a case where the agent has knowledge but lacks the right to assert and *eo ipso* means KSA is not refuted.

condition over another, knowing details about the size and nature of the cancer, and so on” (Lackey 2011, 254). Lackey believes that the aforementioned epistemic duties relate only to assertion. However, they seem much more like evidential requirements on knowledge, on the one hand—reviewing test results, possessing reasons. On the other hand, “knowing details about the size and nature of the cancer” speaks to not only the evidential requirements but also the previous issue about the complexity of what is involved in a doctor knowing that someone has cancer (Lackey 2011, 254).

While the above considerations are not definitive, they are at least some *prima facie* reasons for doubting that the agent in Lackey’s counterexamples, in general, and Matilda in DOCTOR, in particular, does actually have the knowledge required for her argument to work. I now turn to another consideration that suggests that it is far from clear that Matilda knows that Derek has cancer. What was argued above basically amounts to the fact that doctors need stronger justification than the word of a student—no matter how reliable it might be—to make a diagnosis. Thus, isolated secondhand testimony is insufficient for Matilda to know that Derek has cancer. More obviously, however, is the fact that Matilda is unable to make the same appropriate inferences from the reliably produced testimony of Nancy, which she would be if she actually *knew* that Derek had cancer—from looking at test results and so forth.

Imagine that when Matilda tells Derek he has cancer and he starts asking her questions: “Is his case terminal, is he going to need surgery or chemotherapy or both, how large is the tumor, is there a danger of the cancer spreading?” Based on Nancy’s testimony, Matilda cannot answer any of these questions. The reason she is unable to answer the questions is because she really does not know that Derek has cancer. In fact, the most natural response to Derek’s questions would be for Matilda to say “I don’t know”, and the reason she does not know the answers to his questions is because she does not really know he has cancer. Matilda may have a true belief based on Nancy’s testimony, but one of the things that makes knowledge more valuable than mere true belief is the ability to make appropriate inferences.

Lackey implicitly relies on these inferential considerations in a similar example in order to argue that someone similarly situated to Matilda lacks the epistemic right to assert. Lackey states:

Suppose, for instance, that he asks [Matilda] what exactly the ultrasound and MRI revealed, or how large his tumor is, or why she thinks it is pancreatic cancer. All she can say [...] is

that she had been told that he has pancreatic cancer [...] that she hadn't actually seen any of the test results herself, and that she has no additional information to offer about his particular diagnosis. Wouldn't [Derek] be entitled to resent [Matilda] under such circumstances, to feel that he has been epistemically cheated by his doctor who owes him more than a diagnosis grounded purely in isolated secondhand knowledge (Lackey 2013, 38)?

What the above quote suggests is a few things; first it underscores what was said above regarding the justification, and evidence, needed for a doctor to know someone has cancer. Second, it suggests that part of Derek's resentment involves Matilda's inability to make appropriate inferences based on Nancy's testimony. Finally, the reason that Derek is entitled to feel "epistemically cheated" is not merely because Matilda lacks the epistemic right to assert, but, more importantly, the information that Matilda is conveying to Derek does not meet the standards for knowledge. Derek is being cheated because Matilda is trying to pass off true belief as knowledge—i.e. Matilda is giving Derek something less epistemically valuable (a true belief) when what he deserves is something more epistemically valuable (knowledge).

Now one could possibly object by claiming something along the lines that DOCTOR is a peculiar case because of the context or stakes involved raises the demands for knowledge. However, even in low-stakes scenarios Matilda still does not know that Derek has cancer, based solely on Nancy's testimony. Later in her paper, in responding to possible objections to her thesis Lackey gives the following modified DOCTOR case. "Suppose, for instance, that instead of flat out asserting to Derek that he has pancreatic cancer in DOCTOR, she [Matilda] casually asserts this fact to her husband over dinner" (Lackey 2011, 272).<sup>9</sup>

Lackey thinks that in the modified DOCTOR case that Matilda actually has knowledge and the epistemic right to assert—at least intuitively. If what I have been arguing thus far is right, though, even when talking to her husband, Matilda still does not know. Imagine that when Matilda mentions Derek's cancer to her husband, he says to her "Is there anything you'll be able to do for him?" Matilda will be forced to answer "I don't know", because even in that context Matilda does not really know that Derek has

---

<sup>9</sup> Certainly, it would be inappropriate for Matilda to tell her husband that Derek had cancer based on the norms governing doctor-patient privilege. However, as I mentioned above those norms are not of concern for this article—only the epistemic norms governing assertion are of concern here.

cancer. Even if the low-stakes lowers the evidential requirements, the fact that she is unable to draw appropriate inferences is enough to justify the claim that Matilda does not know even when talking to her husband.

### 3.2. Edward Craig and Virtue Epistemology

Having given some more general reasons for doubting that the agent in Lackey's counterexamples does, in fact, have knowledge, I now turn to some more theoretical considerations. It seems that a virtue theoretic account of knowledge and Edward Craig's practical explication of knowledge give one reason to think that the agent in Lackey's counterexamples does not know what is being asserted.<sup>10</sup> Again, for simplicity, I will focus on DOCTOR. However, the same reasoning can be applied to her other cases.

#### 3.2.1. Virtue Epistemology

According to John Greco on a virtue theoretic account of knowledge  
S knows that *p* if and only if

1. *p* is true;
2. S believes that *p*; and
3. S believes the truth because S's belief is produced by intellectual ability. (Greco 2010, 12)

It seems that Matilda, in DOCTOR, does not satisfy all three of these conditions. Even if 1 and 2 are granted, it does not seem that 3 is satisfied.<sup>11</sup> In a sense, Matilda's intellectual abilities are involved—e.g. her ability to hear. However, on a virtue theoretic account it is not simply that any intellectual ability needs to be involved in order to attain knowledge, but the relevant intellectual ability needs to be involved in the right way.

---

<sup>10</sup> This line of argument is not definitive, but it does the work that it is supposed to do. For Lackey's counterexample to work it needs to be the case that it is obvious that Matilda has knowledge but lacks the right to assert. Presenting multiple theories that would indicate that Matilda does not have knowledge switches the burden to explain why, beyond the intuition. Of course, one might disagree with Greco or Craig, but that is actually fine. Simply the existence of such theoretical accounts undermine Lackey's claim. My thanks to an anonymous reviewer for pointing out that my argumentative strategy might not have been clear on this point. The same reviewer also pointed out that from the perspective of "speech-act theory" there is something odd about Lackey's counterexample. As I am unfamiliar with the particulars of the theory, I have decided to leave considerations of that perspective to someone better suited to the task.

<sup>11</sup> I am not even sure that 1 and 2 are satisfied because of what was said above about the complexity of knowing something like someone has cancer. However, even granting 1 and 2 still does not get Matilda knowledge. Therefore, I will focus on 3.

In order to clarify 3, and why Matilda has not satisfied it, consider that 3 really amounts to something more like this: “when we attribute knowledge to someone we mean to give the person credit for getting things right. Put another way, we imply that the person is *responsible* for getting things right” (Greco 2003, 111; emphasis added). In other words, in order to be attributed knowledge one has to have done something creditworthy, or one has to have done something that makes one responsible for having the true belief. What creditworthiness, and responsibility, amount to can of course vary. However, it does not seem that Matilda has done anything creditworthy, nor is she really responsible for having a true belief regarding Derek’s cancer.<sup>12</sup>

If anyone deserves credit, it would be Nancy. Nancy used her excellent visual, reasoning and other abilities in deciding that Derek has pancreatic cancer. Those are the type of abilities relevant for a cancer diagnosis. Matilda used no abilities like that; she simply believed what Nancy told her. Thus, Matilda is neither creditworthy nor responsible for the true belief, and therefore does not have knowledge.

To underscore the point that Matilda is neither creditworthy nor responsible reduce Nancy’s reliability for cancer diagnosis. Even if it is intuitively plausible that Matilda could have knowledge based on Nancy’s highly reliable testimony, as one reduces Nancy’s reliability there will come a point where Matilda does not have knowledge. Further, other than a naïve evidentialist, that point will probably far exceed a .5 probability.<sup>13</sup> Here is the idea. Nancy’s testimony is too unstable to transfer credit or responsibility to Matilda. If Nancy’s testimony is too reliable, the credit that accrues to Nancy will swamp any credit that could be assigned to Matilda. If Nancy’s testimony is too unreliable, then a true belief is not transferred at all.

More importantly, what the variability of Nancy’s reliability illustrates, is precisely what the relevant abilities are, and how a doctor can be responsible for knowing someone has cancer. When Nancy is highly reliable it is because of her education, specialized visual abilities needed for reviewing test results and excellent reasoning abilities, for example. When Nancy is less reliable, it is because those same abilities are not as good. Further, when Nancy is less reliable one does not give credit to Matilda because Matilda did not exercise the abilities—education, visual

---

<sup>12</sup> Again, it is not even clear to me that Matilda has a true belief, but, for the sake of argument, it will be granted.

<sup>13</sup> Actually, to me even 100% reliability would not meet the standards necessary. For precisely the reasons which will be explained.

and reasoning abilities—and in fact would be blamed for not using them. By the same reasoning then, even when Nancy is highly reliable, Matilda does not deserve credit nor is she responsible precisely because Matilda did not exercise the relevant abilities.

Another way to think about the issue under consideration here—that Matilda is not creditable—is to note that Matilda’s abilities are not a salient part of the causal story leading to the true belief. Greco discusses two ways that something can be a salient part of the causal story. Only one of which is relevant here. According to Greco, a “major factor governing salience is our interests and purposes” (Greco 2003, 118). The interests and purposes involved in a cancer diagnosis, predominately, all relate to knowing what to do with the diagnosis. Based on what was said above regarding Matilda’s inability to make appropriate inferences, and Derek feeling epistemically cheated, it seems clear that Matilda’s abilities—e.g. hearing—do not serve the relevant interests and purposes.

Alternatively, and relatedly, one can think of the creditability along more Sosaian lines and see that Matilda is not creditable because she has not manifested her abilities *qua* oncologist (Sosa 2007). Remember that Matilda is an expert oncologist giving a cancer diagnosis. Therefore, even if Matilda has a true belief her coming to hold that true belief does not manifest her abilities. It does not manifest her abilities because, on the one hand, the nature of what is to be known, demands certain abilities—like the ones mentioned above—and, on the other hand, it does not manifest her abilities as an expert oncologist. Thus, again, Matilda is not creditable for the true belief, and therefore does not have knowledge.

Finally, consider Linda Zagzebski’s virtue epistemology. Unlike Greco and Sosa who understand epistemic virtues more along the lines of a type of reliabilism—i.e. cognitive abilities like reasoning, and perception—Zagzebski understands virtues more along the lines of traditional virtue ethics, and she refers to them as “intellectual virtues”. Intellectual virtues are motives or dispositions like intellectual courage—standing by one’s beliefs—intellectual humility—being open to contrary evidence—and so forth. For Zagzebski, “[k]nowledge is belief arising out of acts of intellectual virtue” (Zagzebski 1999, 109). In DOCTOR, Matilda, by basing her diagnosis on Nancy’s testimony, was exercising a kind of intellectual vice—something like intellectual laziness. Thus, Matilda fails to know that Derek has pancreatic cancer if knowledge is understood along Zagzebski’s line because her belief did not arise out of an act of intellectual virtue. In fact, Matilda’s belief arose out of an act of intellectual vice, which explains, in part, why Derek would feel epistemically cheated.

### 3.2.2. Edward Craig and the Practical Explication of Knowledge

Having discussed why Matilda fails to have knowledge according to a virtue theoretic account of knowledge, I now turn to Edward Craig's practical explication of knowledge. According to Craig, "the concept of knowledge is used to flag approved informants" (Craig 1986-7, 215). While that is Craig's, so to say, definition of knowledge, what is helpful for present purposes is "that the principal candidates [for] the analysis of the everyday concept of knowledge all lie very close to the concept constructed by adopting the point of view of the inquirer" (Craig 1986-7, 225).

By adopting the point of view of the inquirer, if one wants to know whether Matilda knows Derek has pancreatic cancer, then one needs to adopt Derek's perspective. Fortunately, Lackey has explained what Derek's perspective is, or would be.

Derek reasonably has the right to expect his doctor to fulfill such a duty [viz. reviewing test results firsthand, possessing reasons for choosing one condition over another, knowing details about the size and nature of the cancer]. [...] Wouldn't Derek be entitled to resent Matilda under such circumstances [i.e. not fulfilling here epistemic duties], to feel that he has been epistemically cheated by his doctor who owes him more than a diagnosis grounded purely in isolated secondhand knowledge (Lackey 2011, 6-7).

The issues in the above quote have been discussed already, but Craig gives them some more theoretical substance.

If one, again, imagines that Derek begins asking Matilda questions about the diagnosis, and Matilda is unable to respond, at first Derek would probably be a bit confused. Matilda then "reveals to Derek that she had been told that he has pancreatic cancer from her student Nancy, that she hadn't actually seen any of the test results herself, and that she has no additional information to offer about his particular diagnosis" (Lackey 2011, 6). It seems the most natural response that Derek might give—beyond resenting Matilda and feeling epistemically cheated—would be something like: "So, you really don't know I have cancer. Your student believes I may have cancer. That's what you're telling me".

What Craig's practical explication of knowledge gives is yet another reason to doubt that Matilda both has knowledge and lacks the epistemic right to assert. By adopting the perspective of the inquirer, namely Derek,

it becomes abundantly clear that what is epistemically problematic in DOCTOR is the fact that Matilda is asserting something she does not know, as if she does know it. It is not, as Lackey maintains, merely that Matilda lacks the epistemic right to assert—which she does—but the reason why she lacks the epistemic right to assert is, in part, that KSA, or something like it, holds and Matilda lacks knowledge. At a minimum, the way Lackey explains the fact Matilda lacks the epistemic right to assert seems much more like she is explaining why Matilda does not know, but if that is right then she clearly has not shown that KSA is false.

A further implication of Craig's view is that knowledge flags good informants for actionable information.<sup>14</sup> Thus, part of the problem in DOCTOR is that merely asserting to Derek that he has pancreatic cancer is not an actionable piece of information. He does not know if he needs more tests, or needs to start taking medicine, or change his diet, or prepare for surgery. What would make the information actionable is answers to the types of questions he would ask that Matilda cannot answer. Therefore, yet again, in order to know Matilda would have to have looked at the test results and so forth, in order to make the information she was asserting actionable for Derek, and thereby knowledge on her part.

All of the considerations put forth in section 3 may not, individually, tell against Lackey's counterexamples.<sup>15</sup> However, by taking everything together it seems clear that the assertor in Lackey's counterexamples does not actually have knowledge, and therefore does not disprove KSA.<sup>16</sup> Lackey is correct that the assertor in her counterexamples lacks the epistemic right to assert, but she is wrong about why they lack that right. It is unclear why Lackey would think they do not have the epistemic right to assert, but the real reason is, in part, because of KSA, or something like it, and the fact that the assertor does not know what they are asserting. Still, even if Lackey is not relying implicitly on KSA it does not seem that she has shown that KSA is false.

### 3.3. A Fall Back Position

All the reasons discussed above seem to point to the fact that the agent in Lackey's counterexamples does not have knowledge, demonstrating that Lackey has failed to disprove KSA. However, some still might not think that anything that has been said actually *proves* that the agent lacks

---

<sup>14</sup> John Greco has brought this to my attention by.

<sup>15</sup> I actually think many of them do, but the strength of the argument against Lackey is the force of everything taken together.

<sup>16</sup> Recall, I have focused on DOCTOR for simplicity. The same reasoning can be equally applied to all of her counterexamples.

knowledge. In this sub-section of the article, I present a, kind of, fallback position. The idea is that even if I have not established that the agent does not know, the reasoning above makes it plausible that the agent might not know. If the agent might not know, that is sufficient to undermine the force of Lackey's counterexamples, which, in turn, gives one good reason to think that she has not really established that KSA is false.

Lackey has made a very bold claim. The problem is that bold claims need to be well established. Lackey is not unaware of the fact that she is endorsing a rather strong position. In order to falsify KSA Lackey needs to, in her words, establish "that there are various kinds of cases in which a speaker asserts that *p*, clearly knows that *p*, and yet does not have the proper epistemic authority or credentials to make such an assertion" (Lackey 2011, 253). The crux is that the speaker, or agent, asserting "*clearly knows that p*". It could be argued that the above considerations do not establish that the speaker, or agent, in her counterexample clearly does not know that *p*. However, they certainly show that the speaker, or agent, does not *clearly* know that *p*. Yet, clearly knowing that *p* is exactly what Lackey needs to draw the conclusion that she does. Therefore, the onus is on Lackey to do much more work to establish that it is clear that the agent has knowledge in the cases she describes. Until she has done that, there is little motivation to accept her conclusion.

Another fallback position is to point out that Lackey's counterexamples, especially DOCTOR, are too unrealistic to really have enough force to undermine a general norm of assertion, like KSA. At least intuitively, it does not seem reasonable that an expert oncologist, like Matilda, would base a cancer diagnosis solely on the testimony of a student, no matter how reliable the student might be. At least intuitively, it does not seem that Matilda would take herself to know that Derek has cancer, or at least know in the right way necessary for assertion, based simply on the word of her student.

If Nancy were not a student, but rather another competent oncologist, perhaps Matilda would accept that as sufficient for knowing that Derek has cancer. However, then the intuition that Matilda lacks the epistemic right to assert is not so clear. On the other hand, if Nancy were another expert oncologist it does not seem that the scenario of DOCTOR, *mutatis mutandis*, would play out exactly the way Lackey describes. It seems reasonable to think that instead of flat out asserting to Derek that he has cancer, Matilda might add the preamble that her colleague had looked at Derek's test results and concluded that Derek had cancer, or she might bring Nancy in to meet with Derek as well. But by adding the preamble or bringing Nancy with her to the meeting, which seems like at least

something that Matilda might do, points to the fact that Matilda might not take herself to actually know that Derek has cancer, and hence might just be more evidence that it is not clear that Matilda knows.<sup>17</sup>

Many of Lackey's other counterexamples seem a bit unrealistic for the same type of reasons. Consider two others, PROFESSOR and FOOD.

PROFESSOR: Judith is a professor at one of the best law schools in the country, and today's lecture is on U.S. copyright law. While she is generally quite knowledgeable of this topic, she has failed to keep up with some recent developments in this area. Over lunch yesterday, one of her colleagues briefly expressed his belief that it is extremely improbable that the Supreme Court will consider a case challenging the addition of 20 years to the original copyright protection of 50 years after the death of authors. Though Judith does not know any of the reasons or considerations underlying this claim, she asserts to her students in class, "The Supreme Court is unlikely to hear the upcoming challenge to the recent extension of U.S. copyright protections to 70 years after the author's death". While this assertion is in fact true, it is based purely on the basis of the reliable testimony of Judith's colleague (Lackey 2011, 254).

FOOD: My neighbor Ken is a connoisseur of fine dining. As we were leaving Starbucks this afternoon, he told me that the food at a new local restaurant about which I was previously quite unfamiliar, Quince, is exquisite, though being in a hurry prevented him from offering any details or evidence on behalf of this claim. While talking to my friend Vivienne later in the day, she was fretting over where to take her boyfriend to dinner for Valentine's Day. I promptly relieved her stress by truly asserting, "The food at Quince is exquisite" (Lackey 2011, 257).

It seems that it is more realistic for the speakers in each of the cases to preface their assertion with something like "I have been told that [...]" or "According to [...]", than for the speakers to just flat out assert what Lackey has them assert. If it is likely that the speakers would preface their

---

<sup>17</sup> Actually, it is not unreasonable that Matilda might not do something like this even when Nancy is just a student. Instead of flat out asserting to Derek that he has pancreatic cancer, she would instead reschedule his appointment until she has looked at the test results, or preface the assertion with something like "It appears that [...]" and then add "[...] but I will need to review the results more carefully".

assertion, then it seems reasonable to think that the speaker does not take themselves as clearly knowing the propositions under consideration.

Here, I wanted to just draw attention to the fact that even if the arguments put forward in 3.1-2, are not as convincing as I take them to be, they are still sufficient to undermine Lackey's argument. Lackey needs it to be the case that the agent in her counterexamples clearly knows, but she has not demonstrated that the agent clearly knows. Since the agent does not clearly know her conclusion does not follow; at least it does not follow in the right way to truly undermine KSA. I also drew attention to the fact that Lackey's counterexamples, as presented, seem highly unrealistic. Making them more realistic seems to make the intuition that the agent in her counterexamples both has knowledge and the epistemic right to assert less clear, which is sufficient to undermine the force of her argument.

#### 4. What Lackey's Argument Might Show

Despite the fact that Lackey has failed to prove that KSA is false, her paper is instructive. What Lackey's paper points to is the inadequacies of various epistemological theories. Before concluding, then, I will briefly discuss the positive upshot of Lackey's paper. It will be suggested that what Lackey's counterexamples demonstrate are problems with simplistic evidentialist and process reliabilist theories of knowledge.

For evidentialism, knowledge is, roughly, simply justified true belief. Further, "S is justified in believing *p* at [time] *t* if and only if S's evidence for *p* at *t* supports believing *p*" (Mittag 2020).<sup>18</sup> What counts as "support" varies according to different theories of evidentialism, and can be context sensitive. Further, support basically amounts to something like the evidence makes it highly likely that *p* is true. Imagine, as Lackey does in responding to some objections, that Nancy and her testimony might be quite truth conducive. Here is the problem then for the evidentialist if Nancy's testimony is extremely truth conducive, then Matilda is going to get knowledge. Yet, if what has been said above is right, then Matilda does not have knowledge. Thus, Lackey has not shown that KSA is false, but that a straightforward evidentialism does not do justice to how one understands whether or not someone can know—at least not in all cases.

The same type of argument can be made against a simplistic form of process reliabilism. According to process reliabilism, "S knows that *p* if

---

<sup>18</sup> Mittag's Encyclopedia entry is used here for efficiency. For more complete accounts of evidentialism, see, for example, Conee and Feldman (2004).

and only if *S* believes that *p*, *p* is true, and *S*'s belief that *p* is formed by a reliable process" (Becker 2020).<sup>19</sup> What counts as a reliable process can vary, but is, roughly, a process that is highly truth conducive. Therefore, if Nancy and her testimony are highly reliable, then Matilda will get knowledge based on Nancy's testimony, but that is the wrong result. The problem is thus with process reliabilism, not KSA.

What both evidentialism and process reliabilism have in common is their focus on truth-conduciveness, and an implicit acceptance of knowledge being about simple propositions. Because of this—i.e. the focus on truth conduciveness and acceptance of knowledge being about simple propositions—evidentialism and reliabilism, at least simplistic forms of them, are not fine grained enough to notice that things like cancer diagnoses are more complex than a simple true or false proposition. The upshot is that something like a virtue theoretic account, whether a virtue reliabilism like Greco and Sosa or a virtue responsibilism like Zagzebski, has more theoretical power, and thus, is plausibly a superior epistemological theory. That, then, is what the real lesson of Lackey's paper is.

Now, if one did not find the arguments put forth in 3.1.-2. to convincingly show that the agent in Lackey's counterexamples lacked knowledge, then what has been said in this section might not be convincing either. On the other hand, what has been said in this section can still be instructive. At most what Lackey has shown with her counterexamples is that KSA and simplistic evidentialist, or process reliabilist, epistemology, are at least *prima facie* incompatible, or more strongly KSA is perhaps false only if one holds a simplistic evidentialist or process reliabilist epistemology. That in itself—KSA being perhaps false, conditionally—is telling against Lackey. Remember, Lackey needs it to be the case that the agent in her counterexample clearly does not know, but if her argument only works on certain assumptions about what counts as knowledge, then she has failed to make the case that KSA is actually false.

## 5. Conclusion

In this article, I have argued that Jennifer Lackey has failed to prove that the KSA is false. According to the KSA, if a subject *S* knows that *p*, then *S* has the epistemic right to assert that *p*. Lackey's argument proceeded by

---

<sup>19</sup> As with Mittag (2020), Becker's Encyclopedia entry is used here for the sake of efficiency. The literature on Reliabilism is vast, its foremost proponent being Goldman (1979), but see also Heller (1995), and Becker (2006).

putting forth various counterexamples where it seemed, at least intuitively, that a subject both knew that  $p$  but lacked the epistemic right to assert that  $p$ , thereby falsifying KSA. It was shown that, in fact, the subject under consideration actually did not know. Since the subject lacked knowledge, the subject's lack of an epistemic right to assert was not proof that KSA is false, but rather, perhaps, that KSA is true. The idea is that if the agent lacks the epistemic right to assert, then the agent fails to know, by *Modus Tollens* and KSA. However, at a minimum, it seems that Lackey is implicitly working with something that corresponds to KSA, where one can assert that  $p$  only if one knows that  $p$ . Since the subject in Lackey's counterexample lacks the epistemic right to assert then the subject does not know; this seems to be the case because the reasons that Lackey gives for claiming the subject in her counterexamples does not have the epistemic right to assert are better understood as reasons for thinking the agent does not know. Notice that, even if Lackey is not relying on KSA, if the agent in the counterexample does not know then she has not falsified KSA.

To avoid the quagmire of intuition bumping, it was demonstrated that from a variety of perspectives, and for a variety of reasons the agent in Lackey's counterexamples does not know. First, it was shown that, for example, cancer diagnoses are complex and require special evidence. Second, it was shown that part of what distinguishes knowledge from mere true belief is the ability to make appropriate inferences. Then, it was shown that from the perspective of a virtue theoretic account of knowledge and Edward Craig's practical explication of knowledge the agent in the counterexamples lacks knowledge.

Finally, it was concluded that although Lackey failed to successfully prove that KSA is false, there were some valuable lessons. It was suggested that a simplistic evidentialism and process reliabilism, are theoretically less powerful than a virtue epistemology. Thus, what Lackey truly gives in her paper is a *prima facie* argument in favor of, perhaps, something like a virtue epistemology.

## REFERENCES

- Becker, K. 2020. Reliabilism. *Internet Encyclopedia of Philosophy*. <https://www.iep.utm.edu/reliabil/>. Accessed January 4, 2020.
- Becker, K. 2006. Reliabilism and safety. *Metaphilosophy* 37: 691-704.
- Conee, E., and R. Feldman eds. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Clarendon Press.
- Craig, E. 1986-7. The practical explication of knowledge. *Proceedings of the Aristotelian Society* 87: 212-226.

- DeRose, K. 2002. Assertion, knowledge, and context. *The Philosophical Review* 111: 167-203.
- Goldman, A. 1979. What is justified belief? In *Justification and Knowledge*, ed. G. Pappas, 1-24. Boston: D. Reidel.
- Greco, J. 2003. Knowledge as credit for true belief. In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, eds. M. DePaul and L. Zagzebski, 111-134. Oxford: Oxford University Press.
- Greco, J. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hawthorne, J., and J. Stanley. 2008. Knowledge and action. *The Journal of Philosophy* 105: 571-590.
- Heller, M. 1995. The simple solution to the generality problem. *Nous* 29: 501-515.
- Lackey, J. 2007. Why we don't deserve credit for everything we know. *Synthese* 158: 345-361.
- Lackey, J. 2011. Assertion and isolated secondhand knowledge. In *Assertion: New Philosophical Essays*, eds. J. Brown and H. Cappelen, 251-276. Oxford: Oxford University Press.
- Lackey, J. 2013. Deficient testimonial knowledge. In *Knowledge, Virtue, and Action: Putting Epistemic Virtues to Work*, eds. T. Henning and D. P. Schweikard, 30-52. New York: Routledge.
- Mittag, D. M. 2020. Evidentialism. *Internet Encyclopedia of Philosophy*. <https://www.iep.utm.edu/reliabil/>. Accessed January 4 2020.
- Sosa, E. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford: Oxford University Press.
- Simion, M. 2016. Assertion: Knowledge is enough. *Synthese* 193: 3041-3056.
- Zagzebski, L. 1999. What is knowledge? In *The Blackwell Guide to Epistemology*, eds. J. Greco and E. Sosa, 92-116. Malden, MA: Blackwell Publishers.

## RIGHT TO BE PUNISHED?

Adriana Placani  
University of Graz

Stearns Broadhead  
University of Graz

Original scientific article – Received: 12/02/2020 Accepted: 7/6/2020

### ***ABSTRACT***

*It appears at least intuitively appropriate to claim that we owe it to victims to punish those who have wronged them. It also seems plausible to state that we owe it to society to punish those who have violated its norms. However, do we also owe punishment to perpetrators themselves? In other words, do those who commit crimes have a moral right to be punished? This work examines the sustainability of the right to be punished from the standpoint of the two main theories of rights—the will and the interest conceptions. The right to be punished is shown to be largely indefensible on both accounts: on the will theory, the right to be punished conflicts with autonomy, and it can neither be claimed nor waived by a perpetrator; on the interest theory, a perpetrator's interest in punishment, inasmuch as it exists, is not sufficient to ground a duty on the part of the state.*

**Keywords:** *Punishment; right to be punished; Duff; communicative theory of punishment; rights; will theory; interest theory*

### **Introduction**

It seems at least intuitively right to claim that we owe it to victims to punish those who have wronged them. It also seems plausible to say that we owe it to society to punish those who have flouted its norms. However, can it be said that we owe punishment to perpetrators themselves? That is, and

this might sound odd at first, can we assert that those who commit crimes have a moral right to be punished? The idea of a right to be punished has a long history<sup>1</sup> that stretches into the present. Contemporary advocates of the right include Christopher Bennett (2008), Thom Brooks (2013), Gabriel Hallevy (2013), and Antony Duff (1986; 2000; 2001; 2003).

This article considers the defensibility of the right to be punished from the standpoint of the will and interest theories of rights, and shows the right to be largely indefensible on both theories. On the will theory, the right to be punished conflicts with autonomy and a perpetrator can neither claim nor waive it. On the interest theory, a perpetrator's interest in punishment, inasmuch as it exists, is not sufficient to ground a duty on the part of the state. Although this work primarily focuses on Duff's conception of punishment and perpetrators' right to it (to the exclusion of other conceptions), its analysis brings much needed clarification to an idea oftentimes plagued by loose and vague language.

The structure of this essay is as follows. The first section explicates the right to be punished and the role of this particular right in Duff's communicative theory of punishment, which is selected because of its considerable influence in the contemporary literature. This section also clarifies other main concepts and definitions used throughout. The second section argues against the idea that perpetrators have a right to be punished. It shows that the will-based theory cannot plausibly sustain the right. It also briefly addresses aspects of Duff's communicative theory of punishment and points out how arguments against the right to be punished affect it. Much like the second section, the third argues against the right to be punished, but it does so in accordance with the interest-based theory. This section also connects these arguments to Duff's broader account. Finally, a short conclusion ends this work.

## **1. Anthony Duff's Theory and the Right to be Punished**

The right to be punished seems like the height of humanity to some and the pinnacle of preposterousness to others. Those in the first camp often claim that the right to be punished arises out of the idea that all persons, including criminals, ought to be treated as responsible moral agents to whom respect is owed (Dubber 1998). On this view, then, punishment broadly amounts to a way of showing respect for fellow human beings. Those in the second camp often begin their attacks by pointing out that the right to be punished

---

<sup>1</sup> See Dubber (1998) for an overview on the philosophical history of the right to be punished.

makes very little sense (see Bagaric and Amarasekara 2000; Deigh 1984; Feinberg 1974). This is because it seems implausible that a right-holder would exercise a right that would deprive her of liberty and cause her suffering. It seems more sensible that there would be a right *against* such things.<sup>2</sup>

A contemporary advocate of the right to be punished is Antony Duff, whose communicative theory of punishment includes this right. His theory can be summarized as follows. For Duff, punishment is and ought to be a communicative process aimed at engaging the perpetrator in a moral dialogue about her crime (2011, 372). Punishment endeavors to communicate to the perpetrator a proper understanding of the wrongdoing she committed and bring her (i.e., the offender) to communicate that repentant understanding to her fellow citizens (Duff 2000, 412).

In *Trials and Punishment*, Duff writes that a

sane offender has a right to be punished: a right to be punished rather than be subjected to some kind of manipulative or preventive treatment which would not address him as a rational agent. (Duff 1986, 283)

In more recent works, Duff's emphasis on this right is lessened, but remains a part of his theory. In his article, "In Defense of One Type of Retributivism", Duff clarifies his previous elaborations on the right to be punished:

To talk of a 'right' to be punished is to talk of punishment as something that is owed to the offender (not just, for instance, to the victim or the wider community), and as something that is supposedly for her own good (not just as something that she deserves). Perhaps it is also to imply that punishment is something that the offender would claim for herself, if she realised the truth. (Duff 2000, 418)<sup>3</sup>

Duff seems to argue in favor of the right to be punished from a number of vantage points. First, we owe punishment to a perpetrator out of respect for her as a responsible moral agent. Society owes it to a perpetrator to treat

---

<sup>2</sup> See Husak (2008, 92-103) arguing about a right *not* to be punished.

<sup>3</sup> Note that Duff attempts to persuade the reader further by comparing his reliance on the right to be punished to alternatives such as ignoring perpetrators' crimes or subjecting perpetrators to measures aimed at preventing them from repeating their wrongdoings. To this, it is unclear both why the right should be considered only in light of alternatives and why it should be considered only in light of the alternatives suggested by Duff.

her as a responsible member of the political community (Duff 2003, 305). Second, punishment is also owed to the perpetrator because it is supposedly for her own good (as opposed to being something that she deserves) (Duff 2000, 417-418). Finally, punishment is something that the perpetrator would claim for herself (if she realized that it was for her own good) (Duff 2000, 418). These standpoints can be generalized so as to fit with the two main theories of rights on the basis of which the right to be punished will be assessed.

Before going further, however, the purpose of punishment can be clarified. After all, whether or not there is a right to be punished depends in part on what that right amounts to—a right to what exactly? In keeping with Duff as well as others who have focused on this topic, it is only legal punishment that is of concern here (i.e., certain negative treatment that can and does follow from legal proceedings) (Williams 1955, 123). Punishment arising in family life, school, or other personal relationships falls outside the general purview.

Since Duff's conceptions will be in focus, it is useful to put forward his definition of punishment. According to Duff, punishment is (typically) something intended to be burdensome and painful, and it is imposed on a supposed offender for a supposed offense by someone with the supposed authority to do so (i.e., the state, state-constituted institutions, agents of the state) (Duff 2001, xiv-xv).<sup>4</sup>

With this definition of punishment in place, the answer to the above question—'a right to what?'—can be phrased in the following way: The right to be punished is the right of perpetrators to the imposition by state-constituted institutions of what are generally regarded as burdensome and/or painful consequences for an offense. Burdensome and/or painful consequences typically range from community service, forced incarceration, all the way to the death penalty.

Having provided the contours of Duff's conception of punishment alongside its embracement of the right to be punished, we may now consider the right from the standpoint of the two main theories of rights: the will and interests conceptions.

---

<sup>4</sup> Duff acknowledges that whether pain is a necessary feature of punishment is a controversial matter (see Adler 1992).

## 2. The Will Theory of Rights

On the will theory, rights serve the function of protecting and fostering the autonomy of the right-holder. According to H. L. A. Hart, whose writings are the *locus classicus* of the will theory, rights count as such only when the right-holder has the option of enforcing or waiving the duty correlative to the right (Hart 1982, 174-193). Famously, the right-holder is a “small-scale sovereign to whom the duty is owed” (Hart 1982, 183). Exercising a right is the measure of the right since “it is hard to think of rights except as capable of exercise” (Hart 1982, 183). One has a right only when one has a degree of control or when one may exercise a choice over the performance of the duty that correlates with the right. On the will-based theory, rights are justified and should be recognized if and only if they protect the exercise of autonomous choice (Hart 1982, 183).

Rights on the will theory serve to foster the autonomy of right-holders. This involves the right-holders’ capacity to exercise their rights (e.g., to claim, enforce, leave unenforced, extinguish or waive them) (Hart 1982, 184). Furthermore, following Joel Feinberg, “having rights, of course, makes claiming possible; but it is claiming that gives rights their special moral significance” (1980, 195). In other words, what makes an agent a holder of a right on the will theory is the agent’s degree of control over the correlative duty that she is owed.

### 2.1. Against the Right to be Punished on the Will Theory of Rights

This section shows that we have good reasons to doubt that the right to be punished can be sustained on the will theory. First, as a coercive practice under the control of the state, punishment does not foster the autonomy of those onto whom it is imposed. Given the coerciveness of punishment, the right cannot be exercised by perpetrators. They can neither claim nor waive their purported right because they lack control over the infliction of punishment. Second, punishment as a threat against would-be perpetrators does not ground a duty on the part of the state to punish perpetrators. This is not to deny that the state has a duty to punish those who violate the law, rather it is to say that such a duty is not owed to law-breakers. It should be borne in mind that the following advances a two-pronged argument against the will theory, which means that the two critiques are meant to be considered as parts of a single, complex argument, highlighting different aspects of examination.

To see why punishment is said to be within the control of the state, as opposed to being within the control of offenders, consider the following: Punishment has been defined as the imposition by someone with

presumptive authority (i.e., the state, state-constituted institutions, agents of the state) of something intended to be burdensome, unpleasant or painful on a supposed offender for a supposed offense (Duff 2001, xiv-xv). In accordance with the definition, the source of legal punishment is the state, since the state holds the *de facto* and (presumptive) *de jure* authority to inflict punishment.

Through its various institutions, the state formulates and enacts rules that, when abrogated, can make one liable to punishment. Once legal rules are in place, their violation can set in motion the criminal process. This process typically involves various stages: arrest; arraignment; pleading; pre-trial motion; criminal trial; sentencing; and, so on. At the trial stage, the judge or jury finds the defendant to be guilty or not guilty (excluding other possibilities, such as mistrials). At the sentencing stage, the appropriate punishment for the convicted is determined. This indicates that agents of the state exercise control over the entire criminal process from arrest to punishment. This, however, does not imply that perpetrators are entirely shut out of legal processes (e.g., they can be heard at trial). Nonetheless, it does suggest that both the liability to incur punishment as well as the actual incurrence of punishment lie within the power of the state.

The kind of control that the state possesses makes punishment a coercive practice, which is to say that punishment is inflicted onto offenders regardless of their will (whether they wish it or not). Furthermore, legal punishment is the most severe form of state coercion, a fact not debated even by those who think it justified (Dubber 2007, 2597). As Hans Kelsen wrote, the coercive nature of the law lies in the fact that it inflicts

an evil—such as deprivation of life, health, liberty, or economic values—which, if necessary, is imposed upon the affected individual even against his will by the employment of physical force. (Kelsen 1967, 33)

Given that the imposition of legal punishment does not take into account the will of the perpetrator, such coerciveness is at odds with perpetrators' autonomy. This represents the first mark against the right to be punished: perpetrators lack relevant control. Even if perpetrators had control over their own punishment, it is hard to imagine circumstances in which an agent would ever claim her right to be punished. Perhaps cases in which some sort of extrajudicial treatment (e.g., lynch mobs) would be more fearsome than legal punishment are plausible examples. In general, though, self-interest would advocate against demanding legal punishment, which would likely be harmful in the short and long term. It is more likely that

perpetrators would want to waive their right to be punished but any such attempt would be vacuous.

Still, let us consider in what sense a perpetrator could claim the right to be punished and whether the claim would be valid. This takes us to the next prong of the argument. According to the correlativity thesis, claiming this right would be claiming that the state has a duty to punish the perpetrator. With respect to the right to be punished, at least, it is very difficult to discern why such a duty would be owed by the state. The advocate of the right to be punished, abiding by will-based precepts, might answer that a perpetrator who chooses to commit a crime is entitled to punishment because of that choice—that is, because punishment would respect the offender’s will and further her autonomy. However, willfully breaking the law does not *necessarily* mean that a perpetrator wills her own punishment. It is possible (and likely) for an agent to choose to commit a crime all the while not choosing the negative consequences of this (i.e., punishment). That is, in such cases, an agent would simply accept the risk of being punished, not choose to be punished. Accepting the risk that X may occur is not the same as choosing, consenting to, or willing the actual occurrence of X (Thompson 1986, 188). In the case of punishment, it seems plausible (if contentious<sup>5</sup>) that offenders are simply hoping not to get caught. The agent is also likely to engage in calculations about the possibility of getting caught and go ahead with the perpetration only if the magnitude of the risk (likelihood of getting caught coupled with the severity of punishment) does not outweigh her expected reward from engaging in the perpetration.

In this respect, having an autonomy-furthering right to choose (even among unlawful options) does not necessarily confer a right to be punished for those choices. If a perpetrator freely chooses to violate a legal rule, she does so knowing that punishment *can* follow. However, it seems erroneous to conclude that because punishment follows from an individual’s wrongful conduct that the state thereby owes a duty to that individual because of her wrongful conduct. This is the case even when such wrongful conduct was freely performed or omitted, and it was known to be wrong at the time of its commission.

The above is in line with a particular view about the nature of legal punishment, which construes the latter as a threat made by the state in order to provide agents with prudential reasons not to break the law. This view is advocated not just by those with utilitarian inclinations (regarding the justification of the institution of punishment), but also by expressivists,

---

<sup>5</sup> See Nino (1983) and Alm (2018) for alternative views about consenting to the liability to being punished.

such as Joel Feinberg (1974, 73). If this view is correct,<sup>6</sup> then advocates for the right to be punished misconstrue the nature of punishment itself by (implicitly) regarding the latter as a promise whose fulfillment is owed to perpetrators by the state. Notice about threats that there is no moral obligation to carry them out. As such, the state's threat to punish may not ground a duty (to perpetrators) to punish (Downie 1985, 266).

Again, one peculiarity of the right to be punished is that it is putatively held by a perpetrator, and the correlative duty is owed by a state to that individual. This putative right–duty nexus can easily be misinterpreted as the more plausible thesis that a right is held by the aggrieved. In this sense, the right is rendered as ‘victims have the right to have their victimizers punished’. Clearly this version, which provides justice for victims as right-holders, is not the one that is under consideration.

Although Duff believes that not punishing a perpetrator would treat her as less than an autonomous and morally responsible agent, there are other forms of treatment (Duff 2000, 418). These could aim to restore the agent and help guide her future conduct without appeal to suffering. There is nothing in some alternatives to punishment that renders them necessarily incapable of treating agents with due respect. Whether punishment, too, can treat agents with due respect is at the heart of its justifiability and must be shown to be the case. The case of alternative responses (other than punishment) will be addressed in the next section as well.

The two sources of criticism advanced in this section show that the right to be punished encounters serious difficulties on the will theory. The first—the coerciveness of punishment seriously diminishes the autonomy of perpetrators—coupled with the second—there seems to be no duty owed to perpetrators by the state—offer good reasons to doubt that the right can be sustained. However, this does not yet spell doom for the right to be punished. The interest theory might be able to provide it with proper grounding. This possibility is considered next, and it is rejected.

---

<sup>6</sup> Here, Lipkin's (1988, 92–93) criticism of Duff's communicative theory is helpful in pointing out the advantages of prudential theories. Lipkin criticizes Duff's view on the grounds that it fails to provide a general motive to obey the law, and asserts the superiority of prudentially-based theories that are able to draw in every rational person regardless of their morality by providing them with prudential reasons for obedience.

### 3. The Interest Theory of Rights

Within the interest theory, rights are grounded in the interests of the right-holder. On one of the most prominent versions of the theory, Joseph Raz's account,

'X has a right' if and only if X can have rights, and, other things being equal, an aspect of X's well-being (his interest) is a sufficient reason for holding some other person(s) to be under a duty. (Raz 1986, 166)

Broadly speaking, rights on the interest theory serve to benefit right-holders. If the right to be punished is to be sustained on this theory, it is by virtue of perpetrators' interests in being punished. Notice that such interests must be sufficient to ground a duty in the state to inflict punishment. There are, then, two distinct conditions for establishing the right to be punished on the interest theory: first, punishment advances the interest(s) of perpetrators; and, second, the interest(s) are sufficient to ground a duty in the state. This section aims to show that the right to be punished on the interest theory fails because of the many and serious doubts surrounding the satisfaction of both conditions. It should be borne in mind that although the first and second conditions are treated separately for the sake of clarity, the two must hold in conjunction in order for the right to be punished to be properly supported by the interest theory. This is important because the success of only one of our criticisms (i.e., either against the first or second condition) is sufficient for our purposes.

#### 3.1. Against the View that Punishment Advances Perpetrators' Interests

To speak of interests, and then to connect these to rights, some notion of interests is needed. Duff makes a case in favor of the right to be punished on the grounds that it is for the good of the perpetrators and accords with their autonomy.<sup>7</sup> His view is that punishment is beneficial to perpetrators because it is a way for them to repent for their crimes, which achieves self-reformation and reconciles them with the communities whose norms they violated and to which they belong (Duff 2001, 107).<sup>8</sup> This description

---

<sup>7</sup> In the preceding section, the commitment to autonomy was analyzed in connection with the will theory. However, it can also be construed as being in the interest of a perpetrator. Thus, the right to be punished may also be grounded in perpetrators' interest in autonomy, which is purportedly advanced through punishment. Part of the preceding criticism would speak against such a claim.

<sup>8</sup> The aims of punishment on Duff's theory will henceforth be referred to as the three Rs.

succinctly captures what Duff regards as the ideal aims of punishment, which he also calls the three Rs of punishment.

Given the above, the right to be punished can be said to protect the interest that perpetrators have in being punished. To this, it seems fitting to rely on that which punishment can and ought to secure, *inter alia*, for the good of the perpetrators (at least according to Duff's theory): repentance, reform and reconciliation. Thus, arguably, perpetrators could be said to have a right to be punished because punishment can protect and secure the interests that they have in the three Rs. Given this interpretation of Duff's conception, it seems fair to assess the right to be punished in relation to the specific interests that it should serve—the three Rs. We will continue to operate with this conception of interests in being punished and consider the plausibility of the first condition—roughly, that punishment advances perpetrators' interests—in light of it.

For the first condition to be true, it must be the case that being punished advances the three Rs of perpetrators. Firstly, this means that the three Rs must be in the interest of perpetrators in the first place. Secondly, this means that the benefits of securing the three Rs outweigh the cost of punishment for perpetrators.<sup>9</sup>

The first, determining whether the three Rs are interests of (all) perpetrators, is hardly a straightforward matter. Duff's discussion of different types of ideal offenders, such as the morally persuaded, repentant, and defiant offenders, proves useful in highlighting the different perspectives on punishment that individuals may take (Duff 2001, 116–124). One such ideal type is of the defiant or principled offender. This class of perpetrators affords the clearest source of doubt for the claim that all perpetrators have an interest in the three Rs (Duff 2011, 374).

Principled offenders are individuals who commit an offense on the basis or for the purpose of expressing their moral convictions or conscientious objection. The problem here stems from the fact that if such offenders reject community's values (as some would), then they would have little reason to pursue the path towards repentance, reform or reconciliation—in other words, they would have little interest in being punished. In such cases, Duff believes that there will be other grounds that will still justify

---

<sup>9</sup> While there is room for debate on this, benefit is here taken to be an overall increase or improvement relative to costs. Where  $\beta$  = benefits, and any integer above 0 expresses such an increase, then  $\beta > 0$ . Again, how much of an increase and determinations about whether there has been (in fact) or will be an increase is debatable. Part of the aim in treating interests in this way is to highlight the extent to which such contingencies need definition in order to justify the right to be punished, and punishment itself (at least on Duff's account).

their punishment, such as the community's and victim's interests (Duff 2001, 123). However, this would mean perpetrators' interests, which would, in turn, ground offenders' right to be punished, might be off the table. Not so, however, because Duff insists that punishment is owed even to such principled offenders out of respect for them as moral agents, as members of the moral community (Duff 2001, 123).

However, the insistence that punishment is owed to principled offenders is difficult to square with the idea of respect for persons and their autonomy, which Duff rightly treasures. At least in the case of some principled offenders (e.g., whistleblowers, dissidents, and political activists), it is hard to see how their autonomy could be respected and a charge of paternalism avoided. This is because on Duff's view, autonomy appears divorced from perpetrators' actual system of practical reasons, their autonomous decisions or desires (Lipkin 1988, 97). Instead, as Robert Justin Lipkin (1988, 97) argues, Duff seems to be portraying autonomy as the autonomy to be moral whether one wants to be or not. Lipkin (1988, 97) contends that this is a peculiar view to advance given that autonomy must be connected to individual wants. Moreover, if this were not the case, then one could claim to be acting out of respect for and on the basis of offenders' autonomy without taking into account any of their desires and, then, any kind of punishment could be justified on such bases (Lipkin 1988, 97).

Having noted these concerns, let us nevertheless suppose that the three Rs *are* interests of perpetrators. This brings us to the second issue about the first condition. The question becomes whether punishment can advance the three Rs in such a way as to offset its costs. The modest claim is that punishment does not *necessarily* do so; contingencies pertaining both to perpetrators and the modes of their punishment affect whether punishment is able to meet this condition. Particularly when it comes to punishment such as incarceration the thought that the satisfaction of the three Rs will necessarily outweigh the costs incurred in their pursuit is far from obvious. Furthermore, since success in achieving the communicative aims of punishment is not built into Duff's theory, it is the benefit of making the attempt to secure the three Rs that should be weighed against the costs of making the attempt. It is far from clear how to go about calculating the expected benefits of the three Rs against the, sometimes devastating, costs of such punishment.

Moreover, accepting that the right to be punished exists for all perpetrators *qua* perpetrators (and that being a perpetrator confirms an interest in this right), would require us to set aside and disregard the unique moral situation of individual perpetrators. Regardless of character, severity of crime, mode of punishment, and so on, perpetrators would seem to have an

interest in and a right to be punished. However, the case of principled or defiant offenders speaks against a blanket generalization of this kind.

Duff believes that a perpetrator has a right to that degree of punishment sufficient to advance her three Rs. He writes:

If we care, as we should, both about the burdens that punishment of its nature imposes and about the danger that those burdens will be excessive or harmful, we should accept a principle of penal parsimony, which requires us to impose punishments no harsher (in mode or amount) than is strictly necessary for the aims that punishment is to serve. (Duff 2001, 134)

Thus, the mode and the degree of punishment that should be administered is qualified and appears to be that which would seem necessary to secure the three Rs of the perpetrator (under proportionality constraints). This view, however, assumes an otherwise questionable harmony between the three Rs. There is the possibility that for any given perpetrator the three Rs could conflict when it comes to the mode or extent of punishment that could secure them. What should we say of a perpetrator whose interest in self-reformation would be sufficiently advanced by a lenient sentence (e.g., because of facts about her character) but whose interest in reconciliation with the community would only be sufficiently advanced by incarceration (e.g., because of the severity of her crime or because the community's norms are particularly demanding)? The above does not refute the existence of the right to be punished, but gives rise to doubts regarding its sustainability on the basis of the interests of perpetrators that take the form of the three Rs.

The right to be punished faces yet another problem. Perpetrators may be said to have an interest in the three Rs (e.g., making amends and being reintegrated into society), but why that should *necessarily* involve punishment is decidedly unclear. That being said, Duff believes that punishment is necessary. He writes:

But it is not a contingent matter that punishment—the infliction of suffering on an offender for her offense—is the appropriate way of trying to achieve the kind of penitential reform which is its justifying aim: that aim, can of its nature, be achieved only by bringing the offender to suffer for what she has done. If someone suggested, for instance, that we might hope to develop some kind of drug or psycho-surgical technique which would provide a more efficient and less

painful method of securing the kind of reformatory change at which punishment aims, we would not need to question the empirical plausibility of her suggestion: for the suggestion itself is incoherent. No such technique could, logically, produce the results at which punishment aims: such techniques do not address the criminal, as punishment must address her, as a responsible moral agent who can and should come to understand the moral implications of what she has done; they are not, as punishment must be, attempts to solicit and arouse her repentant understanding of her crime; and the acquisition of such an understanding must of its nature be painful to the criminal. (Duff 1986, 262)<sup>10</sup>

One can reasonably interpret Duff as advocating for a rather unsavory view that might be captured as ‘persuasion through pain’. Duff’s definition and conception of punishment necessarily involves pain, and pain is regarded by him as indispensable if perpetrators are to come to understand their crimes. The above excerpt is not only illuminating, but still fairly consistent with Duff’s more recent writings at least when it comes to retaining the element of pain. In particular, in *Punishment, Communication and Community*, he preserves the emphasis on the pain and suffering perpetrators should feel as inducement to repentance:

*Repentance* is [...] an aim internal to censure. [...] Repentance is necessarily painful, since it must pain me to recognize and admit (to myself and others) the wrong I have done. In aiming to induce repentance, punishment thus aims to bring offenders to suffer what they deserve to suffer—the pains of repentance and remorse. (Duff 2001, 107)

But, why should persuasion involve pain? Surely there are other ways to persuade someone to understand the gravity of her actions. It is true that pain might be a part of atonement in the sense that once a perpetrator understands her wrongdoing (e.g., the pain she has caused victims) she should feel remorse, maybe even psychological pain. However, there are at least two ways in which such an understanding of pain seems to differ from Duff’s and from the rationale of punishment generally. First, on this view, pain would come as a result of atonement, not as a means to it. Second, pain would not be inflicted upon a perpetrator to make her

---

<sup>10</sup> The view presented in the paragraph comes from Duff’s earlier work. However, even in more recent works such as Duff (2001, 107) and Duff (2003, 186), he maintains that some forms of punishment are intrinsically appropriate as censure and “censure is intended to hurt”.

understand her crimes; rather, pain would be something a perpetrator comes to feel on her own.

Some of Duff's reasons for seeing pain as necessary seem to revolve around the fact that without it we would not address the criminal as a responsible moral agent. He points out that other ways of addressing perpetrators fail in this way. However, it is strange that the alternatives to punishment that Duff chooses to highlight are extreme ones. Drugs and psycho-surgical techniques are hardly the first things that come to mind as plausible substitutes. Agents could come to understand and make amends for their crimes in myriad other ways that would not involve punishment (e.g., therapy, financial reparations, or a combination of these).

Therapy, for instance, could achieve the same three Rs without the pain and suffering that is necessarily involved in punishment. Here, it is not sufficient to claim that other treatments would deny the status of perpetrators as morally responsible and autonomous beings; alternatives to punishment, such as therapy, would be presented to perpetrators *because* they committed crimes and potentially with the same aims that Duff has in mind.<sup>11</sup> Further, the autonomy of individuals could be better protected if they were left to choose one of these alternatives for themselves, with prior and/or subsequent approval by a state body.

The fulfillment of the first condition was the first hurdle for establishing the right to be punished on the interest theory. In accordance with Duff's account, meeting this condition requires showing that the right advances the three Rs of a perpetrator. However, the above showed that there are reasons to doubt that the right to be punished can satisfy this condition. Furthermore, it is worth remembering that the first condition must be met *in conjunction* with the second, which means that success in undermining either is sufficient to dispel the notion of a right to be punished on this view.

### **3.2. Against the View that Perpetrators' Interests in Punishment are Sufficient to Ground a Duty in the State**

We now turn to the final piece of the argument. Given the preceding and the nature of claims in this subsection, the following will be brief but conclusive. The second condition for establishing a right to be punished requires that perpetrators' interests in being punished are sufficient reasons

---

<sup>11</sup> See Lacey and Pickard (2013) for a thoughtful discussion and defense of the clinical model of responsibility that takes the therapeutic ends of punishment seriously without thereby effacing moral agency.

to ground a duty in the state to punish them. As discussed at the beginning of this section, an agent has a right if and only if a certain aspect of her well-being (i.e., an interest) is a sufficient reason for holding others to be under a duty (Raz 1986, 166-172).

Setting aside problems about determining whether punishment advances interests of perpetrators from the preceding subsection, claiming that such interests (e.g., in reconciliation, in self-reformation, in being treated as moral agents, and so on) can ground a duty to punish on the part of the state is based on a confusion about the kinds of reasons that these interests can provide. That is, interests in being punished can provide reasons in favor of punishment, but they cannot provide *sufficient* reasons in favor of punishment. To see why, consider the following.

Rights on the interest theory can be justified on the basis of their instrumental value in securing or promoting individuals' interests. However, not all interests can give rise to rights. Determining which interests are sufficient to ground rights necessitates, among other things, probing into the correlative of the purported right — the duty that it would establish. In other words, for our purposes, the question is: Are perpetrators' interests, namely the three Rs, sufficient to establish a duty on the part of the state? The following will seek to show that this is not the case.

To start, the current practice of legal punishment demonstrates that the state does not act as if it were under a duty. To see this, consider that legal punishment is a result, a form of treatment, which follows from legal judgment. Before judgment, the state is not considered to owe perpetrators punishment because it is not clear whether this legal response ought to be brought against them.

Throughout the legal process the state retains the power to impose punishment or not. If the state were considered to be under a duty to punish or indeed had a duty to punish, then not punishing perpetrators would amount to wronging them. In such cases, it is possible that a perpetrator could demand to be compensated for the wrongdoing suffered. In those instances when the state chooses not to punish it is clear, however, that no wronging against the perpetrator is considered to have been committed.

To put matters differently, the claim that perpetrators have a right to receive a result (i.e. punishment) is dubious because, *ex ante*, it is not clear whether this *should* be imposed on them. The matter of punishment is determined after deliberation, which involves weighing considerations that include, but are not limited to, a perpetrator's interests. If, however,

perpetrators' interests were sufficient to determine the duty of the state to inflict punishment, then the state would, *ex ante*, find itself under a duty to punish.

The nature of punishment is such that organs of the state may choose not to punish a perpetrator (e.g., offer her pardon or some form of response other than punishment) or punish her to a greater or lesser extent (e.g., symbolic punishment, long sentence). Even though much is known about the likelihood of a perpetrator suffering punishment, whether or how much punishment is to be applied remains a case-by-case matter. For instance, whether to pursue punishment or how much punishment to inflict is within state agents' discretionary powers. Discretion means that cases may be dropped, charges changed, or punishment forsaken for a variety of reasons (e.g., preserving social order, victim's preference).<sup>12</sup>

Although the interests of perpetrators might serve as *pro* or *contra* reasons when deciding whether to inflict punishment, the preceding subsection underscored that the advancement of interests that take the form of the three Rs through the infliction of punishment is questionable. Furthermore, even if we can determine that perpetrators' interests will be advanced, such considerations are only some out of the multitude of reasons that organs of the state ought to take into account. Thus, assuming that such interests obtain, they do not constitute sufficient reasons in favor of or against punishment.

Recall that for the right to be punished to exist on the interest theory, it must be the case that perpetrators' interests in being punished constitute *sufficient* reasons for grounding a state's duty to punish them. However, examples that show the interests of perpetrators to be *insufficient* considerations are easy to find. For instance, a dissident might have an interest in being punished, but the state knows that punishing her would result in widespread violence. The state might then rightly decide not to punish the dissident because, all things considered, the right course of action dictates against punishing her. In order to make the point we need not rely on hypotheticals alone. Real and even high-profile examples are also available. Gerald Ford offered Richard Nixon a full pardon on the grounds that Nixon had already suffered enough by relinquishing his office (see Ford 1974). Other considerations, such as age, health, suffering of dependents, may also affect the extent (if at all) of sentencing. These considerations point to the fact that interests of perpetrators might constitute reasons to punish (or not), but they are certainly not determinant ones.

---

<sup>12</sup> We wish to thank an anonymous reviewer for these helpful points.

The above is more than descriptive of a state of affairs: perpetrators' interests are among the many reasons to be considered, but they should never be more than that. This is because claiming that the advancement of interests of those who commit crimes are sufficient to establish a duty for the state to punish would mean that it is not necessary for the state also to consider other interests (e.g., victims' interests, societal interests, or yet other interests) for it to find itself under a duty.

Take Raz's theory, where an agent may be said to have a right if and only if some aspect of her well-being (some interest of hers) is sufficiently important in itself to justify holding some other person or persons to be under a duty. It should be noted that because of how counter-intuitive it would be to suppose that agents have an interest in punishment in itself, we assume that any right in punishment would be based on other interests (e.g., the three Rs). This means that the right would not be a core right, but rather one derived from other aspects of individual well-being.

Having specified what it means to have a right in accordance with the above version of the interest theory, it is also important to identify the conditions that determine the absence of a right. According to Raz, if conflicting considerations show that the interests advanced or protected by a purported right are not enough to justify subjecting anyone to any duty, then the right does not exist.

Raz writes: "where the conflicting considerations altogether defeat the interests of the would-be right-holder, or when they weaken their force and no one could justifiably be held to be *obligated* on account of those interests, then there is no right" (Raz 1986, 183-184). Furthermore,

only where one's interest is a reason for another to behave in a way which protects or promotes it, and only when this reason has the peremptory character of a duty, and, finally, only when the duty is for conduct which makes a significant difference for the promotion or protection of that interest does the interest give rise to a right. (Raz 1986, 183-184)

One of the most damaging marks against the existence of the right to be punished is that if the right exists, then the interests protected by it would have to be sufficient to ground a duty on the part of the state. Furthermore, the nature of duties is such that they provide not just reasons for action of significant weight. Duties give rise to exclusionary reasons that have a special peremptory force (Raz 1986, 195). Thus, the state would have an exclusionary reason to inflict punishment (see Raz 1975; 1979; 1986; 1989; 1990). This means that the state would have a second order reason

to punish offenders that excludes acting on and perhaps even considering first order reasons that go against punishment.

The above could be a rather unwelcome conclusion as other considerations (such as the interests of dependents, victims, or society) ought to play a role in deliberation and state action. It is not unusual for interests to come into conflict, which means that to regard perpetrators' interests as grounding duties that altogether displace other reasons is problematic.

We do not have good reasons to correlate the interests of perpetrators with a duty to punish of exclusionary force *if* this would mean that victims' and society's interests would, thereby, have to be put to the side. One might think that interests of victims, perpetrators, and society are compatible, but that is not always the case. Furthermore, even if one assumes these interests to be compatible there remains something uneasy about inflicting punishment for reasons that by definition could exclude the interests of victims, society, and others.

This is not to say that perpetrators' interests ought not to be considered or that they ought to weigh less than other reasons. Perpetrators' interests should be considered during deliberation and judgment. However, if we grant the existence of the right to be punished, then it follows that the interests of perpetrators (e.g., the three Rs) will ground the preemption of other reasons. It is this latter ramification that is challenged.

The above considerations capture the second condition's main failings. Since interests of perpetrators cannot and should not be sufficient considerations in favor of punishment, the state is not under a duty to punish. Furthermore, this conclusion would hold even if we were to cash out perpetrators' interest in punishment in a different way (i.e., other than the three Rs). Thus, it becomes clear that the state does not owe punishment to perpetrators based on their rights to it.

#### 4. Conclusions

As noted in the introductory paragraphs of this article, the idea of a right to be punished has been around for quite some time. Its longevity clearly does not represent an argument for favoring it. This article has shown that even in its more recent iterations, notably from Duff, the idea of a right to be punished is far from secure. In fact, it has been shown that on the will-based and the interest-based theory of rights, there are strong reasons to disregard the right to be punished as anything but an outmoded or misguided notion. On the will theory the right to be punished was shown

to be at odds with the autonomy of (purported) right-holders, while on the interest theory serious doubts were presented regarding the (alleged) right's ability to serve its function as promotor of offenders' interests that are sufficient to ground a duty to punish on the part of the state.

One final point is about how this work advances the state of the discussion. The ongoing conversation about legal punishment continues unimpeded and many of the long-standing debates remain as contentious as ever. This article ultimately helps to show that while there are many questions about punishment worth examining, the question about a (purported) right to be punished might be best laid to rest. Clearly, the other questions about punishment were left unexamined in any thoroughgoing way because the aim of this paper was narrower. We examined one question—right to be punished?—so that those others may be more profitably investigated elsewhere.

## Acknowledgments

The authors wish to thank two anonymous reviewers for their careful readings and helpful suggestions and comments.

## REFERENCES

- Adler, J. 1992. *The Urgings of Conscience: A Theory of Punishment*. Philadelphia: Temple University Press.
- Alm, D. 2018. Punishment, consent and value. *Ethical Theory and Moral Practice* 21: 903–914.
- Bagaric, M., and K. Amarasekara. 2000. The errors of retributivism. *Melbourne University Law Review* 24: 124-189.
- Bennett, C. 2008. *The apology ritual: A Philosophical Theory of Punishment*. Cambridge: Oxford University Press.
- Brooks, T. 2013. The right to be punished: A defense. *Legal Theory in China* 3: 21-39.
- Deigh, J. 1984. On the right to be punished: Some doubts. *Ethics* 94: 191-211.
- Downie, R. S. 1985. Three accounts of promising. *The Philosophical Quarterly* 35: 259-271.
- Dubber, M. D. 1998. The right to be punished: Autonomy and its demise in modern penal thought. *Law and History Review* 16: 113-146.
- Dubber, M. D. 2007. Punishment, morality, and the state: Legitimizing penal law. *Cardozo Law Review* 28: 2597-2613.

- Duff, R. A. 1986. *Trials and Punishments*. Cambridge: Cambridge University Press.
- Duff, R. A. 2000. In defense of one type of retributivism: A reply to Bagaric and Amarasekara. *Melbourne University Law Review* 24: 411-426.
- Duff, R. A. 2001. *Punishment, Communication, and Community*. Oxford: Oxford University Press.
- Duff, R. A. 2003. Penance, punishment and the limits of community. *Punishment & Society* 5: 295-312.
- Duff, R. A. 2011. In response. In *Crime, Punishment, and Responsibility: The Jurisprudence of Antony Duff*, eds. R. Cruft, M. H. Kramer, and M. R. Reiff, 351-379. Oxford: Oxford University Press.
- Feinberg, J. 1974. *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press.
- Feinberg, J. 1980. *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy*. Princeton: Princeton University Press.
- Ford, G. R. 1974. Proclamation 4311 - Granting Pardon to Richard Nixon. September 8, 1974.  
<https://www.fordlibrarymuseum.gov/library/speeches/740061.asp>.
- Hallevy, G. 2013. *The Right to Be Punished: Modern Doctrinal Sentencing*. Berlin: Springer-Verlag.
- Hart, H. L. A. 1982. *Essays on Bentham: Studies in Jurisprudence and Political Theory*. Oxford: Clarendon Press.
- Husak, D. 2008. *Overcriminalization: The Limits of the Criminal Law*. New York: Oxford University Press.
- Kelsen, H. 1967. *Pure Theory of Law*. Berkeley: University of California Press.
- Lacey, N., and H. Pickard. 2013. From the consulting room to the court room? Taking the clinical model of responsibility without blame into the legal realm. *Oxford Journal of Legal Studies* 33: 1-29.
- Lipkin, R. J. 1988. Punishment, penance and respect for autonomy. *Social Theory and Practice* 14: 87-104.
- Nino, C. S. 1983. A consensual theory of punishment. *Philosophy & Public Affairs* 12: 289-306.
- Penner, J. E. 1997. The analysis of rights. *Ratio Juris* 10: 300-315.
- Raz, 1975. *Practical Reason and Norms*. Princeton: Princeton University Press.
- Raz, J. 1979. *The Authority of Law*. Cambridge: Cambridge University Press.
- Raz, J. 1986. *Morality of Freedom*. Clarendon: Oxford University Press.
- Raz, J. 1989. Facing up: A reply. *Southern California Law Review* 62: 1153-1235.

- Thomson, J. J. 1986. *Rights, Restitution, and Risk: Essays in Moral Theory*. Cambridge: Harvard University Press.
- Williams, G. 1955. The definition of crime. *Current Legal Problems* 8: 107-130.



## CYBERBULLYING, MORAL RESPONSIBILITY, AND SOCIAL NETWORKING: LESSONS FROM THE MEGAN MEIER TRAGEDY\*

Raphael Cohen-Almagor  
University of Hull

Invited article – Received: 15/10/2019 Accepted: 07/07/2020

### ***ABSTRACT***

*This paper addresses the concepts of moral and social responsibility on the Internet in considering the most troubling phenomenon of cyberbullying that results in loss of life. Specifically, I probe the moral and social responsibilities of Internet users (agents), of the education system in fighting cyberbullying, and of Internet intermediaries. Balance needs to be struck between freedom of expression and social responsibility. The tragic story of Megan Meier serves as an illustrative example and some further incidents in which this ugly phenomenon of cyberbullying had cost young life are mentioned. It is argued that all relevant stakeholders need to think of the consequences of their conduct, that Internet abusers should be accountable for their wrongdoing, and that people who have the ability to stop or at least reduce the risk of cyberbullying should take proactive steps, exhibiting zero tolerance to cyberbullying.*

**Keywords:** *Aristotle; bullying; cyberbullying; Internet; Megan Meier; moral and social responsibility; social networking*

\* This article builds upon and expands on material published in Cohen-Almagor (2011, 2015, 2018).

## 1. Introduction

In October 2018, 13-year-old Ben McKenzie committed suicide after he was subjected to online threats and bullying on social media and on his mobile phone (Hendry 2018). Member of Parliament Paul Masterton raised the issue of his death during Prime Minister's Question time. Prime Minister Theresa May responded that cyberbullying is an "extremely serious issue" that needed to be tackled by the Internet social networks as well as the education system. PM May noted that Internet safety was and remains a major concern and that despite some progress in improving Internet users' safety, cyberbullying has remained a serious worry. May promised that the British government will continue to address this issue (Seith 2018).

Cyberbullying and bullying are highly upsetting and exasperating issues. They are distressing because at times they result in suicide, and any loss of life is sad. These phenomena are particularly distressing because often the life that are lost as a result of online and offline forms of bullying are those of young people, often in their teens. Cyberbullying and bullying deprived them of their future. These phenomena are frustrating because most of those misfortunes could have been avoided and/or prevented if relevant stakeholders were to conform to basic norms of social responsibility. Indeed, cyberbullying and bullying are social tragedy. These phenomena are a sombre testimony of the dark side of human nature. The Internet has exacerbated the problem of bullying and made it into a constant nightmare. The Internet has equipped bullies with a powerful weapon that enables them to torment victims relentlessly with no reprieve.

The Internet has affected all aspects of society. Digital platforms are increasingly where we meet new people and maintain older contacts. This became very apparent during the recent coronavirus crisis, where many countries went into a lockdown and people were forced to conduct their affairs online. People work, study and make phone calls; conduct business, video conferencing and social campaigns; search for information; shop, socialize and flirt; share photos and experiences; listen to music, watch movies and explore the world online. The world population is nearing 8 billion people. Of them, more than 4.6 billion people are using the Internet. In Europe and North America, the Internet penetration rate is more than 87% (Internet World Stats 2020).

In the Internet age, people have active life on social networking platforms and have far more virtual "friends" than genuine, true friends on which they could rely at challenging times. Facebook alone has a staggering number of almost 2.6 billion monthly active users (Clement 2020). Many

people have more than one identity. People create fake identities for all kind of purposes, kosher and non-kosher, legitimate and illegal. During 2019, Facebook removed 5.4 billion fake accounts. In 2018, Facebook removed roughly 3.3 billion fake accounts (Fung and Garcia 2019; Segarra 2019). In other words, the number of fake accounts exceeds the number of true accounts. The ease of opening new accounts has significant consequences which until now have not been adequately addressed.

This Essay discusses moral and social responsibility on the Internet. Section 2 explains the concepts of moral, legal and social responsibility by focusing on the writings of Aristotle. Section 3 elucidates the cyberbullying phenomenon. Section 4 discusses the responsibility of people who are using the Internet, Internet agents. Section 5 probes the Megan Meier suicide, a tragedy that illustrates an immoral use of the Internet on a social networking website, abusing the functions of the Internet without regard to the potential tragic consequences. Section 6 discusses the responsibilities of the education system, and, finally, Section 7 is concerned with responsibility of Internet intermediaries. Internet companies have a vital role in making cyberbullying part of our lives and in helping to redeem this painful social challenge.

## **2. Moral and Social Responsibility**

*Moral responsibility* relates to the agent's conscience, one's ethical conduct and the moral compass that guides one's life that is normally within certain moral codes of society in a particular period of time. *Social responsibility* relates to these societal norms, and to the broader implications of people's moral conduct. *Legal responsibility* refers to the conduct of agencies of state power in legislating and enforcing laws that enable living together and that demand people to be accountable for their conduct (Cohen-Almagor 2011).

The philosophical foundations of the concept of moral responsibility lie in Aristotle's *Nicomachean Ethics*, where Aristotle discussed what acting in accordance with one's best interests means. Aristotle (1962) discussed human virtues and their corresponding vices. He distinguished between *voluntary* action and *coercive* action. Individuals who are coerced to do something cannot be held accountable for their conduct. Their coercer is the responsible agent. I have discussed the issue of coercion elsewhere (Cohen-Almagor 2006). Of relevance to the discussion here are agents who act of their own free will. Individuals are responsible for their conduct when they are competent, well informed, and aware of what they are doing (Aristotle 1962, 1110B15-25).

A voluntary action must have its origin in the agent (Aristotle 1962, 1110a-1111b4). Agents' conduct expresses their conception of the good. For Aristotle (1962, 1111b15-1113b22), competency, deliberation, choice and moral agency are important in evaluating one's conduct. Deliberation precedes choice. People who have failed to deliberate are led by their emotions and/or passions. Choice between meaningful alternatives is important. People set for themselves desirable ends and secure relevant means to achieve them. When people choose to act unjustly from choice, they are vicious (Aristotle 350 BCE, Book V). Aristotle explained (*Ibid*): "But if a man harms another by choice, he acts unjustly; and these are the acts of injustice which imply that the doer is an unjust man, provided that the act violates proportion or equality. Similarly, a man is just when he acts justly by choice" (for further discussion, see Sauve Meyer 2012; Erginel 2016; Talbert 2019).

A just person is a moral person, and a moral person avoids three kinds of behavior: vice, incontinence and brutishness (Aristotle 1962, Book VII). Vice (*kakia*) is concerned with pain and pleasure. It is an excess or deficiency of virtue and is a matter of choice (Aristotle 350 BCE, Book II). This means that competent and free willed agents are responsible for their state of mind and for the choices they make. They bear responsibility for acquiring and exercising virtues and they bear responsibility for acquiring and exercising vices. Incontinence (*akrasia*) means lack of self-restraint (or lack of mastery) and therefore it is contrary to choice. An akratic person is acting without sufficient reason which is the result of some *pathos*, such as emotions and feelings. In turn, brutishness (*thēriotēs*) "is found chiefly among barbarians, but some brutish qualities are also produced by disease or deformity; and we also call by this evil name those men who go beyond all ordinary standards by reason of vice" (Aristotle 1962, Book VII). Brutish people include cannibals, people who devour their infants, or who lend "their children to one another to feast upon" (*Ibid.*).

Competent people who act voluntarily choose whether they wish to be virtuous and noble, or evil and bad. We all should know to distinguish between good and evil. Ignorance will not absolve us of responsibility. Society imposes penalties on people who harm others even when they did not intend to. Aristotle wrote that we punish people for their very ignorance, if they are deemed responsible for the ignorance, "as when penalties are doubled in the case of drunkenness" (Aristotle 350 BCE, Book III). People have the power of not getting drunk which made them idle and brought them to make the harmful mistake (Aristotle 350 BCE, Book III). And we punish those who are ignorant of anything in the laws that they should know and that is not too difficult to grasp. We also punish those who do bad things because they are careless. We assume that it is in

their power to act with care. Thus, people who act against their better judgment, termed akratic people, are morally blameworthy for their harmful conduct (FitzPatrick 2008; Kraut 2018). If an agent does something bad with knowledge that the action is bad, knowing full well that she should not be doing it, then the agent is said to be acting with clear-eyed akrasia (FitzPatrick 2008, 590; Lawrence 1988).

The concept of social responsibility refers to the responsibilities of individuals, the public sector, the private sector and the government to society. Our actions have some bearing on others and we should strive that this bearing will be positive. Responsible conduct is a caring conduct; it is acting with foresight while we are cognizant that actions have consequences, and we aim that the consequences will affect us and others for the better. Responsible people proactively do good and avoid harm (Buntun 1998; Christians and Nordenstreng 2004; Kaliski 2001; Marshall 1994; Rivers, Schramm, and Christian 1980; Cohen-Almagor 2015).

In the context of the professions, social responsibility is especially important because professionals are trained to hold a specific skill that requires autonomous judgment and expertise. Professionals are duty bound to serve their clients. Often a broader responsibility is attributed and expected (McQuail 2003: 191). Conduct is dictated and evaluated in accordance with a given set of standards. Professional standards may change with time; therefore, they are carefully monitored, and corporations are expected to bear responsibilities to their clients and to society at large. Adopting social responsibility norms and adhering to moral codes of conduct is the right way to behave (Novak 1996; Trevino and Nelson 1999; Cohen-Almagor, Arbel-Ganz, and Kasher 2012; on Corporate Social Responsibility [CSR], see Carroll 2015; Wan-Jan 2006; Goodpaster 2010; Carroll and Shabana 2010; Abend 2014; Kerr Janda, and Pitts 2009; Gawu and Inusah 2019).

### **3. Agent's Responsibility**

Cyberbullying involves the use of digital technologies to target people in order to harass, offend, threaten, degrade, ridicule, or humiliate them (Alipan et al. 2020; Kowalski, Limber, and Agatston 2008, 1). It involves targeting victims via computers, smart phones and any other electronic device. Some forms of cyberbullying involve electronic stalking, identity theft, password theft, the spread of malicious rumours or exposing private or privileged information without the victims' consent. Cyberbullying might also involve the distribution of photos and video clips of sexual or violent nature that would lower the victims' status in the eyes of peers or

society at large, damage their reputation and cause them great embarrassment. Extremely harmful forms of cyberbullying include circulating rape footages, blackmail and online death threats (Gerson and Rappaport 2011; Pesta 2013; Lallitto 2017; Petrov 2019).

The motivation for bullying online and offline is varied. Bullies wish to gain a feeling of power, purpose and control over others. Some bullies suffer from low self-esteem and engage in this activity in order to mask how they feel about themselves or wish to receive recognition from their peers (Salmivalli 2010; Ditch the Label 2018). Bullies engage in this sort of activity because they are bored, angry, or because they seek some twisted sense of entertainment. Some are motivated by revenge or frustration (Salmivalli, Huttunen, and Lagerspetz 1997; Perren and Alsaker; Duffy and Nesdale 2009; Doehne, Grundherr, and Shafer 2018). Bullies are likely to have experienced considerable stress or trauma. Many bullies feel that their parents/guardians do not spend enough time with them. They do not trust their relationships with friends and families and wish to gain attention and appreciation by showing their influence/power over others (Ditch the Label 2018).

Sometimes, bullies exploit anonymizing tools to assault victims. Those bullies are not likely to utter those offensive statements in one's face, but with the Internet as a filter and facilitator they have no qualms harassing their victims, pushing them to intolerable and most troubling state of mind (Hinduja and Patchin 2009; Smith, Mahdavi, Carvalho, and Tippett 2006). Studies estimated that between 13% and 46% of young victims of cyberbullying did not know their harasser's identity. 22% of the bullies did not know the identity of their victim (Kowalski and Limber 2007; Wolak, Mitchell, and Finkelhor 2007; Ybarra, Diener-West, and Leaf 2007; Cohen-Almagor 2018).<sup>1</sup>

The structure of social networking sites makes it easier to state rude, intrusive and offensive words that one would be hesitant to state face to face. The offence is exacerbated as often the victim is alone and hesitant to inform others about the aggression s/he is facing. Not knowing the identity of the electronic bully leaves the victim guessing who the person behind the aggression is: is it someone whom s/he knows, or a complete stranger? This creates a suspect and unsafe environment for the bullied.

Bullying is certainly not new. It has been part of life for many generations. In every class there are always children who become the target for some of

---

<sup>1</sup> It is reiterated that cyberbullying does not necessarily relate only to young people; it is just that most of the research in this field has tended to focus on the young.

their classmates who enjoy ridiculing them and exposing their vulnerabilities in order to have a “good laugh”. At school, students who are somehow different attract the attention of bullies as they seem to be more vulnerable and defenceless, easy to pick on, humiliate and intimidate. Children with disabilities, youth with confused sexuality, students with special needs, and socially isolated adolescents attract the bully’s attention (Department for Children, Schools and Families 2007, 2009; Riggio 2013). Also ethnic minorities and homosexuals are disproportionately targeted (Beaty and Alexeyev 2008; Berlan et al. 2010; Kahle 2017).

Cyberbullying has desensitizing effect on bullies and bystanders (Steffgen et al. 2011; Pabian et al. 2016) and it can be relentless. It can take place simultaneously on multiple online forums, employing multiple technologies. Tormenting images of bullying can be posted on many social networking sites and cause victims prolonged suffering. Indeed, technology has the potential to exacerbate wrongdoing. Before the age of the Internet and smart phones, bullying stopped as the victim entered the shelter of the home. Today the harassment follows the victim wherever she goes, without a reprieve.

A study among European children aged 9-16 showed that one in twenty children was bullied online more than once a week, and one in ten was bullied a few times during the past year (Livingstone et al. 2011, 61). 12% reported that they bullied others during the past year (Livingstone, Haddon, Görzig, and Ólafsson 2011, 64; see also Görzig and Frumkin 2013). A British study that surveyed children aged six to nine reported that 20% children were the victims of “aggressive or unpleasant” behaviour online. In Denmark, 21% of the teens reported that they experienced cyberbullying. The study shows that Danish parents talk less to their children about Internet safety than before (Livingstone, Mascheroni, and Ólafsson 2014; for further discussion, see Navarro, Larrañaga, and Yubero 2018, 122-125). This serious problem may explain, at least in part, the rise in cyberbullying. According to the Megan Meier Foundation, approximately 34% of students report experiencing cyberbullying during their lifetime.<sup>2</sup>

At times, bullying accompanies another kind of evil doing. In Canada, 17-year-old Rehtaeh Parsons was raped by four boys while she was at a house party. Those boys made the trauma worse for Parsons by photographing her and then circulating Parson’s brutal ordeal. Parsons hanged herself amid months of persistent online bullying, with peers insulting her by calling her a slut, circulating her photos, messaging her and harassing her

---

<sup>2</sup> Megan Meier Foundation, <https://meganmeierfoundation.org/cyberbullying>.

both online and offline (Newton 2013; Pesta 2013; Arthur 2014). Canadian Prime Minister Stephen Harper said that as a father he was “sickened” by the alleged events that led to Parsons’ death (CBC 2013). Those boys acted with full knowledge that what they were doing was wrong, and they still relished the opportunity to act immorally, which is flagrant clear-eyed akrasia. Parsons fell victim to a string of failures: of the boys’ families and of the education system to equip those boys with values of compassion and social responsibility.

#### **4. Agent’s Responsibility**

Taking responsibility has a significance in virtue of the role that this act plays in maintaining our self-conceptions as agents (Bero 2020). Per Aristotle, an agent would be held responsible for speech that has directly led to harm when she voluntarily chose to engage in that activity. While establishing a direct link between speech and harm is not always easy, undoubtedly some forms of speech, such as those that urge victims to kill themselves and murder threats are inciteful in nature and in no way can be regarded as protected speech (Mill 1948, chap. 3; Cohen-Almagor 1994, 2017). Agents who utter such words are blameworthy and should be responsible for their harmful consequences. Indeed, it must be acknowledged that words can inflict a great deal of pain. Words can upset and hurt. Words can move people to action.

Furthermore, anonymity plays an important role in linking traditional and cyber forms of bullying and harassment (Walters and Espelage 2020). Victims persistently worry about the perpetrator’s identity. We all feel anxious in the face of the unknown. We all worry about our reputation. Many people care about their public image and wish to be perceived by others in a positive light. Many are concerned about their social status and the way they are perceived by their peers and by other people who are of significance to them. Information posted on the Internet can enhance careers and contribute to one’s social status. But information can also ruin careers and lives.

In an earlier article (Cohen-Almagor 2011), I described how JuicyCampus.com was used to ruin the name of young people. Behind the shield of anonymity, agents dusted away all responsibility and inflicted great harm on their victims “for fun”. JuicyCampus closed down on February 5, 2009 after it gained deserved notoriety that caused users and businesses to shun. But JuicyCampus was soon replaced by other no less intrusive and damaging forums. In 2010, Ask.fm was established to enable people the posting of anonymous questions in the most offensive and

degrading way. In 2013 alone, nine teenagers were driven to commit suicide after they were subjected to cyberbullying on Ask.fm (Edwards 2013). Ask.fm is still alive and kicking. And in 2013, Tyler Droll and Brooks Buffington established an anonymous gossip app called Yik Yak. On this platform as well, people were able to say whatever they wanted without accountability. In 2017, after four years filled with scandals, harassment and irresponsible gossip, Yik Yak had shut down (Kircher 2017). One of the scandals was concerned with Tysen Campbell, a student at Western Washington University who, in 2015, was charged with a hate crime for posting “Let’s lynch her” directed at a black student leader (Green 2015; see also Larimer 2015; Diehl 2015).

Many of the cyberbullying cases that led to suicide involve adolescents tormenting peer victims and pushing them to death (Kaplan 2014). However, one of the early cases of cyberbullying in the USA involved a mother who recklessly brought about the suicide of her teenage neighbour because she suspected that that teenager did not behave kindly to her daughter.

## **5. The Megan Meier Tragedy**

Sarah Drew and Megan Meier were both 13-year olds. They used to be good friends but then had a falling out. Sarah was concerned that Megan had bad-mouthed her behind her back. Sarah’s mother, Lori Drew, 49, together with Ashley Grills, 19, a family friend and employee, created a fake account on MySpace, which was the most popular social networking site in the USA until the birth of Facebook (Sawyer and Roberts 2008). One day Megan received an invitation to connect with “Josh Evans” who presented himself as a 16-year-old from a nearby school. Megan’s parents were reluctant for her to approve Josh’s friendship request as she did not know him (Stossel, Vargas, and Roberts 2007) and because Megan was a vulnerable girl. She received treatment for attention deficit disorder and depression and had been in counselling since third grade (Deutsch 2008; Jones 2008). Her parents were, therefore, understandably concerned about Megan’s wellbeing. Megan insisted to approve that “hot guy”. Her parents complied as they understood this issue was important for Megan and did not wish to upset her. For the next six weeks Megan and Josh, under the watchful eye of Megan’s mother, embarked on an online relationship that became the center point of Megan’s life. Lori Drew later explained that the communication between “Josh” and Megan intended to gain Megan’s confidence in order to find what Megan felt about her daughter and other mutual acquaintances (Grohol 2018). But in October 2006 “Josh” wrote to Megan “I don’t know if I want to be friends with you anymore because

I've heard that you are not very nice to your friends" (Pokin 2007; McFadden and Fulginiti 2008). Megan wished to understand the reasons for Josh's sudden negative twist but "Josh"'s response was even more upsetting and insulting (Jones 2007; Pokin 2007). Megan's father, Ron, found after Megan's death what he believed to be the last message Megan read from "Josh" which said that everybody hated Megan, and that the world "would be a better place without you" (Pokin 2007; Collins 2008). Megan responded: "You're the kind of boy a girl would kill herself over" (Steinhauer 2008). Megan committed suicide that same day.

Lori Drew and her co-conspirators are blameworthy and morally culpable for masterminding the events that led to Megan's suicide. They played on Megan's emotions in a crude and cynical way. They did not act under compulsion. They were responsible for being unjust and self-indulgent, in deciding to cheat a young, vulnerable girl for their selfish and petty interests. They failed to exhibit fair judgment and did not consider how their careless and heartless game might come to a sad conclusion. They exhibited a strong form of clear-eyed akrasia, acting against their adult better judgment.

Lori Drew was reported saying that she felt her prank contributed to Megan's suicide, but that she did not feel "as guilty" because she found out that "Megan had tried to commit suicide before" (Pokin 2007). Somehow, instead of feeling more responsible for what she did because she pushed a vulnerable girl to her death, Drew felt less responsible. She felt no guilt or remorse (Lauer and Lewis 2007). Drew did not desire the apparent good, had control over what she did, and succumbed to the flaws of her character. She was responsible for the state of mind that brought her to concoct the fake account, and she needed to be accountable for her evildoing. Aristotle (350 BCE, Book III) wrote: "every one does evil acts through ignorance of the end, thinking that by these he will get what is best", but one "must be born with an eye, as it were, by which to judge rightly and choose what is truly good, and he is well endowed by nature who is well endowed with this".

## **6. Responsibility of the Education System**

Slađana Vidović was 16-year-old when she committed suicide. She was the daughter of a Bosnian family who immigrated to Ohio. Slađana was subjected to continued bullying and harassment. At school, she was ridiculed for her thick accent. Classmates insulted her repeatedly (Crimesider 2010). Phone callers threatened her, told her to return to Croatia, and that they would harm her after school. "Slađana did stand up

for herself, but toward the end she just kind of stopped”, said her best friend, Jelena Jandrić. “Because she couldn’t handle it. She didn’t have enough strength” (Barr 2010). Slađana left a suicide note in which she described the prolonged harassment she endured at Mentor High School (Krouse 2019). Slađana’s parents implored the school to intervene and enforce anti-bullying policy. The school managers promised to take care of the young girl (Barr 2010). They did not. Slađana was one of no less than five students at Mentor schools who committed suicide between July 2005 and October 2008. The problem of bullying was evident. Yet the education system directors failed to see it. They suffered from clear-eyed akrasia and did not rush to act responsibly and address the problem head-on.

Schools need to adopt a policy of Zero Tolerance to bullying and cyberbullying both on and off campus. Parents and psychologists should be involved in such programs (Williams and Godfrey 2011). Educational programs that tackle bullying and cyberbullying should include extensive discussions that address the problems of silence, explaining that silence is not a solution. Quite the opposite. Silence only helps the bullies continue with the harassment. Such programs should also explain that thrill seeking should not include bullying. Tragedies, such as the painful story of Megan Meier, should be explained at schools. Clear procedures to report and investigate bullying and cyberbullying on and off school should be established (Hong, Espelage, and Lee 2018, 359-374; Schargel 2014). Effective intervention programs decrease cyberbullying and significantly improve classroom atmosphere (Cioppa et al. 2015; Cross et al. 2016; Aizenkot and Kashy-Rosenbaum 2018).

Furthermore, teachers and school administrators need to familiarise themselves with the use of new technologies. Schools should have digital citizenship classes in which students learn the basics of ethical and legal conduct on the Internet, mobile phone and other electronic devices (Wakefield 2017). School administrators should also acquaint themselves with the available mental health programs that could assist students in need. In September 2019, the Anna Freud National Centre for Children and Families in the United Kingdom has embarked on providing training sessions to 22,000 schools and colleges, bringing together education and National Health Service (NHS) professionals in order to ensure that pupils will receive the support they require in a timely fashion. In 2017, it was reported that one in nine young people aged 5 to 15 had a diagnosable mental health condition (Department of Education and Hinds 2019; for further discussion, see Farrington et al. 2017; Long, Gardani, McCann et al. 2020). This explains why bullying and cyberbullying are such pressing problems. Vulnerable pupils attract the attention of bullies and are unable

to cope with the continued harassment. In some cases, when young people are feeling trapped, they might start thinking of suicide as a way out.

## 7. Responsibility of Internet Intermediaries

The issue of responsibility of Internet intermediaries is urgent and pressing. Their actions and inactions directly affect the information environment. They have discretion whether their services are opened for all or limited in one way or another. Most Internet intermediaries adopt some form of moral and social responsibility. They opt for some standards of self-regulation by adopting codes of practice. Internet Service Providers (ISPs) have guidelines regarding what users are not allowed to post on their servers. They have the right and the duty to report potentially criminal activities. They may pre-screen, filter and remove content at their discretion. For instance, Facebook's Community Standards includes a clause on safety: "We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook".<sup>3</sup> In 2020, the #StopHateForProfit Campaign forced Facebook to announce new content policies that would include tighter restrictions on advertising and flagging harmful posts published by public figures in violation of Facebook's rules. The change in policy was done under mounting public pressure and effective advertisement boycott that demanded Facebook to impose tighter restrictions on false news, bigotry and incitement to violence. The boycott by more than 100 advertisers, including some of the largest companies in the world, reportedly reduced Facebook's market value by \$56 billion and caused a heavy loss of \$7.2 billion to Zuckerberg's personal fortune (Sharp and Griffith 2020). Zuckerberg, the champion of free speech who believes racism, bigotry and hate speech are all protected under the First Amendment to the US Constitution, declared that he stands against hate, or "anything that incites violence or suppresses voting, and we're committed to removing that no matter where it comes from" (Sharp and Griffith 2020). Of course, declarations, codes and standards make sense only if they are appropriately enforced. They should not merely serve as a fig leaf to hide *kakia* or *akrasia* such as self-indulgence or dogmatism.

Internet intermediaries are gatekeepers. As such they have a duty to protect vulnerable third parties. Anti-social activities are most prevalent on three American social networking platforms committed to the First Amendment that holds, *inter alia*, that Congress shall make no law abridging freedom

---

<sup>3</sup> <https://www.facebook.com/communitystandards/>

of speech and of the press. These companies are Instagram (42%),<sup>4</sup> Facebook (37%) and Snapchat (31%)<sup>5</sup> (Petrov 2019; Kao et al. 2019).

The managers of Facebook wish to be all inclusive, believe in freedom of expression, and wish to promote merchandise by subjecting “friends” to subtle and not-so-subtle advertisement. Internet intermediaries exist to make money. Censorship contradicts their *raison d’être*. Richard Allan (2018), Facebook Vice President of Policy, explains: “free expression is key to a thriving society. So, barring other factors... we lean toward free expression. It’s core to both who we are and why we exist”. Indeed, for Facebook, Instagram, Snapchat and other digital social platforms, freedom of expression is of utmost importance to the extent that Facebook initially did not have rules on what speech violated its terms of service (Rosen 2013). Allan (2018), in the same quoted paper, goes on to outline the exceptions to free expression. Strangely and revealingly, the words “bullying” and “cyberbullying” are not mentioned.

Presently, Facebook managers should be well aware of the harms that their platform facilitates. They have the ability to limit anti-social activities but until now they are not sufficiently proactive. It is possible to devise an algorithm that would flag abuse, especially continuous abuse and then a human eye would inspect the flagged content and make a decision. Internet intermediaries failed to fight cyberbullying to the extent they are able to and should. The question that they themselves need to grapple with is whether they have prioritized human life over and above all other considerations.

Freedom of use is not freedom to abuse. Freedom of speech is not unlimited. It needs to be within certain confines of security so that people of all ages would feel comfortable while surfing the Internet and enjoy the wealth of information that it contains. Gatekeeping equips Internet intermediaries with great powers, and practicing these powers requires great responsibility. Moral and social responsibilities are no less important than freedom of expression. A delicate balance needs to be maintained between having a wide forum for discussion and ascertaining that free speech does not instigate violence. Internet companies should assume responsibility as the buck stops with them. They should be committed to

---

<sup>4</sup> Instagram (also known as IG or Insta), was launched by Kevin Systrom and Mike Krieger in 2010, and in 2012 it was bought by Facebook for \$1bn. It is a photo and video-sharing social networking service.

<sup>5</sup> Snapchat is also an image and video messaging application created in 2011 by Evan Spiegel, Bobby Murphy and Reggie Brown while they were studying at Stanford University. Facebook tried to buy Snapchat in 2013 for \$3 billion to boost its appeal with younger users but its offer was declined.

safety considerations. They should carefully think about the trade-offs resulting from their decisions and conduct.

I have been studying evil on the Internet for a long time. Cyberbullying is a heart-breaking phenomenon because it can be avoided. Attentiveness, care, responsibility, appropriate monitoring and support for the victims and also for the bullies are all important in the fight against bullying and cyberbullying. The bullies might have been themselves subjected to bullying and domestic abuse. With a better understanding of the reasons that make individuals bullies, it is possible to reduce the harms of cyberbullying (Hinduja and Patchin 2009; Ditch the Label 2018; Oakes 2019). Moreover, the technology that enables cyberbullying can be used against it. For instance, applications that would tell victims if someone who is threatening to them is nearby. The ability exists. It is a question of will and investment.

## 8. Conclusion

Following Aristotle, and per the data cited here and in many other studies about the harms of bullying and cyberbullying, relevant stakeholders cannot claim ignorance as basis for inaction. Since the Internet entered its commercial phase in the 1990s, they have acquired the understanding of the harms of bullying and cyberbullying; therefore, stakeholders are expected to take measures that would have corrected or avoided those social wrongs. It is very troubling that schools and Internet intermediaries failed to tackle bullying and cyberbullying that are often related and supplement each other to the extent they should have, either due to *akrasia* or due to vices such as dismissiveness, laziness, dogmatism, self-indulgence and lack of moral and social responsibility.

I opened with Prime Minister May and I will close with Sir Tim Berners-Lee, one of the forefathers of the Internet, who initiated a global campaign to save the Internet from political manipulation, fake news, privacy violations and other threatening forces that might bring about a “digital dystopia”. Emily Sharpe, the director of policy at the Sir Tim’s Web Foundation, said:

The web’s power to be a force for good is under threat and people are crying out for change. We are determined to shape that debate using the framework that the Contract sets out [...]. Ultimately, we need a global movement for the web like we now have for the environment, so that governments and

companies are far more responsive to citizens than they are today. (Sample 2019)

People, when acting collectively, have power. Word of mouth travels fast in the digital age: People can send and receive information to family, friends and colleagues through social media and are able to influence others by launching online petitions and campaigns. People mobilize crowds to challenge corporations and bring about change. Activists have, for example, evoked awareness to the 2015 Paris Agreement on climate change and urged people to choose tap water over bottled water, recycle their waste, or purchase fair-trade products. I have mentioned the 2020 #StopHateForProfit Campaign at the background of the Black Lives Matter Movement that twisted Zuckerberg's hand to reconsider his stand on sheltering and facilitating hate speech.<sup>6</sup> Companies are more attentive and responsive to people power in the digital age. They are well aware that organised campaigns aimed at increasing awareness to particular problems can rally communities and bring about behavioural change. It is time for a social campaign to move Facebook to do all that it can to curb the challenge of cyberbullying. After all, nothing short than human lives are at stake. And if Internet intermediaries will not be proactive and responsible, then governments should step in to see that people, especially young people, could use the Internet without being subjected to abuse.

## REFERENCES

- Abend, G. 2014. *The Moral Background: An Inquiry into the History of Business Ethics*. Princeton, N. J.: Princeton University Press.
- Aizenkot, D., and G. Kashy-Rosenbaum. 2018. Cyberbullying in WhatsApp classmates' groups: Evaluation of an intervention program implemented in Israeli elementary and middle schools. *New Media and Society*, 20: 4709–4727.
- Alipan, A., J. L. Skues, S. Theiler, et al. 2020. Defining cyberbullying: a multifaceted definition based on the perspectives of emerging adults. *International Journal of Bullying Prevention* 2:79–92.
- Allan, R. 2018. Hard questions: Where do we draw the line on free expression? *Facebook* (August 9, 2018), <https://about.fb.com/news/2018/08/hard-questions-free-expression/>
- Aristotle. 350 BCE. *Nicomachean Ethics*, Book II,

---

<sup>6</sup> For further discussion on the power of social campaigns, see Mavrodieva et al. 2019; Moblab.

- <https://resources.saylor.org/wwwresources/archived/site/wp-content/uploads/2011/10/PHIL103-5.8.pdf>
- Aristotle. 350 BCE. *Nicomachean Ethics*, Book III, <http://classics.mit.edu/Aristotle/nicomachaen.3.iii.html>
- Aristotle. 350 BCE. *Nicomachean Ethics*, Book V, <http://classics.mit.edu/Aristotle/nicomachaen.5.v.html>
- Aristotle. 1962. *Nicomachean Ethics*. Indianapolis: Bobbs-Merrill. Book VII. Or W.D. Ross online edition, <http://classics.mit.edu/Aristotle/nicomachaen.7.vii.html>
- Arthur, G. 2014. Rehtaeh Parsons: Father of cyberbully victim speaks out, *BBC* (April 1, 2014) <https://www.bbc.co.uk/news/magazine-26723618>
- Barr, M. 2010. 1 Ohio school, 4 bullied teens death at own hand. *NBC News* (October 8, 2010).
- Beaty, L. A., and E. B. Alexeyev. 2008. The problem of school bullies: What the research tells us. *Adolescence* 43: 1-10.
- Berlan, E., H. Corliss, A. Field, E. Goodman, and S. Austin. 2010. Sexual orientation and bullying among adolescents in the growing up today study. *Journal of Adolescent Health* 46: 366-371.
- Bero, S. 2020. Holding responsible and taking responsibility. *Law and Philosophy* 39: 263–296.
- Bunton, K. 1998. Social responsibility in covering community: A narrative case study. *Journal of Mass Media Ethics* 13: 232-246.
- Carroll, A. B. 2015. Corporate social responsibility: The centerpiece of competing and complementary frameworks. *Organizational Dynamics* 44: 87-96.
- Carroll, A. B., and K. M. Shabana. 2010. The business case for corporate social responsibility: A review of concepts, research and practice. *International Journal of Management of British Academy of Management* 12: 85-105.
- CBC. 2013. Stephen Harper ‘sickened’ by Rehtaeh Parsons story. *CBC* (April 11, 2013).
- Christians, C., and K. Nordenstreng. 2004. Social responsibility worldwide. *Journal of Mass Media Ethics* 19: 3-28.
- Cioppa, V., A. O’Neil, and W. Craig. 2015. Learning from traditional bullying interventions: A review of research on cyberbullying and best practice. *Aggression and Violent Behavior* 23: 61–68.
- Clement, J. 2020. Number of Facebook users worldwide 2008-2020, *Statista* (April 30, 2020), <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Cohen-Almagor, R. 1994. *The Boundaries of Liberty and Tolerance*. Gainesville, FL: The University Press of Florida.

- Cohen-Almagor, R. 2006. On compromise and coercion. *Ratio Juris*, 19: 434-455.
- Cohen-Almagor, R. 2011. Responsibility of net users. In *The Handbook of Global Communication and Media Ethics*, vol. 1, ed. Robert S. Fortner and P. Mark Fackler. Oxford: Wiley-Blackwell.
- Cohen-Almagor, R. 2015. *Confronting the Internet's Dark Side: Moral and Social Responsibility on the Free Highway*. New York and Washington DC.: Cambridge University Press and Woodrow Wilson Center Press.
- Cohen-Almagor, R. 2017. J. S. Mill's Boundaries of Freedom of Expression: A Critique. *Philosophy*, 92: 565-596.
- Cohen-Almagor, R. 2018. Social responsibility on the internet: Addressing the challenge of cyberbullying. *Aggression and Violent Behavior* 39: 42-52.
- Cohen-Almagor, R., O. Arbel-Ganz, and A. Kasher, eds. 2012. *Public Responsibility in Israel*. Tel Aviv and Jerusalem: Hakibbutz Hameuchad and Mishkanot Shaananim, Hebrew.
- Collins, L. 2008. Friend game. *New Yorker* (January 21, 2008): 34-41.
- Crimesider staff. 2010. Sladjana Vidovic: Four bullied teens dead by their own hands at Ohio School. *CBS News* (October 8, 2010).
- Cross, D., T. Shaw, and K. Hadwen. 2016. Longitudinal impact of the cyber friendly schools program on adolescents' cyberbullying behavior. *Aggressive Behavior* 42: 166-180.
- Department for Children, Schools and Families. 2007. *Safe to Learn: Embedding Anti-Bullying Work in Schools*. Nottingham: DCSF Publications.
- Department for Children, Schools and Families. 2009. *Safe from Bullying*. Nottingham: DCSF Publications.
- Department of Education and Damian Hinds. 2019. National mental health Programme between Schools and NHS, *Gov.uk*, <https://www.gov.uk/government/news/national-mental-health-programme-between-schools-and-nhs>
- Deutsch, L. 2008. Woman indicted in Missouri MySpace suicide case. *Associated Press Online* (May 16, 2008).
- Diehl, C. 2015. Students protest new threats posted on Yik Yak, this time at Lewis & Clark, *USA Today* (November 18, 2015), <https://eu.usatoday.com/story/college/2015/11/18/students-protest-new-threats-posted-on-yik-yak-this-time-at-lewis-clark/37409191/>
- Ditch the Label. 2018. Why do people bully? The Scientific Reasons. *Ditch the Label*, (November 14, 2018), <https://www.ditchthelabel.org/why-do-people-bully/>

- Doehne, M., M. von Grundherr, and M. Schafer. 2018. Peer influence in bullying: The autonomy-enhancing effect of moral competence. *Aggressive Behavior* 44: 591-600.
- Duffy, A. L., and D. Nesdale. 2009. Peer groups, social identity, and children's bullying behaviour. *Social Development* 18: 121-139.
- Edwards, J. 2013. Users on this web site have successfully driven nine teenagers to kill themselves. *Business Insider* (September 16, 2013),  
<https://www.businessinsider.com/askfm-and-teen-suicides-2013-9?r=US&IR=T>
- Erginel, M.M. 2016. *Akrasia* and Conflict in the *Nicomachean Ethics*. *British Journal for the History of Philosophy* 24: 573-593.
- Farrington, D. P., H. Gaffney, F. Losel, et al. 2017. Systematic reviews of the effectiveness of developmental prevention programs in reducing delinquency, aggression, and bullying. *Aggression and Violent Behavior* 33: 91-106.
- FitzPatrick, W. J. 2008. Moral responsibility and normative ignorance: Answering a new skeptical challenge. *Ethics* 118: 589-613.
- Fung, B., and A. Garcia. 2019. Facebook has shut down 5.4 billion fake accounts this year, *CNN* (November 13, 2019),  
<https://edition.cnn.com/2019/11/13/tech/facebook-fake-accounts/index.html>
- Gawu, P. S., and H. Inusah. 2019. Corporate social responsibility: An old wine in a new gourd. *Journal of Philosophy and Culture* 7: 1-6.
- Görzig, A., and L. A. Frumkin. 2013. Cyberbullying experiences on-the-go: When social media can become distressing. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 7.
- Gerson, R., and N. Rappaport. 2011. Cyber cruelty: Understanding and preventing the new bullying. *Adolescent Psychiatry* 1: 67-71.
- Goodpaster, K. E. 2010. Corporate responsibility and its constituents. In *The Oxford Handbook of Business Ethics*, eds. G. G. Brenkert and T. L. Beauchamp, 126-157. New York: Oxford University Press.
- Green, S. J. 2015. WWU student charged with hate crime for alleged online threats. *The Seattle Times* (December 10, 2015),  
<https://www.seattletimes.com/seattle-news/crime/wwu-student-charged-with-hate-crime-for-alleged-online-threats/>
- Grohol, J. M. 2018. The power of deception online: The Megan Meier story. *Psych Central* (July 8, 2018).
- Hendry, M. 2018. TEEN'S TORMENT. Tragic Scots schoolboy, 13, took his own life after suffering cruel online abuse from cyberbullies as Prime Minister urged to act. *The Scottish Sun* (December 14, 2018).

- Hinduja, S., and J. W. Patchin. 2009. *Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying*. Thousand Oaks, CA: Sage.
- Hong, J. S., D. L. Espelage, and J. M. Lee. 2018. School climate and bullying prevention programs. In *The Wiley Handbook on Violence in Education: Forms, Factors, and Preventions*, ed. H. Shapiro, 359-374. New York: Wiley.
- Internet World Stats. 2020. World Internet Usage and Population Statistics, <https://www.internetworldstats.com/stats.htm>
- Jones, T. 2007. Cyber-bullying by classmate's parents ends with teen's life. *Chicago Tribune* (November 16, 2007).
- Jones, T. 2008. A deadly web of deceit. *The Washington Post* (January 10, 2008).
- Kahle, L. 2017. Are sexual minorities more at risk? Bullying victimization among lesbian, gay, bisexual, and questioning youth. *Journal of Interpersonal Violence* (July): 1–19.
- Kaliski, B. S. ed. 2001. *Encyclopedia of Business and Finance*. New York: Macmillan.
- Kaplan, K. 2014. Teens taunted by bullies are more likely to consider, attempt suicide. *LA Times* (March 10, 2014), [http://www.latimes.com/science/sciencenow/la-sci-sn-bullying-cyberbullying-suicide-risk-20140310,0,3990497.story?utm\\_medium=email&utm\\_source=newsletter&utm\\_campaign=cheatsheet\\_morning&cid=newsletter%3Bemail%3Bcheatsheet\\_morning&utm\\_term=Cheat%20Sheet#ixzz2ve0zz1f5](http://www.latimes.com/science/sciencenow/la-sci-sn-bullying-cyberbullying-suicide-risk-20140310,0,3990497.story?utm_medium=email&utm_source=newsletter&utm_campaign=cheatsheet_morning&cid=newsletter%3Bemail%3Bcheatsheet_morning&utm_term=Cheat%20Sheet#ixzz2ve0zz1f5)
- Hsien-Te, K., S. Yan, D. Huang, et al. 2019. Understanding cyberbullying on Instagram and Ask.fm via social role detection. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion)*, May 13–17, 2019, San Francisco, CA, USA.
- Kerr, M., R. Janda, and C. Pitts. 2009. *Corporate Social Responsibility – A Legal Analysis*. Markham, Ontario: LexisNexis.
- Kircher, M. M. 2017. Yik Yak goes extinct. *Intelligencer* (April 28, 2017), <http://nymag.com/intelligencer/2017/04/yik-yak-anonymous-gossip-app-dies.html>
- Kowalski, R. M., and S. P. Limber. 2007. Electronic bullying among middle school students. *Journal of Adolescent Health* 41: S22–S30.
- Kowalski, R., S. Limber, and P. W. Agatston. 2008. *Cyber Bullying*. Malden, MA: Blackwell.
- Kraut, R. 2018. Aristotle's Ethics. *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/aristotle-ethics/>

- Krouse, P. 2019. 5 student suicides put focus on bullying issue in Mentor. *Cleveland.com* (January 12, 2019).
- Lallitto, B. 2017. 15 Shocking cases of cyberbullying that led to suicide. *The Clever* (August 15, 2017), <https://www.theclever.com/15-shocking-cases-of-cyberbullying-that-led-to-suicide/>
- Larimer, S. 2015. University of Missouri police arrest suspect in social media death threats. *The Washington Post* (November 11, 2015), <https://www.washingtonpost.com/news/grade-point/wp/2015/11/11/university-of-missouri-police-arrest-suspect-in-social-media-death-threats/>
- Lauer, M., and G. Lewis. 2007. Teenager, Megan Meier, takes her own life after falling victim to a cruel Internet hoax; Megan's parents discuss importance of monitoring children's online activities. *NBC News Transcripts* (November 19, 2007).
- Lauer, M., and J. Larson. 2007. Tina Meier talks about her daughter, Megan, who committed suicide over MySpace relationship that turned out to be hoax by adult neighbour. *NBC News Transcripts* (November 29, 2007).
- Lawrence, G. 1988. Akrasia and clear-eyed akrasia in Nicomachean ethics 7. *Revue de Philosophie Ancienne* 6: 77-106.
- Livingstone, S., L. Haddon, A. Görzig, and K. Ólafsson. 2011. *Risks and Safety on the Internet: The perspective of European children. Full Findings*. LSE, London: EU Kids Online.
- Livingstone, S., G. Mascheroni, and K. Ólafsson et al. 2014. *Children's Online Risks and Opportunities*. London: LSE.
- Long, E., M. Gardani, M. McCann et al. 2020. Mental Health Disorders and Adolescent Peer Relationships. *Social Science and Medicine*, 253 (May).
- Malamuth, N. M., D. Linz, and M.Z. Yao. 2005. The Internet and Aggression: Motivation, Disinhibitory and Opportunity Aspects. In *The Social Net: Understanding Human Behavior in Cyberspace*, ed. Y. Amichai-Hamburger, 163-190. Oxford: Oxford University Press.
- Marshall, M. L. 1994. Ensuring social responsibility. *Thrust for Educational Leadership* 23: 42-43.
- Mavrodieva, A.V., O. K. Rachman, V. B. Harahap, et al. 2019. Role of social media as a soft power tool in raising public awareness and engagement in addressing climate change. *Climate* 7: 122.
- McFadden, C., and M. Fulginiti. 2008. Searching for justice; online harassment. *ABC News Transcript* (March 24, 2008).
- McQuail, D. 2003 *Media Accountability and Freedom of Publication*. New York: Oxford University Press.
- Mill, J. S. 1948. *Utilitarianism, Liberty and Representative Government*. London: J. M. Dent.

- Moblab. 10 Ways People Power Can Change the World, <https://mobilisationlab.org/resources/10-ways-people-power-can-change-the-world/>
- Navarro, R., E. Larrañaga, and S. Yubero. 2018. Cyberbullying. In *SAGE Encyclopedia of the Internet*, ed. Barney Warf. Thousand Oaks, CA.: SAGE Publications Inc.
- Newton, P. 2013. Canadian teen commits suicide after alleged rape, bullying. *CNN* (April 10, 2013).
- Novak, M. 1996. *Business as a Calling: Work and the Examined Life*. New York: Free Press.
- Oakes, K. 2019. Why children become bullies at school. *BBC* (September 16, 2019), <https://www.bbc.com/future/article/20190913-why-some-children-become-merciless-bullies>
- Pabian, S., H. Vandebosch, K. Poels, et al. 2016. Exposure to cyberbullying as a bystander: An investigation of desensitization effects among early adolescents. *Computers in Human Behavior*, 62 (September): 480-487.
- Perren, S., and F. D. Alsaker. 2005. Social behavior and peer relationships of victims, bully-victims, and bullies in kindergarten. *Journal of Child Psychology and Psychiatry* 47: 45-57.
- Pesta, A. 2013. Fighting back: One girl's mission to end cyberbullying. *Cosmopolitan* (June 4, 2013).
- Petrov, C. 2019. Cyberbullying Statistics 2019. *TechJury* (February 28, 2019), <https://techjury.net/stats-about/cyberbullying/#gref>
- Pokin, S. 2007. Megan's story. *Megan Meier Foundation*, <https://meganmeierfoundation.org/megans-story>
- Riggio, R. E. 2013. Are you an easy target for bullies? *Psychology Today* (January 3, 2013), <https://www.psychologytoday.com/gb/blog/cutting-edge-leadership/201301/are-you-easy-target-bullies>
- Rivers, W. L., W. Schramm, and C. G. Christian. 1980. *Responsibility in Mass Communication*. New York: Harper and Row.
- Rosen, J. 2013. The delete squad. *New Republic* (April 29, 2013), <http://www.newrepublic.com/article/113045/free-speech-internet-silicon-valley-making-rules>
- Salmivalli, C. 2010. Bullying and the peer group: A review. *Aggression and Violent Behavior* 15: 112-120.
- Salmivalli, C., A. Huttunen, and K. M. J. Lagerspetz. 1997. Peer networks and bullying in schools. *Scandinavian Journal of Psychology* 38: 305-312.
- Sample, I. 2019. Tim Berners-Lee unveils global plan to save the web. *The Guardian* (November 24, 2019),

- [https://www.theguardian.com/technology/2019/nov/24/tim-berners-lee-unveils-global-plan-to-save-the-internet?utm\\_term=RWRpdG9yaWFsX0d1YXJkaWFuVG9kYXIVS19XZWVrZGF5cy0xOTExMjU%3D&utm\\_source=esp&utm\\_medium=Email&utm\\_campaign=GuardianTodayUK&CMP=GTUK\\_email](https://www.theguardian.com/technology/2019/nov/24/tim-berners-lee-unveils-global-plan-to-save-the-internet?utm_term=RWRpdG9yaWFsX0d1YXJkaWFuVG9kYXIVS19XZWVrZGF5cy0xOTExMjU%3D&utm_source=esp&utm_medium=Email&utm_campaign=GuardianTodayUK&CMP=GTUK_email)
- Sauve Meyer, S. 2012. *Aristotle on Moral Responsibility: Character and Cause*. New York: Oxford University Press.
- Sawyer, D., and R. Roberts. 2008. The MySpace suicide: Ashley Grills tells her story. *ABC News Transcript* (April 1, 2008).
- Schargel, F. 2014. Bullying: What schools, parents and students can do. *Huffington Post* (January 23, 2014), [https://www.huffpost.com/entry/bullying-what-schools-par\\_b\\_4103901?guccounter=1](https://www.huffpost.com/entry/bullying-what-schools-par_b_4103901?guccounter=1)
- Segarra, L. M. 2019. Facebook removed 2.2 billion fake accounts this year. It Only Has 2.38 Billion Active Users. *Fortune* (May 23, 2019), <https://fortune.com/2019/05/23/facebook-fake-accounts-transparency-report/>
- Seith, E. 2018. Calls to tackle cyberbullying following teen suicide. *Tes* (December 12, 2018), <https://www.tes.com/news/calls-tackle-cyberbullying-following-teen-suicide>
- Sharp, R., and K. Griffith. 2020. Mark Zuckerberg finally promises action on hate speech as Facebook shares drop \$56 BILLION after Unilever and Coca-Cola pulled their advertisements. *The Daily Mail* (June 26, 2020), <https://www.dailymail.co.uk/news/article-8464369/Unilever-latest-company-pull-advertising-Facebook-Twitter.html>
- Smith, P., J. Mahdavi, M. Carvalho, and N. Tippett. 2006. *An Investigation into Cyberbullying, Its Forms, Awareness and Impact, and the Relationship between Age and Gender in Cyberbullying*. A Report to the Anti-Bullying Alliance, Unit for School and Family Studies, Goldsmiths College, University of London.
- Steffgen, G., A. Konig, J. Pfetsch, et al. 2011. Are cyberbullies less empathic? Adolescents' cyberbullying behavior and empathic responsiveness. *Cyberpsychology, Behavior, and Social Networking*, 14(11) (November).
- Steinhauer, J. 2008. Verdict in MySpace suicide case. *New York Times* (November 27, 2008).
- Stossel, J., E. Vargas, and D. Roberts. 2007. The hoax; MySpace suicide. *ABC News Transcript* (December 7, 2007).
- Talbert, M. 2019. Moral responsibility. In *Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta: <http://plato.stanford.edu/>

- Trevino, L. K., and K. A. Nelson. 1999. *Managing Business Ethics: Straight Talk about How to Do It Right*. New York: John Wiley.
- Wakefield, J., 2017. Instagram tops cyber-bullying study. *BBC* (July 19, 2017),  
<http://www.bbc.co.uk/news/technology-40643904>
- Walters, G. D., and D. L. Espelage. 2020. Assessing the relationship between cyber and traditional forms of bullying and sexual harassment: Stepping stones or displacement? *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 14, Article 2.
- Wan-Jan, W. S. 2006. Defining corporate social responsibility. *Journal of Public Affairs* 6: 176-184.
- Williams, S. G., and A. J. Godfrey. 2011. What is cyberbullying and how can psychiatric-mental health nurses recognize it? *Journal of Psychosocial Nursing and Mental Health Service* 49: 36-41.
- Wolak, J., K. J. Mitchell, and D. Finkelhor. 2007. Does on-line harassment constitute bullying? An exploration of on-line harassment by known peers and on-line only contacts. *Journal of Adolescent Health* 41: S51-8.
- Ybarra, M. L., M. Diener-West, and P. J. Leaf. 2007. Examining the overlap in internet harassment and school bullying: Implications for school intervention. *Journal of Adolescent Health* 41: S42-S50.



## INFERENCE TO THE BEST EXPLANATION: THE CASE OF POTENTIAL ENERGY

Peter J. Riggs  
Australian National University

Original scientific article – Received: 14/4/2020 Accepted: 21/7/2020

### ***ABSTRACT***

*It has been claimed that kinetic energy is an objective physical quantity whilst at the same time maintaining that potential energy is not. However, by making use of the method of ‘inference to the best explanation’, it may be readily concluded that potential energy is indeed an objective physical quantity. This is done for an example drawn from the foundations of modern chemistry. In order to do so, the criteria of what counts as ‘most probable’ and ‘most reasonable’ are defined and then employed for choosing the best explanation.*

**Keywords:** *Potential energy; inference to the best explanation; objectivity of energy; instrumental hypothesis*

### **1. Introduction**

Any physics textbook will apprise its readers that potential energy is the energy stored in physical systems and is central to the law of the conservation of energy (see e.g. Serway and Jewett 2008, chap. 7-8; Halliday et al. 2013, chap. 8; Young and Freedman 2016, chap. 7; Shankar 2019, chap. 5-6). Physics, engineering and chemistry textbooks are full of examples showing the operation of potential energy in all sorts of physical situations. Potential energy as a principal feature of physical phenomena has been accepted wholeheartedly for more than a century by the vast majority of physicists, chemists, and engineers worldwide.

It is somewhat surprising then, that the respected American physicist Eugene Hecht made the claim in 2016 that potential energy does *not* physically exist. He wrote:

Although PE [potential energy] describes a significant aspect of the state of a system and is therefore indispensable theoretically, it is *no longer required to be a physical actuality* [...]. (Hecht 2016, 2, italics added)

It seems here that Hecht is declaring potential energy to be “indispensable theoretically” so that his conclusion about potential energy not having “physical actuality” would not be immediately dismissed as ridiculous by most scientists and by many philosophers of science. It was not that Hecht was claiming energy *per se* does not exist as he explicitly maintained the “physical actuality” of kinetic energy (i.e. he affirms that kinetic energy exists in the physical world). Hecht, in effect, claimed that potential energy is only an instrumental hypothesis (albeit an indispensable one), i.e. only a theoretical ‘device’ or ‘instrument’, used for tracking changes in physical systems and for making predictions. Such theoretical ‘instruments’ are also referred to by some philosophers of science as ‘conceptual fictions’ (see Stace 1934; Quine 1951; Smart 1968, 152; Giere 1988, 26).

Hecht is pursuing his own realist agenda claiming that kinetic energy exists in the physical world and denying this for potential energy. We’ll briefly summarise, in the next section, Hecht’s justification for this position and why it fails. However, we need to acknowledge upfront that because scientific realism conjures up all sorts of issues, problems and images (see e.g. Giere 1988, chap. 4–5; Okasha 2002, chap. 4; Psillos 2009; McCain 2016, 219–223), we cannot do justice to it in the space of this article. Instead, we shall take the (minimal) realist ontological perspective of energy which holds that energy is a quantity (as it can be given a numerical value) having a physical existence and which is an essential attribute of physical systems. Such a realist ontology is rejected by instrumentalists who view energy as only a theoretical ‘device’ for making predictions and a kind of hypothetical ‘ledger’ to describe changes in physical systems. The expressions ‘physically objective quantity’ and ‘physical objectivity’ will be used as a shorthand for the minimal realist ontology in which energy has an essential physical existence (rather than using Hecht’s terms ‘physical actuality’ and ‘objective reality’). The relevant issues may then be discussed without having to venture into the wider morass of philosophical realism.

The aim of this article is to present a case for potential energy being a physically objective quantity. The rest of the article will proceed as

follows. Section 2 discusses the failure of Hecht's physical argument and highlights the philosophical implications to be examined. Section 3 discusses the importance of potential energy and its explanatory role in contemporary physics. Section 4 outlines the method of 'inference to the best explanation' and presents a concrete example of how to find the best explanation. Section 5 applies this method to the case of potential energy. Section 6 delivers a verdict on the ontological status of potential energy. Section 7 answers the question about potential energy's theoretical indispensability and summarises the article's conclusions.

## **2. Failure of Hecht's Physical Argument and its Philosophical Implications**

Hecht's physical argument against potential energy is based on his claims about what is and what is not directly measurable. He assumed that if a quantity could be directly measured then it has physical objectivity (Hecht 2016, 8). Hecht concluded that potential energy is not a physically objective quantity chiefly because he maintained that it cannot be directly measured. On the other hand, he claimed that kinetic energy is directly measurable and hence is a physically objective quantity. Hecht's approach has some features in common with the view of prominent instrumentalist Bas van Fraassen (1980), except for Hecht's acceptance of the physical objectivity of kinetic energy. Nonetheless, Hecht is excessively instrumental as his view:

- must accept that the quantitative formalism of potential energy has enormous utility, i.e. the widespread usefulness and accuracy of the (mathematical) formalism for description and prediction of phenomena in physics, engineering and chemistry, but does so without explaining why potential energy has this immense utility; and
- concedes that the potential energy hypothesis cannot be done away with (i.e. is "indispensable theoretically"), again without explaining why this is so.

In addition to these shortcomings, it has become evident that Hecht's view ultimately 'derails' for, most critically, kinetic energy is no more directly measurable than is potential energy. The lack of direct measurability of kinetic energy may be seen as follows.

In order to 'measure' the kinetic energy of an object, a number of steps are required and this usually includes the empirical determination of the object's speed. Speed may be determined by performing measurements of successive positions of an object over specified time intervals. The object's

speed is then calculated using these values of position and time and this result is substituted into an equation for kinetic energy to obtain a numerical value (Riggs 2019, 3). Furthermore, it turns out that detailed analyses of alternative empirical methods for ‘measuring’ an object’s kinetic energy reveal that the operations involve direct measurements only of position and/or time (Riggs 2019, 4). In other words, there simply cannot be any direct measurement of kinetic energy. Therefore, if Hecht’s assumption regarding how to establish physical objectivity is accepted, then neither kinetic energy nor potential energy could be accepted as objective quantities! Hecht’s argument thereby fails to achieve his goal of establishing that potential energy is not physically objective whilst still affirming the objectivity of kinetic energy.

There are obviously some pertinent philosophical issues arising in this context. Consider the following broad questions:

- What is implied if potential energy does *not* have physical objectivity?  
and
- What follows if potential energy *is* an objective physical quantity?

In particular, what issues immediately stand out from these questions? In respect to the first question, if potential energy is not an objective physical quantity then we have the seemingly inexplicable situation where, despite not having physical objectivity, potential energy nevertheless quantitatively describes a significant aspect of the state of *any physical system*, irrespective of the constitution of the system (e.g. being composed of ‘dark’ matter) and of any and all extreme conditions in its vicinity. Such conditions could include being subject to the pressure in the core of a planet, or the temperature inside a star, or the gravitational ‘tidal forces’ exerted near a stellar-mass black hole. In respect to the second question, if potential energy is an objective physical quantity then it should be possible to validly infer its objective status. The method of ‘inference to the best explanation’ will be applied below as part of a case which concludes that potential energy is a physically objective quantity. This conclusion will also resolve the above described seemingly inexplicable situation.

### **3. Potential Energy in Contemporary Physics and its Explanatory Significance**

There is no *general* expression for energy, as articulated by the French physicist and philosopher of science, Henri Poincaré:

[...] In every particular case we clearly see what energy is, and we can give it at least a provisory definition: but it is impossible to find a general definition of it (Poincaré 1905, 132).

Our observations of the natural world have led to the stipulation that there are only two *fundamental* types of energy. Needless to say, these are kinetic energy (energy of motion) and potential energy (stored energy). We typically illustrate these two types of energy with reference to individual physical situations. There are, of course, countless numbers of very common examples including: falling objects, ferrous metal fragments pulled towards magnets, pieces of paper attracted to plastic rubbed on wool, etc. Such instances also show the working of the law of energy conservation, i.e. energy may be transformed from potential to kinetic (and vice versa) thereby conserving total energy.

Potential energy in contemporary physics is understood as an aspect of physical systems, as stated in a leading physics textbook:

[...] [i]f the energy change of the system is not in the form of kinetic energy [...] we call the energy storage mechanism [...] potential energy [...] [and] find that the potential energy of a system can only be associated with specific types of forces acting between members of a system. (Serway and Jewett 2008, 178)

The quantification of a system's potential energy is expressed in terms of the relative configuration of the parts of the system, e.g. positions of particles making up the system. It is also well established that each force is mediated by a physical field which ensures causal connection and conservation of energy. Consequently, potential energy may be characterised as the energy stored in physical fields. An electrically charged particle such as an electron, for example, placed inside an external electric field will gain kinetic energy and accelerate by drawing on some of the potential energy in the electric field enclosing it.

The enormous utility of the potential energy hypothesis allows for both qualitative and quantitative descriptions of dynamical phenomena (i.e. of the changes that occur in physical systems). In this context, it should be pointed out that phenomena explained by the hypothesis of potential energy are *exceedingly familiar* in our homes, workplaces, and in research laboratories and industrial facilities. The most commonplace of such phenomena include:

- conversions of practical energy modes as observed every day, e.g. electrical to light, chemical to mechanical, solar to heat; and
- the self-restoration of deformed elastic materials with accompanying motion, e.g. compressed or extended springs.

In its quantified expressions, the hypothesis of potential energy is a *crucial part* of descriptions of the changes in physical systems in accordance with known laws of nature in specific areas of science, e.g. with the laws of electromagnetism, nuclear reactions, gravitation, materials science, and chemical reactivity. The quantitative expressions of potential energy, not surprisingly, are different for each of the fundamental forces of nature. The expression for the electrostatic potential energy in a given spatial region, for example, depends on the number, polarity, and distribution of electric charges in that region. This is totally different from say, the expression for the potential energy of an atomic nucleus due to the Strong Nuclear force (i.e. the force which holds the nucleus together). The various expressions for potential energy reflect the different natures of the fundamental forces.

Subject to the law of energy conservation, quantitative changes in potential energy appear as kinetic energy in its various forms, e.g. heat (as increased kinetic energy of surrounding particles). Indeed, the (factual) outcomes of a staggering number of physics, chemical, and biological experiments and also engineering processes (see Jaffe and Taylor 2018, esp. chap. 9) which are predicted and explained by the hypothesis of potential energy testifies to it being crucial to describing changes in physical systems. Although these empirical outcomes highlight the utility of potential energy, enquiring into their basis inevitably leads back to the questions of whether potential energy is an objective physical quantity and why it is that potential energy (in Hecht's words) "describes a significant aspect of the state of a system [...]" (Hecht 2016, 2). We shall provide suitable answers to these questions.

#### 4. Finding the Best Explanation

Arguments for and against the method of 'inference to the best explanation' are easily found in the philosophical literature (e.g. Harman 1965; Vogel 1998; Okasha 2002; Lipton 2004; Psillos 2009; Mackonis 2013; McCain and Poston 2017). It is beyond the scope of this article to review these arguments. Instead, we shall accept (as many philosophers do) that this method yields explanations which are true (or at least very likely to be true), when based on accurate premises and properly conducted (see Psillos 2009, chap. 10; Brössel 2013, 53) as the method "exploits the truth-conducive virtues of explanation" (Kosso 1992, 98). In order to assist

in a determination of the ontological status of potential energy (i.e. to conclude that it is a physically objective quantity), we shall perform an inference to the best explanation.

The schema for making an inference to the best explanation has the general form (McCain 2016, 158):

- (1) There is a set *F* of related facts (e.g. observation statements, measurements, etc.) requiring an explanation.
- (2) A particular explanation *E* accounts for all the facts in *F*.
- (3) *E* accounts for *F* better than any other known explanation.

Yet, the schema (1) – (3) is just the ‘bare bones’ and we need to ‘flesh out’ an inference to the best explanation by initially adding the following to this schema (cf. Schick and Vaughn 1995, chap. 5):

- (4) Any acceptable explanation must not be logically inconsistent.
- (5) Any acceptable explanation must be compatible with relevant, established theories or confirmed data (i.e. with background knowledge).
- (6) Any acceptable explanation must not postulate entities or activities of dubious kinds, e.g. violations of known natural laws, speculative (and unverified) physical effects, animated cadavers, magical spells, etc.

A domestic example will serve to demonstrate the operation of the schema (1) – (6) and how we ought to decide which explanation counts as best, before this schema is applied to the case of potential energy. Suppose that I arrive at my (locked) house one night to find that the pieces on my chessboard have been orderly rearranged from where they were earlier that same day. I observe that no one is in the house, there are no indications of forced entry, no items appear to be missing, and nothing seems to have been disturbed except for the chess pieces. How then might this orderly rearrangement of the chess pieces be explained? I begin my deliberations with bringing to mind aspects which are relevant to this situation by:

- attempting to recall all of my actions before leaving my house this morning;
- noting that no visitors nor deliveries were expected or scheduled today;
- noting that house burglaries are quite common in my city;
- noting who has a key to my house;
- noting that several people have previously told me that the chess pieces should be rearranged on aesthetic grounds; and

- discounting any dubious entities and processes which might be postulated as causes of the movement of the chess pieces (e.g. psychic levitation).

I thereby incorporate conditions (4) – (6) into the process of formulating suitable explanations from which the best one may be inferred.

Using both my observations and thoughts on the situation, I am led to the formulation of four possible explanations:

- a) It was myself who repositioned the chess pieces before leaving the house in the morning but, as I had several pressing issues on my mind needing immediate attention, I simply forgot that I had moved them and have not been able to recall this.
- b) There has been a ground shaking event at my house's location during the day which caused the chess pieces to shift position.
- c) An unknown individual broke into my house in a way yet to be discovered, moved the chess pieces on motives unknown and then departed without taking anything.
- d) A particular friend who has the only other key, let herself into my house, rearranged the chess pieces, removed nothing, and locked the house upon leaving.

Since I am not prone to having memory lapses and there are no apparent signs of a break-in, or of robbery, nor any obvious indications of a ground shaking event, I would tend to accept explanation (d). However, just because I have not found any evidence of a break-in, or of ground shaking, or of definite forgetfulness does not, by themselves, eliminate explanations (a), (b) and (c), i.e. all four explanations still account for the movement of the chess pieces and satisfy conditions (4) – (6) above. In the absence of evidence to the contrary, the four explanations remain in contention and I need something more to decide which of the four explanations is the best explanation and why it is best.

Additional factors have to be taken into account to make and justify this decision. There is (as a minimum) one relevant factor which applies to each of the explanations (a) – (d). First, on the basis of my medical history and current medical state, my personal physician assures me that any loss of memory of recent experienced events is extremely unlikely. Second, the local geology is so stable that a ground shaking event would be highly improbable. Third, given what usually occurs in house burglaries in my city, it would also be improbable that a stranger should go to the trouble of breaking into my house and then take nothing when there are valuable items inside. Fourth, knowing the character of my friend with the house

key, it is quite likely that she would let herself into my house when I am not there so that she can ‘play a joke’ on me.

My analysis of the situation and its additional factors leads me to arrive at the following deductions. The possibility of any continuing inability to recall recent events on my part may safely be dismissed. A ground shaking event strong enough to shift the chess pieces would not leave them as found, i.e. all upright and orderly. A burglar would not be bothered to orderly shift the chess pieces (or anything else). My friend with the house key would move the chess pieces if she was alone in my house.

The criteria for choosing which particular explanation is best out of a competing set of explanations has been argued over in the philosophical literature (at least) since the publication of Gilbert Harman’s seminal papers on the topic (Harman 1965, 1968) and remains the subject of debate (cf. Glass 2012, 412; McCain 2016, 159-160). This debate is obfuscated by the situation that the meanings of some of the terms used in discussions of the criteria vary. Most prominent amongst criteria deemed suitable is the criterion of coherence which is considered central to determining the best explanation (Kosso 1992, 100). Adolfas Mackonis, for example, draws attention to the term ‘coherence’ sometimes being used to mean ‘consistency with background knowledge’ and on other occasions to mean ‘plausibility with respect to background knowledge’ (Mackonis 2013, 980). We shall avoid adding to the confusion over ‘coherence’ by not utilising the term at all. How then shall we decide which explanation is *best*?

Although explanations (a), (b) and (c) are not logically excluded, in light of the facts, the additional factors and my deductions, I infer that explanation (d) is the *best* explanation. Why? There are two clear reasons for reaching this conclusion. Given how the argument developed following the schema (1) – (6), these reasons are that explanation (d) is:

- ❖ the most probable of the four explanations as it has likely circumstances in its favour and the other three explanations do not; and
- ❖ the most reasonable of the four explanations as it stands up better to rational analysis than the other three explanations do.

These are sufficient for deciding which of the explanations (a) – (d) is *best*. Therefore, the criteria for choosing the best explanation may be limited to ‘most probable’ and ‘most reasonable’ (as defined). This outcome vindicates our use of the same criteria in the case of potential energy and we need look no further for suitable criteria.

## 5. Applying ‘Inference to the Best Explanation’ to the Potential Energy Case

It was stated in Section 2 that potential energy (and also kinetic energy) are not directly measurable. The method of ‘inference to the best explanation’ is the appropriate form of inference when dealing with quantities that are not directly measured, as pointed out by Adolfas Mackonis:

*Any argument for the truth or reality of a theoretical term, concept, entity or theory in general is an instance of IBE [Inference to the Best Explanation]. IBE is a fundamental component of theoretical reasoning in general and of scientific practice in particular. (Mackonis 2013, 975–976, italics in original)*

We shall now apply the schema for finding the best explanation to the potential energy case. Returning to the schema (1)–(6), suppose explanation E has both kinetic energy and potential energy as components. Let the set F in condition (1) be the huge number and assortment of both commonplace and scientific facts as mentioned in Section 3. Let condition (2) hold for E being the potential energy explanation, i.e. all the facts comprising this set F are explained by E. On the basis of both empirical and theoretical findings made over more than a century, conditions (4)–(6) also all hold for E. The question which then emerges is whether condition (3) holds. If so, then this might be considered enough justification for taking E to be correct.

Just as in the domestic example, we need to stipulate the relevant alternatives to E to answer this question. These alternatives may be denoted, for current purposes, as explanations which accept kinetic energy as a component but *not* potential energy. Assume that conditions (4)–(6) apply to the alternatives to E and that these alternatives can explain (by various means not including potential energy) the facts in set F. Given the depiction of E and its alternatives, we have a situation paralleling the domestic example as, in order to decide whether condition (3) holds, other factors are required. Fortunately, a decision regarding condition (3) is readily ascertainable by contemplating an example drawn from the foundations of modern chemistry. In doing so, we will proceed in a similar manner to the domestic example and employ the same criteria of ‘most probable’ and ‘most reasonable’ for choosing which explanation is *best*.

Much of chemistry and biochemistry is based on an understanding of the bonds between atoms/molecules, i.e. on chemical bonds and their

reactivity (Luo 2007, 1; Kolasinski 2017, 570; Sagan and Mitoraj 2019, 4616). There can be little doubt that chemical bonds (of some kind) do have physical objectivity or else macroscopic matter (including biological organisms) would not exist. Moreover, the objectivity of chemical bonds is now well established by the empirical data collected from a variety of experiments (see Shin et al. 2002; Friedrich 2018; Wilson et al. 2019; Hu et al. 2019; Cao et al. 2020), as frankly expressed by Valerio Magnasco: “Experimental evidence shows that molecules [...] have a structure made of bonds [...]” (Magnasco 2010, xi).

There are three main classifications of chemical bonds: covalent, ionic, and metallic. These bonds are constituted, to some extent, by the forces between parts of atoms/molecules (i.e. between parts of microscopic physical systems). It was already noted in Section 3 that it is generally accepted that the potential energy of any physical system is associated with specific types of forces between parts of the system. This association of potential energy does, of course, apply to the forces acting on atoms/molecules (Housecroft and Constable 2006, 113–114) and accordingly, applies to chemical bonds. The physical fields which mediate each force not only ensure causal connection and conservation of energy but also (in chemical reactions) ensure the contiguity of bonding.

How are chemical bonds made? Let’s consider a standard account of their formation. The most common bond in molecules is the covalent bond where the electrons from individual atoms are shared in a molecule. The simplest illustration is the single covalent bond between two hydrogen (H) atoms in the hydrogen molecule ( $H_2$ ). When two hydrogen atoms initially separated by a large distance (in comparison to their size) approach each other, the electrons and protons in each atom have kinetic energy and each atom has potential energy. When the atoms become sufficiently close, each will contribute an electron which are then shared between the two atoms forming a covalent bond (Kolasinski 2017, 570–571). Why should these electrons get into a shared arrangement? The answer is straight-forward in terms of potential energy and because natural processes always tend (other things being equal) towards the lowest available energy state (Zumdahl 2009, 595). The standard account for the creation of chemical bonds is that when the atoms closely approach each other, there is a *lowering* of the total potential energy in the course of forming the molecule (Levine 2009, 457; Silberberg 2012, 329). The amount by which the potential energy is reduced appears as (i.e. is converted into) heat which disperses into the surrounding environment (Zumdahl 2009, 411). In general, the lower energy state that arises when atoms bond together creates stability and permits the growth of elaborate physical structures to proceed. We shall see that examining two factors concerning energy and chemical bonds

brings out the issues which are important for making an inference to the best explanation in this example.

The first factor is about the decreases in the energy of the bonding atoms. These decreases actually happen when chemical bonding occurs, e.g. heat is released during the formation of chemical compounds. Such decreases are confirmed by measurements of temperature changes in numerous chemical reactions of relevant kinds showing, independent of specific theoretical models, that decreases in energy do occur when atoms bond. The potential energy explanation offers a mechanism which quantitatively accounts for the energy released when bonds form as the amount of heat measured correlates with the calculated decreases in potential energy (cf. Luo 2007; Zumdahl 2009, 361; Silberberg 2012, 345; Gupta 2016, 391-392). Conversely, reactions in which chemical compounds are dissociated, i.e. reactions that break bonds, require precise energy inputs (e.g. by applying heat or an electric current) for the reactions to proceed (see Luo 2007). Note that chemical reactions will not proceed and no structures will grow unless energetically possible (Gupta 2016, 387).

These energy correlations *tie* potential energy to chemical bonding and therefore strongly support the potential energy explanation. Alternative explanations, i.e. ones without potential energy and for which conditions (4) – (6) apply, are not generally supported by these energy considerations since they must postulate (rather than calculate) some other means to account for the heat released/absorbed in chemical reactions. These alternative explanations lack the very specific *correlations* between the heat released/absorbed and the quantitative changes that are calculable from the formalism of potential energy. This indicates a higher probability for the potential energy explanation than for its alternatives, i.e. the potential energy explanation is the most probable explanation for the formation of chemical bonds.

The second factor concerns potential energy and bond characteristics. Chemical bonds have characteristics such as bond length, bond angle and bond strength, which are quantifiable. Bond length in the hydrogen molecule, for example, is the distance between the nuclei of the two H atoms when the energy of the molecule as a whole is a minimum (Zumdahl 2009, 595). The bond angle is the angle formed by the bonds in a molecule consisting of three or more atoms (Housecroft and Constable 2006, 200). Bond strength is defined in terms of the energy needed to break a particular bond (Silberberg 2012, 339). Potential energy is intimately linked to the characteristics of chemical bonds. Consider the changes in the energy of a molecule which occur when its structure is altered, e.g. when there are changes in bond length. Such changes are quantified by chemists using an

extremely powerful analytical tool called the ‘potential energy surface’ (PES) which gives a molecule’s energy as a function of the positions of its atoms (Gupta 2016, 216). The PES allows molecular shapes (e.g. bond lengths and angles) and reaction rates to be determined, as succinctly stated by chemist V. P. Gupta:

During a reaction process, the molecules undergo structural changes that change their energies. The way the energy of a molecule changes with small changes in its structure is specified by its potential energy surface. (Gupta 2016, 390)

The PES displays potential energy linkages to chemical bonds in the context of their characteristics and demonstrates that potential energy is integral to molecular structure and the conduct of chemical reactions (Wales 2003, 1; Gupta 2016, 218).

These potential energy linkages are vital to the *consistency* of accounts of the stability of chemical compounds, their reactivity, and their resulting structures in conjunction with the forces acting within and between atoms/molecules. This is *not just a matter of its utility* for the following reason. The extent to which the potential energy linkages are essential to the characteristics of (empirically verified) chemical bonds and the chemical structures which arise from them is such that chemical reactions and structure building *does not make sense* without the potential energy linkages. Accordingly, the potential energy explanation stands up to rational analysis in a way that its alternatives do not. Therefore, the potential energy explanation is also the most reasonable explanation for the formation of chemical bonds.

Since the potential energy explanation not only accounts for the relevant facts about chemical bonds but is more probable and more reasonable than its alternatives, the set criteria are met for choosing the best explanation. Therefore, the potential energy explanation is the *best explanation* for the formation of chemical bonds.

## 6. The Ontological Status of Potential Energy

Does potential energy have the ontological status of being a physically objective quantity? Since it has been shown that the potential energy explanation is both the most probable and the most reasonable explanation for the formation of chemical bonds and that the criteria of ‘most probable’ and ‘most reasonable’ are sufficient for making an inference to the best explanation, it has also been inferred that the potential energy explanation

is the *best* explanation for the formation of chemical bonds and, by extension, the chemical structures which subsequently stem from them.

Is this enough to justify concluding that potential energy is an objective physical quantity? Those who think that best explanations are true would answer affirmatively. Although the best explanation argument presented here does offer very compelling grounds for accepting the physical objectivity conclusion, it remains contestable for even best explanations cannot *guarantee* the truth of a conclusion. Therefore, we are arguably still a step removed from conferring physical objectivity on potential energy. What is needed to bridge the gap in this case is one or more instances where the denial of potential energy's physical objectivity would have outcomes contrary to established results. Instances of this kind would allow the argument to advance past the terminal point achieved by 'inference to the best explanation'.

There is at least one such instance relevant to chemical bonds. If potential energy were not an objective physical quantity, then what would follow in light of the energy linkages outlined in the previous section? Let's consider this issue. We have seen that potential energy is intimately linked to chemical bonds in a manner that goes beyond the utility of the potential energy formalism. It was especially emphasised that the potential energy linkages are so essential to chemical bonds that the characteristics and structure of bonds would not make sense without these linkages. Yet, if potential energy was not an objective physical quantity then the potential energy linkages *could not be physically objective either*. In the absence of these linkages, there would be an absurd situation where molecules would not have the physical conditions needed for their existence. Consequently, it would follow from potential energy *not* being physically objective that chemical bonds would also *not* have an objective physical existence, contrary to the experimental evidence. We conclude then, that this finding in conjunction with the potential energy explanation being the best explanation, does indeed warrant the status of physical objectivity for potential energy. More poetically, we might say that *potential energy is no fiction!*

## 7. Final Remarks

The conclusion reached with the aid of the method of 'inference to the best explanation' is that potential energy is a physically objective quantity and not just a theoretical 'instrument'. Acceptance of potential energy as physically objective provides an explanation which extends further than mere theoretical utility can. This conclusion also provides an answer to the

question of why potential energy “describes a significant aspect of the state of a system”. It is precisely because potential energy is an objective physical quantity which is essential to the workings of any physical system that the potential energy formalism provides precise descriptions of aspects of a system’s state. This is why potential energy proves to be theoretically indispensable.

## Acknowledgment

The author thanks the anonymous reviewers and the *EuJAP* Editor-in-Chief for their helpful suggestions.

## REFERENCES

- Brössel, P. 2013. Correlation and Truth. In *EPSA11: Perspectives and Foundational Problems in Philosophy of Science*, eds. V. Karakostas and D. Dieks. Cham, Switzerland: Springer International.
- Cao, K. et al. 2020. Imaging an Unsupported Metal–Metal Bond in Dirhenium Molecules at the Atomic Scale. *Science Advances* 6 (3): eaay5849.
- Friedrich, J. 2018. *Metal–Polymer Systems: Interface Design and Chemical Bonding*. Weinheim, Germany: Wiley-VCH.
- Giere, R. N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Glass, D. H. 2012. Inference to the best explanation: Does it track truth? *Synthese* 185: 411–427.
- Gupta, V. P. 2016. *Principles and Applications of Quantum Chemistry*. Amsterdam: Academic Press.
- Halliday, D., R. Resnick and J. Walker. 2013, 10th edn. *Fundamentals of Physics*. Hoboken, N. J.: Wiley.
- Harman, G. H. 1965. The inference to the best explanation. *Philosophical Review* 74: 88–95.
- Harman, G. H. 1968. Enumerative induction as inference to the best explanation. *Journal of Philosophy* 65: 529–533.
- Hecht, E. 2016. Relativity, potential energy, and mass. *European Journal of Physics* 37: 065804-1–065804-21.
- Housecroft, C. E., and E. C. Constable. 2006, 3rd edn. *Chemistry: An Introduction to Organic, Inorganic and Physical Chemistry*. Upper Saddle River, N.J.: Pearson Prentice Hall.

- Hu, F., L. Shi, and W. Min. 2019. Biological Imaging of Chemical Bonds by Stimulated Raman Scattering Microscopy. *Nature Methods* 16: 830–842.
- Jaffe, R. L., and W. Taylor. 2018. *The Physics of Energy*. Cambridge: Cambridge University Press.
- Kolasinski, K. W. 2017. *Physical Chemistry: How Chemistry Works*. Chichester, West Sussex: Wiley.
- Kosso, P. 1992. *Reading the Book of Nature: An Introduction to the Philosophy of Science*. Cambridge: Cambridge University Press.
- Levine, I. N. 2009, 6th edition. *Quantum Chemistry*. Upper Saddle River, N.J.: Pearson Prentice Hall.
- Lipton, P. 2004, 2nd edition. *Inference to the Best Explanation*. London: Routledge.
- Luo, Y. -R. 2007. *Comprehensive Handbook of Chemical Bond Energies*. Boca Raton, FL.: CRC Press.
- McCain, K. 2016. *The Nature of Scientific Knowledge: An Explanatory Approach*. Cham, Switzerland: Springer.
- McCain, K., and T. Poston, eds. 2017. *Best Explanations: New Essays on Inference to the Best Explanation*. Oxford: Oxford University Press.
- Mackonis, A. 2013. Inference to the Best Explanation, Coherence and Other Explanatory Virtues. *Synthese* 190: 975–995.
- Magnasco, V. 2010. *Models for Bonding in Chemistry*. Hoboken, N.J.: Wiley.
- Okasha, S. 2002. *Philosophy of Science: A Very Short Introduction*. Oxford: Oxford University Press.
- Poincaré, H. (trans. G. B. Halsted). 1905. *Science and Hypothesis*. New York: Science Press.
- Psillos, S. 2009. *Knowing the Structure of Nature: Essays in Realism and Explanation*. Basingstoke, Hampshire: Palgrave Macmillan.
- Quine, W. V. O. 1951. Two Dogmas of Empiricism. *Philosophical Review* 60: 20–43.
- Riggs, P. J. 2019. Comment on ‘Relativity, potential energy, and mass’. *European Journal of Physics* 40: 028001-1–028001-5.
- Sagan, F., and M. P. Mitoraj. 2019. Kinetic and potential energy contributions to a chemical bond from the charge and energy decomposition scheme of extended transition state natural orbitals for chemical valence. *The Journal of Physical Chemistry A* 123: 4616–4622.
- Schick, T. and L. Vaughn. 1995. *How to Think About Weird Things: Critical Thinking for a New Age*. Mountain View, Calif.: Mayfield Publishing.

- Serway, R. A., and J. W. Jewett Jr. 2008, 7th edition. *Physics for Scientists and Engineers with Modern Physics*. Belmont, Calif.: Brooks Cole.
- Shankar, R. 2019. *Fundamentals of Physics I: Mechanics, Relativity, and Thermodynamics*, expanded edition. Yale: Yale University Press.
- Shin, S. -H., T. Luchian, S. Cheley, O. Braha, and H. Bayley. 2002. Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level. *Angewandte Chemie International Edition* 41: 3707–3709.
- Silberberg, M. S. 2012, 6th edition. *Chemistry: The Molecular Nature of Matter and Change*. New York: McGraw-Hill.
- Smart, J. J. C. 1968. *Between Science and Philosophy: An Introduction to Philosophy of Science*. New York: Random House.
- Stace, W. T. 1934. The Present Dilemma in Philosophy. *Journal of Philosophy* 31: 365–372.
- van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Clarendon Press.
- Vogel, J. 1998. Inference to the Best Explanation. In *Routledge Encyclopedia of Philosophy*, ed. E. Craig [online]. London: Routledge.  
<[www.rep.routledge.com/articles/thematic/inference-to-thebest-explanation/v-1](http://www.rep.routledge.com/articles/thematic/inference-to-thebest-explanation/v-1)>. Accessed: 6 March 2020.
- Wales, D. J. 2003. *Energy Landscapes*. Cambridge: Cambridge University Press.
- Wilson, T. R. et al. 2019. Observing the 3D chemical bond and its energy distribution in a projected space. *ChemPhysChem* 20: 3289–3305.
- Young, H. D., and R. A. Freedman. 2016, 14th edn. *Sears and Zemansky's University Physics with Modern Physics*. London: Pearson Education.
- Zumdahl, S. S. 2009. *Chemical Principles*. Boston: Houghton Mifflin.



## BOOK REVIEW

John McMillan

***THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS***

**Oxford University Press, 2018**

**ISBN-13: 978-0199603756**

**ISBN-10: 0199603758**

IVA MARTINIĆ

Faculty of Humanities and Social Sciences, University of Rijeka

With the advancement of bioethics, its formation as an interdisciplinary field and its separation from applied ethics, the question of its methodology comes to the forefront. John McMillan's book *Methods of Bioethics: An Essay in Meta-Bioethics* focuses on this question.

The book has 3 main parts: Bioethics—explaining the definition of bioethics and what “good” bioethics should entail; The Specter of Bioethics—explaining methods based on a theory-driven approach and fact/value distinction; and The Methods of Bioethics—in which McMillan gives various examples of explanations of empirical ethics, what ethical argument should be, the connection between speculative argument and bioethics, and an important way of accessing bioethics—Drawing Distinctions, which involves defining and analyzing moral concepts. In what follows, I will provide a summary of the chapters that, in my opinion, bring to the forth the main theses advanced by McMillan.

In the first chapter of Part One, McMillan does not define bioethics by necessary and sufficient conditions, but defines it by listing the essential goals, questions, and characteristics of bioethics. He introduces bioethics as a broad field that includes areas such as research ethics, philosophical bioethics, empirical bioethics, and public health-medical ethics. He also emphasizes the connection between bioethics and policy making and the fact that bioethicists in public policy making should be sensitive to moral implications that ethical views have when applied to law. In addition, as far as academic bioethics is concerned, McMillan warns of its possible inapplicability to bioethics because of the methodology and topics covered. In doing so, McMillan questions ways of doing bioethics that rely on standard normative theories such as utilitarianism, virtue ethics,

deontology, including those based on theological considerations, thereby diminishing a possible interdisciplinary approach to moral phenomena in practical and complex situations. McMillan argues that the main goal of bioethics is to apply moral reasoning to practical issues with the goal of devising normative recommendations that will address the problems confronted by patients, doctors, and the interested public.

In Chapter Three, McMillan seeks to distinguish the definition of bioethics as an interdisciplinary field from what constitutes “good” bioethics. McMillan proposes that moral reasoning, understood as a process in which practical moral questions and relevant facts are extrapolated and subjected to normative arguments, should be the main method of good bioethics. Moral reasoning in bioethics involves not only philosophical theorizing about ethics, but it also relies on norms grounded in legal systems and professional contexts that are essential for solving practical problems. Thus, McMillan emphasizes practical normativity as a necessary feature of good bioethics because it contributes to understanding and problem solving.

McMillan summarizes the history of debates in bioethics by reflecting on the main ideas brought by these debates. With respect to the discussions of bioethics, McMillan concludes that bioethics involves: interdisciplinary research, the application of moral reasoning, the structure of different fields of work, and most importantly, the education of the reader on methodological skills in direct practice, which is also the purpose of this book.

In the second part of the book, McMillan looks at different views on bioethics – The Moral Mantra and the Tedious Theory Tendency, The Ethics Sausage Machine, Philosopher Kings and Other Queens of the Sciences. The moral mantra signifies the tendency of bioethics to be laden with the shadows of different approaches. That is, McMillan believes that normative moral theories are not so useful in applied ethics because concentrating only on action that are motivated by moral principles can be restrictive of bioethics. Thus, in the section, The Moral Mantra and the Tedious Theory Tendency, McMillan emphasizes the privileged status of normative moral theories as a major problem. In addition to utilitarianism, deontological ethics, theology, and virtue ethics, more recent theories such as empirical, feminist, and narrative ethics also do not facilitate the pursuit of bioethics. McMillan draws such conclusions primarily from the four principles enacted in the 1970s by Tom Beauchamp and James Childress - who sought to unify the moral theories into 4 fundamental principles (autonomy, justice, beneficence, and non-maleficence) that would serve as a framework for delivering ethical justification.

According to McMillan, theory-driven approaches, such as deontological ethics, utilitarianism, and theology, pose a threat to bioethics because they might include other premises that contradict our normative intuitions. This is a problem McMillan calls the “Ethics Sausage Machine”. If we do not accept the whole “machine” (e.g. utilitarianism) we will not accept the “sausage” (arguments based on the theory).

McMillan believes that too much theory in bioethics can weaken our moral intuitions, and as examples of this he criticizes utilitarianism as expounded by Peter Singer, Pope Paul VI’s *Humanae Vitae*, and different deontological approaches. McMillan argues that theoretical approaches rely on premises that are not practically useful for specific bioethical issues. The next problem includes philosophers who think they are always right because they think they possess the methodological key to solving bioethical issues. In this regard, McMillan points out that bioethics is not only about philosophy because philosophy is not the only discipline that can deal with bioethics and the usefulness of moral theory is often over-emphasized. An approach without a specific, privileged demand for a particular theory (without special pleading) is required, and one must strive for progressive interdisciplinary work rather than isolated work of different disciplines.

McMillan’s further discussion extends to the fact/value distinction in which he covers logical positivists, Hilary Putnam, the Humean distinction between is and ought statements, and Moore’s naturalistic fallacy. McMillan builds on Putnam, pointing out that bioethical concepts ignore the distinction between facts and values, because sometimes they are taken as descriptive while at other times they are used as normative concepts. McMillan agrees with this and criticizes the passing of laws based on what people think and concludes that if ethics is nothing but subjective feeling of an individual and says nothing about the world then public opinion survey is suitable for bioethics. The weakness of McMillan’s claim is that he does not explain the reason why only experts should do bioethics, nor does he say who the experts are. McMillan mentions the conceptual approach as wanting in certain respects. When it comes to the conceptual approach, as opposed to the empirical approach to bioethics, McMillan favors the latter because the conceptual approach as a purely theoretical endeavor might not be taken seriously enough by all the relevant parties involved in a bioethical issue. This also calls for the empirical involvement in bioethics as the way forward.

In Chapters 6 through 10, McMillan introduces the methods of Speculative Reason and Drawing Distinctions. Given these methods, McMillan suggests that bioethics should be rigorous and systematic and that it must

be practically normative and empirical. Drawing Distinctions is a fundamental method in law and sociology and in any discipline that relies on distinguishing and clarifying concepts that play an important role in discussions or in problem-solving. By combining Socratic speculative reason and conceptual analysis, McMillan's view becomes more akin to the methodology of analytic philosophy. Thus, it may be unclear to non-philosophers but also non-bioethics practitioners how exactly all of these disciplines will contribute to producing a "good bioethics" argument. It seems that the explanation of exactly how in practice bioethics embraces different methodologies of medicine, sociology, law and philosophy, and what exactly is their job, seems to be missing. This methodology does not necessarily tell us what to do but to act with caution in how we act. In addition to these methods, McMillan cites examples of constructing ethical syllogisms, counterexamples, introducing moral concepts, transcendental distinction, slippery slope arguments, and so on.

In chapter nine, McMillan uses a series of examples to demonstrate how his way of doing bioethics might be applied in practice. This chapter emphasizes cases where interdisciplinary bioethical skills might be usefully applied. To illustrate this point, McMillan uses as examples the concept of "futility" of treatment in the euthanasia debate and a 2013 New Health Inc. case where there was a complaint about unlawful adjustment of intake of fluoride into drinking water. With these examples, McMillan emphasizes the importance of drawing distinctions that signify important moral differences. As one of the more sophisticated argumentative strategies, McMillan points out that sometimes a concept does not imply what one thinks it does. Clarifying an ambiguous concept usually involves considering possible ways in which the concept might be interpreted, and then checking whether the concept is performing the needed normative work.

In general, McMillan indicates that the book was written for beginners who want to study bioethics as well as those who are more familiar with bioethics and are interested in expanding their knowledge. I tend to agree with these views. I would like to add that the detailed descriptions of empirical case studies in the book provide a good overview for those who wish to pursue bioethics without a background in philosophy. In addition, many examples of McMillan's arguments can serve as a template for teachers to work with students of philosophy because they put emphasis on problematic research and argumentation that are based on real-life situations. This book can also be used to connect students of philosophy with students of medicine, law, sociology and history, with the aim of jointly exploring problematic practical situations. Thus, I highly

recommend it to anyone who is interested in learning more and/or teach about the methodologies of bioethics.



## BOOK REVIEW

Dražen Pehar

***PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST-  
DAYTON***

**CEU Press, 2019**

**ISBN-13: 978-963-386-302-2**

**VIŠESLAV RAOS**

Faculty of Political Science, University of Zagreb

This is Pehar's fourth book, and the second one published by an international academic publisher. The author, originally a philosopher, has over the course of his career as researcher and university lecturer evolved into a political scientist with keen interest in diplomacy, especially regarding the role of international actors in brokering a peace treaty in Bosnia and Herzegovina and maintaining the political order that arose from the Dayton Peace Treaty of 1995.

In his new book, Pehar offers a discursive analysis of what he calls, in Foucauldian terms, “the condition” of Bosnia and Herzegovina, of “peace as war”. In this sense, Pehar follows Foucault’s inversion of von Clausewitz. From the perspective of political science and security studies, one would speak of “negative peace”, a societal state in which the bellicose factors are still present and thus lasting, “positive peace” cannot be established, i.e. achieved. Thus, the author’s main argument is that the legal and political framework of post-Dayton Bosnia and Herzegovina creates conditions for reproduction of political conflict and war-like relations between the country’s main political actors.

The book is divided in two parts, with four chapters dedicated to the first part, three to the second, and an interlude chapter inserted between the two parts. It is worth noting that the individual chapters originally appeared as essays in the *TransConflict* online journal, but were since revised and expanded as to be included and collated in this book.

In the first part, the author offers his interpretation of political and legal reading of the minutiae of the Dayton Peace Treaty, both by domestic and international actors. The second part is a critical examination of the role of

political actors in enabling the state of “peace as war”. In this regard, the author is especially focused on the contested role of the United States in the upholding and expanding of the Dayton political order in Bosnia and Herzegovina. The hiatus chapter five serves as a bridge that connects his main argument about “continuation of war by other means”.

The first chapter is dedicated to the concept of “dediscoursification” in the context of the post-Dayton order. Pehar presents an argument about the inability of ethno-political elites (formerly war parties) in Bosnia and Herzegovina to maintain a common discursive ground. By failing to reach a mutual way of communication about and around the implementation of the peace treaty, each of the main political actors in the country is no longer an *ens loquens*, but an *ens belli*. The author specifically lays the guilt for the process of dediscoursification at the feet of late President Alija Izetbegović, leader of the main Bosniak party, the Party of Democratic Action (SDA).

In the second chapter, Pehar deals with the *Dayton Agreement on Implementing the Federation of Bosnia and Herzegovina*, i.e. the “lesser Dayton treaty”, which deals with political, institutional, legal, and, economic relations inside the Federation and Bosnia and Herzegovina, primarily regarding the power-sharing and consociational arrangements between the Bosniak and Croat ethno-political communities. The author claims that this treaty, largely forgotten both by the academia and the general public, contains provisions which maintain some sense of political agency for the Croat Republic of Herzeg-Bosnia (HRHB).

The third chapter is an endeavor in refuting the political and legal arguments concerning ethnic and linguistic provisions contained in the entity constitutions (Republic of Srpska and Federation of Bosnia and Herzegovina respectively). The chapter is devoted to a counter-argumentative undertaking regarding the decision of the Bosnian-Herzegovinian Constitutional Court in 2000, which, after a petition filed by Alija Izetbegović, ruled that the three peoples legally endowed with constitutionality – Bosniaks, Croats, and Serbs, enjoy that status on the whole territory, without further qualifications.

Further, in chapter four, Pehar opens the well-known and much-debated topic of application and implications of electoral law in Bosnia and Herzegovina, specifically the case of Željko Komšić and elections for the Croatian member of the three-headed presidency of the country. The author presents arguments about the Komšić case in light of the spirit of the Dayton constitution, as well as general principles of democratic representation and electoral linkage between voters and parties and/or

candidates. At the end of this chapter, he draws a rather stark and somewhat hyperbolic parallel between Komšić and Vidkun Quisling, the Norwegian World War II collaborator.

The fifth, interpolated chapter, is a summary of the debate Pehar had with Jasmin Mujanović, a young political scientist, currently based at Elon University, North Carolina. Their debate was led through texts and responses of the two authors in the *TransConflict* journal. Pehar accuses Mujanović of misrepresentation of his arguments and positions on the role of Izetbegović in the early days of the 1992-1995 war in Bosnia and Herzegovina. He also reiterates his positions on the ethnic character of the post-Dayton political system and juxtaposes those with a Bosniak “unitarianism”, i.e. a political position that wishes to dismantle the ethnic (and federal) foundations of the political system. While advancing his arguments, he labels Mujanović as a proponent of such a “unitarian” position.

While the first part of the book is mostly concerned with various types of legal arguments revolving around the nature and spirit of the Bosnian-Herzegovinian constitution, as well as the entity constitution of the Federation of Bosnia and Herzegovina, the second party deals with discourse linked with concrete political action.

Thus, in chapter six, Pehar offers his views on the role of the High Representative, a key feature of the international protectorate that has defined the political system and the policy process in Bosnia and Herzegovina ever since the war. Although Pehar does touch upon concrete persons that have over the years held the post, such as Carlos Westendorp or Paddy Ashdown, this chapter is much more a general assessment of the role of the Office of the High Representative, especially concerning the self-expansion of political capacity by the High Representative himself, as laid down in the so-called “Bonn powers”.

The last two chapters deal with the role of the United States in the post-Dayton political order. In chapter seven, Pehar presents his views of specific narratives about the three ethnic groups in Bosnia and Herzegovina and their relation to US foreign policy towards that country. In chapter eight, he deals specifically with a notion of something that he deems misrepresentation of the Bosnian-Herzegovinian case and the political situation within the Dayton institutional framework in the debates in the US Congress. Once again, he forwards the argument that these misrepresentations and misunderstandings tend to favor the Bosniak discourse, while they tend to be detrimental for the Bosnian-Herzegovinian Croats.

Ever since the signing of the Dayton Peace Agreement, liters of ink have been spilled by numerous authors, both academic and non-academic, about the relative merits and shortcomings of this curious institutional framework. Pehar's new book is a summation and culmination of his previous work and thus, in a way, reiterates his previous observations on the dysfunctionalities of political life in Bosnia and Herzegovina. This book adds a new dimension by highlighting the discursive-legal dimension of the state of "negative peace" or, as the title of the book puts it – "peace as war". This book might from the onset receive negative reactions in Bosniak intellectual circles and among those researchers in Croatia and elsewhere which see Bosnian-Herzegovinian Croats as non-constructive actors in the post-Dayton order. Nevertheless, this book should be read and judged on its own merits. Pehar does not shy away from his obvious preconceptions and political positions, and tries to present them and support them with legal, political, historical, logical, and philosophical arguments in a clear and coherent way. For experts on Bosnia and Herzegovina, much of what is presented in the book might not be that novel, yet the way Pehar presents it rather readable and useful, while the broader, less-informed audience will surely find it very informative.

## ABSTRACTS (SAŽECI)

### DOUBLE VISION, PHOSPHENES AND AFTERIMAGES: NON-ENDORSED REPRESENTATIONS RATHER THAN NON-REPRESENTATIONAL QUALIA

Işık Sarıhan  
Central European University

#### ABSTRACT

Pure representationalism or intentionalism for phenomenal experience is the theory that all introspectible qualitative aspects of a conscious experience can be analyzed as qualities that the experience non-conceptually represents the world to have. Some philosophers have argued that experiences such as afterimages, phosphenes and double vision are counterexamples to the representationalist theory, claiming that they are non-representational states or have non-representational aspects, and they are better explained in a qualia-theoretical framework. I argue that these states are fully representational states of a certain kind, which I call “automatically non-endorsed representations”, experiential states the veridicality of which we are almost never committed to, and which do not trigger explicit belief or disbelief in the mind of the subject. By investigating descriptive accounts of afterimages by two qualia theorists, I speculate that the mistaken claims of some anti-representationalists might be rooted in confusing two senses of the term “seeming”.

**Keywords:** Perception; representationalism; qualia; non-conceptual content; afterimages; double vision

### DVOSTRUKI VID, FOSFENI I PASLIKE: NEPRIHVAĆENE REPREZENTACIJE PRIJE UMJESTO REPREZENTACIJSKIH KVALIJA

Işık Sarıhan  
Central European University

#### SAŽETAK

Čisti reprezentacionalizam ili intencionalizam u pogledu pojavnog iskustva teorija je prema kojoj se svi introspektivno dostupni kvalitativni aspekti svjesnog iskustva mogu analizirati kao kvalitete svijeta koje iskustvo na nepojmovni način reprezentira. Neki filozofi su argumentirali da iskustva poput paslika, fosfena i dvostrukog vida predstavljaju

protuprimjer reprezentacionalističkoj teoriji, tvrdeći da su to nereprezentacijska stanja ili da posjeduju nereprezentacijske aspekte, te da ih je bolje objasniti u okviru teorije kvalija. Argumentiram da su ova stanja u potpunosti reprezentacijska stanja određene vrste koju nazivam „automatski neprihvaćenim reprezentacijama“. To su iskustvena stanja koja najčešće ne prihvaćamo kao istinita i koja ne aktiviraju eksplicitna vjerovanja u umu subjekta. Kroz istraživanje opisnih objašnjenja paslika koje nudi dvoje teoretičara kvalija, predlažem da je moguće da su pogrešne tvrdnje nekih antireprezentacionalista ukorijenjene u miješanju dvaju smisla riječi „naizgled“.

**Ključne riječi:** Percepcija; reprezentacionalizam; kvalija; nepojmovni sadržaj; paslike; dvostruki vid

## KNOWLEDGE AND ASSERTION: A CRITIQUE OF LACKEY

Joshua Anderson  
Virginia State University

### ABSTRACT

In the literature on assertion, there is a common assumption that having the knowledge that *p* is a sufficient condition for having the epistemic right to assert that *p*—call this the Knowledge is Sufficient for Assertion Principle, or KSA. Jennifer Lackey has challenged KSA based on several counterexamples that all, roughly, involve isolated secondhand knowledge. In this article, I argue that Lackey's counterexamples fail to be convincing because her intuition that the agent in her counterexamples both has knowledge and do not have the epistemic right to assert is wrong. The article will progress as follows: In section 2, I present Lackey's argument. In section 3, I suggest some more general reasons for doubting that the agent in her counterexamples actually has knowledge. I then show that from a virtue theoretic and Edward Craig's practical explication of knowledge perspectives the agent in Lackey's counterexamples does not know. Since the agent in Lackey's counterexamples does not have knowledge, she has failed to convincingly prove that KSA is false. In section 4, I conclude by suggesting that, at most, what Lackey's counterexamples demonstrate is a problem with a simplistic evidentialist and/or process reliabilist epistemology.

**Keywords:** Assertion; Jennifer Lackey; secondhand knowledge; virtue epistemology

## ZNANJE I TVRDNJA

Joshua Anderson  
Virginia State University

### SAŽETAK

U literaturi o tvrdnji, uobičajena je pretpostavka da je posjedovanje znanja da p dovoljan uvjet za posjedovanje epistemičkog prava da se tvrdi da p— nazovimo taj princip Znanje je dovoljno za tvrdnju, ili ZDT. Jennifer Lackey dovela je u pitanje ZDT koristeći nekoliko protuprimjera koji, ugrubo, uključuju izolirano znanje iz druge ruke. U ovom članku, argumentiram da Lackeyini protuprimjeri nisu uvjerljivi zbog pogrešne intuicije da djelatnik iz njezinih protuprimjera ujedno posjeduje znanje i epistemičko pravo tvrdnje. Članak se razvija na sljedeći način: U odjeljku 2, predstavljam Lackeyin argument. U odjeljku 3, dajem općenitije razloge za sumnju da djelatnik u njezinim protuprimjerima stvarno posjeduje znanje. Nakon toga pokazujem iz perspektive teorije epistemičkih vrlina i praktičke eksplikacije znanja Edwarda Craiga da djelatnik u Lackeyinim protuprimjerima ne posjeduje znanje. Budući da djelatnik u Lackeyinim protuprimjerima ne posjeduje znanje, ona ne uspijeva uvjerljivo dokazati da je ZDT pogrešan. U odjeljku 4, zaključujem predlažući da Lackeyini protuprimjeri u najboljem slučaju ukazuju na problem s pojednostavljenim varijantama epistemološkog evidencijalizma i/ili procesnog relijabilizma.

**Ključne riječi:** Tvrdnja; Jennifer Lackey; znanje iz druge ruke; epistemologija vrlina

## RIGHT TO BE PUNISHED?

Adriana Placani  
University of Graz

Stearns Broadhead  
University of Graz

### ABSTRACT

It appears at least intuitively appropriate to claim that we owe it to victims to punish those who have wronged them. It also seems plausible to state that we owe it to society to punish those who have violated its norms. However, do we also owe punishment to perpetrators themselves? In other words, do those who commit crimes have a moral right to be punished? This work examines the sustainability of the right to be punished from the

standpoint of the two main theories of rights—the will and the interest conceptions. The right to be punished is shown to be largely indefensible on both accounts: on the will theory, the right to be punished conflicts with autonomy, and it can neither be claimed nor waived by a perpetrator; on the interest theory, a perpetrator's interest in punishment, inasmuch as it exists, is not sufficient to ground a duty on the part of the state.

**Keywords:** Punishment; right to be punished; Duff; communicative theory of punishment; rights; will theory; interest theory

## PRAVO NA KAZNU?

Adriana Placani  
University of Graz

Stearns Broadhead  
University of Graz

## SAŽETAK

Intuitivno se čini prikladnim tvrditi da žrtvama dugujemo kažnjavanje onih koji su im naštetili. Čini se također plauzibilnim tvrditi da dugujemo društvu kaznu za one koji narušavaju njegove norme. Međutim, dugujemo li kaznu i samim počiniteljima? Drugim riječima, imaju li oni koji čine zločine moralnu dužnost biti kažnjeni? Ovaj rad istražuje održivost prava na kaznu iz točke gledišta dviju teorija prava—konceptije volje i interesa. Pokazuje se da je pravo na kaznu uvelike neobranjivo prema obje teorije: prema teoriji volje, pravo na kaznu je u konfliktu s autonomijom te ga se ne može niti tražiti niti ga se odreći; prema teoriji interesa, interes počinitelja da bude kažnjen, ukoliko postoji, nije dovoljan da utvrdi dužnost države.

**Ključne riječi:** Kazna; pravo na kaznu; Duff; komunikativna teorija kazne; prava; teorija volje; teorija interesa

## **CYBERBULLYING, MORAL RESPONSIBILITY, AND SOCIAL NETWORKING: LESSONS FROM THE MEGAN MEIER TRAGEDY**

Raphael Cohen-Almagor  
University of Hull

### **ABSTRACT**

This paper addresses the concepts of moral and social responsibility on the Internet in considering the most troubling phenomenon of cyberbullying that results in loss of life. Specifically, I probe the moral and social responsibilities of Internet users (agents), of the education system in fighting cyberbullying, and of Internet intermediaries. Balance needs to be struck between freedom of expression and social responsibility. The tragic story of Megan Meier serves as an illustrative example and some further incidents in which this ugly phenomenon of cyberbullying had cost young life are mentioned. It is argued that all relevant stakeholders need to think of the consequences of their conduct, that Internet abusers should be accountable for their wrongdoing, and that people who have the ability to stop or at least reduce the risk of cyberbullying should take proactive steps, exhibiting zero tolerance to cyberbullying.

**Keywords:** Aristotle; bullying; cyberbullying; Internet; Megan Meier; moral and social responsibility; social networking

## **KIBERVRSNJAČKO ZLOSTAVLJANJE, MORALNA ODGOVORNOST I DRUŠTVENE MREŽE: POUKE IZ TRAGIČNOG SLUČAJA MEGAN MEIER**

Raphael Cohen-Almagor  
University of Hull

### **SAŽETAK**

Ovaj se članak bavi pojmovima moralne i društvene odgovornosti na internetu razmatrajući najproblematičnije pojave vršnjačkog kiberzlostavljanja koje rezultira gubitkom života. Posebice, razmatram moralne i društvene odgovornosti korisnika interneta (djelatnika), sustava obrazovanja u suzbijanju vršnjačkog kibernasilja i internetskih posrednika. Mora se postići ravnoteža između slobode izražavanja i društvene odgovornosti. Tragična priča o Megan Meier služi kao ilustrativan primjer te se spominje još nekoliko incidenata u kojima je ovaj ružan fenomen

vršnjačkog kiberzlostavljanja doveo do gubitka života. Argumentira se da svi relevantni dionici trebaju razmišljati o posljedicama svojih činova, da zlostavljači na internetu moraju biti odgovorni za svoje loše djelovanje te da bi ljudi koji su u mogućnosti zaustaviti ili barem smanjiti rizik vršnjačkog kiberzlostavljanja trebali poduzeti proaktivne korake, pokazujući nultu toleranciju na vršnjačko kiberzlostavljanje.

**Ključne riječi:** Aristotle; vršnjačko zlostavljanje; kibervršnjačko zlostavljanje; internet; Megan Meier; moralna i društvena odgovornost; društveno umrežavanje

## INFERENCE TO THE BEST EXPLANATION: THE CASE OF POTENTIAL ENERGY

PETER J. RIGGS  
Australian National University

### ABSTRACT

It has been claimed that kinetic energy is an objective physical quantity whilst at the same time maintaining that potential energy is not. However, by making use of the method of 'inference to the best explanation', it may be readily concluded that potential energy is indeed an objective physical quantity. This is done for an example drawn from the foundations of modern chemistry. In order to do so, the criteria of what counts as 'most probable' and 'most reasonable' are defined and then employed for choosing the best explanation.

**Keywords:** Potential energy; inference to the best explanation; objectivity of energy; instrumental hypothesis

## ZAKLJUČAK NA NAJBOLJE OBJAŠNENJE: SLUČAJ POTENCIJALNE ENERGIJE

PETER J. RIGGS  
Australian National University

### SAŽETAK

Neki autori istovremeno tvrde da kinetička energija jest, a potencijalna energija nije objektivna fizička količina. Međutim, koristeći metodu 'zaključka na najbolje objašnjenje', može se zaključiti da je potencijalna energija također objektivna fizička količina. To se čini na temelju primjera

iz temelja suvremene kemije. Kako bi se to postiglo, definiraju se kriteriji koji određuju što vrijedi kao ‘najvjerojatnije’ i ‘najrazumnije’ koji se u nastavku koriste za odabir najboljeg objašnjenja.

**Ključne riječi:** Potencijalna energija; zaključak na najbolje objašnjenje; objektivnost energije; instrumentalna hipoteza

**BOOK REVIEW: John McMillan *THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS*, Oxford University Press 2018**

IVA MARTINIĆ

Faculty of Humanities and Social Sciences, University of Rijeka

**ABSTRACT**

BOOK REVIEW John McMillan *THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS* Oxford University Press, 2018 ISBN-13: 978-0199603756 ISBN-10: 0199603758

**RECENZIJA KNJIGE: John McMillan *THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS*, Oxford University Press 2018**

IVA MARTINIĆ

Filozofski fakultet, Sveučilište u Rijeci

**SAŽETAK**

Recenzija knjige John McMillan *THE METHODS OF BIOETHICS: AN ESSAY IN META-BIOETHICS* Oxford University Press, 2018 ISBN-13: 978-0199603756 ISBN-10: 0199603758

**BOOK REVIEW: Dražen Pehar *PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST-DAYTON*, CEU Press, 2019**

Višeslav Raos  
Faculty of Political Sciences, University of Zagreb

**ABSTRACT**

BOOK REVIEW Dražen Pehar, *PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST-DAYTON*, CEU Press, 2019 ISBN-13: 978-963-386-302-2

**RECENZIJA KNJIGE: Dražen Pehar *PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST-DAYTON*, CEU Press, 2019**

Višeslav Raos  
Fakultet političkih znanosti, Sveučilište u Zagrebu

**SAŽETAK**

RECENZIJA KNJIGE Dražen Pehar, *PEACE AS WAR: BOSNIA AND HERZEGOVINA, POST-DAYTON*, CEU Press, 2019 ISBN-13: 978-963-386-302-2

Translated by Marko Jurjako and proofread by Martina Blečić  
(University of 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.

EuJAP subscribes to Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Users can freely copy and redistribute the material in any medium or format, remix, transform, and build upon the material for any purpose. Users must give appropriate credit, provide a link to the license, and indicate if changes were made. Users may do so in any reasonable manner, but not in any way that suggests the licensor endorses them or their use. Nonetheless, users must distribute their contributions under the same license as the original.



## **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: ZABHR 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)

