Voluntariness, Consent, and Justification



Maximilian Kiener

St. Peter's College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy in Philosophy

Trinity Term 2019

To my parents Elvira and Siegfried

Acknowledgments

I wish to thank *The German National Academic Foundation (Studienstiftung des deutschen Volkes e.V.)* and *The Arts and Humanities Research Council UK* (grant reference number AH/L503885/1) for their generous funding of my doctoral studies at the University of Oxford. Moreover, I am particularly indebted to my supervisors Prof John Hyman and Prof Dominic Wilkinson for their outstanding and invaluable academic support.

Abstract

When we consent to something (e.g. a medical procedure, sex, a tattoo), we permit people to act in ways that would otherwise be legally or morally impermissible. However, in order for consent to be valid, certain conditions must be met. In this thesis, I discuss the condition that consent needs to be *voluntary*, I focus specifically on *consent to medical procedures*, and I argue that such consent is voluntary unless it is motivated by an influence that cannot be *interpersonally justified*. I call my view *Interpersonal Consenter-Consentee Justification (ICCJ).* ICCJ introduces the concept of interpersonal justification, known from the debate on contractualism, to the debate on consent and most clearly differs from other views in the literature by (i) focusing on the reasons or motivation for which a person consents, instead of mere causal links or counterfactuals, (ii) allowing the moral legitimacy and not just the strength of an influence to have a bearing on the voluntariness of consent, and (iii) making the behaviour of the person receiving consent (the consentee) central to its approach.

Word count: 72 196

Table of Contents

INTRODUCTION **7**

PART ONE: THE THEORY OF VOLUNTARY CONSENT **10**

1. THE HISTORY AND DEFINITION OF VOLUNTARINESS AND CONSENT 10

1.1 The Philosophical Sources of the Debate on Consent *10*

1.2 Defining the Voluntariness and Validity of Consent *30*

1.3 Looking Ahead *41*

2. DESIDERATA OF AN ACCOUNT OF VOLUNTARY CONSENT 42

2.1 Voluntariness and Causation *43*

2.2 Voluntariness and Morality *61*

2.3 Voluntariness and Coercion *90*

3. MY PROPOSAL: *INTERPERSONAL CONSENTER-CONSENTEE JUSTIFICATION (ICCJ)* 122

3.1 Treatment *124*

3.2 Clinical Research *140*

3.3 Satisfying the Desiderata *144*

PART TWO: THE PRACTICE OF VOLUNTARY CONSENT **147**

4. NON-PERSUASIVE INFLUENCES: NUDGING AND MANIPULATION 148

4.1 Behavioural Psychology *150*

4.2 Easy Resistibility, Transparency, and Rationality *155*

4.3 ICCJ on Non-Persuasive Influences *169*

5. PAYMENT FOR RESEARCH PARTICIPATION 186

5.1 Clinical Trials and Payment for Participation *188*

5.2 Aversion, Alternatives, and Risk *190*

5.3 ICCJ on Paid Trial Participation *210*

6. LIVING ORGAN DONATION 217

6.1 The Prospective Donor's Predicaments *219*

6.2 Willingness, Relationships, and Moral Agency *222*

6.3 ICCJ on Living Organ Donation *240*

CONCLUSION **249**

REFERENCES **252**

Introduction

Rambha Gajre is a young mother from India who is worried about her son. He is seriously ill and needs life-saving treatment, but Rambha cannot pay for it. She has so little money that she struggles to pay her rent and has been threatened with eviction. To earn enough money for the rent and her son's treatment, Rambha consents to paid participation in a clinical trial. Such participation may be highly intrusive, involve exploitation, and be the consequence of unjust circumstances. Yet, Rambha enrols, and explains her situation as follows:

"Many people commit suicide and I didn't want to become one of those and I didn't want people to think I did it to avoid repaying [i.e. paying the rent]. I have two young kids, 10 and 12 years old. What would become of them? (…) I am helpless, I have to do this [i.e. paid trial participation]. (...) They don't really force us, but I don't have a choice."[1]

Is Rambha's consent to trial participation voluntary? Is it valid? It is true that she deliberately chose to consent. But she is also highly reluctant to undergo these procedures and lacks any other option to earn the money for the rent and her son's treatment.

In medical ethics, people are acknowledged to have a right of self-determination, i.e. a right to decide how their own body will be treated. This means that competent individuals, i.e. individuals possessing the required mental capacities, have the right to decide which available medical procedures to undergo. Consent is a tool for exercising such a right and one by which individuals grant their physicians or clinical investigators

[1] Reference from Wertheimer, 2014, 280-281. Wertheimer listed the NBC News Open Channel (2012) as the source, which longer makes Rambha's story available. Here are two further websites giving information about Rambha: https://mcquaig.typepad.com/blog/2012/03/in-india-oversight-lacking-in-outsourced-drug-trials.html; https://www.firstpost.com/living/indias-poor-pharma-industrys-favourite-guinea-pigs-537682.html

permission to perform a medical procedure, which it would otherwise be illegitimate to perform.

But in order for a competent person's consent to be valid, i.e. make an otherwise impermissible procedure permissible, it has to be both informed and voluntary. Consent is *informed* if the individual knows the relevant facts about the medical procedure. The quantity and nature of these facts has been the focus of a rapidly growing debate. *Voluntariness*, on the other hand, has been neglected. In contrast to most historical documents on informed consent, as well as the law on informed consent,[2] bioethicists have given it short shrift.[3]

However, it is impossible to understand consent as a tool for self-determination without understanding voluntariness. It is not enough for an individual to be competent and informed. Being subjected to coercion or manipulation, being placed in a situation with no acceptable alternative other than to consent to a procedure, or being in an abusive relationship are all potential obstacles to self-determination and highly pertinent to voluntariness, but not to the requirements of competence or information. Rambha's case is only one among hundreds where people consent under pressure. An inquiry into the voluntariness of consent is urgently needed.

In this thesis, I aim to advance the understanding of voluntary consent, in particular voluntary consent to medical procedures. I will argue that consent is voluntary

[2] With regards to research, see *The Nuremberg Code*, §1. *The Declaration of Helsinki*, §§25-26. *The Belmont Report,* C.1. With regards to treatment, see for instance *Freeman v Home Office* [1984]. *Mrs U v Centre for Reproductive Medicine* [2002].

[3] No discussion of voluntariness in Dworkin, 1988. Wear, 1998. Little discussion of pressure in: O'Neill, 2002. Manson and O'Neil, 2007. Young, 2009. Brief mention of voluntariness in: Berg, 2001, 67, 88, 266. Beauchamp and Childress, 2013, 137-140. Maclean, 2009, 133, 156-159. McLean, 2010, 91, 185, 189. Bowman, 2012, Chapter 4. Explicit discussion of "control" in Faden and Beauchamp, 1986, 256-26, 337-381. Review of some parts of the scarce literature on voluntariness in Nelson and Merz, 2002. Only recently did Appelbaum (2009), Nelson (2011), and Wertheimer (2012) present explicit accounts of voluntariness in the context of informed consent.

unless it is motivated by an influence that cannot be interpersonally justified, and I will call this view *Interpersonal Consenter-Consentee Justification (ICCJ)*.[4]

Three features will set ICCJ apart from other views in the literature: (i) ICCJ will focus on the reasons or motivation for which a person consents, instead of mere causal links or counterfactuals, (ii) it will allow the moral legitimacy and not just the strength of an influence to have a bearing on the voluntariness of consent, and (iii) it will make the behaviour of the person receiving consent (the consentee) central to its approach.

I will further explain and support ICCJ with a two-step method. In *Part One*, I will employ a *Top-Down* approach and defend ICCJ on the basis of how it satisfies general desiderata for any satisfactory account of voluntary consent. In *Part Two*, I will employ a *Bottom-Up* approach and support ICCJ in its application to practical challenges, including Rambha's case (Chapter 5).

[4] Formulated positively and without the double negative, my view reads as follows: consent is voluntary if and only if there are either no influences on a person's decision-making or the influences present can be interpersonally justified. I chose the negative formulation and the 'unless' as it signals most clearly that my view will consider the voluntariness of consent to be the default.

Part One: The Theory of Voluntary Consent

1. The History and Definition of Voluntariness and Consent

When is consent to a medical procedure voluntary? Before I answer this question, I first need to explain the philosophical sources of my inquiry (1.1) and provide a definition of its main terms (1.2). The former is critical because questions about voluntary consent are not part of an ahistorical episode of practical ethics but have grown from 20th-century philosophy of action. The latter is critical because the multidisciplinary literature on consent has produced several diverging definitions and thereby removed a common terminological basis.

1.1 The Philosophical Sources of the Debate on Consent

Questions about voluntary consent are not solely concerned with current practical ethics. They are also concerned with aspects that were crucial to various views in 20th-century philosophy of action. Unfortunately, people have largely ignored the important connections between these two philosophical areas. In this section 1.1, I will elucidate these connections and thereby reveal the broad ramifications involved in the investigation of voluntary consent.

1.1.1 Voluntariness and 20th-Century Philosophy of Action

20th-century philosophy of action widely rejected theories according to which voluntary action is action caused by a distinctive mental event.[1] Such theories were predominant in early modern philosophy and found a paradigmatic defender in John Locke, who argued that

"we find in ourselves a *power* to begin or forbear, continue or end several actions of our minds and motions of our bodies, barely by a thought or preference of the mind ordering or, as it were commanding, the doing or not doing such or such a particular action. This *power* (…) is that which we call the *will.* The actual exercise of that power, by directing any particular action, or its forbearance, is that which we call *volition* or *willing.* The forbearance of that action, consequent to such order or command of the mind, is called *voluntary.* And whatsoever action is performed without such a thought of the mind is called *involuntary.*"[2]

According to Locke, voluntary action requires the presence of a particular mental event, i.e. "volition" or "willing", which he explains as "nothing but that particular determination of the mind, whereby, barely by a thought the mind endeavours to give rise, continuation, or stop, to any action which it takes to be in its power."[3]

Although Locke says that acts are voluntary if they are "consequent to" a volition, his view requires a *causal* rather than merely *temporal* relation between volitions and voluntary action, as becomes clear from the following passage:

[1] In this thesis, I will use 'act' as a count noun and 'action' as a mass noun. Others, whom I will quote, depart from this usage. I think my terminology enhances clarity but none of my claims depend on it.
[2] Locke, 2009, 2.21.5. Original emphasis.
[3] Locke, 2009, 2.21.30.

"My right Hand writes, whilst my left Hand is still: What *causes* rest in one, and motion in the other? Nothing but my Will, a Thought of my Mind, [a volition]."[4]

Hence, on Locke's view, action is voluntary if it is *caused by* a volition, and involuntary otherwise. Thus, blinking, reflexes, or digestion will be involuntary because these bodily movements are not caused by a volition. But deliberately handing over one's wallet to a robber will still be caused by a volition and therefore voluntary on Locke's view.[5]

Locke's view is a paradigmatic example of theories postulating volitions because, firstly, he assumes that voluntary action is action that has a particular type of cause and, secondly, because he does not distinguish between volition and intention, or between voluntary action and intentional action. In fact, Locke seems to think that all human action is caused by a volition.

During the 20th century, theories postulating volitions came under severe attack. Especially the criticisms from Ryle and Wittgenstein were so influential that the rejection of 'volitions' became common ground in 20th-century philosophy of action and the treatment of voluntariness mainly a post-volitional project.[6]

Once philosophers rejected theories postulating volitions, they could reconsider several questions that had seemingly received fixed answers, e.g. whether voluntary

[4] Locke, 2009, 4.10.19. Emphasis added.

[5] There are still some challenges for how to interpret Locke here as Locke remained slightly ambiguous about the causal relata: one could interpret him as saying that the volition causes the act as a whole (e.g. the process of moving, raising one's arm), or only the result of the act (e.g. the state of motion, the rise of one's arm). Lowe opted for the latter interpretation on the grounds that otherwise a volition, being part of the act as a whole, would implausibly have to cause itself. Therefore, Lowe considers Locke to claim that an agent acts voluntarily if and only if he wills a certain act and such willing causes the result of the act. Applied to an example: "An agent A raised his arm voluntarily if and only if A willed to raise his arm and A's willing to raise his arm caused A's arm to rise." (Lowe, 2002, 112). For the purpose of illustrating Locke's view as paradigmatic for theories postulating volitions, I do not need to take sides. I will rely on a general picture of such theories without discussing the many differences that exist between different accounts of them.

[6] See Ryle, 2009, chapter 3. Wittgenstein: 1958, p. 151-152. 1980, § 726. 1981, § 577. 2013, §628. Arguably, also the works of James and Russell were particularly important in this regard. See James, 1981, particularly chapter 26. Russell, 1995, particularly chapter 14. For a discussion, see Hyman, 2015, chapters 1 and 4.

action requires causation by some distinctive mental event, or whether there are differences between intentional action, voluntary action, and action as such. Although the 20th century did not pursue these questions in great depth and often gave little attention to voluntariness in particular, two main approaches arose: (I) *Ethical Models* and (II) *Anscombian Models.* In the rest of this sub-section, I will describe them and thereby provide the basis for highlighting parallels to the debate on consent.

I. Ethical Models

The first approach in the 20th century can be called *Ethical Models* as its proponents considered voluntariness to be of general relevance to ethical questions and they were interested in voluntariness precisely for that reason. For instance, Feinberg claimed that voluntariness was generally relevant when considering whether one may paternalistically interfere with a person's action. Feinberg, defending soft paternalism, argued that "the state has the right to prevent self-regarding harmful conduct (…) *when but only when* that conduct is substantially nonvoluntary, or when temporary intervention is necessary to establish whether it is voluntary or not."[7] Voluntary conduct, on the other hand, "preclude[s] interference."[8]

Ryle claimed that voluntariness was generally relevant to the question of whether someone can be excused for his wrongdoing. He argued: "[A person] is accused of making a noise, and the guilt is his, if [and only if] the action was voluntary, like laughing; he has successfully excused himself, if he satisfies us that it was involuntary, like a sneeze."[9] Ryle argued that "[i]n their most ordinary employment 'voluntary' and 'involuntary' are used, with a few minor elasticities, as adjectives applying to *actions*

[7] Feinberg, 1986, 12. Original emphasis.
[8] Feinberg, 1986, 143.
[9] Ryle, 2009, 56. Emphasis added.

which ought not to be done"[10] and thereby tied the concept of voluntariness closely to the ethical context of guilt. Outside of this context, he considered it "absurd to discuss whether *satisfactory, correct or admirable performances* are voluntary or involuntary."[11]

Finally, Nozick claimed that voluntariness is generally relevant to the validity of contracts and whether the transfer of goods satisfies principles of justice. He argued that contracts are valid and "justice in transfer"[12] is obtained if and only if there was a "*voluntary* exchange."[13] And like Ryle, Nozick ties the concept of voluntariness tightly to the ethical context under consideration.

On this basis, *Ethical Models* were able to reconsider questions to which theories postulating volitions had provided seemingly fixed answers. *Ethical Models* rejected the view that voluntary action is action caused by a particular mental event. However, they retained a *negative causal* proposal, i.e. the view that voluntary action is any action *not* caused by certain influences, e.g. coercion.

This negative causal proposal made the voluntariness of a person's behaviour the default: unless a person was under the influence of coercion or a similarly controlling influence, his action would be voluntary; and making voluntariness the default assumption was used to explain other default assumptions made by proponents of *Ethical Models*, e.g. the assumptions that, due the voluntariness of a person's behaviour, (1) it is illegitimate to interfere with his acts on paternalistic grounds (Feinberg), (2) this person can be blamed for his wrongdoing (Ryle), and (3) a transaction with this person is valid (Nozick).

[10] Ryle, 2009, 56. Emphasis added.
[11] Ryle, 2009, 56.
[12] Nozick, 1974, 150.
[13] Nozick, 1974, 262. Emphasis added.

Furthermore, *Ethical Models,* unlike theories postulating volitions, distinguished between intentional action and voluntary action. Almost all of them accepted the following two claims:

(1) Intention without Voluntariness One can act intentionally without acting voluntarily.

(2) Voluntariness without Intention One can act voluntarily without acting intentionally.

Nozick, Ryle, and Feinberg supported (1) *Intention without Voluntariness* with reference to cases of coercion: if someone is severely coerced into handing over his wallet to a robber, he may do so intentionally. Yet, coercion negates the voluntariness of his acts. Hence, according to *Ethical models*, voluntary action and intentional action are different: whereas 'intentional action' is about the beliefs and desires of an agent, the voluntariness of an action depends on whether the agent was forced to act as he did.

On the other hand, Nozick, Ryle, and Feinberg supported (2) *Voluntariness without Intention* in two types of cases. Firstly, there are cases of smiling or frowning, i.e. acts that people do not do intentionally, yet do voluntarily.[14] Secondly, there are cases of negligence where a person neither intends nor foresees that his acts harm someone else, but they do. If the person was not forced to act as he did, he acted voluntarily but not intentionally.[15]

---

[14] Admittedly, Ryle tied the concept of voluntariness closely to the ethical context of guilt. However, he also acknowledged that the use of the terms 'voluntary' and 'involuntary' is still intelligible when applied to acts like smiling or frowning.

[15] It is only at the margins of the debate of the 20th century that people reject *(2) Voluntariness without Intention.* Only Williams stated that "an agent does X fully voluntarily if X-ing is an *intentional* aspect of an action he does which has no inherent or deliberative defect" (Williams, 1990, 4. Emphasis added), and thereby rejected the view that one can act intentionally without acting voluntarily. Williams highlights the methodological advantages of such an understanding of voluntary action: he argues that focusing on the class of (in)voluntary acts so understood facilitates a more fine-grained evaluation of responsibility for intentional acts. Williams focuses on a particular aspect pertinent to responsibility, i.e. the state of mind with which intentional actions are done, and claims that doing so "helps to serve the purposes of justice" (Williams, 1993, 67).

In addition, almost all *Ethical Models* claimed that voluntariness requires the absence of force *and ignorance*, and thereby adopted a broadly Aristotelian view.[16] According to Aristotle's account, only voluntary acts warrant praise or blame, and force as well as ignorance can render these inappropriate. Just as a person may not be blamed for what she was forced to do, a person revealing a secret while ignorant that it is a secret may also be immune from blame if her ignorance was non-culpable. Thus, voluntariness requires the absence of force and ignorance.[17]

However, *Ethical Models* were divided over whether the voluntariness of one person's action depends on the morality of another person's action. Nozick answered affirmatively:

"Whether a person's actions are voluntary depends on what it is that limits his alternatives. If facts of nature do so, the actions are voluntary. (…) Other people's actions place limits on one's available opportunities. Whether this makes one's resulting action non-voluntary depends upon whether these others had the *right to act as they did*."[18]

Nozick claims that there are only certain limitations that negate the voluntariness of action, i.e. those which have been imposed by other people, as opposed to natural limitations, and which are illegitimate because the people imposing them *did not have the right to do so.* Thus, on Nozick's view, the voluntariness of one person's action depends on the morality of other people's action: only morally illegitimate, i.e. limitations that exceed rights, negate voluntariness. Natural facts and legitimate constraints do not negate voluntariness, however great their impact.

[16] See Ryle, 2009. Feinberg, 1986. An exception seems to be Nozick, 1974.
[17] See Aristotle, 1999 (EN), III.
[18] Nozick, 1974, 262. Emphasis added.

On the other hand, Feinberg denied that the voluntariness of one person's action depends on the morality of another person's action:

"One can think of voluntariness as a matter of degree. At one end of a spectrum are the acts (…) which I shall call (…) perfectly voluntary: actions of normal adult human beings *in full control* of their deliberative faculties (…) fully informed of all relevant facts and contingencies, with their eyes wide open, so to speak, and in the absence of all coercive pressure. In the ideal case, there must be calmness and deliberateness (…) no distracting or unsettling emotions, no neurotic compulsion, no misunderstanding. (…) At the very opposite end of the spectrum are [situations] (…) when one *lacks all muscular control*, or when one is knocked down, or pushed, or sent reeling by a blow or an explosion—or when through ignorance one chooses something other than what one means to choose, for instance thinking arsenic powder is table salt, and thus choosing to sprinkle it on one's scrambled eggs."[19]

On Feinberg's view, the voluntariness of a person's action depends on how much control he has over it, i.e. the extent to which he can exercise the physical and mental power required to determine his action and the extent to which he is free of pressure from others. As a person can have more or less control over his action, voluntariness becomes a matter of degree too. Thus, Feinberg thinks that something diminishes voluntariness *to the extent to which* it reduces a person's control and he understands control as a morally neutral concept. Therefore, the moral legitimacy of an influence lacks independent significance. Accordingly, and *pace* Nozick, Feinberg holds that something can undermine the voluntariness of a person's action even if it was *not* imposed by other people and therefore *cannot* be assessed in terms of rights or moral legitimacy, and something can fail to undermine voluntariness even if it was the influence of someone exceeding his rights.

[19] Feinberg, 1986, 104. Emphasis added.

However, Feinberg believes that moral considerations play a role in the *application* of voluntariness. For instance, "depending on (…) the moral or legal purpose to be served"[20], Feinberg explains, we need to make a moral judgment about whether an act was "'voluntary enough' to be immune from [paternalistic] restriction"[21], i.e. whether a person was sufficiently in control of his action, or whether "'consent' is (…) 'voluntary enough' to be valid."[22] In each context, a different degree of voluntariness may be required and moral considerations may help to determine which degree is the correct one.

Ryle seems to be closer to Feinberg than Nozick on the question of whether the voluntariness of one person's action depends on the morality of another person's action. Claiming that someone acts involuntarily if he "could not help doing it"[23] or was "trying hard enough"[24] to avoid the performance of the act, Ryle – like Feinberg – refers to the control of the agent and refrains from any reference to the morality of other people's behaviour. However, when Ryle talks about "external coercion and (…) duress"[25], he seems to allow for the possibility that the illegitimacy of creating coercion or duress is relevant to whether doing so negates the voluntariness of another person's action.

II. Anscombian Models

The second main approach to voluntariness in the 20th century I will term *Anscombian Models* after their most prominent proponent, G.E.M. Anscombe. Anscombe and her followers[26] focused on voluntariness to provide a comprehensive understanding of human agency rather than to answer ethical questions.

[20] Feinberg, 1986, 117.
[21] Feinberg, 1986, 118.
[22] Feinberg, 1986, 174.
[23] Ryle, 2009, 56.
[24] Ryle, 2009, 57.
[25] Ryle, 2009, 58.
[26] Apart from Anscombe, the second most prominent proponent is Anthony Kenny.

Anscombe's objective was "to delineate such concepts as the voluntary and the intentional."[27] Concerning the voluntary, Anscombe explicitly rejected theories postulating volitions. She claims that "[a] positive act of mine is voluntary, not because it is accompanied or preceded by an act of will."[28] Rather, an act is voluntary if and only if

a. "it is in the power of the agent to do it or not (…) [and this is] a matter of physical possibility"[29], and
b. "it is done (…) either for its own sake or for the sake of something else."[30]

Condition (a) requires the agent's physical control to determine his action and condition (b) requires that the agent is in some sense aware of what he is doing, i.e. that he is not in complete ignorance about what his bodily movements amount to. Anscombe's condition (b) may depart slightly from a very similar condition that Kenny defended. Kenny argued that a person acts voluntarily only if he, in a minimal sense, *wants* to act as he did, and added: "to say that an agent wants to Φ in this sense is merely to say that he knows that it is in his power to refrain from Φ-ing if only he will give up one of his chosen purposes."[31] Fortunately, the potential slight difference between Anscombe's 'doing something for its own sake or for the sake of something else' and Kenny's 'wanting to do something in a minimal sense' will not affect my further arguments. Thus, I will assume that they both embrace roughly the same condition (b).

Anscombe then explains that intentional actions are

[27] Anscombe, 2000, 10 (§5).
[28] Anscombe, 1981, 107. See also Kenny, 1975.
[29] Anscombe, 2008, 178. See also Anscombe, 2000, 12-13 (§7). Kenny, 1975, 53.
[30] Anscombe, 1981, 107.
[31] Kenny, 1975, 58.

"the actions to which a certain sense of the question 'Why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting."[32]

For instance, suppose someone orders a taxi and is asked why he did so. If he answers that he needed a taxi in order to arrive at his appointment in time, the aforementioned Why-Question is given application and his positive answer provides a specific reason for his act. Therefore, he acted intentionally. By contrast, if he answers that he accidentally dialled the number of the taxi company, the aforementioned Why-Question is rejected, no reason for the act is given, and his act was therefore unintentional.

On this basis, Anscombe claims that physical movements are *voluntary rather than intentional* when the person is aware of performing them but lacks a specific reason for doing so, e.g. when someone fidgets (Case I), or if "the movements are not considered by the agent, though he can say that they are if he does consider them"[33], e.g. in most instances of smiling or frowning (Case II). Furthermore, she claims that an act is voluntary rather than intentional "if it is the antecedently known concomitant result of one's intentional action"[34] rather than what one intended to do (Case III).

In these cases, a person acts voluntarily because he is in control and the act is done either for its own sake or for the sake of something else. However, the person does not act intentionally because he would reject Anscombe's Why-Question. In Cases I and II, there was no specific reason for the act at all, and in Case III the person did not intend to perform the act under the description of producing the concomitant result.

[32] Anscombe, 2000, 9 (§5).
[33] Anscombe, 2000, 89 (§49).
[34] Anscombe, 2000, 89 (§49).

Anscombe then gives two examples of *involuntariness.* Firstly, she considers peristaltic movement and reflexes of the knee involuntary and, secondly, she considers certain acts due to ignorance involuntary, giving the following example:

"If I do something out of revenge which is in fact advantageous rather than harmful to my enemy, my action, in its description of being advantageous to him, is involuntary."[35]

Peristaltic movement and reflexes of the knee are involuntary because condition (a) of voluntariness is not satisfied: a person cannot directly control these bodily movements. Certain acts done in ignorance, e.g. benefitting my enemy, are involuntary because condition (b) is not satisfied: benefitting my enemy is neither done for its own sake nor for the sake of something else. For this reason, Anscombe claims that "voluntariness is relative to description of action."[36] However, such relativity to a description must be distinguished from the relativity to a description that is required for intentional action. Anscombe allows that children and animals can act voluntarily even if they lack the linguistic skills required to act intentionally.[37] Hence, *description* in cases of voluntary action must be understood so broadly that agents who lack certain skills are still capable of having such a description in their minds when setting out to act.[38]

Anscombe's account can distinguish between intentional action, voluntary action, and action as such. Action is intentional if a person performs it for a specific reason. It is voluntary rather than intentional if it is under the control of the agent and wanted but not

[35] Anscombe, 2000, 21-22 (§14).
[36] Anscombe and Morgenbesser, 1963, 399.
[37] See Anscombe, 1979, 221. Anscombe, 2000, 12-13 (§7).
[38] At this point, I leave aside the difficult exegetical question of how Anscombe distinguishes between types of ignorance cancelling voluntary action and those types of ignorance cancelling intentional action. Doing so would also require taking into account Anscombe's ethical work, where she claims that ignorance does not negate the voluntariness of an act if the state of being ignorant is itself voluntary, which is the case if the person could have known and should have known. See Anscombe, 2008.

performed for a specific reason, e.g. in cases of fidgeting or bringing about foreseen but unintended side-effects. Finally, action can be involuntary and unintentional: if someone mistakenly benefits his enemy, he acts neither intentionally nor voluntarily, but he acts nonetheless.[39]

Following this view, *Anscombian Models* had an ambivalent position on

(1) Intention without Voluntariness One can act intentionally without acting voluntarily.

Kenny clearly rejected (1) *Intention without Voluntariness,* claiming that "[i]ntentional actions appear to be a subclass of voluntary conscious actions."[40] And most of Anscombe's remarks support Kenny, e.g. most notably when she says that "[e]very intentional action is also voluntary."[41] In these passages, Anscombe and Kenny imply that if a person's movements satisfy the requirements for intentional action, they will thereby also satisfy the requirements for voluntary action.

However, it is noteworthy that Anscombe, unlike Kenny, also made room for the opposite view that one can act intentionally without acting voluntarily. Towards the end of *Intention,* Anscombe says:

"*From another point of view*, however, such things [i.e. certain acts] can be called involuntary, if one regrets them very much, but feels 'compelled' to persist in the intentional actions in spite of that."[42]

---

[39] It is worth highlighting that Anscombe's main competitor, Donald Davidson, is not relevant in this respect. Davidson thought that all action is intentional action: "a man is the agent of an act if what he does can be described under an aspect that makes it intentional" (See Davidson, 2001, 46). In addition, Davidson mentions voluntary action only very few times and seems to equate it with intentional action, most clearly in Davidson, 2001: 25, 28 together with note 10, 230, 247. Hence, even though Davidson also rejects theories postulating volitions, he does not distinguish between voluntary action, intentional action, and action as such.

[40] Kenny, 1975, 53.

[41] Anscombe, 2000, 89 (§49).

[42] Anscombe, 2000, 89 (§49). Emphasis added.

And in a lecture entitled *Sin*, Anscombe claimed that

"*'voluntary' is being used in a specially broad way* when we say that (…) If someone does something under threat from another man, say of pain or death or eviction, (…) it was forced on him. 'I did it quite voluntarily' implies that I acted without such compulsion."[43]

Thus, these passages suggest that Anscombe thinks voluntariness can be understood in two different ways. In one "specially broad way", a person *can* act intentionally without acting voluntarily, e.g. when he acts very reluctantly or under compulsion. And in another way, one *cannot* act intentionally without also acting voluntarily. As the latter sense finds greater support in Anscombe's work and is also supported by Kenny, I will assume that *Anscombian Models* adopt the view that one *cannot* act intentionally without acting voluntarily, i.e. they reject (1) *Intention without Voluntariness.*

On the other hand, *Anscombian Models* clearly accept

(2) Voluntariness without Intention One can act voluntarily without acting intentionally.

Relying on examples like fidgeting or frowning, Kenny and Anscombe support (2) *Voluntariness without Intention* and explain that in order for an act to be done intentionally, it has to satisfy requirements other than those of voluntariness. Furthermore, Anscombe also thinks that someone acts voluntarily rather than intentionally when he brings about an unintended but foreseen consequence.[44]

[43] Anscombe, 2008, 178. Emphasis added.
[44] Gordon (1996), like *Anscombian Models,* did not make ethical considerations key to her discussion of voluntariness. However, unlike *Anscombian Models,* she rejected (2) and claimed that an analysis of ordinary language shows that voluntary action implies intentional action.

In addition to these claims, Anscombe also led her inquiry of voluntariness beyond the category of agency and referred to passivity:

"Things may be voluntary which are not one's doing at all, but which happen to one's delight, so that one consents and does not protest or take steps against them: as when someone on the bank pushes a punt out into the river so that one is carried out, and one is pleased."[45]

In this example, it is not about "one's doing at all", i.e. activity or agency, but rather about being acted upon by others, i.e. passivity. Hence, Anscombe shows that voluntariness is not restricted to agency but also applies to passivity. Moreover, in her ethical writings Anscombe also allows for *states of affairs* to be voluntary, e.g. she suggests that being in ignorance can be voluntary.[46]

Thus, *Anscombian* and *Ethical Models* both accept that (i) voluntary action is *not* action caused by a distinctive mental event; (ii) one can act voluntarily without acting intentionally; (iii) ignorance can cancel voluntariness. *Anscombian* and *Ethical Models* diverge as only *Anscombian Models* (i*) hold that one cannot act intentionally without acting voluntarily; (ii*) apply voluntariness to passivity and states, not just to agency; (iii*) focus on voluntariness independent of any ethical context.

[45] Anscombe, 1957, 89.
[46] See Anscombe, 2008, 189.

1.1.2 The Debate on Consent

I claimed that questions about the voluntariness of consent have grown from 20th-century philosophy of action. I will now explain this claim in greater detail.

Like 20th-century philosophy of action, the debate on consent rejects theories that postulate volitions. Voluntary consent is never seen as consent caused by a volition. Like *Ethical Models,* the debate on consent considers voluntariness to be of general relevance to an ethical question, which – for the debate on consent – is the question as to when consent is valid and thereby makes an act permissible that would otherwise be impermissible.[47] Like *Ethical Models,* the debate on consent adopts a negative causal proposal: consent is considered voluntary unless it is caused by certain influences, e.g. coercion. Like *Ethical Models,* the debate on consent accepts

1. **Intention without Voluntariness** One can act intentionally without acting voluntarily.

People in the debate on consent unanimously accept (1) because they think that consent will be given intentionally, but not voluntarily, if obtained by means of coercion.

Finally, like *Ethical Models,* the debate on consent is divided over the question of whether the voluntariness of one person's action depends on the *morality* of another person's action. In fact, almost exactly the same positions are defended again. Like Feinberg, Nelson claims that "whether an external influence is morally legitimate is conceptually and morally distinct from whether the action taken in response to that influence is voluntary or involuntary"[48], and thereby denies that the voluntariness of one

[47] See for instance Faden and Beauchamp, 1986, 256- 26, 337-381. Nelson and Merz, 2002. Appelbaum, 2009. Nelson, 2011. Wertheimer, 2012. Beauchamp and Childress, 2013, 137-140.
[48] Nelson, 2011, 13.

person's behaviour depends on the morality of another person's behaviour. Moreover, like Feinberg, Nelson focuses on control and degrees of voluntariness when he says:

"There are degrees of control and noncontrol, and, in this respect, degrees of voluntariness as a function of the continuum from total control to total noncontrol. Each point on the continuum can be conceived as corresponding to a different degree of voluntariness, ranging from lowest to highest."[49]

Finally, like Feinberg, Nelson also claims that deciding on the voluntariness of an act in a specific context requires setting "threshold lines"[50] on an otherwise morally neutral scale of control.[51]

On the other side of the debate, Appelbaum adopts Nozick's approach by claiming that consent is involuntary if and only if it is given under the influence of another person and "the person exerting the influence does not have the *right* to act in this way."[52] Thus on Appelbaum's view, as on Nozick's view, the voluntariness of one person's behaviour depends on the morality of another person's behaviour. It is striking that these parallels between Feinberg and Nelson, and between Appelbaum and Nozick, have not been acknowledged in the debate on consent at all.

However, unlike *Ethical Models* and *Anscombian Models,* the debate on consent is divided over:

2. **Voluntariness without Intention** One can act voluntarily without acting intentionally.

[49] Nelson, 2011, 11.
[50] Nelson, 2011, 11.
[51] Unlike Feinberg, however, Nelson does not include a requirement of information in the concept of voluntariness.
[52] Appelbaum, 2009, 33.

Appelbaum and Wertheimer implicitly accept (2). Although they do not discuss (2), their accounts of voluntariness are consistent with the view that one can act voluntarily without acting intentionally.[53] However, on Appelbaum's and Wertheimer's view, intentional action is a requirement for an act to qualify as *consent* in the first place. Thus, any token of consent and *a fortiori* any instance of *voluntary* consent will be intentional anyway.

On the other hand, Nelson rejects (2) by claiming that the "two conditions of voluntariness (…) [are] noncontrol *and intentional action*."[54] Nelson defends the condition of noncontrol, i.e. the absence of controlling influences, to ensure that a person's consent is, negatively speaking, not interfered with; and he introduces the condition of intentional action to ensure that a person is, positively speaking, "*in control* [which] is likewise essential [to voluntariness]."[55] Thus, on Nelson's account, intentional action is a requirement for the *voluntariness* of consent, whereas on Appelbaum's and Wertheimer's view, intentional action is a requirement for an act to qualify as consent in the first place, independent of whether such consent will then also be voluntary.

Furthermore, unlike *Ethical Models* and *Anscombian Models,* the debate on consent is divided over whether ignorance cancels voluntariness. Whereas 20th-century philosophy of action widely accepted that some types of ignorance cancel voluntariness, most people in the debate on consent rejected this view and claimed that only force, but not ignorance, negates the voluntariness of consent.

To be clear, they accept that ignorance can vitiate consent, but assume that it does so by preventing requirements relating to *intention* or *information* from being met, as opposed to the requirement that consent be *voluntary*. This is certainly true of Appelbaum and Wertheimer as they clearly separated the requirements of information and intention

[53] Appelbaum and Wertheimer did not introduce intentional action as a requirement of voluntariness. See Appelbaum, 2009. Wertheimer, 2012.
[54] Nelson, 2011, 11. Emphasis added.
[55] Nelson, 2011, 10. Original emphasis

from the requirement of voluntariness. Nelson, however, seems to adopt a different view after introducing intentional action as a condition of *voluntary* consent. On Nelson's account, if ignorance means that a person no longer consents intentionally, i.e. that he no longer acts under the description of 'consenting',[56] his consent will therefore cease to be *voluntary*. Thus, for Nelson, certain types of ignorance may prevent the requirement of intention from being met and thereby *also* render consent involuntary.

These historical connections between the debate on consent and 20th-century philosophy of action reveal the broad ramifications of investigating the voluntariness of consent. Accordingly, this thesis is not only a contribution to practical ethics but also to a modern understanding of voluntariness more generally and part of the attempt to explain voluntariness independently of theories that postulate volitions.

Furthermore, these historical connections raise the following important questions for the debate on consent:

- Does voluntary consent still require the presence of some mental states, e.g. willingness, or at least the absence of some other mental states, e.g. reluctance?
- Is *acting intentionally* a requirement for the *voluntariness* of consent, or is it a requirement for consenting in the first place, independent of whether such consent is then also voluntary?
- Do force *and* ignorance negate the voluntariness of consent?
- How do influences like coercion *cause* and thereby vitiate consent?
- Does the voluntariness of a person's consent depend on the morality of another person's influence?

[56] As Nelson later emphasises that the "[t]he *action* must be (…) intentional" (Nelson, 2011, 14. Emphasis added), I will interpret him to claim that, in order for consent to be voluntary, the act *itself* must be intentional (under the description of consenting) rather than the act being the *product* of another intentional act.

Finally, these historical connections also provide important resources. The criticisms against Feinberg and Nozick also apply to Nelson and Appelbaum and can therefore be used to test and refine the positions in the debate on consent. For instance, Feinberg's view was criticised for arbitrarily basing a binary verdict (voluntary versus involuntary) on a scalar property (degrees of control),[57] for being unable to account for the ethical context of voluntariness without making moral distinctions between different influences on a person's decision-making,[58] for conflating voluntariness with rational choice,[59] for misconceiving the relation between voluntariness and autonomy,[60] and for his particular take on autonomy as a so-called exercise concept.[61]

On the other hand, Nozick has been criticised for making his view on voluntariness redundant and circular, as well as for falsely suggesting that an influence needs to be illegitimate in order to negate voluntariness.[62]

Moreover, since the debate on consent closely follows *Ethical Models* rather than *Anscombian Models,* the latter provide a contrasting account against which the debate on consent can be evaluated. Moreover, *Anscombian Models* also offer further insights, e.g. their take on how voluntariness applies not only to agency but also to passivity and states of affairs.

---

[57] See Shafer-Landau, 2005, 171-172, 190-191. Alexander, 2008.
[58] See Wertheimer, 2012.
[59] See Arneson, 1980, 484, 487.
[60] See Olsaretti, 2004, chapters 4 through 6.
[61] See O'Neill, 2002, chapters 2 and 4.
[62] The most important discussions can be found in Miller D., 1983, 72 footnote 10. Cohen G.A., 1995, 60. Olsaretti, 2004, 125. See also Smith, 1976, 89. Ryan, 1977. Wertheimer, 1987, 307-310. Waldron, 1991, 308. Kristjánsson, 1996, 72. Wertheimer, 2012, 249. Shnayderman, 2013, 729-730. Bader, 2016. Olsaretti, 2016. Bader, 2018. These criticisms often target Nozick's view on freedom but also apply to his account of voluntariness.

1.2 Defining the Voluntariness and Validity of Consent

Having explained the philosophical sources of the debate on consent, I will now define the main terms of my inquiry. More precisely, I will adopt the following working definitions: *Consent* is a communicative rather than a purely mental act. When consent is *valid*, it has its characteristic transformative effect, i.e. the effect, roughly, that an act that was impermissible becomes permissible. Finally, *voluntariness* is a necessary condition for the validity of consent and requires the absence of controlling influences.

1.2.1 Consent

When a person consents, he permits another person to perform a certain kind of act. Being an act of *giving permission*, consent is *transformative* and, in medical settings, it is *revocable.* People can withdraw their consent at any time.

In addition, being an act of giving permission *to another person*, consent is always directed towards a recipient. In most cases, the recipient is a particular individual, e.g. a particular sexual partner or physician. In other cases, it may be a group of people, e.g. the members of a clinical team. Who exactly and how many people receive consent will depend on the consenter's decision, but the permission granted will be restricted to that person or that group of people.[63]

Finally, giving permission for the *performance of a certain kind of act,* consent is always restricted to the kind of act specified by the consenter, e.g. a medical procedure, sex, or transfer of property.

[63] The law clearly acknowledges the recipient as an integral part of consent, not only in medical contexts but also in contract law and the criminal law concerning sexual relations. See *Grabowski v Quigley* [1996] and *Taylor v Albert Einstein Medical Center* [2000] (medical consent), *Boulton v Jones* [1857] (contract law), and *R v Elbekkay* [1995] (criminal law on sexual relations).

So far, this is common ground. Disagreement arises over what consent actually consists in. Three main views have emerged.

(1) According to a *subjectivist view*, consent is a mental act or state. The *expression* of consent is not part of consent itself and must not be confused with consent, just as the expression of belief is not belief itself and must not be confused with it. As a result, subjectivist views claim that the expression of consent plays a purely evidential role. It informs others that consent has been given and, therefore, the content of the expression is truth-apt. It can correctly or falsely report on the presence of consent. Most of the subjectivist accounts focus on a mental *act,* i.e. an intention, decision, or choice, rather than on a mental *state,* i.e. something that is merely felt or perceived in some way. However, opinions differ as to what the relevant mental act is. Consenting has been defined as "specifically *intend[ing]*"[64] that another person do something, "[mentally] *waiving* [one's] (…) *right* that the conduct not occur,"[65], "an internal choice (…) - a decision that 'this is okay with me'"[66], or (heading more towards a mental *state*) as a "felt willingness to agree with - or to choose - what another person seeks or proposes."[67]

Opposing the subjectivist view, other views define consent as a *communicative* act, i.e. something people do publicly rather than just mentally. Two versions of this view need to be distinguished.

(2) On the one hand, there is a *hybrid view.* In this view, "both a communication of consent *and* the existence of a certain mental state [respectively a mental act] are required in order for an agent to consent."[68] The hybrid view thereby adds something to the subjectivist view, namely the requirement of communication, while still accepting

[64] Hurd, 1996, 146. Emphasis added.
[65] Alexander, 2014, 107. Original emphasis. See also Alexander, 1996.
[66] Ferzan, 2015, 398.
[67] Westen, 2016, 4.
[68] Healey, 2015, 354. Original emphasis. See also Sherwin, 1996.

that mental acts are part of consent and that communicating consent plays at least a partly evidential role.

(3) On the other hand, there is the *performative view,* according to which consent is exclusively a communicative act and not constituted by any mental acts or states whatsoever. Hence, communication does not *report* consent but *constitutes* it. The required communicative act may take various forms and does not have to be verbal communication. What counts as consent will depend on the context and could consist of a mere gesture, e.g. nodding.

Opposing various characterisations, the performative view does not consider the intentions and beliefs of the person consenting altogether irrelevant.[69] But it has remained unclear how exactly proponents of the performative view account for them without incorporating at least some aspects of the subjectivist view.[70] It is here that I wish to go beyond reporting on existing accounts and specify the performative view by drawing two contrasts with the subjectivist and hybrid views.

Firstly, all subjectivist and hybrid views identify something like 'permitting' as the relevant mental act, whereas the performative view identifies 'intending to communicate permission.' In doing so, the performative view refrains from making any further demands on whether the person internally intends another person's conduct (Hurd), acquiesces to it (Westen), or mentally waives a particular right (Alexander). 'Intending to communicate permission' can be understood as a Gricean communicative intention.[71] Consent is given with the required communicative intention when a person

[69] Hurd, for instance, characterises such a view as holding that "a victim's mental state is *irrelevant* to assessing whether she has consented to a defendant's conduct" (Hurd, 1996, 136. Emphasis added). Similar statements can be found in Kazan, 1998, 37. Alexander, 2014, 103-104. Ferzan, 2015, 412.

[70] See for instance the rather tentative statements in: Archard, 1998, 4-5. Wertheimer, 2003, 150-151. Kleinig, 2010, 11.

[71] I here draw on Grice's account of what he called non-natural meaning. See Grice, 1957. For a useful discussion of Gricean communicative intentions, see Recanati, 1986.

(i) intends to make the recipient believe that he is permitted to perform a certain kind of act, (ii) intends the recipient to recognise her intention of giving him permission, and (iii) intends the recipient's recognition of her intention to provide him with reasons for believing that he is now permitted to perform a certain kind of act. Note that such a communicative intention does not imply that the person consenting wishes or desires the recipient to do the kind of act consent is given for. For instance, I may invite my colleague Daniel to my party (thereby consenting to him entering my house) even if I strongly prefer him not to be there. I may invite him only because I feel obligated to reciprocate an invitation or because I want to avoid an awkward situation at work.[72] Moreover, I may consent to my friend taking my car to go on holiday while strongly preferring that he does not do so and helps me with my gardening work instead.[73]

Secondly, the performative view takes a different stance on *why* a person's intentions are relevant. It holds that 'intending to communicate permission' is relevant because executing the communicative act of consent is something one can only do intentionally. In this respect, the communicative act of consent is similar to lying: one can only lie if one intends to do so.[74] But this does not mean that lying, and also the communicative act of consent, consist of *two* acts: a mental act and a communicative act. On the contrary, there is just *one* act in each case. The subjectivist and hybrid views diverge here and claim that there is a mental act *distinct* from the communicative act.

In the inquiry to follow, I will adopt a performative view and consider consent to be a communicative act performed with the Gricean intention to communicate permission. However, I will do so *not because* it may ultimately be the correct view. In fact, I do not need to decide which of the three views is ultimately the correct one because

[72] Alexander (2014: 108) mentions a similar example.
[73] Tadros (2016: 209) mentions a similar example.
[74] This is the majority position in the literature on deception and should therefore serve the purpose of an example here, even if it has not gone unchallenged that one can only lie intentionally.

an inquiry into voluntariness does not need to take a stance on the ontology of consent, i.e. what consent ultimately consists in. Questions about the ontology of consent are independent from questions about the voluntariness of consent. A person can be pressured into performing the act of consent, whether it is a mental or communicative act, and the different ontological views have to deal with questions about voluntariness in the same way. None of them implies any further assumption about voluntariness at this point and my proposal on the voluntariness of consent in Chapter 3 will be compatible with all three views.

Nevertheless, I will adopt the performative view because, firstly, settling on one particular account enhances clarity and, secondly, because there are the following two strategic reasons in favour of a performative view:

(1) A performative view, unlike a subjectivist view, does not commit this inquiry to a very narrow notion of permissibility. To illustrate, suppose a patient performs the mental act of consent but has not told the physician that he did so. The physician, being impatient, simply performs the procedure without talking to the patient. On the subjectivist view, the person consented and performing the procedure was therefore permissible. Thus, the subjectivist view commits an inquiry to a very narrow notion of permissibility that only tracks a person's mental acts but is insensitive to the serious moral failure of the impatient physician. By contrast, performative and hybrid views afford a wider notion of permissibility and track communication. On these views, since consent requires communication, the physician lacked the patient's consent and therefore acted impermissibly. It is this wider notion of permissibility that I am interested in.

(2) Only the performative view avoids the implausible position that one can trick others into wrongdoing. Subjectivist and hybrid views imply that people can *lie* about whether they consented, just as they can lie about what they believe. People lie about

consent when they secretly refrain from performing the mental act required for consent, hence do not consent, but still tell others that they consent, e.g. by deliberately signing a consent form and communicating permission. But if they do so, they can trick others into performing an impermissible act and thereby turn them into wrongdoers. I take this to be an implausible position. The deceived person is not a wrongdoer at all, not even a non-culpable wrongdoer. As I cannot see how subjectivist and hybrid views could respond here, the performative view of consent seems preferable. Once consent is a communicative act, one can no longer deceive others about one's consent and trick them into performing impermissible acts.

Thus, in order to avoid a narrow notion of permissibility, I need to reject the subjectivist view and in order to avoid the position that one can trick others into wrongdoing, I need to reject the subjectivist as well as the hybrid view. Therefore, I will rely on the performative view from now on.

1.2.2 Validity

At the beginning of this section 1.2, I claimed that when consent is *valid*, it has its characteristic transformative effect of making a previously impermissible act permissible. Now I wish to expand on this statement and claim that when consent is valid it "changes the normative landscape so that an act that would have been impermissible *for some reason* is no longer impermissible *for that reason*."[75] For example, prior to a competent person's valid consent, performing surgery on her would be impermissible *because* it would violate her right to bodily integrity. Subsequent to her consent, however, surgery is no longer impermissible *for that reason*. Her consent permits the surgeon to perform the operation.

[75] Walker, 2018, 131. Emphasis added.

The permission that consent creates is only a pro tanto permission as opposed to an *all-things-considered* permission. Consent ensures that an act no longer wrongs the person consenting in a particular way (i.e. by violating the right to bodily integrity). However, this does not imply that the act has thereby become permissible all things considered. A woman who consents to an abortion makes it the case that the performance of the abortion no longer violates her right to bodily integrity, but she cannot thereby settle the question concerning the permissibility of her abortion all things considered.

On this basis, my inquiry will be concerned with the *moral* as opposed to the *legal* validity of consent. *Morally* understood, consent is valid if it makes an act *morally* permissible which would otherwise be *morally* impermissible. In this way, consent regulates the extent to which people interacting with the consenter act immorally and, in the absence of an excuse or justification, become liable to moral criticism. Legally understood, consent is valid if it makes an act *legally* permissible which would otherwise be *legally* impermissible. In this way, consent regulates the extent to which people interacting with the consenter act illegally and, in the absence of justification, become liable to legal action (as available in tort or criminal law).[76] Legal and moral validity are different and can diverge from each other because the law, unlike morality, is "a blunt and expensive instrument"[77] and finds itself under constraints that morality is free from, e.g. the criminal law's constraint to focus on the most serious wrongdoings only. As my concern will be the moral validity of consent, I will take into account the legal literature only to the extent that the considerations therein also apply to moral validity.

In addition, I will consider validity to be a binary notion. Consent is either valid or invalid. This is because validity determines permission and permission is itself a binary concept: one can be permitted or not permitted to do something, but not half-permitted.

[76] See Herring, 2016, 156-158.
[77] Wertheimer, 2003, 5.

Hence, in order to maintain the link between validity and permission, the former must be considered a binary concept too. This view is generally accepted.[78] But it also poses a challenge. Many factors determining the validity of consent may be scalar (e.g. whether consent is valid may depend on *the extent to which* a person's decision to consent was controlled by others[79]). Therefore, any account of consent must explain how to arrive at a binary verdict if such a verdict is based on scalar properties.[80]

Finally, I will interpret *'valid'* as the *normative* property of consent, i.e. the property that determines when consent has its transformative effect. I contrast this normative property of consent with the *constitutive* properties of consent, i.e. the properties that determine whether there is a token of consent in the first place, independent of whether it is also transformative.[81] Hence, according to my terminology, the phrase 'valid consent' is not pleonastic, nor is the phrase 'invalid consent' contradictory.[82]

1.2.3 Voluntariness

Voluntariness partly determines the validity of consent: voluntariness is a necessary condition for valid consent and, together with competence and information, provides sufficient conditions.

---

[78] To the best of my knowledge, there is no published article rejecting the binary view of validity. However, Dougherty rejected a binary view of validity in favour of a scalar view in his presentation at the Moral Philosophy Seminar at the University of Oxford on 4th June 2018.

[79] See Nelson, 2011.

[80] After all, the binary distinction will have to make a categorical differentiation between two cases which are only minimally different with reference to what is considered the morally relevant property (e.g. the degree of control). This problem is not recognised in the debate on consent (an exception is Alexander, 2008) but is prominent in other areas of philosophy: see the so-called spectrum arguments. Parfit, 1987, Part III. Mulgan, 2004.

[81] The use of *validity* in Miller and Wertheimer, 2010, is different. Miller and Wertheimer argue that the transformative effect of consent does not require what is usually associated with the term 'validity.' However, this is only a difference in terminology and will not affect the arguments to follow.

[82] This use of terms may differ from the legal terminology in the area of sexual relations where consent is often seen as *valid* by definition. I separate constitutive from normative properties in order to focus specifically on the normative realm where voluntariness becomes important.

I will assume that the condition of voluntariness requires the absence of controlling influences on a person's decision-making and I will interpret 'influence' widely as anything that changes the eligibility or availability of options for a person, or in some other way affects a person's decision-making about whether to undergo a medical procedure. Such influences include threats, offers, forms of manipulation, certain ways in which medical information is presented or made available, oppressive circumstances, and psychological states that exert pressure on a person's decision-making. An influence is controlling and therefore negates the voluntariness of consent if and only if it compels or forces a person to consent in a way yet to be specified. This view of voluntariness requires some explanation.

Firstly, it takes the voluntariness of consent to be the default. Unless a person's decision-making is controlled or pressured in some way, his consent will be voluntary. Considering voluntariness as the default aligns with the negative causal view of *Ethical Models* in 20th-century philosophy of action.

Secondly, this view acknowledges that mental capacities are relevant to both competence and voluntariness. Mental capacities are relevant to competence because competence is *defined* by the possession of such capacities: the possession of certain capacities simply *is* what competence consists in. By contrast, mental capacities are relevant to voluntariness because questions about the voluntariness of a person's acts usually *presuppose* that this person possessed these capacities, i.e. that he had the ability of being guided by reasons. Accordingly, I will stipulate that a person consents *voluntarily* if, in addition to being competent, he satisfies the requirements of voluntariness (i.e. the absence of controlling influences) and consents *involuntarily* if, despite being competent, he fails to satisfy these requirements. On the other hand, a person consents *non-voluntarily* if, independent of whether he satisfies the requirements

of voluntariness, he lacks the basic mental capacities that define competence. But this conclusion is consistent with the view that competence and voluntariness are still distinct conditions of valid consent: competence is defined by the possession of mental capacities and voluntariness is defined by the absence of controlling influences.

Thirdly, I will reject Nelson's view, according to which intentional action is a condition of the *voluntariness* of consent. As one can only consent intentionally if one is to consent at all, the requirement of intentional action is already satisfied once there is a *token of consent*. Therefore, introducing intentional action again as a requirement of *voluntary* consent is superfluous. Moreover, Nelson's condition of intentional action also seems too narrow when applied to voluntariness more generally: if voluntariness requires intentional *action,* it is inconsistent with Anscombe's important insight that voluntariness not only applies to agency but also to passivity.

Finally, the voluntariness of consent, as I understand it, does not require the absence of ignorance. I thereby follow the majority view of voluntariness in the debate on consent but depart from the view which was prevalent in the 20th century. But to be clear: I accept that ignorance can vitiate consent, but I will assume that it does so by preventing the requirements relating to *intention* or *information* from being met, as opposed to the requirement that consent be *voluntary*.

Some types of ignorance prevent the requirement relating to intention from being met, i.e. they undermine the intention constitutive of consent. For instance, suppose a person signs a consent form believing he is merely acknowledging receipt of information. Such ignorance means there is no consent in the first place. To consent, a person needs to *intend* the act to be one of granting permission. Other types of ignorance prevent the requirement of information from being met. Suppose a patient is ignorant about important

risks of a procedure. Although he still gave his consent, his consent was not informed and potentially invalid for this reason.

There is no universal formula for deciding whether ignorance of some aspect cancels the very act of consent or prevents the requirement of information from being met. It will depend on the context of the situation, the importance of certain aspects for a person's self-determination, and potentially also the asymmetry between the consenting parties. Usually, performing an act without consent is a more serious wrong than performing an act subsequent to uninformed consent, and the law uses the distinction between different types of ignorance to account for these different wrongs.[83]

In this thesis, I will not attempt to analyse types of ignorance and rely on the assumption that the requirements of *intention* and *information* already provide comprehensive protection against ignorance. Therefore, there is no need for the requirement of voluntariness to include the absence of ignorance.

Hence, when investigating the voluntariness of consent, my central question will be which influences on a person's decision to consent are consistent with his consent being voluntary and which are not. In the remainder of this thesis, I will assume that the person giving consent is competent, fully informed, and acts with the required intentions. Hence, *if consent is invalid or vitiated* in my examples, it will be due to the negation of voluntariness, and *if consent is valid* in my examples, it must also be voluntary.

[83] See Herring, 2016, 156-158.

1.3 Looking Ahead

To conclude: This chapter aimed to render my thesis historically informed (1.1) and terminologically defined (1.2). The historical context showed that investigating the voluntariness of consent has broad ramifications, highlighted important questions for the further inquiry, and provided critical resources. The definitions of consent, validity, and voluntariness clarified important terminology and provided a clear focus amidst the multidisciplinary debates on consent. On this basis, I can now proceed to develop the desiderata which I argue any satisfactory account of voluntary consent must satisfy and which will then support my own proposal of voluntary consent in Chapter 3.

2. Desiderata of an Account of Voluntary Consent

In this Chapter 2, I will develop the following three clusters of desiderata, i.e. claims that I argue any satisfactory account of voluntary consent must adopt:

2.1 Voluntariness and Causation: A seriously illegitimate influence vitiates consent only if it *causes* consent, i.e. only if it prompts a person to consent. Counterfactual models fail to explain when an influence causes consent. Moreover, not every causal link between an influence and consent will suffice to vitiate consent. In order to cause and thereby vitiate consent, an influence must induce in a person certain reasons or motivations for consenting.

2.2 Voluntariness and Morality: Whether an influence vitiates consent does not solely depend on the strength of the influence, i.e. the extent to which it affects a person's decision-making. It also depends on the moral illegitimacy of the influence, i.e. whether exerting an influence violates specific norms. Such moral illegitimacy is best explained in terms of violating obligations owed to the person giving consent.

2.3 Voluntariness and Coercion: The widely accepted *Coercion Principle,* according to which coercion vitiates consent, must be disambiguated. Coercion vitiates consent only if consent was *obtained by* coercion, i.e. if the person receiving consent exerted the coercion or at least failed to discharge obligations owed to the person consenting; coercion does *not* vitiate consent if consent was *merely motivated by* coercion, i.e. when a third-party exerts the coercion and the person receiving consent did not violate any obligations owed to the person giving consent.

As far as possible, I will develop these claims with reference to consent in general rather than the medical context in particular. In Chapter 3, I will then restrict my focus to medical procedures and defend my proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* on the basis that, unlike other views, it aligns with all these desiderata.

2.1 Voluntariness and Causation

The debate on consent follows *Ethical Models* from the 20th century in adopting a negative causal view: Consent is voluntary unless it is caused by certain influences, e.g. coercion. However, the debate on consent has given very short shrift to the question of when and how an influence actually causes and thereby vitiates consent. In this section 2.1, I wish to fill this gap.

2.1.1 Introduction

Suppose Alfred threatens to hit Berta unless she invites him to her party. But Berta just laughs and falsely believes that Alfred is joking. Later, Berta invites Alfred only because she likes him to be there. Is Berta's consent to Alfred entering her house invalid? After all, Alfred issued a violent threat.

Surely, consent is invalid if elicited by a violent threat. But Berta's consent was not *elicited* by Alfred's threat because it did not affect Berta's decision at all. Thus, Berta's consent was valid. But if this is true, when exactly does a threat or another seriously illegitimate influence affect a person's decision-making in the way necessary to vitiate consent?

'Affecting' a person's decision-making is causal terminology and refers to a necessary yet not sufficient condition for an influence to vitiate consent, i.e. the condition that an influence vitiates consent only if it *causes* consent. This condition requires a causal link between an influence and the reasons for which a person consents. Hence, the causal *relata* are the influence and the reasons for consent, rather than the influence and the reasons *and* the act of consent. Whether the act of consent is caused by one's reasons, e.g. in an account similar to Davidson's, or whether the act of consent must be understood teleologically with regard to one's reasons, e.g. in an account similar to Anscombe's, is

beyond the scope of this condition of causation. Hence, we can separate the debate on causing consent from the debate on reasons and causes in the theory of action.

In what follows, I will focus on the condition of causation so understood and I will assume that the influences under consideration already satisfy other potential conditions for an influence to vitiate consent, i.e. that they are already clearly illegitimate and sufficiently severe. Hence, if an influence satisfies the condition of causation, it will also vitiate consent.

In the literature on consent, most people endorse a simple counterfactual view according to which an illegitimate influence causes and, other conditions being met, vitiates consent if and only if the person had not consented in the absence of the influence. Applied to my example, Alfred's threat did not vitiate Berta's consent because Berta would have consented even if Alfred had not threatened her.

In this chapter, I will claim that counterfactual views, simple or more sophisticated, are false. Instead, I will argue that a seriously illegitimate influence causes and thereby vitiates a person's consent if and only if it induces in a person a certain reason for consent, i.e. either the reason it purports to induce (e.g. to avoid the harm specified in the threat) or, in cases of deviant causal chains, it induces another reason that would also vitiate consent when intentionally induced. Accordingly, Berta's consent was valid *not* because she would have consented anyway, but rather because her reason for consenting was neither the one Alfred intended to induce (i.e. the fear of bodily harm) nor any other reason that would vitiate consent when intentionally induced (e.g. the fear of harm to another person).

To explain and defend my view, I will proceed as follows: after explaining Appelbaum's view as a paradigmatic example of the simple counterfactual view (2.1.2),

I will reject it alongside all other counterfactual models (2.1.3) and propose my own account (2.1.4). Finally, I will summarise my results (2.1.5).

2.1.2 The Simple Counterfactual View

Recall the question at hand: when exactly does a threat, or a seriously illegitimate influence more generally, affect a person's decision-making in the way necessary to vitiate consent? Appelbaum, paradigmatic for many others in the literature on consent, answers as follows:

> "to invalidate a decision, an influence must be *causally* linked to the subject's choice. If a person exerts illegitimate external pressure on a subject to consent to research participation when the person was inclined to participate *anyway* and intended to accept the offer, the voluntariness of the subject's decision is not negated. Applying such pressure may be reprehensible, but since it is not *causally* related to the decision, it does not render the decision involuntary."[1]

Appelbaum's answer has two components. Firstly, he invokes the concept of *causation* to explain when an influence vitiates consent. On Appelbaum's view, an influence vitiates consent only if it is *causally* linked to consent, and such a causal link – as explained above – must obtain between an influence and the reasons why a person consents. Secondly, he explains the required *causal link* in terms of *counterfactual dependence*. Appelbaum claims that there is no causal link if the person had consented "anyway"[2], i.e. in the absence of the influence. Formulated positively, there *is* a causal link if the person had *not* consented in the absence of the influence. Hence, on Appelbaum's view, a seriously

[1] Appelbaum, 2009, 33-34. Emphasis added. See also Nelson (2011), who follows Appelbaum's view in this respect.
[2] Appelbaum, 2009, 34.

illegitimate influence causes and thereby vitiates consent if and only if consent counterfactually depends on it.

It is striking how closely Appelbaum follows the view defended by David Lewis in the debate on causation in metaphysics. Lewis, aiming to explain the concept of causation as such and without any reference to consent, claimed:

"We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects - some of them, at least, and usually all - would have been absent as well."[3]

Hence, on Lewis's view, just as on Appelbaum's, an event C cannot be the cause of another event E if E would have occurred *anyway* without C.

The parallel between Appelbaum and Lewis is an important asset for the debate on consent. This is because the debate around Lewis's view contains a rich discussion of counterfactual dependence highlighting challenges for any counterfactual view on causation (including Appelbaum's) as well as various tools designed to address them. Yet, to the best of my knowledge, this parallel has not been exploited. This chapter aims to do so.

2.1.3 Rejecting Counterfactual Models

In the literature on consent, counterfactual views are widely adopted, usually without argument.[4] But despite their popularity, they cannot explain when a seriously illegitimate

[3] Lewis, 1973, 557.
[4] See for instance Nelson 2011, 13. An exception is Tadros, 2016, chapters 12-13. But even Tadros does not engage in detail with the arguments for and against counterfactual views on the general debate on causation. The situation is slightly different with regards to the question as to which mistakes or kinds of deception vitiate consent. See Archard, 1998, 46. Wertheimer, 2003, chapter 9. Dougherty, 2013.

influence causes and thereby vitiates consent. The following two cases, adapted from the discussion of Lewis's account, show why:

Preemption Berta considers enrolling in a clinical trial. Although she is currently undecided, she would consent if asked by her physician Wayne. Wayne does not know this and thinks Berta will never consent freely. Being afraid of insufficient recruitment, Wayne threatens Berta with removing all her medical care if she does not enrol. After Wayne's threat, Berta consents to enrolling in the trial merely out of fear of losing her medical care.[5]

Anonymous Blackmail Berta urgently needs medical treatment to overcome some disease but is unwilling to consent to it. Her physician Wayne really cares about Berta's health and, seeing no other option, he sends Berta an anonymous blackmailing letter threatening to report Berta's drug crimes to the police if she does not consent. Berta thinks the letter is a bad joke from one of her friends, but the letter still prompts her to search for the name of her disease again online. She finds a video of another ill woman who became infertile as a result of having the disease. Being frightened that this may happen to her too, Berta now consents, but only for reasons concerning her own health. [6]

[5] Cases like *Preemption* have been discussed in Lewis, 1973. Paul, 1998. Coady, 2004. Collins, 2004. Hall, 2004. Lewis, 2004. Paul, 2013. In such cases, it is always about two causes, c1 and c2, which could cause an event e on their own. But as it stands, c1 causes e and thereby prevents c2 from causing e.

[6] Anonymous Blackmail is similar to those cases in the metaphysical literature that question whether causation is transitive. See Lewis, 2004. Hall, 2004. Paul, 2004. Maslen, 2004. Hitchcock, 2007.

Preemption shows that simple counterfactual dependence is not *necessary* for an influence to vitiate consent. In *Preemption,* Berta would have consented anyway. Hence, according to the simple counterfactual view, Wayne's threat did not cause and vitiate Berta's consent. However, Wayne's threat *did* cause and vitiate Berta's consent because, as a matter of fact, Wayne forced Berta to consent and Berta consented only out of fear.

Anonymous Blackmail shows that simple counterfactual dependence is not *sufficient* for an influence to vitiate consent. In *Anonymous Blackmail,* Berta would not have consented had Wayne not threatened her. It was Wayne's threat that initiated Berta's second deliberation about the treatment, which eventually led to her decision to consent. Hence, the simple counterfactual view yields the result that Wayne's threat caused and vitiated Berta's consent. It may seem correct that Wayne's threat *caused* Berta's consent in some way (and I will return to this point later), at least if *Anonymous Blackmail* can be seen as an unproblematic case of transitive causation.[7] But it is incorrect that Wayne's threat therefore also *vitiated* Berta's consent. After all, Berta's decision to consent was not motivated by the desire to escape what Wayne threatened to do, but only by considerations of her own health.

Hence, *Preemption* and *Anonymous Blackmail* show that the simple counterfactual view fails. To deflect such criticism, Appelbaum could resort to two strategies, as explored in the metaphysical literature on causation.

[7] If the transitivity of causation is correct, then the following claim is true: if A causes B, and B causes C, then A causes C. Applied to the cases at hand: Wayne's blackmailing letter causes Berta's online search, which in turn causes Berta's consent. Transitivity has been challenged in various ways, but at least a somewhat restricted version of transitivity remains common ground in the literature on causation. See Lewis, 2004. Hall, 2004. Paul, 2004. Maslen, 2004. Hitchcock, 2007.

Strategy 1: Ancestral Dependence

Lewis initially tried to save his counterfactual view against cases like *Preemption* by re-defining causation as "the ancestral of dependence."[8] He claimed that "[e]vent C causes event E if and only if there is a *chain of dependencies* running from C to E."[9] Hence, Lewis not only talks about the cause C and the effect E but also about intermediary steps leading from C to E.

Thus, let us assume a standard case of preemption: There are two potential causes, c1 and c2, both of which would be sufficient to cause event e on their own. However, c1 preempts c2, and causes event e through an intermediate event d. As a result, c2 does not cause e.[10] Focusing on such a case, Lewis says:

"As far as causal dependence goes, there is no difference: e depends neither on c1 nor on c2. If either one had not occurred, the other would have sufficed to cause e. So the difference must be that, thanks to c1, there is no causal chain from c2 to e; whereas there is a causal chain of two or more steps from c1 to e. Assume for simplicity that two steps are enough. Then e depends causally on some intermediate event d, and d in turn depends on c1. Causal dependence is here intransitive: c1 causes e via d even though e would still have occurred without c1."[11]

Applied to my example *Preemption,* Wayne's threat is c1, Berta's counterfactual willingness to consent is c2, Berta's fear of Wayne carrying out his threat is the intermediate step d, and Berta's consent is e. At first, it seems plausible that Berta's consent (e) depends on Berta's fear (d), which in turn depends on Wayne's threat (c1)

[8] Lewis, 2004, 79.
[9] Lewis, 2004, 93. Emphasis added.
[10] To clarify: I am using capitalised and non-capitalised letters (e.g. c and C, e and E) interchangeably to align with Lewis's own usage in the respective quotations. There is no difference between the capitalised C and the non-capitalised c.
[11] Lewis, 1973, 567.

and not on Berta's willingness to consent (c2). Thus, there is a causal chain between c1 and e but not between c2 and e. However, there is an obvious objection to this line of reasoning and Lewis anticipates it:

"It remains only to deal with the objection that e does not depend causally on d, because if d had been absent then cl would have been absent and c2, no longer preempted, would have caused e. We may reply by denying the claim that if d had been absent then c1 would have been absent. That is the very same sort of spurious reverse dependence of cause on effect that we have just rejected in simpler cases. I rather claim that if d had been absent, c1 would somehow have failed to cause d. But c1 would still have been there to interfere with c2, so e would not have occurred."[12]

Lewis does not give an example of the instance of causation he has in mind. But what he says here is certainly not true in *Preemption.* It is not true that "if d had been absent, c1 would somehow have failed to cause d. But c1 would still have been there to interfere with c2, so e would not have occurred."[13] Admittedly, c1 (Wayne's threat) could have been there without causing d (Berta's fear), e.g. because Berta might not have been afraid of the threat or might have falsely thought Wayne was only joking. However, even if c1 (Wayne's threat) does no longer lead to d (Berta's fear), it is still an open question whether c1 (Wayne's threat) interferes with c2 (Berta's willingness to consent): subsequent to the threat, Berta may lose the confidence in Wayne and withhold consent, in which case c1 would have interfered with c2; or alternatively, Berta may consider this incident ridiculous or unimportant and consent to the medical procedures only for reasons related to her health. Hence, there is no guarantee that c1 (Wayne's threat) always interferes with c2 (Berta's willingness to consent). Therefore, e (Berta's consent) does not depend on d

[12] Lewis, 1973, 567.
[13] Lewis, 1973, 567.

(Berta's fear) and there seems to be no other candidate for an intermediate step d that could fare any better.

Hence, employing "the ancestral of dependence"[14] cannot deflect the criticism illustrated by *Preemption.* It still fails to highlight what is actually relevant in *Preemption,* namely that Berta consented out of fear.

Strategy 2: Co-Variation

So far, I have relied on Appelbaum's version of counterfactual dependence and focused on a *whether – upon – whether* dependence: *whether* E occurs depends upon *whether* C occurs. But this is certainly not the only option.

Why not think about different alterations of E? An alteration of E would be any event that shares with E its essential features but differs from it in its contingent ones, e.g. the time at which it occurs or the manner in which it occurs. Accordingly, we could then formulate a *whether/when/how – upon – whether* dependence: *whether* or *when* or *how* E occurs depends upon *whether* C occurs. And once we differentiate between alterations of E, we can do the same with C. Accordingly, we will arrive at a *whether/when/how – upon – whether/when/how* dependence: *whether* or *when* or *how* E occurs depends upon *whether* or *when* or *how* C occurs.

In fact, Lewis chose this strategy to deal with yet another set of pre-emption cases. He framed his account of counterfactual dependence in terms of *causal connections* and *influences,* and summarised his preferred counterfactual as follows:

"Where C and E are distinct actual events, let us say that C *influences* E iff there is a substantial range C_1; C_2; . . . of different not-too-distant alterations of C (including the actual alteration of

[14] Lewis, 2004, 79.

C) and there is a range $E_1;E_2$; . . . of alterations of E, at least some of which differ, such that if C_1 had occurred, E_1 would have occurred, and if C_2 had occurred, E_2 would have occurred, and so on. Thus we have a pattern of counterfactual dependence of whether, when, and how on whether, when, and how."[15]

On this basis, Lewis developed a test for causal connections. If cause and effect relate to each other as outlined, alterations of the cause lead to alterations in its effect. Therefore, "seeing what else 'wiggles' when you 'wiggle' one or another event tells you which ones are causally connected to which."[16] Thus, in order to find out whether X is a cause of Y, one alters X while holding everything else fixed and one looks at whether anything is happening in Y. This modification of Lewis's account renders what used to be a simple counterfactual *dependence* of C and E into a counterfactual *co-variation* of C and E.

This account of counterfactual co-variation can now handle my example *Preemption.* Admittedly, *whether* Berta consents does still not depend on *whether* Wayne threatens her. But *how (or why)* she consents, i.e. from fear and with particular motives, depends on *how* Wayne threatened her. Had Wayne not threatened her, Berta would not have consented from fear and with the intention of avoiding the loss of her medical care. Had Wayne threatened Berta differently, e.g. with violence instead of removing medical care, Berta's intention to consent would have changed accordingly.

Thus, changes in Wayne's threat, while holding everything else fixed, effect changes in the reasons for which Berta consents. By contrast, changing Berta's initial willingness to consent (e.g. assuming she is *not* willing to consent before being threatened, or assuming she is willing to consent on different grounds), while holding

[15] Lewis, 2004, 91. Original emphasis.
[16] Lewis, 2004, 91.

everything else fixed, does not change Berta's reasons for consenting: she would still consent because of Wayne's threat.

Hence, there is counterfactual co-variation between Wayne's threat and Berta's consent, but not between Berta's willingness and Berta's consent. Therefore, Wayne's threat causes and thereby vitiates Berta's consent in *Preemption.* The account of counterfactual co-variation arrives at the right result.

However, there are two problems. Firstly, counterfactual covariation still falsely suggests that Wayne vitiates Berta's consent in *Anonymous Blackmail.* After all, there is still a *whether-upon-whether* dependence: had Wayne not threatened Berta, she would not have consented; and the view on counterfactual co-variation still accepts that, once there is such a dependence between an influence and consent, the influence vitiates consent.

Secondly, there are cases where co-variation seems unable to capture a causal link in the first place. Consider the following example:

Controlled Letter Investigator Charlotte writes a letter to Harry threatening to report him to the police for his drug crimes if Harry does not participate in her clinical trial. Unbeknownst to Charlotte, her colleague George watches her and thinks that Charlotte is doing exactly the right thing. In fact, if Charlotte had not written the letter, slightly deviated from writing this particular letter, or written it at a different time, George would have intervened and written it himself. Harry receives Charlotte's letter without George's interference and consents to enrolment in order to avoid prosecution.[17]

[17] Schaffer presented an analogous critique of Lewis's account and also claimed that causation does not require co-variation. However, Schaffer's critique focused on physical causation and he concluded that causation required a physical link, i.e. what he calls *effluence*. See Schaffer, 2001.

In *Controlled Letter,* let us assume, there is no counterfactual co-variation between Charlotte's action and Harry's consent: no matter what Charlotte does, Harry will still receive the same letter at the same time. This is because George would intervene and send the letter himself if Charlotte writes it only slightly differently or at another time. However, there is great counterfactual co-variation between George's action and Harry's consent: if George destroys Charlotte's letter or writes another one, Harry will receive no letter at all or a very different one. Yet, in the actual case, it is Charlotte's action and not George's that causes and vitiates Harry's consent. Therefore, an influence can cause and vitiate consent even if there is no co-variation (e.g. Charlotte's letter) and it can fail to cause and vitiate consent even if there is great co-variation (e.g. George's behaviour).

Lewis was aware of such examples. He would call Charlotte's letter a *pre-empting cause* and George's behaviour a *pre-empted backup*. But he cannot offer a convincing reply to defend his view of counterfactual co-variation. Lewis could only require *pre-empted backups* to be removed from an example before the test of counterfactual co-variation is applied. However, not only would this be an *ad hoc* reply, it is sometimes also unclear what should count as a pre-empted backup, or even whether there is a pre-empted backup at all. Most importantly, however, cases like *Controlled Letter* highlight that causation requires something that counterfactual dependence will sometimes be unable to capture, i.e. an *actual link* between cause and effect, or between an influence and consent. Therefore, counterfactual dependence, i.e. the second component in Appelbaum's view on consent and influences, needs to be abandoned.[18]

In the next section, I will show that also Appelbaum's first component, i.e. a *causal link* between an influence and consent, requires substantial modification because not every causal link will ensure that a seriously illegitimate influence vitiates consent.

[18] See Lewis, 2004.

2.1.4 Searching for the Required Causal Link

Let me return to *Anonymous Blackmail:* Wayne threatens Berta and unintentionally initiates Berta's second deliberation, which then led to her consent. Focusing on *Anonymous Blackmail*, I claimed that counterfactual models falsely imply that Wayne vitiates Berta's consent. Yet, if we understand these models as explanations of *causation*, they correctly imply that Wayne's threat *causes* Berta's consent. After all, Wayne's threat caused Berta to research her illness, which in turn caused brought about her consent. Thus, if causation is transitive, Wayne's threat caused Berta's consent.[19]

However, if this is correct, *Anonymous Blackmail* shows that not just any causal link between a seriously illegitimate influence and consent will ensure that the influence vitiates consent. Further specification is needed.

One may argue that consent was not vitiated in *Anonymous Blackmail* because there was a *deviant* causal chain: Wayne intentionally exerts an influence on Berta to make her consent, succeeds in doing so, yet the process that led to her consent unfolded in a way that Wayne did not intend, anticipate, or control.[20] Hence, one might argue that a seriously illegitimate influence vitiates consent if and only if there is a *non-deviant causal link* to consent. However, the following case illustrates why such a view is also false:

---

[19] As outlined earlier in footnote 7 in this Chapter 2.1, the transitivity of causation is sometimes called into question. Yet, it is universally accepted that causation is at least sometimes transitive and *Anonymous Blackmail* seems to be a good example of a situation where it is.

[20] These cases of deviant causal chains are often called non-basic and are contrasted with basic deviant causal chains. They are called non-basic because a person still acts intentionally under some description. In cases of basic deviance, a person fails to act intentionally under any description, as illustrated by Davidson's famous case of the climber. See Davidson, 2001, 79: "A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally." For further discussion of different cases of deviant causal chains in the philosophy of action, see for instance Keil 2007. Stout, 2010. Mayr, 2011. It is worth emphasising that the interest in deviant causal chains in the philosophy of action and my interest here are very different. Whereas the philosophy of action exclusively focuses on the person initiating the deviant causal chain and whether he can be seen as acting intentionally, my sole concern is whether *the person affected by a deviant causal chain still consents voluntarily.*

Misread Story Investigator Wayne threatens Berta with violence unless she participates in his trial. Berta thinks this is a joke and, being quite amused, researches Wayne to find out more about him. Berta finds a website where someone different, but with the same name, is reported to have committed violent assaults. Berta mistakes this other person for Wayne, now believes Wayne's threat to be sincere, and consents to participate in order to avoid serious harm.

In *Misread Story*, there is a deviant causal chain. Wayne intends to make Berta consent, succeeds in doing so, yet the process leading to her consent unfolded in a way that Wayne did not intend, anticipate, or control. Nevertheless, Berta's consent is vitiated because the deviant chain simply delayed the effect that Wayne's threat would have had anyway and thereby still constrains Berta's autonomy in the same way as initially intended. After all, the threat Berta perceives is identical to the one Wayne actually issues. Hence, the absence of deviant causal chains is not required for a seriously illegitimate influence to vitiate consent.

Alternatively, one may argue that consent was vitiated in *Misread Story* because, unlike in *Anonymous Blackmail,* Berta still consented for the reason that Wayne intended to induce. But once the deviant causal chain changes the reason for consenting, an influence may no longer vitiate consent. Hence, one could argue that a seriously illegitimate influence vitiates consent if and only if there is a *reason-preserving causal link* to consent. However, this view is also false, as the following case illustrates:

Pain Wayne threatens to kill Berta unless she participates in his trial. Berta is shocked and researches Wayne. She finds a video of one of his lectures where he talks about how giving certain drugs to people in Berta's condition can put them in agonising pain. Berta does not believe that Wayne would kill her, but she now believes that he might inflict agonising pain on her if she declines to participate.

Wayne causes Berta to consent but Berta's reason for consenting is not the one Wayne intended to induce. Therefore, Berta's perceived threat and the actual threat differ. Nevertheless, Berta's consent is invalid because there is a real constraint on Berta's autonomy, Berta urgently needs protection, and the change of reasons does not mean that Wayne no longer commits an assault if he performs invasive procedures on Berta. Requirements of consent would indeed have a significant blind spot if they could not detect Berta's consent as clearly invalid.

It is immaterial whether Berta's belief in the different threat is true or reasonable. Admittedly, an influence normally vitiates consent due to the features it actually has rather than due to the features someone falsely and unreasonably attributes to it. This is, firstly, because the main purpose of the requirements of consent is to protect against real threats rather than purely imaginative ones. Secondly, it is because the recipient of consent would otherwise be treated unfairly: if false and unreasonable beliefs about influences could be sufficient to vitiate consent, a recipient may, without any fault of his own and without any chance of anticipating such beliefs, be turned into a wrongdoer. However, in *Pain* these reasons do not apply. Firstly, Berta needs protection given the actual violent threat she faces. This is not a situation where there is *just* an imaginative threat. Secondly, it does not seem unfair to consider Wayne a wrongdoer. After all, had

he been successful, he would have vitiated consent anyway. Thus, Berta's mistaken beliefs must not exculpate Wayne from the wrongdoing he would otherwise commit.

Therefore, Berta's consent is invalid, even though there was no *reason-preserving causal link*. That is to say, it is invalid even though Wayne did not mention the infliction of pain and did not even intend to inflict pain on Berta when issuing the threat.

Thus, it seems that the critical aspect is the *reason why* a person consents. In *Anonymous Blackmail,* the reason was innocuous and would not vitiate consent if intentionally induced by other means, whereas the reason for which Berta consented in *Misread Story* and *Pain* was different and would vitiate consent if intentionally induced via a non-deviant causal chain. Therefore, I claim that a seriously illegitimate influence vitiates consent if and only if it either induces the reason it purports to induce, or induces another reason that would also vitiate consent when intentionally induced via a non-deviant causal chain.

Two clarifications about this view are needed. Firstly, as in *Misread Story* and *Pain,* my proposal relies on the assumption that the initial threat is upheld. It is in such cases that a threat vitiates consent even if there is a reason-changing and deviant causal chain. If the initial threat is withdrawn, but still induces the reason it purported to induce or another similar reason, consent will be vitiated only if the assumption of a persisting threat is reasonable. This is because there is no longer any actual constraint on a person's autonomy in these cases, but only a perceived one. Hence, if Wayne threatens Berta, but shortly afterwards apologises and withdraws the threat, making it clear that Berta can make her decision freely, then Berta's unreasonable belief in a persisting threat may not be sufficient to vitiate her consent.

Secondly, although I mainly focused on threats, my account could also apply to other types of influences. It can apply to offers, where the reasons for consenting are no

longer to avoid harm but to attain some good. Moreover, it can also apply to cases where an influence no longer provides a person with *any reason* for consenting, but changes the motivation in some other way. For instance, in cases of nudging, a person's psychological mechanisms are targeted in a way that makes certain behaviour more likely, but do not give a person a *reason* to act in a particular way. In these cases, my account would hold that an influence causes and (other conditions being met) vitiates consent if and only if it either triggered the psychological mechanism it purported to trigger, or triggered another mechanism that would vitiate consent when triggered intentionally.

2.1.5 Conclusion

This Chapter 2.1 focused on the question of when a seriously illegitimate influence (e.g. a serious threat) affects a person's decision-making so as to vitiate her consent. I thereby focused on what is widely considered a necessary yet insufficient condition for an influence to vitiate consent, i.e. the condition that an influence vitiates consent only if it *causes* consent. I argued that no version of counterfactual dependence can answer this question. Instead, I claimed that a seriously illegitimate influence causes and thereby vitiates a person's consent if and only if it induces the reason for consent it purports to induce (e.g. to avoid loss of medical care) or, in cases of deviant causal chains, it induces another reason that would also vitiate consent when intentionally induced (e.g. to avoid infliction of pain).

My view is a novel contribution to the debate on consent not only because it shows why the widely adopted counterfactual models fail, but also because my alternative proposal highlights that not every causal link between a seriously illegitimate influence and consent suffices to vitiate consent. Although it is beyond the scope of this Chapter 2.1 to explain how my arguments also apply to Ryle's, Feinberg's, and Nozick's accounts

of voluntariness, my view will also be of relevance to these. They also seem to employ counterfactual reasoning and therefore fall prey to the same objections as Appelbaum's account.

Lastly, my results showed that an influence can vitiate consent via a deviant causal chain. Therefore, those views according to which influences vitiate consent only if they are intentionally exerted, e.g. Appelbaum's view, must be understood so as to allow for deviant causal chains to vitiate consent or otherwise be inaccurate.

2.2 Voluntariness and Morality

One of the central questions of 20th-century philosophy of action was whether the *voluntariness* of A's action can depend on the *morality* of B's influence on A. In this section, I will answer affirmatively for cases where A's action is consenting and make the following three claims:

(2.2.1) Whether an influence vitiates consent does not solely depend on the strength of the influence, i.e. the extent to which it affects a person's decision-making.

(2.2.2) Whether an influence vitiates consent also depends on the moral legitimacy of the influence.

(2.2.3) Moral legitimacy (in this context) is best explained in terms of the obligations owed to the person giving consent.

To remind the reader again, when an influence vitiates consent in my examples, it will always do so by negating the voluntariness of consent, unless stated otherwise.

2.2.1 Consent and the Strength of an Influence

According to the majority view in the literature on consent, the extent to which an influence affects an individual's decision-making determines whether it vitiates consent: strong influences vitiate consent, whereas weak influences do not. The moral quality of an influence on the other hand, i.e. whether exerting the influence violates moral norms, lacks independent significance for the voluntariness and validity of consent. Hence, the majority view thereby exclusively focuses on the actual psychological impact that an influence has on a person's decision-making.

The extent to which an influence, e.g. a threat, affects a person's decision-making often corresponds to the degree of the harm a person is threatened with, e.g. a threat of bodily violence may have a high impact, whereas a threat to have one's flowers trampled may have a low impact. However, the strength of an influence and the degree of the threatened harm can come apart, e.g. when a threat that would normally have a considerable impact happens to have only a low impact on a particular person. In such cases, the majority claims it is the strength of an influence, i.e. the actual psychological impact, and *not* the presence of an objectively significant threat that determines the voluntariness of consent.

In the following, I aim to show that the majority view is false and claim that the strength of an influence does not determine the voluntariness of consent. More precisely, I will argue that a significant influence, e.g. a threat, vitiates consent even if it happens to have only a low impact on a particular person's decision-making, I will proceed as follows: after (I) setting out the majority view in greater detail, I will (II) present a novel objection against it, based on an analysis of the right to withdraw consent, and thereby defend the claim just mentioned.

I. The Majority View

Nelson, endorsing the majority view, claims that

"[t]here are degrees of control and noncontrol, and in this respect, degrees of voluntariness as a function of the continuum total control to total noncontrol. Each point on the continuum can be conceived as corresponding to a different degree of voluntariness, ranging from lowest to highest."[21]

[21] Nelson, 2011, 11.

Accordingly, an influence negates the voluntariness of consent when it reaches a sufficiently high point on this continuum of control: informed consent is voluntary if there is "a *substantial* satisfaction of the condition of noncontrol [by others]"[22] and involuntary if there is a *substantial* impediment to its satisfaction.[23] Nelson understands control in terms of the strength of an influence (i.e. the actual psychological impact on a person's decision-making) and, therefore, it is the strength of an influence that determines whether voluntariness is negated. Weak influences cannot, or are very unlikely to, negate the voluntariness of consent. The moral legitimacy of an influence in contrast is irrelevant to voluntariness. As Nelson explains: "whether an external influence is morally legitimate is conceptually and morally distinct from whether the action taken in response to that influence is voluntary or involuntary."[24]

Nelson does not define the threshold above which a controlling influence vitiates consent. He simply assumes that such threshold can be found.[25] For the moment, I will ignore the difficulty faced by an account that bases a binary property (valid versus invalid consent) on a scalar property (degrees of control).[26] I am here only interested in Nelson's claim that the degree to which conduct by another influences a person's decision-making is the only relevant consideration when evaluating voluntariness.[27]

---

[22] Nelson, 2011, 11.
[23] Note that it is here about the control *other people* exert over the decision-making of the person consenting, and not about the control which the *person consenting* has. However, nothing depends on this perspective and one could equally explain Nelson's view in terms of the control that the person consenting has.
[24] Nelson, 2011, 13.
[25] See Nelson, 2011, 11.
[26] See Alexander, 2008.
[27] In the general debate on informed consent, this majority view, focusing on the strength of an influence, can also be found in Faden and Beauchamp (1986), Feinberg (1986), Beauchamp and Childress (2013). In the more specific debate on informed consent and nudging, the majority view can be found in Saghai (2013). And in the debate on informed consent and payments to research participants, it can be found in Grant (2004), Vanderpool (2005), McGregor (2005). Finally, in the debate on third party coercion, most notably Millum (2014) still claims that weak influences cannot vitiate consent, even though Millum concedes that the moral quality is also a relevant consideration.

II. An Objection to the Majority View

The majority view, i.e. Nelson's view, is inconsistent with a person's right to withdraw consent and ought therefore to be rejected. To support this claim, I first need to recall a distinction from Hohfeld's analysis of rights.

Hohfeld distinguished between different types of rights, two of which are especially relevant for understanding how consent works. On the one hand, Hohfeld explained what he called a *claim-right.* A claim-right to do X makes it permissible for a person to do X and imposes an obligation on everyone else not to interfere. The rights over one's property are usually claim-rights: people can permissibly use their property and everyone else has a correlative obligation not to interfere with such use. On the other hand, Hohfeld explained what he called a *privilege* or *mere permission.* A *privilege* or *mere permission* also makes it permissible for a person to do X but does *not* impose an obligation on everyone else not to interfere. For instance, when being on a train, I may be permitted to look at other people. However, these people are under no obligation to let me look at them. If they decide to leave or put up a screen, I will not have any claim against them. One cannot demand that one be given a privilege; nor can one demand that it be guaranteed over time.[28]

Hohfeld's distinction is crucial to understanding that consent creates *rights* for others.[29] Consent in business contracts usually creates claim-rights: consent to what has been agreed to in a contract can normally be enforced and, therefore, provides each party to the contract with a claim-right to the other party's performance of what has been consented to. In contrast, consent in the context of health care only creates permissions.[30] By consenting to a medical procedure, a person does not confer a claim-right upon his

[28] See Hohfeld, 1913. Feinberg, 1973.
[29] Hurd, 1996, 121. Simmons, 2010, 305-306.
[30] I use "permission" and "privilege" interchangeably.

physician. At any time before (and if possible even during) the procedure, the person remains entitled to change her mind and end a procedure. If she spontaneously decides not to undergo a certain procedure, the physician cannot insist upon it.

This restriction to mere permission is equivalent to the patient's right to withdraw consent. When patients withdraw their consent, they remove the permission for a medical procedure previously given to their physician. The question as to whether one chooses the phrase 'right to withdraw consent' or 'restriction to permission' depends solely on one's perspective. From the patient's perspective it is a right to withdraw consent, whereas from the physician's perspective it is a restriction to permission. To keep philosophical jargon to a minimum, I will only use the more common term of the right to withdraw consent henceforward. However, the 'right to withdraw consent' can be substituted with the 'restriction to permission' at any point.[31]

With this explanation on the creation of rights at hand, I will first argue that an influence that violates the right to withdraw consent also violates the requirement that consent be voluntary (i). I will then argue that the extent to which an influence affects a person's decision-making does not determine whether it violates the right to withdraw consent (ii). Finally, I will argue that this result can be generalized, i.e. that the extent to which an influence affects a person's decision-making does not determine whether it violates the requirement that consent be voluntary, whether the influence is on a person's decision to withdraw consent, or to grant it in the first place (iii).

[31] To clarify: this argument applies not only to patients and physicians in the context of treatment but also to trial participants and investigators in the context of clinical research.

(i) An influence that violates the right to withdraw consent also violates the requirement that consent be voluntary.

The principle of self-determination, according to which "[e]very human being of adult years and sound mind has a right to determine what shall be done with his own body"[32], is axiomatic in medical ethics. The right to withdraw consent follows from this principle. At any time after initial consent has been given, a person is entitled to determine what shall be done with his own body and stop a procedure. Hence, the protection afforded by the right to withdraw consent is implied by the principle of self-determination.[33]

Self-determination, as valued in the eponymous principle, requires that - if a person decides to undergo a medical procedure - this person must be *competent* and give his *voluntary* and *informed* consent. Each of these requirements (i.e. competence, voluntariness, and information) affords a specific type of protection. Voluntariness, in particular, protects against controlling influences and can be defined with reference to its scope: the requirement that consent be voluntary affords the type of protection which the requirements of information and competence alone cannot afford, but which is still necessary for the validity of consent.

The right to withdraw consent also protects against controlling influences and has the very scope that is attributed to the requirement of voluntariness. To see this, consider the following example. Suppose Peter enrols in a clinical trial. He is competent to make medical decisions and is well informed about the procedures involved. However, Peter lacks the right to withdraw consent. Researchers can enforce his initial valid consent and insist that he remain enrolled or otherwise be penalised (e.g. by losing his medical care). As Peter can no longer freely determine what shall be done with his body in such a case,

[32] *Schloendorff v New York Hospital* [1914], 126.
[33] Justice Cory claimed that a right to determine what shall be done with one's body "must *include* the right to stop a procedure." See *Ciarlariello v Schacter* [1993], 121. Emphasis added. See also Grubb, 2010, 449-452.

Peter's situation is clearly incompatible with the principle of self-determination and, therefore, his consent to continued participation (if maintained to avoid penalties) is no longer valid. However, Peter's consent is no longer valid precisely because he lacks the type of protection which the requirements of information and competence alone cannot afford, but which is still necessary for the validity of consent. And he lacks such protection because of his lack of a right to withdraw consent.

Hence, the protection afforded by the right to withdraw consent is not only implied by the principle of self-determination. It is, more specifically, implied by the requirement that consent be voluntary, because it is within the scope of the requirement of voluntariness. Therefore, claim (i) holds: an influence that exerts pressure on a person's decision-making and thereby violates the right to withdraw consent also violates the requirement that consent be voluntary.[34] As a result, it becomes possible to investigate the requirement of voluntariness by investigating the right to withdraw consent.

(ii) The extent to which an influence affects a person's decision-making does not determine whether it violates the right to withdraw consent.

The *WMA Medical Ethics Manual* claims that a person must be able to withdraw consent "without any sort of reprisal."[35] The *Universal Declaration on Bioethics and Human Rights* requires that a person must be able to withdraw "without any disadvantage."[36] And the *CIOMS Guide* holds that a person must be free to withdraw consent "without penalty or loss of benefits to which he or she would otherwise be entitled."[37]

---

[34] One can, of course, also violate the right to withdraw consent in ways that do not involve influencing an individual's decision-making. However, my discussion in this thesis is concerned with influences only.
[35] *WMA Medical Ethics Manual,* 2005, page 106.
[36] The *Universal Declaration on Bioethics and Human Rights,* 2005, article 6.
[37] The *CIOMS Guide,* 2002, guideline 9.

These official documents, and others like them, support claim (ii): the extent to which an influence affects a person's decision-making does not determine whether it violates their right to withdraw consent. What matters is the *kind* of influence: any significant "reprisal" or "disadvantage" or the loss of benefits a person "would otherwise be entitled to" violates an individual's right to withdraw consent to a medical procedure. Thus, although a threatened harm may have to reach a certain level of seriousness in order to qualify as 'reprisal' or 'disadvantage' in the relevant sense, it is not required that such reprisals or disadvantages *also* affect a person's decision-making to a great extent. Even if the impact remains low, they would violate a person's right to withdraw.

The support for claim (ii) can be further strengthened by looking at two paradigm examples of violating a person's right to withdraw consent.

Firstly, a person's right to withdraw consent is violated if she is put under an obligation to undergo treatment after having consented to it or pay significant damages if she withdraws. But the reason for this is not that the existence of such an obligation would exert a powerful influence on an individual's decision-making. Some people may not be bothered at all by being placed under such an obligation, or even welcome it as an opportunity to signal their commitment to physicians.[38] The reason is that having an unqualified right to withdraw consent is inconsistent with its exercise leading to one's liability to pay damages. If one has such a right, one must not be legally or morally disadvantaged for exercising it. Hence, the extent to which an obligation to undergo treatment affects a person's decision-making does not determine whether it violates their right to withdraw consent.

[38] For instance, consider the context of clinical research: some people may consider participation a moral obligation and acknowledge the effect that attrition can have on the scientific validity of a trial, as well as its economic costs. For this reason, they may welcome an obligation to undergo a procedure.

The law and the majority position in the philosophical literature indeed hold (and for the reasons just mentioned) that individuals are *not* obliged to continue to undergo a procedure they have consented to, and are not liable to pay damages if they choose to end it.[39] However, it is most noteworthy that even those who oppose the right to withdraw, or wish to qualify it, do not base their arguments on considerations concerning the extent to which such an obligation would affect a person's decision-making. One of the major arguments against the unqualified right to withdraw consent is based on considerations concerning the autonomy of the person assuming an obligation. It is argued that binding oneself (e.g. in promises or contracts) is itself a genuine exercise of autonomous agency. The unqualified right to withdraw without penalty is therefore said to be "patronizing to those who want to assume a responsible and active role in research from its beginning to its conclusion."[40] At least sometimes, people should be able to waive their right to withdraw consent if they want to, *even if* being under such an obligation to continue with some medical procedure may later be experienced as coercive.[41] Other arguments against the restriction to permission in the literature do not refer to the strength of influences either. Instead they refer to the consequences for the scientific validity of a trial,[42] the way in which drop-outs negatively affect the contribution from other participants of a study,[43] the economic costs,[44] or consequences for public health.[45] Hence, the extent to which a person's decision-making is affected lacks significance not only for the arguments *against* an obligation to undergo a medical procedure after having consented to it, but also for the arguments *in favour of it.*

[39] See McConnell, 2010. Holm 2011. Schaefer and Wertheimer, 2011. Hug, 2012.
[40] Edwards, 2005, 121.
[41] See Chwang, 2008, 372-373. Chwang uses "coercive" in the sense of being substantially controlling, but not necessarily also in the sense of vitiating consent.
[42] See Edwards, 2005.
[43] See Holm, 2011.
[44] See Chwang 2008.
[45] See Spillman and Sade, 2007.

Secondly, a person's right to withdraw consent is also violated in the following case. Suppose a physician threatens to withdraw medical care (and terminate their patient-physician relationship) if the patient does not continue to undergo certain procedures as part of a clinical trial, which the patient has previously consented to. If the physician issues such a threat, he violates a patient's right to withdraw consent. But the reason for this is not that such a threat will always exert a powerful influence on a patient's decision-making. It may sometimes only make a small contribution to their decision. The reason is that such a threat removes an option to which the patient is normally entitled and which is of paramount importance: receiving medical care and continuation of the patient-physician relationship without having to undergo a particular research procedure. Hence, the extent to which threats in such cases affect a person's decision-making does not determine whether they violate the right to withdraw consent.

Thus, official documents and these two examples support claim (ii): the extent to which an influence affects a person's decision-making does not determine whether it violates the right to withdraw consent.

(iii) Hence: the extent to which an influence affects a person's decision-making does not determine whether it violates the requirement that consent be voluntary.

In (i), I argued that the requirement that consent be voluntary implies the right to withdraw consent and, therefore, an influence that violates the right to withdraw consent also violates the requirement that consent be voluntary. In (ii), I argued that the extent to which an influence affects a person's decision-making does not determine whether it violates the right to withdraw consent. Hence, it follows that (iii) the extent of an influence cannot determine whether the influence violates the requirement that consent be given and

maintained voluntarily. However, there is reason to assume that if certain influences vitiate a person's consent to the continuation of a procedure (by violating the right to withdraw consent), such influences will also vitiate consent if exerted at the time before initial consent has been given. For instance: a physician's threat to remove medical care seems consent-vitiating irrespective of whether it is intended to obtain initial consent to a procedure or whether it aims to prevent the patient from withdrawing initial consent. In both scenarios, the threat vitiates consent because it deprives the patient of an option to which he is otherwise entitled: continuation of the patient-physician relationship without undergoing a particular procedure. Hence, the exclusive focus on the strength of an influence is misplaced. The extent to which an influence affects a person's decision-making does not determine whether it vitiates consent. An influence may vitiate consent even if the extent to which if affects a person's decision is rather low. Hence, my claim in (2.2.1) is correct: whether an influence vitiates consent does not solely depend on the strength of the influence. Therefore, in cases of threats for instance, we need to distinguish between the extent to which a threat happens to affect a particular person's decision-making on the one side and the significance of the threatened harm on the other. Nelson falsely suggested that the former determines the validity of consent.

2.2.2 Consent and the Morality of an Influence

In the previous section, I challenged Nelson's claim that weak influences cannot vitiate consent. In this section, I wish to continue my criticism of Nelson and focus more explicitly on his related claim that "whether an external influence is morally legitimate is conceptually and morally distinct from whether the action taken in response to that influence is voluntary or involuntary."[46] *Pace* Nelson, I will argue that the voluntariness of consent can depend on the moral legitimacy of an influence, but I will add that such moral legitimacy must be further specified.

I will proceed in three steps: (I) Presenting a *Dilemma of Equal Thresholds,* I will claim that the moral legitimacy of an influence is key to the voluntariness of consent; (II) arguing for a *Mirror Thesis,* I will claim that an influence vitiates consent only if it has three specific features; (III) developing an argument against blackmailing, I will explain why both threats *and* offers can vitiate consent.

I. The Dilemma of Equal Thresholds

Suppose Peter considers, but is undecided about, participation in a clinical trial. It would require some influence to push him in either direction. Let us assume that the trial does not pose great risks or inconvenience. The following two scenarios seem possible:

(1) Peter is offered moderate compensation and values this monetary offer as an expression of respect for people's participation. Although he does not really need the money, the offer tips the balance and makes him consent to participating in the trial. Such payments are current practice and seem morally legitimate.

(2) The investigator of the trial is also Peter's physician and he threatens Peter to abandon him as a patient if he does not participate. Although this threat does not exert a

[46] Nelson, 2011, 13.

strong influence over Peter, it tips the balance and makes him consent to participate in the trial. The investigator's threat is morally illegitimate.

Let us assume that in both cases the influences are controlling to the same degree. Hence, according to Nelson's account, either both influences render subsequent consent involuntary or neither of them does. The fact that one influence is morally legitimate and the other illegitimate is immaterial for Nelson.

This equal verdict over both scenarios leads to difficulties. On the one hand, Nelson is certainly justified in allowing for some influences on a person's decision-making. Not only would a no-influence-at-all requirement lead to voluntary decisions being the absolute exception (if not impossible), it would also be based on a caricature of the individual person who needs to be free of any influence to decide voluntarily. Hence, voluntariness must be compatible with some margin of influence. Scenario (1) seems to be a good example where an influence is compatible with voluntary consent.

On the other hand, however, there are good reasons against morally illegitimate influences like those in scenario (2). In the last section 2.2.1, I aimed to show that such threats violate important norms in the medical context, conflict with the protection afforded by the requirements of consent, and therefore vitiate consent.

This situation creates a *Dilemma of Equal Thresholds*. If Nelson allows for some margin of influence as compatible with voluntariness, this margin will also apply to illegitimate influences and allow for the violation of important norms concerning a patient's autonomy. If Nelson, on the other, hand aims to avoid such conflict, his account will rule out too many influences as being inconsistent with voluntary consent. Only by making moral distinctions between kinds of influence can one ensure sufficient protection of a person's autonomy *and* allow for a reasonable margin of influence.

This conclusion becomes even stronger once we go beyond influences exerted *by people* and also include *circumstantial* influences. Suppose a person is in circumstances where he has to choose between debilitating surgery and dying from terminal illness without surgery. Surely, this person retains his ability to give voluntary and valid consent, even if such circumstances have a considerable impact on the person's decision and even if the situation is such that a person would face great harm when he declines surgery. By contrast, if this person is faced with the alternatives of undergoing similar surgery or being shot, consent is clearly involuntary and invalid, even when the psychological pressure on the person's decision-making and the impending harm are identical to those in the previous case of circumstantial pressure. To account for the differences between the two cases, one needs to make moral distinctions between influences: only the other person's violent threat but not a person's circumstantial pressure is a *morally illegitimate* influence and therefore only the former but not the latter vitiates consent.

Hence, the two scenarios of Peter's situation as well as the example of circumstantial pressure show that the voluntariness of consent depends on the moral legitimacy of an influence.

II. The Mirror Thesis

There is an obvious challenge to the general claim that the moral legitimacy of an influence determines whether it vitiates consent, namely the well-established fact (and examples will soon confirm this) that not all morally illegitimate influences vitiate consent. Thus, I need to specify my claim further by outlining which morally illegitimate influences vitiate consent, which do not, and why.

In this sub-section, I will claim that influences vitiate consent only if they share three features with another wrong, i.e. the wrong committed when, for instance, a physician performs surgery on a competent patient without that patient's consent.

I will call this claim the *Mirror Thesis* because it states that the features of the wrong which is required to vitiate consent, e.g. blackmail, mirror certain features of the wrong of performing a certain act without consent, e.g. assault. I will present the *Mirror Thesis* as the conclusion of the following argument:

1. Performing certain acts without consent is morally wrong. For instance, when a physician operates on a competent patient without his consent, he commits the wrong of assault.[47] Such a wrong displays the following three features:
    a. It is a 'wronging' rather than 'wrongdoing' more generally;
    b. It violates a person's right to bodily integrity rather than other rights;
    c. It is non-trivial.
2. The validity conditions of consent (i.e. competence, information, and voluntariness) have a distinctive *function:* they are supposed to ensure that acting on the basis of a token of consent is no longer the wrong described in (1), e.g. they are supposed to ensure that when a physician operates on a competent patient *with his valid consent*, he no longer commits the wrong of assault.
3. The validity conditions can fulfil their function only if they adhere to the *Mirror Thesis,* i.e. the claim that influences vitiate consent only if it they share three features characteristic of wrongs like assault or battery, as described in (1).
4. Hence, in order for the validity conditions to fulfil their function, the *Mirror Thesis* must be adopted.

[47] I assume that this is not an emergency or any other kind of exceptional situation.

In what follows, I will explain and defend premises (1) and (3). I will take premise (2) for granted as it is universally accepted in the debate on consent.

Premise (1)

Premise (1) claims that performing a medical procedure on a competent person without her valid consent is a wrong and that such a wrong displays three features:

(a) It is *wronging* rather than *wrongdoing*. Wronging means violating a right of someone, which in this context is a right of the person consenting. It differs from wrongdoing as the latter is more general and does not require any link to the person consenting. To illustrate: suppose a physician performs an abortion on a competent woman. Whether or not the physician *wrongs* the woman depends on whether she gave valid consent to the abortion. Whether or not the physician commits a *wrongdoing,* or acts wrongly more generally, is a much wider question and also depends on further considerations, e.g. on whether the foetus could already live outside the uterus.

(b) Furthermore, performing a medical procedure without valid consent specifically violates a person's right to bodily integrity rather than other rights. Therefore, the presence of valid consent also only precludes the violation of that right, but cannot preclude the violation of other rights too. To illustrate, suppose someone consents to participating in a clinical trial in exchange for money. His valid consent makes it the case that performing certain procedures no longer violates his right to bodily integrity. But it cannot guarantee that performing those procedures will not violate his other rights. If, half-way through the trial, the investigator embezzles the money that was supposed to be transferred to trial participants and thereby deprives participants of their compensation, the continuation of the trial without informing the participants may violate the

participant's contractual rights, even though it will not turn the performance of medical procedures into bodily assault.

(c) Finally, performing a medical procedure without valid consent is a severe wrong and not just a trivial slip in moral behaviour. It severely undermines a person's self-determination.

Premise (3)

The last section explained three features characteristic of the wrong in performing an act without the required valid consent. This section argues that influences vitiate consent only if they share these three features. I will defend this view by discussing three examples, each of which supports one of the three characteristics as a necessary condition for an influence to vitiate consent.

I will take it to be intuitively plausible that consent is not vitiated in the examples to follow. However, to strengthen my case and avoid relying on intuitions only, I will use a further test to decide whether an influence vitiates consent, one which focuses on a sound balance of positive and negative autonomy.

Requirements of consent ought to protect a person's negative and positive autonomy. Protection of negative autonomy is the protection against external interference. It consists of those requirements that declare a person's consent invalid if it was obtained by means of certain influences, e.g. coercion. Hence, negative autonomy is about being *free from* something. On the other hand, protection of positive autonomy is the protection of a person's exercise of the normative power of consent, i.e. the power to create permissions by giving consent. It consists of those requirements that declare consent valid even if it was given under less than ideal conditions. Hence, positive autonomy is about being *free to do* something. As important as the protection of negative

autonomy is, it still always curtails positive autonomy. The more restrictive the requirements for valid consent, the fewer the situations that allow a person to give valid consent.

Using these considerations, one can devise the following test: Influences vitiate consent only if they are such that a person requires the protection of negative rather than positive autonomy. If there is no need for the protection of negative autonomy, influences cannot vitiate consent.

On this basis, consider the following example:

Promissory Obligations Clinical investigator Charlotte promised her employer not to offer financial compensation to potential trial participants. Nevertheless, Charlotte offers Peter financial compensation and Peter enrols for that reason.

By offering compensation Charlotte breaks her promise and, other things being equal, acts wrongly. However, if Peter enrols, Charlotte will not violate any of Peter's rights when subjecting him to medical procedures, no matter how morally or legally severe the breach of the promise is. (Let us assume for the sake of argument that the financial compensation is not problematic apart from being a breach of promise.) Hence, Peter must have consented validly. Charlotte's illegitimate influence on Peter's decision-making, i.e. the monetary offer, did not vitiate his consent. Why?

The reason is that it did not *wrong* Peter. The wrongness of the influence was exclusively due to a violation of a promissory obligation owed to a third party. But if Peter was not wronged at all, his negative autonomy does not require protection. Hence, as a first condition, I claim that a morally illegitimate influence vitiates consent only if it *wrongs* the person tokening consent as opposed to being wrong on other grounds.

Now, consider this second example.

Exploitation Clinical investigator Daniel offers Peter free lunches for his participation in a clinical trial and Peter subsequently consents to participation. But unbeknownst to him, Daniel offered other participants both free lunches and additional financial compensation. Daniel offers Peter less because he is racist and Peter is a person of colour.

Daniel treats Peter unjustly by giving him too little in comparison to what others receive and thereby wrongs Peter: he violates Peter's right to equal respect. Nevertheless, Peter's consent to participation in the trial still seems to be valid. Why?

The reason is that Daniel's influence on Peter is not wrong due to a conflict with Peter's right to bodily integrity. The right to bodily integrity entitles people to decide what happens to their body. In order to conflict with this right, an influence must pass two tests. Firstly, as the right to bodily integrity requires that a person decide free from pressure, an influence must be wrong *due to the pressure* it exerts on a person's decision-making or at least due to how it renders the option of consenting *more* attractive, as opposed to being wrong on other grounds. Secondly, since the right to bodily integrity does not include reference to third parties, one must be able to explain the wrongness of the influence without reference to others.

Both tests fail in *Exploitation.* Firstly, Daniel's offer is not wrong due the pressure it exerts on Peter. If Daniel did not offer Peter any money at all, i.e. if the influence on Peter's decision-making were taken away altogether, the wrong would not disappear. There would still be unequal compensation conflicting with norms of justice or at least an instance of deceiving Peter about what others receive. But if the influence were increased, i.e. if Peter were offered *more* money so that he received the same amount as

everyone else, the wrong would *not* persist. Secondly, one cannot explain what is wrong with Daniel's influence without referring to third parties. Peter is wronged precisely because he is treated differently to *others* without justification and, therefore, one cannot explain the wrongness of Daniel's influence without taking into account what others receive in comparison to Peter.

Therefore, Daniel's influence does not conflict with Peter's right to bodily integrity and does not vitiate consent for that reason. This conclusion is further supported by the fact that Peter does not require protection of his negative autonomy, i.e. keeping him free from monetary offers altogether and considering the offer in *Exploitation* as vitiating consent. Rather, he needs protection of his positive autonomy, i.e. protecting his opportunity to have access to equal compensation.

Thus, a second necessary condition for an illegitimate influence to vitiate consent is that it wrongs due to a conflict with the right to bodily integrity.

Finally, consider the following example:

Trivial Threat Clinical Investigator Susanne threatens to steal £1 from Peter's wallet unless he consents to participating in her clinical trial.

A threat to steal a person's money unless he consents to a medical procedure is morally illegitimate. Such a threat wrongs the person (first condition) and one can avoid such wronging simply by not issuing the threat in the first place (second condition). However, in *Trivial Threat* consent still seems valid.[48] Why?

[48] I assume that part of the person's reason for undergoing the procedures was to avoid the theft of £1. Admittedly, it is very unlikely that someone competent will consent to an unpleasant medical procedure in order to avoid losing £1. However, this is an empirical assumption that cannot replace an ethical assessment for the potential cases in which such a threat did make a contribution.

Peter's consent still seems valid because the harm threatened is not sufficiently severe to undermine Peter's self-determination. Any account of voluntariness of consent must have some kind of non-triviality condition. People make their decisions under the impact of many different influences, some of which may be morally illegitimate. But if *any* morally illegitimate influence were to vitiate consent, a person's power to give consent would be drastically reduced, which cannot be conducive to their autonomy and would unduly prioritise negative over positive autonomy.

This non-triviality condition is consistent with my previous claim that weak influences *can* sometimes vitiate consent. The non-triviality condition focuses on the degree of *harm a person is threatened with*, whereas the previously discussed majority view (see 2.2.1) focused on the actual *psychological impact* on a person's decision. As mentioned earlier, these can come apart. A significant threat (e.g. removal of medical care) may affect a particular person only to a low degree.

This non-triviality condition requires further specification, which I will aim to provide when developing my own view in Chapter 3. At this point, I only wish to state it in its general form: *a morally illegitimate influence vitiates consent only if it is non-trivial.*

The three examples (*Promissory Obligations, Exploitation, Trivial Threat)* show something surprising. An influence vitiates consent only if it shares three features with the wrong of performing an act without consent: wronging, conflict with the right to bodily integrity, and non-triviality.

Conclusion (4)

If my claims about premises (1) and (3) are correct, the *Mirror Thesis* must be adopted. The *Mirror Thesis* is a new contribution to the debate on consent and departs from existing views, which all fail on different grounds.

To begin with, Appelbaum claimed that an influence vitiates consent if the person exerting the influence *lacks the right* to do so. Thus, since the people exerting the influences in my examples (*Promissory Obligations, Exploitation, Trivial Threat)* all lacked the right to do so, Appelbaum's view yields the result that their influences vitiated consent. But I hope to have shown that this conclusion is false and that consent is still valid in *Promissory Obligations, Exploitation,* and *Trivial Threat*.[49] Therefore, also Nozick's theory of voluntariness, which also focuses on the influencer's rights, cannot be correct when applied to consent.[50]

On the other hand, Wertheimer suggested that an influence vitiates consent if it *violates a right of the person consenting*.[51] However, this account is still too broad. As *Exploitation* showed, the violation of the right to equal respect does not vitiate consent.

Finally, Hyams argued for a No-Gains principle, according to which any *illegitimate rights violation leading to a gain for the recipient*, e.g. increasing the number of trial participants, vitiates consent.[52] But again, as *Exploitation* shows, one may violate the consenter's rights *and* make a gain but still not vitiate consent.

Thus, the *Mirror Thesis* provides a more plausible account of when influences vitiate consent and is therefore superior to other views currently defended in the literature. However, it also requires further qualification, concerning both what counts as a relevant wronging and when an influence is non-trivial.

[49] See Appelbaum, 2009.
[50] See Nozick, 1974.
[51] See Wertheimer, 2012.
[52] See Hyams, 2011.

III. Voluntariness and Types of Influence

The *Mirror Thesis* holds that an influence vitiates consent only if it *wrongs* a person. Threats normally wrong people and therefore also vitiate consent. But how about *offers?* Can they ever really wrong someone? In this sub-chapter, I will claim that both threats and offers can in principle vitiate consent and, therefore, the *Mirror Thesis* must be understood so as to allow both threats *and* offers to fall under its definition of wronging.[53]

On the most widespread understanding of threats, a person is threatened if someone proposes to violate his rights unless he does X, but this is not the only way. A person can also be threatened in cases of blackmail. Blackmailing threats differ from ordinary threats as they no longer propose to violate a person's rights unless he does X. As I have explained elsewhere:

"The blackmailer proposes to do what is *per se* a legitimate act. What makes the blackmailing proposal a threat and problematic is, firstly, that the proposed action would still infringe the blackmailed person's interests and, secondly, that there is an undue connection between the proposed act and what is demanded from the blackmailed person."[54]

Consider the following example:

Blackmail Roger threatens Peter that he will report him to the police for his drug crimes unless he participates in his clinical trial.

Roger has the right to report Peter's crime, but he must not threaten to do so unless Peter participates in his trial. This is because Roger would thereby misuse his power to report

[53] The arguments in this sub-chapter have been published in my article Kiener, 2019.
[54] Kiener, 2019, 67. Original emphasis.

crimes and exert control over Peter's decision-making which he is not entitled to have, particularly over decisions about intrusive medical procedures. Since the harm that would await Peter when he declines participation is severe, e.g. imprisonment, and Roger's unwarranted demand is clearly illegitimate, Roger's threat vitiates Peter's consent. Roger's threat is inconsistent with the protection against undue control and abuse of asymmetries in power as afforded by the requirements of informed consent.

The law follows this conclusion.[55] It acknowledges that blackmailing vitiates consent on the grounds that a person's "power of considering whether he ought to do it or not [e.g. trial participation], whether it is prudent to do it or not, is altogether taken away from [him]"[56] and because "(a) the defendant [e.g. Roger] has no reasonable grounds for making the demand and (b) the threat would not be considered by reasonable and honest people to be a proper means of reinforcing the demand."[57]

But now, consider the following variation of *Blackmail*:

Last Minute Rescue The police started investigating Peter on alleged drug crimes. Roger, who had no prior contact with the police on this matter, offers to hide important evidence against Peter if Peter participates in his trial.

Elsewhere, I explained the difference between the two cases as follows:

"The difference between *Blackmail* and *Last Minute Rescue* is that, in the former, Roger proposes to make Peter worse off by his own acts whereas, in the latter, Roger simply proposes not to

[55] The references that follow come from UK Law. However, a very similar position can be found in continental law, e.g. German Civil Law (BGB). Moreover, it is worth mentioning at this point that the validity of consent in blackmailing cases can be challenged on two grounds: (i) illegality or moral illegitimacy (doing something illegal or morally illegitimate, e.g. hiding crime, cannot be part of a valid agreement), or (ii) undue pressure applied to the person consenting. In the following, I will only be concerned with the latter because it is only the latter that could affect the voluntariness of consent.
[56] *Williams v Baylay* [1866], 219.
[57] *Al Nehayan v Kent [2018],* 188.

prevent the occurrence of harm that is already underway. Therefore, Roger makes a genuine offer: he proposes to make an option available which Peter would not otherwise have and is not entitled to have (i.e. not being prosecuted). But Roger also refrains from proposing to *make* Peter worse off if he declines. Roger would simply do nothing in that case. And doing nothing is what Roger has a right to do (and probably even a duty)."[58]

Thus, *Blackmail* is a case of a threat, whereas *Last Minute Rescue* is a case of an offer. However, this difference does not imply that consent is invalid in *Blackmail* but valid in *Last Minute Rescue*. In fact, the arguments from my discussion of *Blackmail* also apply to *Last Minute Rescue.* In the latter case*,* Roger also exerts control over Peter's decision-making which he is not entitled to have. And since the harm that would await Peter when he declines participation is again very severe and Roger's unwarranted demand clearly illegitimate, Roger's offer vitiates Peter's consent too.

The legal rationale seems to support this conclusion because the "power of considering whether he ought to do it or not [e.g. trial participation], whether it is prudent to do it or not, is [still] altogether taken away from [Peter]"[59] and there are "no reasonable grounds for making the demand."[60]

Hence, both threats and offers can vitiate consent and, therefore, the *Mirror Thesis* must be understood so as to allow both threats *and* offers to fall under its definition of wronging.

[58] Kiener, 2019, 67-68.
[59] *Williams v Baylay* [1866], 219.
[60] *Al Nehayan v Kent [2018],* 188.

2.2.3 Consent and Obligations

So far, my discussion of the strength and moral illegitimacy of influences has aimed to develop an *extensionally* adequate view of consent-vitiating influences: I tried to separate those influences that vitiate consent from those that do not. Furthermore, I claimed that Appelbaum's and Nozick's focus on the rights of the person exerting the influence is inadequate in this respect.

In this section 2.2.3, I wish to show that the focus on rights is also *intensionally* inadequate and that a satisfactory account needs to focus on obligations instead. An account is intensionally inadequate if – independent of whether it correctly identifies those influences that vitiate consent – it gives the wrong explanation of why an influence vitiates consent.

To recall, Appelbaum claims that an influence is illegitimate if the physician does not have a *right* to exert it and legitimate if he does. Alternatively, however, one could have proposed that an influence is illegitimate if the physician violates an *obligation* in exerting it and legitimate otherwise. Of course, Appelbaum acknowledges that physicians have obligations towards their patients and what counts as acting within one's rights will have to take into account one's obligations. However, since these two proposals resort to different concepts, they are not *intensionally* equivalent, i.e. they give different explanations for why an influence is illegitimate or legitimate.

Three differences between rights and obligations show why intensional adequacy requires focusing on obligations rather than rights.

Rights display three features: (i) rights can normally be waived and, therefore, what counts as acting within one's rights is subject to one's discretion. For example, I can waive my right to free speech by signing a confidentiality agreement and, as a result, I no longer have the right to make a certain utterance. (ii) Rights usually grant the freedom to

choose whether to exert them without becoming liable to criticism either way. If I have a right to use my property, I cannot normally be criticised for either exerting or not exerting it.[61] (iii) Finally, rights usually purport to benefit their holder. *My* right to use my property primarily benefits *me,* not anyone else, by protecting me from external interference.

Obligations oppose each of these aspects: (i*) Obligations cannot be waived and are not subject to one's discretion. If I am under an obligation to help someone, I cannot unilaterally absolve myself from it. (ii*) Obligations do not grant the freedom to choose whether to discharge them without becoming liable to criticism. If I am under an obligation to help someone, I normally have to discharge this obligation to avoid criticism. (iii*) Finally, obligations do not normally purport to benefit or protect their holder but rather benefit or protect other people. My obligation to help someone does not primarily aim to benefit me, but rather others.

It is noteworthy that the distinction between legitimate and consent-vitiating illegitimate influences reflects the features of obligations rather than those of rights. (i*): Whether a physician's influence is legitimate or illegitimate clearly does not depend on her discretion. A physician cannot waive her *right* to respect a patient's autonomy (analogous to a right of free speech) and thereby make such respect illegitimate. (ii*): A physician does not have a right to respect a patient's autonomy in the sense that she could choose either to respect autonomy or not respect it and be spared criticism. She *must* respect the patient's autonomy or otherwise she acts illegitimately. (iii*): A physician is required to be trustworthy, not manipulate the patient, and behave in a beneficial way. However, these requirements underlying the distinction between legitimate and illegitimate influences do not aim to benefit the physician but rather protect the patient.

[61] This is not to say that there are no situations where people can be morally criticised for exercising or not exercising a right. For instance, a police officer has a right to employ force and can sometimes be criticised if he does not do so. However, this will be because he not only has a right to employ force but, under certain conditions, also an obligation to do so.

If the legitimacy or illegitimacy of an influence depends on whether physicians discharge their *obligations*, one can immediately understand why the distinction between legitimate and illegitimate influences displays the features (i*) through (iii*). If one follows Appelbaum, however, and assumes that the legitimacy or illegitimacy of an influence depends on whether physicians act within their *rights,* one is left with a puzzle: why does the distinction between legitimate and illegitimate not reflect central features of *rights* if the notion of rights is so central to this distinction? Appelbaum would have to argue that it is about special rights which cannot be waived, do not prevent criticism, and do not aim to benefit their holders. However, by doing so he would introduce what is essential to obligations.

Hence, physicians' obligations are better suited to explain consent-vitiating influences than physicians' rights: whether a physician violates or properly discharges *obligations,* e.g. the obligation to respect a patient's autonomy or promote her well-being (and not whether he acts within his *rights*) determines whether his acts are legitimate or illegitimate in the sense relevant to the validity of consent. Appelbaum's account of rights is therefore intensionally inadequate and the moral legitimacy pertinent to the voluntariness of consent is best explained in terms of the obligations owed to the person giving consent.

Conclusion of 2.2: Voluntariness and Morality

I claimed that whether an influence vitiates consent does not solely depend on the strength of the influence, i.e. the extent to which it affects a person's decision-making, but also on the moral legitimacy of an influence. Developing the *Mirror Thesis,* I claimed that influences vitiate consent only if they have three specific features and I argued that these features must be understood broadly enough to allow both threats *and* offers to fall under their definition. Finally, I argued that an intensionally adequate account requires reference to the obligations owed to the person giving consent, rather than to the rights of the person receiving consent.

2.3 Voluntariness and Coercion

So far, I have focused on two-party cases only, i.e. cases where the person receiving consent exerted the influence on the person giving consent. Now, I will widen the focus and look at three-party cases, i.e. cases where a third person C exerts an influence on person B in order to make her give consent to person A. I will argue that any satisfactory account of the voluntariness of consent must disambiguate what I call the *Coercion Principle* and make the behaviour of the recipient of consent central to its approach.

2.3.1 Introduction

Consent is supposed to protect autonomy, i.e. the autonomy of the person potentially giving consent. By deciding whether or not to give consent, a person controls how others can legitimately interact with her, and the criteria for valid consent (i.e. *voluntary* and *informed* consent by a *competent* person) must ensure that such control is an expression of this person's autonomy.

On the other side, consent is also supposed to regulate liability, i.e. the liability of the person potentially receiving consent. By acting on the basis of consent, a person can avert liability to moral criticism and legal redress, which he would be exposed to had he acted without such consent. Therefore, the criteria for valid consent must also ensure that such liability is appropriately spelled out.

This double focus on the autonomy of the person giving consent and the liability of the person receiving consent constitutes the *Janus face* of consent. Any satisfactory account of the validity of consent must do justice to both of these faces and show how consent can fulfil its double function.

Cases of coerced consent with only two parties – the person giving consent and the person receiving consent – do not usually make it difficult to account for both faces

of consent. If Alfred coerces Berta into giving consent to him, Alfred's influence not only impinges on Berta's autonomy but also justifies legal action and moral criticism against him. Hence, considering Alfred's influence as consent-vitiating at once accounts for an appropriate protection of the autonomy of the person giving consent as well as for an appropriate regulation of the liability of the person receiving consent.

However, things are different in cases involving three parties. Suppose Charles (a third party) coerces Berta into giving consent to Alfred (recipient of consent). It is now Charles who negatively affects Berta's autonomy, but it is Alfred who acts on the basis of Berta's coerced consent. Unlike in cases involving only two parties, coercer and recipient are no longer the same person. As a result, considering coercion from a third party as consent-vitiating may no longer do justice to both faces of consent. For although coercion by a third party still impinges on the autonomy of the person giving consent, it may not automatically justify moral criticism or legal action against the recipient of consent, e.g. when the latter reasonably believed that the former gave her consent freely. So, how are we to deal with such three-party cases in order to protect autonomy *and* regulate liability appropriately?

In this chapter, I will claim that, in cases of third-party pressure, the validity of consent depends on the behaviour of the recipient when obtaining consent. In contrast, a third party's influence on a person's decision-making, e.g. the third party's coercion, does not determine the validity of consent.

I will explain and defend this view through the discussion of the two main approaches to third-party coercion in the philosophical literature. I will firstly reject what I call the *Influence-Focus-View,* which begins with the person giving consent and focuses on the influence exerted on her (2.3.2). I will then defend a version of what I call the *Recipient-Focus-View,* which begins with the person receiving consent and focuses on

his conduct towards the person giving consent (2.3.3). Finally, I will draw conclusions (2.3.4).

Throughout my arguments, I will assume that someone is a 'third party' if and only if he is not the recipient of consent *and* does not collaborate with the recipient. Moreover, I will only be concerned with cases where a consent transaction cannot be reversed, i.e. cases unlike those in the area of equity and changes of title in property. In particular, I will focus on consent to interference with a person's bodily integrity, e.g. in cases of medical procedures, sex, or getting a tattoo.

2.3.2 The Influence-Focus-View

I. Outline of the Influence-Focus-View

According to what I call the *Influence-Focus-View,* there is no relevant difference between two- and three-party cases when identifying those influences that vitiate consent. The validity of consent always depends solely on the features of the exerted influence. Millum, defending a version of this view, argued that "[i]llegitimate controlling third party threats render consent invalid just as they do in two party cases."[62] Millum claims that "valid consent requires that the person who has the authority to make a decision make that decision herself."[63] If someone is under an illegitimate controlling influence, she is "not making the decision (…) herself" and hence her consent is invalid. This holds true irrespective of who exerted the influence.

People defending the *Influence-Focus-View* still diverge on what they think makes an influence consent-vitiating. Whereas some hold that an influence vitiates consent if it is substantially controlling,[64] others argue that an influence vitiates consent if the

[62] Millum, 2014, 114.
[63] Millum, 2014, 118.
[64] See Feinberg, 1986. Nelson, 2011.

influencer lacks the right to exert it.[65] Yet others hold that an influence needs to be both substantially controlling *and* illegitimate.[66] However, all of them think that whether consent is valid depends only on the particular features of an influence, and not on who exerted it; and all of them also think that coercion is a paradigm case for an influence that vitiates consent. Hence, they all accept the following seemingly very plausible principle:

Coercion Principle If a person is coerced into consenting, consent is involuntary and invalid.

The *Influence-Focus-View*, including the *Coercion Principle,* is indeed widely held. Millum[67], Appelbaum[68], Nelson[69] and Feinberg[70] all endorse it. Moreover, versions of *Influence-Focus-View* can be found in most European jurisdictions in civil and criminal law.[71]

The motivation behind the *Influence-Focus-View* is a special concern for the interests and autonomy of the person giving consent. Proponents argue that those giving or withholding consent are best protected if *any* coercive influence, irrespective of its source, is considered to vitiate consent. Whether this is correct remains to be seen. However, the main challenge for the *Influence-Focus-View* lies in how to account for the other face of consent, that is, in how to account for the liability of the person receiving consent. In what follows, I will argue that the *Influence-Focus-View* does not have a convincing answer to this challenge. This is because it fails to provide a convincing solution to a dilemma that arises from the combination of its own claims with another fundamental principle in the debate on consent.

---

[65] See Appelbaum, 2009.
[66] See Millum, 2014.
[67] Millum, 2014, 114.
[68] See Appelbaum, 2009, 33.
[69] See Nelson, 2011, 8.
[70] See Feinberg, 1986, 195.
[71] An exception is the civil law in the Netherlands. See NBW, Book 3, Article 44 (5).

II. A Dilemma of Coercion and Permissibility

People can normally avert blame for what they did in two ways: by means of an *excuse* or by means of a *justification*. An excuse concedes that an act was impermissible but denies that the agent should be blamed for it, either because he was non-culpably ignorant of the nature of his act or was forced to act in the way he did. By contrast, a justification denies that the act was impermissible in the first place and thereby averts blame.

On the *Influence-Focus-View*, someone who acted on the basis of consent elicited by third-party coercion can avert blame only by means of an excuse, but no longer by means of a justification, for the following reason.

According to the *Influence-Focus-View* and the adopted *Coercion Principle,* consent is invalid if a person was coerced into consenting; and according to what we may call the *Permissibility Principle,* there are certain acts (e.g. interference with a person's body) that can be performed permissibly only if the relevant person gave valid consent. If consent is invalid, performing such an act will be impermissible. The *Permissibility Principle* is another highly plausible and widely accepted principle in the debate on consent.

Combining the *Influence-Focus-View* and the *Permissibility Principle* leads to the conclusion that the recipient, who acts on the basis of consent caused by third-party coercion, acts *impermissibly* and, therefore, can avert blame, if at all, only by means of an excuse. A justification for the recipient's conduct is no longer available, since acting on the basis of invalid consent is impermissible.[72]

[72] Note that my focus is on the justification 'internal' to the consent transaction: an act is *internally* justified if and only if (i) a person validly consented to it and (ii) that person's valid consent is necessary to avoid wronging him through the performance of the act. By contrast, I will not be concerned with justification 'external' to the consent transaction: an act is *externally* justified when, despite the absence of the normally required consent of a person, there are other reasons that justify performing it. For instance, someone may face the choice between the wrong of acting without valid consent (i.e. committing an 'internally' unjustified act) and another wrong, e.g. letting ten other people die. It might be that one is 'externally' justified to act without valid consent in such situations as one thereby chooses the lesser of two evils.

The most common excuse in cases of third-party coercion is ignorance. Consider the following passage from Feinberg:

"If the coercion (…) stemmed from the earlier threats of some third party, C, which were unknown to A, then (…) A cannot be fairly punished for what he believed in all innocent good faith he had B's voluntary consent to do."[73]

Feinberg, representing the *Influence-Focus-View,* is correct to point out the relevance of excuses for the liability of the recipient of consent. But how about cases where the recipient was *not* ignorant? Is the *Influence-Focus-View* correct in implying that someone who freely accepts consent while knowing that such consent was elicited by third-party coercion is always blameworthy for doing so? This does not seem to be correct. Consider the following case, taken from a paper by Millum:

Tattoo "Coercer threatens to beat up Victim unless she gets a tattoo from Recipient's tattoo parlor. Recipient has no particular interest in tattooing her, but he cannot otherwise prevent the threat from being carried out and, although she would otherwise not want a tattoo, the beating is clearly worse for Victim than the tattoo. When she explains the situation to Recipient, Victim confirms that she prefers that he tattoo her than take any other action."[74]

In *Tattoo*, Recipient is not ignorant about third-party coercion, nor is Recipient forced to act in the way he did. Thus, there is no excuse available. But Recipient's action is not morally deficient either. Recipient tries to help Victim as much as he can and respects

[73] Feinberg, 1986, 195-196.
[74] Millum, 2014, 118.

Victim by acting as Victim requests. Therefore, it seems that Recipient ought not to be blamed if he executes the tattoo, even if Victim's consent is coerced and there are no excuses for recipient. Otherwise one would have to claim that Recipient should have sent Victim back to the mercy of Coercer. However, if Recipient ought not to be blamed for executing the tattoo and if there is no excuse available, Recipient must have had a justification for acting as he did. And indeed, this seems correct. Recipient acted not only blamelessly but permissibly in executing the tattoo. (I will say more about permissibility in a moment.)

Once we accept that Recipient acts permissibly, a dilemma arises. The *Influence-Focus-View* leads us to the following reasoning:

1. Influence-Focus-View:	The Coercion Principle is correct: if a person is coerced into consenting, consent is invalid. Who exerts the coercion is irrelevant.
2. Facts in Tattoo:	Victim was coerced into consenting.
3. Conclusion:	Victim's consent is invalid.

However, applying the *Permissibility Principle* yields the opposite conclusion:

4. Permissibility Principle:	There are certain acts that can be performed permissibly only if the relevant person gave valid consent. These acts include executing a tattoo.
5. Verdict on Tattoo:	Recipient acted permissibly in executing the tattoo.
6. Conclusion:	Victim's consent is valid.

Conclusions (3) and (6) are inconsistent. Hence, after accepting the facts and verdict in *Tattoo,* we face a dilemma: we need to reject either the *Permissibility Principle* or the *Influence-Focus-View,* at the core of which is the *Coercion Principle*. But the *Permissibility Principle* and the *Coercion Principle* are both highly plausible. So, how can we solve the dilemma?

In the next section, I will show that rejecting the *Permissibility Principle* in its present form is not a defensible option and prepare the basis for modifying the *Coercion Principle* instead. In my discussion, I will use the capitalised term 'Recipient' to refer specifically to the character in *Tattoo,* including the variations of *Tattoo* that are still to come. However, everything I say about Recipient can also be generalised to any person receiving consent. The non-capitalised 'recipient' refers generally to any person receiving consent. Similarly, everything I say about 'Victim' or 'Coercer' can be generalised to any person giving consent and any third party exerting a coercive influence.

III. The Rejection of the Permissibility Principle and Why such Rejection Fails

Millum does not explicitly address the dilemma I set out, but he rejects the *Permissibility Principle* and thereby provides a potential solution. In this section, I will outline his rejection of the *Permissibility Principle* and show why it fails.

(i) Millum's Rejection of the Permissibility Principle

Millum accepts the facts and verdict in *Tattoo:* Victim is coerced into consenting and Recipient acts permissibly. Millum also embraces the *Influence-Focus-View* and the *Coercion Principle*. To recall, he claims that "valid consent requires that the person who

has the authority to make a decision make that decision herself."[75] If someone is under an illegitimate controlling influence, she is "not making the decision (…) herself" and hence her consent is invalid. This holds true, irrespective of who exerted the influence: "[i]llegitimate controlling third party threats render consent invalid just as they do in two party cases."[76] Therefore, if a person is coerced into consenting, consent is invalid.

However, Millum then rejects the Permissibility Principle and thereby offers a potential way to solve the dilemma. But before I discuss Millum's rejection of the *Permissibility Principle*, I need to specify this principle in greater detail.

Strictly speaking, valid consent is neither necessary nor sufficient to render an interference with a person's body permissible, and the *Permissibility Principle* does not claim otherwise. It is not necessary because one can permissibly interfere with a person's body without his consent in cases of medical emergency: if a person temporarily lacks the capacity to consent but still requires invasive medical treatment to survive, one can permissibly subject the person to such treatment. Consent is not sufficient because the permissibility of an interference with a person's body requires more than his consent. A person's consent to having her arms and legs cut off would not make such mutilation permissible. What a person can consent to in the first place depends on certain moral or legal background norms.

However, within certain boundaries, the *Permissibility Principle* claims that valid consent is essential. It is essential if a person has the competence to consent, if what the person consents to can be justified as a legitimate interaction with this person, and if permissibly performing the act under consideration requires a person's consent. In *Tattoo,* these conditions hold: Victim is competent to decide whether to get a tattoo, getting a tattoo is in principle a legitimate interaction, and tattooing requires a person's consent.

[75] Millum, 2014, 118.
[76] Millum, 2014, 114.

Therefore, in cases like *Tattoo,* valid consent is necessary and sufficient to render the interference with a person's body permissible. Note that permissibility in the context of consent always means pro tanto permissibility: consent makes it the case that executing the tattoo no longer wrongs the person in a way it otherwise would, and thereby makes executing the tattoo pro tanto permissible.[77]

Hence, within the limitations of cases like *Tattoo,* the *Permissibility Principle* could be described as a bi-conditional: consent is valid if and only if Recipient acts permissibly (i.e. pro tanto permissibly) by doing what consent is tokened for. This bi-conditional can be dissected into two simple conditionals:

(a) Consent is valid if Recipient acts permissibly by doing what consent is tokened for.
(b) Recipient acts permissibly by doing what consent is tokened for if consent is valid.

Millum still accepts (b), viz. he still accepts that in the cases of valid consent, consent renders the Recipient's acts permissible. However, Millum rejects (a) because, in his view, there are cases where Recipient acts permissibly but consent is invalid. Millum claims that "illegitimate third-party coercion does not entail that it is impermissible to proceed with the act consented to even if it means that the consent is invalid."[78] Whether or not Recipient may proceed on the basis of invalid consent depends, according to Millum, on "the other options available to Recipient, and what Victim wants Recipient to do."[79]

[77] To illustrate this point: as noted earlier, a woman who consents to an abortion makes it the case that carrying out the abortion no longer violates her right to bodily integrity. Hence, she makes it the case that carrying out the abortion no longer wrongs her in some particular way. However, her consent cannot thereby settle the question concerning the permissibility of her abortion all things considered. The abortion may still be wrong all things considered if the foetus could survive outside the woman's uterus.
[78] Millum, 2014, 118-119.
[79] Millum, 2014, 118-119.

Developing these statements, Millum argues that Recipient "may proceed if doing so would constitute a reasonable joint decision of her and the victim of coercion."[80] A "[j]oint decision-making involves each party taking the viewpoint of the other as seriously as her own and deciding on this basis what to do."[81] Following this account, "Recipient thereby respects her autonomy [i.e. the Victim's autonomy] by giving her as much control over her consent decision as he is able."[82] In such cases, "a wrong has been committed by Coercer. But Recipient is not culpable for this wrong: he has done the best he could in non–ideal circumstances."[83] On the basis of this account, Millum claims that in *Tattoo* consent is invalid but it is nevertheless permissible for Recipient to execute the tattoo. Recipient does not commit a wrong because acting as he does is based on a reasonable joint decision and mutual respect.

Hence, contrary to the orthodox view, Millum claims that Recipient acts permissibly not only if he is given valid consent (option 1) but also if he is given invalid consent but his act is based on a reasonable joint decision (option 2). This departure from the orthodox view shows Millum's rejection of the *Permissibility Principle.*

However, it is noteworthy that, in Millum's view, Recipient acts permissibly only if Victim issues a token of consent, that is, only if Victim makes the request that Recipient acts in a certain way. In Millum's view, a token of consent is still necessary to make a recipient's act permissible; and the absence of a token of consent is still sufficient to make Recipient's act impermissible (no matter how well it would accord with the standards of a *reasonable joint decision*).

The relevant difference between Millum's two possibilities in which Recipient can act permissibly is then the following: if consent is valid, Victim's right of self-

[80] Millum, 2014, 126.
[81] Millum, 2014, 121.
[82] Millum, 2014, 119.
[83] Millum, 2014, 119.

determination is not violated and therefore Recipient's action cannot be impermissible in the first place; if consent is invalid, however, Victim's right of self-determination is violated and the permissibility of Recipient's action is to be decided on grounds of further considerations.

In Millum's view, the distinction between valid and invalid consent tracks whether or not the self-determination of the person giving consent is violated. It does *not* track whether or not Recipient acts permissibly. And this is where Millum departs from the orthodox view adhering to the *Permissibility Principle*. Whereas the orthodox view holds that the validity-invalidity distinction determines both (1) whether the self-determination of the person giving consent has been violated and (2) whether Recipient acts permissibly, Millum holds that the validity-invalidity distinction determines only the former.

On the basis of this account, Millum could provide the following solution to the dilemma. Adhering to the *Coercion Principle,* Millum leaves the first set of claims untouched:

1. Influence-Focus-View:	The Coercion Principle is correct: if a person is coerced into consenting, consent is invalid. Who exerts the coercion is irrelevant.
2. Facts in Tattoo:	Victim was coerced into consenting.
3. Conclusion:	Victim's consent is <u>invalid.</u>

However, by modifying the Permissibility Principle, Millum changes the second set of claims:

4.* Modified Permissibility Principle:	There are certain acts that can be performed permissibly either if the relevant person gave valid consent or if performing the act is based on a reasonable joint decision. These acts include executing a tattoo.
5. Verdict in Tattoo:	Recipient acted permissibly in executing the tattoo.
6.* Modified Conclusion:	Victim's consent is <u>either valid or invalid.</u>

On the modified Permissibility Principle, the original conclusion (6), i.e. that consent is valid, does no longer follow. Consent could be either valid or invalid, and as it happens consent is invalid in *Tattoo* and Recipient acts permissibly due to basing his action on a reasonable joint decision. Hence, there is no contradiction anymore. Conclusion (3) and modified conclusion (6*) are consistent with each other.

(ii) Criticism of Millum's Account

Millum is certainly to be applauded for making a genuinely novel and important contribution to the debate on consent. However, as I shall argue in this section, the modification of the *Permissibility Principle* is not a feasible option and therefore ought to be rejected.

To begin with, the arguments Millum himself employs to support his view are not convincing. Firstly, Millum claims that his view "captures more clearly the different ways in which the victim of coercion can be wronged and harmed."[84] He says,

[84] Millum, 2014, 120.

"When Coercer forces Victim to token consent to an act over which she has an autonomy right, he *wrongs* her. This is true even if the act to which she tokens consent does not actually occur. If that act does occur, then she experiences a *harm* over and above the wrong of being coerced."[85]

I agree with Millum's distinction between harms and wrongs but I disagree that this distinction is something that only his view can provide. It is correct to distinguish between the wrong of coercion (done by Coercer) and the harm resulting from whatever consent was tokened for (done by Recipient). Coercer wrongs Victim because he violates Victim's rights by subjecting her to coercion. But Recipient does not violate any of Victim's rights, hence does not wrong Victim but only encroaches on some of her interests (i.e. the interest of not having a tattoo). However, an opposing account, according to which Victim's consent is valid, could equally make these distinctions. This view could acknowledge Coercer's wronging as well as Recipient's harming, and even claim that this is why consent is valid, viz. consent is valid because Recipient only harmed but did not wrong Victim. Hence, Millum's view does not capture the different wrongs and harms "more clearly" than opposing approaches but only 'as clearly' as these.

Secondly, Millum claims that, irrespective of whether a third party or the recipient exerts the coercion, "[Victim's] rights are still violated in the same way"[86], i.e. in both cases Victim's self-determination is undermined so as to vitiate consent. Millum is not explicit here but he seems to rely on Victim's experience, i.e. the claim that since it does not make a difference to Victim who exerts the coercion, consent must be invalid in three-party cases just as it is in two-party cases. However, even if Victim's experience were identical in both types of cases, it does not follow that the rights violations are also the same. John Gardner showed very clearly that a victim's experience alone does not

[85] Millum, 2014, 120. Emphasis added.
[86] Millum, 2014, 119.

determine whether and which of her rights have been violated. He presents a case of an unconscious person who is raped in a way that leaves no traces. The raped person will never find out what happened. Thus, from the perspective of her experience, this situation is identical to the one where she is not raped.[87] However, there is clearly a great difference in terms of violation of this person's rights. Hence, pace Millum, a victim's experience cannot provide conclusions about whether and which of her rights have been violated.

Millum's arguments aside, there are yet further problems. Millum's view puzzlingly implies that there can be rape without a rapist. To see this, consider the following example, which Millum himself mentions and claims to be parallel to *Tattoo:*

Sex "If Coercer forces Victim to agree to sexual relations with someone else, and they have sex, then she will have experienced the harm of having sex against her will, whether or not Recipient is acting permissibly."[88]

By saying "whether or not Recipient is acting permissibly", Millum allows for a scenario where Recipient may permissibly engage in sex with Victim just as he may permissibly execute the tattoo in the original example. Hence, Recipient can have a justification for *knowingly and freely* engaging in non-consensual sex with Victim,[89] that is knowingly and freely committing the actus reus of rape. Hence, the combination of Millum's claims allows for the possibility of rape without a rapist. Coercer cannot be the rapist because he did not touch Victim, and Recipient cannot be the rapist because he was justified in doing what he did. Thus, there can be rape without a rapist in situations where Recipient knows everything about the situation and is not ignorant of Victim's circumstances.

---

[87] See Gardner, 2007, 5.
[88] Millum, 2014, 120.
[89] Sex is non-consensual because of Millum's adherence to the *Coercion Principle*.

Furthermore, there seems to be a conceptual argument against considering consent invalid in cases like *Tattoo* or *Sex.* In these cases, the wrong done to Victim is exerted by a third party and occurs entirely outside the actual consent transaction between Victim and Recipient. However, insofar as (in)validity is supposed to be a predicate *of* a consent transaction, it seems odd to invoke invalidity where there is no flaw *in* the actual consent transaction but only outside of it. For conceptual reasons, therefore, validity as a predicate of consent transactions should only be applied where one can locate a wrong *in* the actual consent transaction.

This point can also be spelled out in more substantial ethical terms. One can describe and evaluate Coercer's wrongs without relying on the moral quality of the consent transaction between Victim and Recipient. Coercer can be ethically and legally charged for subjecting Victim to coercive pressure and, independent of any reference to consent, there is already sufficient protection and means of legal redress available. Hence, invoking the invalidity of consent is ethically and legally superfluous.

However, the main flaw in Millum's account is that he simply accepts the *Coercion Principle,* which lies at the heart of the *Influence-Focus-View* but which requires substantial modification. Modifying the *Coercion Principle* will detract from the *Influence-Focus-View* and instead support the second contender for how to deal with third-party coercion: the *Recipient-Focus-View.* I will therefore use the next section to show why and how the *Coercion Principle* requires modification. At this point, I conclude that Millum's view suffers from several flaws: Millum's own arguments are not convincing, his view puzzlingly implies the possibility of rape without a rapist, and misleadingly invokes the predicate of validity where there is no flaw in the actual consent transaction. On these grounds, I suggest rejecting Millum's view as a solution to the dilemma at hand.

2.3.3. The Recipient-Focus-View

I. Modifying the Coercion Principle as an Alternative Solution to the Dilemma

To see why and how the *Coercion Principle* requires modification, consider the following case:

Poison Coercer wants to break into Victim's house and, therefore, looks for a way to make Victim leave. Coercer decides to poison Victim so that Victim must undergo specific medical treatment in a hospital away from her home, or otherwise suffer serious physical harm. After poisoning Victim, Coercer tells Victim about the poison (but not his burglary plan) and what Victim needs to do to avoid harm. Victim goes to the hospital and consents to treatment.

Can Victim validly consent to treatment? It certainly seems permissible for a physician to treat Victim, and Millum would concede this. But *Poison* is a case of necessity, and not duress, and in cases of necessity consent is valid.[90]

Poison is a case of necessity because it is Victim's circumstances, i.e. being poisoned rather than threatened, which force Victim to consent. *Poison* is a case of necessity even though, just as in *Tattoo,* Coercer deliberately and wrongfully creates a situation that forces Victim to consent to something and even though Coercer also specifies the act required to avoid harm, i.e. specific medical treatment.

Hence, the difference between cases like *Poison* (necessity) and cases like *Tattoo* (duress) is *not* that only the latter involve deliberate human agency. The difference

[90] Calling *Poison* a case of necessity may seem slightly unusual because it departs from paradigm cases of necessity, e.g. *Vincent v Lake Erie Transportation Co.* [1910], which are about violations of property rights in cases of genuine emergency. Nevertheless, *Poison* significantly resembles cases of necessity because the pressure on a person's decision-making comes from a natural process (i.e. the continuing impact of the poison) rather than the conditional threat of another person. And this parallel is what I will be relying on and hope will justify my terminology.

consists in the presence or absence of a *threat:* in *Poison,* Coercer does not threaten Victim. He violates one of Victim's rights by poisoning her. In *Tattoo, however,* Coercer *threatens to violate* one of Victim's rights (e.g. right to bodily integrity) but does not already violate *that* right. Therefore, in *Poison,* harm will occur due to causal necessity following from the administering of the poison, whereas in *Tattoo* harm will occur as a result of another act on the part of Coercer. However, in both cases harm *will* occur: Millum asks us to take for granted that, in *Tattoo,* harm *will definitely occur* if Victim does not consent to the tattoo, even if it does not occur due to causal necessity.

So, if consent is to be invalid in *Tattoo* but not in *Poison,* this could only be because of what distinguishes *Tattoo* from *Poison*: the presence of a threat. However, this difference seems to provide support for the opposite conclusion, namely that consent was *valid.* To see this, consider the following *a fortiori* argument*:* (i) if Coercer's violation of one of Victim's rights in *Poison* does not vitiates Victim's consent (as it is a case of *necessity*), and (ii) if the *violation of a right* is a more serious wrong than a *threat to violate* a right, and (iii) if the illegitimacy of an influence (i.e. being a wrong as well as the seriousness of the wrong) determines validity (what Millum and I both accept), (iv) then it follows *a fortiori* that Coercer's *threat* to violate a right in *Tattoo* does not vitiate Victim's consent, other things being equal. Hence, Victim's consent in *Tattoo* is valid.[91]

If this conclusion is correct, the *Coercion Principle* requires modification. Victim's fear of harm alone cannot vitiate consent because motives are normally not significant for the validity of consent.[92] Coercer's threat cannot vitiate consent given the

[91] Note that this argument relies on the assumptions that harm *will* occur, both in the case of a threat and in the case of necessity, and that the recipient cannot do anything about it except performing the act consent was tokened for. If this situation changes, as it does in two-party cases, my conclusion will no longer follow. In two-party cases, the *recipient himself* could always withdraw the threat and *thereby* prevent harm. Hence, I can still accept the very plausible view that, in two-party cases, consent is invalid in cases of threats but valid in cases of necessity, while also adhering to the argument about three-party cases expressed in the section above.

[92] Jacky may consent to sex for pleasure, to become pregnant, or simply out of boredom. Susanne may consent to a contract to seize a business opportunity, to please her friends, or out of pity for the other party.

a fortiori argument I just outlined. Hence, Victim's consent is *merely motivated* by coercion but not vitiated. Consent is *merely motivated by* coercion if a person's reason for consenting is to prevent a threat from being carried out, but the threat came from a third party and the recipient neither contributed to it nor failed to discharge an obligation towards Victim.

The situation would be entirely different if the recipient exerted the coercion, i.e. if consent were *obtained by* coercion. Consent is *obtained by* coercion if the recipient exerted the coercion himself, collaborated with a third party, or failed to discharge important obligations towards Victim. 'Obtain' implies the involvement of the recipient.[93]

If this distinction is significant, the *Coercion Principle* requires modification. It is false to say that if a person is coerced into consenting, consent is invalid. 'Coerced into consenting' is too general and requires disambiguation: consent is invalid if obtained by coercion, but valid if merely motivated by coercion.

This modification leads to an alternative solution of the dilemma at hand. It is the first set of claims that requires change:

1.* Modified Coercion Principle:	Consent is invalid if obtained by coercion and valid if merely motivated by coercion.
2.* Specified Facts in *Tattoo*:	Victim's consent is merely motivated by coercion.
3.* Modified Conclusion:	Victim's consent is <u>valid.</u>

The second set of claims on the other hand can be left untouched:

None of these motives affects the validity of Jacky's or Susanne's consent. Even fear of death cannot vitiate consent. If Peter consents to a medical procedure because he would otherwise die from serious illness, his consent can still be valid. The rationale behind such a separation of motives and validity is respect for autonomy: people should be able to decide when and why they consent. See Alexander, 1996.

[93] I thank John Hyman for helping me to clarify the terminology here.

4. Permissibility Principle:	There are certain acts that can be performed permissibly only if the relevant person gave valid consent. These acts include executing a tattoo.
5. Verdict in *Tattoo:*	Recipient permissibly executes the tattoo.
6. Conclusion:	Victim's consent is valid.

As a result, there is no contradiction anymore. Modified conclusion (3*) and conclusion (6) are consistent with each other.

This solution not only introduces the necessary modification of the *Coercion Principle* but also avoids the problems Millum's account suffers from. It will consider consent to be *invalid* only if there is an actual flaw *in* the consent transaction rather than outside of it, and thereby avoid the conceptual problem in Millum's view. Moreover, this solution will abandon be the possibility of rape without rapist. If the recipient did not obtain consent by means of coercion, consent will be valid. Furthermore, just as in Millum's view, this solution to the dilemma also accounts for the correct view about harms and wrongs. If consent is *merely motivated* by coercion, there will be Coercer's wrong and Recipient's harm. If consent is *obtained by* coercion, there will be two wrongs. Finally, this view does not falsely suggest that identical victim experience implies identical rights violations.

In the next section, I will explain how modifying the *Coercion Principle* leads away from the *Influence-Focus-View* and demands a different approach to cases of third-party coercion, i.e. the *Recipient-Focus-View.*

II. The Modified Coercion Principle leads to the Recipient-Focus-View

The initial *Coercion Principle* implied the *Influence-Focus-View:* if coercion always vitiates consent, it cannot make a difference who exerts it. The *Modified Coercion Principle* rejects the *Influence-Focus-View*: if consent is invalid only when *obtained* by coercion, but not when merely motivated by coercion, and if 'obtained by' implies the involvement of the recipient, then we need to focus on the *recipient* rather than on the influence alone. Hence, the *Modified Coercion Principle* suggests an account we may call the *Recipient-Focus-View*: in cases of third-party pressure, the validity of consent depends on the behaviour of the recipient.

The core of the *Recipient-Focus-View* is the assumption that consent is genuinely interpersonal, i.e. that the validity of consent depends on the interaction between the person giving and the person receiving consent, and not on what happens outside of this interaction. Therefore, the behaviour of third parties cannot determine the validity of consent. But the behaviour of the recipient of consent is decisive.

To be clear, my focus is on the *recipient's conduct when obtaining consent*, and not on the recipient's performance of what consent was tokened for. Only the former determines the validity of consent. Considering the latter to determine the validity of consent would reverse the correct explanatory order (i.e. (in-)validity explains why doing what consent is tokened for is wrong and not the other way around). Focusing on the recipient's conduct in the process of obtaining consent, the *Recipient-Focus-View* still respects the correct explanatory order: the recipient's conduct in the process of obtaining consent determines the validity of consent, which in turn determines whether doing what consent is tokened for is wrong.[94]

[94] Upholding this explanatory order is important to avoid circularity and sidestep the related criticism that was levelled against views like Nozick's, as mention at the end of Chapter 1.

The *Recipient-Focus View* has prominent historical ancestors. Roman lawyers,[95] early modern natural law theorists,[96] and common law judgments[97] have considered the conduct of the recipient to determine the validity of consent. Paradigmatically for this view, Pufendorf wrote:

"Igitur ille demum metus obligationem producendam elidit, qui in altero ponit vitium, quod ineptum ipsum reddit ad ius acquirendum."[98]

"Thus, fear vitiates the obligation to be produced in the contract if and only if the other party acts in a way that disqualifies him as a recipient of a right."[99]

When Pufendorf talks about "fear" he refers to a coercive threat from a third party, and he argues that such a threat vitiates consent if and only if the recipient rendered himself "ineptum", viz. disqualified himself as a recipient through his own misconduct. Hence, the focus is on the recipient and not the coercive influence taken separately, even in cases

[95] In Roman law, consent was transformative if the recipient acted fairly towards the person giving consent, even if third-party pressure was the condition *sine qua non* for the giving of consent. See for instance Ulpian, Digest 4.2.9.1. Retrieved from: http://www.thelatinlibrary.com/justinian/digest4.shtml

[96] See Pufendorf, 1998, Liber III Cap VI: §X-XI. Wolff, 1740, § 577-581. Wolff here focuses exclusively on the behaviour of the recipient.

[97] In a judgment of 1887, which has been highly influential in American and English law, Judge Holmes claimed: "we know of no distinct adjudication of binding authority that threats by a stranger, made without knowledge or privity of the party, are good ground for avoiding a contract induced by them", from *Fairbanks v Snow* [1887], 599. This case influenced the subsequent adjudication. See *Talbot v von Boris* [1911]. *Kesarmal v Valliappa Chettiar* [1954]. *Barclays Bank v O'Brien* [1994]. In all these cases, the validity of consent to the contract was considered to depend on the behaviour of the contracting parties.

[98] Pufendorf, 1998, Liber III Cap VI: §X. Pufendorf then affirms this view in § XI.

[99] My translation. This is not a literal translation but the one that best captures Pufendorf's claim and takes into account the context in which he makes this claim. I thank Elvira Seibert-Kiener for her help with the Latin translations in this section.

where a third party exerts the influence. Today, it is Mallary,[100] Miller and Wertheimer,[101] and Gerver[102] who come closest to such an approach.

The motivation behind the *Recipient-Focus-View* is to acknowledge the interpersonal dimension of consent. Its focus on the process of obtaining consent renders it well-placed to account for the liability of the person receiving consent. But the challenge for the *Recipient-Focus-View* lies in how to protect the autonomy of the person giving consent. In the next section, I will explain such protection and thereby elaborate on the *Recipient-Focus-View* in greater detail.

III. Requirements for Recipient Behaviour

To recapitulate 2.3.3, section I presented an alternative solution to the dilemma under consideration, based on the *Modified Coercion Principle.* Section II showed how this solution leads to the *Recipient-Focus-View* and highlighted important historical ancestors. In this section III, I will assume that the *Recipient-Focus-View* has provided a convincing solution to the dilemma and turn to two further aspects. Firstly, I will explain in greater detail when consent is 'obtained' in a way that vitiates consent; and secondly, I will show that the *Recipient-Focus-View* so understood can protect the autonomy of the person giving consent, thereby doing justice to the Janus face of consent. Focusing on the autonomy of the person giving consent is necessary because the acceptability of the

[100] See Mallary, 1986.
[101] See Miller and Wertheimer, 2010. Unlike Mallary and others, Miller and Wertheimer separate the conditions of moral transformation from the conditions of validity. Miller and Wertheimer's reasons for doing so are beyond the scope of this chapter. I will therefore not discuss them but continue to assume that consent is transformative if and only if it is valid. However, I want to highlight the main distinction between Millum's separation of validity and permissibility, and Miller and Wertheimer's separation: whereas Millum still assumes that validity is an important aspect because it tracks violations of self-determination, Miller and Wertheimer think that one should depart from the focus on validity altogether because it fails to track what is important in consent transactions in the first place.
[102] See Gerver, 2016.

Recipient-Focus-View will not only depend on a solution to the dilemma set-out earlier but also on whether it accounts for both faces of consent.

When introducing the *Modified Coercion Principle,* I claimed that consent is *obtained by* coercion if the recipient exerts the coercion himself, collaborates with a third party, or fails to discharge an obligation towards Victim. The first two items in this list of disjunctive sufficient conditions (i.e. exerting coercion and collaborating with the third party) make it so that we no longer deal with pure third-party coercion but rather with more complicated two-party cases where the recipient is actively involved in exerting pressure. In contrast, the last item (i.e. failing to discharge an obligation towards the victim of third-party coercion) arises in genuine cases of third-party pressure: the recipient can fail to discharge an obligation without being involved in exerting pressure on a victim. And since my focus in this paper is on genuine third-party cases, this section III will focus on this last item.

This last item highlights the misbehaviour of the recipient in the process of obtaining consent and raises the question as to which kinds of misbehaviour are such so as to vitiate consent. In the rest of this section, I will (i) answer this question by claiming that consent is vitiated if and only if the recipient fails to discharge certain positive obligations. I will then (ii) show how the resulting *Recipient-Focus-View* accounts for the Janus face of consent.

(i) Recipient's Obligations: Influencing Coercer, Offering Alternatives, Respecting Victim's Final Decision

Suppose everything in *Tattoo* stays the same except that Recipient could easily persuade Coercer to remove his threat, or that Recipient could easily fake a tattoo and thereby help Victim to escape. Nevertheless, Recipient cannot be bothered to help. Here, the protection

of Victim's autonomy is certainly flawed, and if the *Recipient-Focus-View* tolerates such a situation, it is worthy of rejection on these grounds. Hence, insofar as protection of autonomy must be included for the *Recipient-Focus-View* to be feasible, and the only way to do so is to impose certain positive obligations on Recipient, such obligations must be included in the *Recipient-Focus-View*.

The *Recipient-Focus-View* I wish to defend holds that consent is invalid if and only if the recipient, when obtaining consent, fails to discharge certain positive obligations towards the victim of third-party coercion. Such obligations must be sensitive to the victim's autonomy and the control she may gain over her situation when being assisted by the recipient. Initially, Victim's control becomes diminished when Coercer removes an option to which Victim is entitled: not being beaten *and* not getting a tattoo. Recipient can enhance and maintain Victim's control in three ways. He can *enhance* it either by influencing Coercer into withdrawing his coercion, thereby restoring the original situation, or by offering Victim alternatives to escape Coercer's threat. He can *maintain* it by respecting Victim as the final decision-maker. As Mallary argued, respecting the victim as the final decision-maker can require that the recipient accept a victim's consent and refrain from influencing a coercing third party if the victim does not want the recipient to do so.[103] These possibilities of enhancing and maintaining a victim's control constitute a recipient's obligations.

However, the recipient of consent can only be under certain obligations, e.g. to influence the coercing third party or to offer alternatives, if discharging these obligations is possible and feasible. It is possible if the recipient is in a position to influence the third party or offer alternatives; and it is feasible if doing so does not impose substantial costs or burdens on the recipient. What counts as *substantial* here and the exact content of the

[103] See Mallary, 1986, 125-126.

recipient's obligations to protect a victim's autonomy will depend on the object of consent, the relation between the consenting parties (patient-physician, spouses, business partners), and the context more generally (medicine, sex, business, etc.). Given the differences across various contexts and their effect on the recipient's obligations, it is not possible to give *one* general elaboration of recipient obligations. It is the particular context that defines the scope of the recipient's obligations, and the *Recipient-Focus-View* has to be adjusted accordingly.

However, there is one further complication. So far, I have assumed that the recipient knows about third-party coercion. But what about cases where he is ignorant of it and therefore fails to discharge his obligations, although it would have been possible and feasible for him to do so? Wolff states the following about the validity of promises:

"Si metu ab alio incusso permotus tibi rei istius minime conscio quit promittam; promissio valida est."[104]

"If I make a promise to you, but have only made this promise out of fear which a third party has instilled in me, this promise is valid if you were ignorant of my situation."[105]

Wolff makes two claims here, both of which can be applied to consent. Firstly, he claims that it is ignorance *simpliciter* that preserves validity and refrains from distinguishing between culpable and non-culpable ignorance. Secondly, he accounts for the significance of ignorance by claiming that it preserves *validity,* rather than by adopting the alternative view, according to which promise and consent are still invalid but the recipient's ignorance provides him with an excuse for acting as he did or makes it the case that, in a

[104] Wolff, 1740, § 577.
[105] My translation.

legal context, he does not possess the mens rea required for a certain offence, e.g. for rape.

Wolff's first claim conflicts with the Janus face of consent. In a case of culpable ignorance, the recipient should have noticed third-party coercion and subsequently assisted the victim. Therefore, Wolff's first claim, i.e. allowing ignorance *simpliciter* to preserve validity, would forgo a chance to strengthen the protection of a victim's autonomy and hold the recipient liable to moral criticism or legal redress when he deserves such liability. In contrast, the rejection of Wolff's first claim, i.e. not allowing ignorance *simpliciter* to preserve validity, could at once account for the protection of a victim's autonomy as well as justify holding the recipient liable for what he did. Hence, we need to reject Wolff's first claim: ignorance should only count when it is non-culpable.

However, Wolff's second claim is justified, at least in cases of non-culpable ignorance. This is because it affords the same advantage (in terms of autonomy protection) as the rejection of this claim does, but it avoids a flaw of the latter.

Wolff's second claim, i.e. the view that non-culpable ignorance preserves validity, affords the same protection of the victim's autonomy as its rejection, i.e. the view that non-culpable ignorance *does not* preserve validity,[106] because the protection of the victim within the consent transaction depends on the recipient's obligations, and both views are on a par with regards to these obligations.

Within the consent transaction, the protection of the victim's autonomy depends on the recipient because it is he alone who interacts with the victim and it is he alone who is in a position to enhance and maintain the victim's control over her situation. Therefore, imposing obligations on the recipient is the only way to protect the victim's autonomy *within* the consent transaction.

[106] I here followed my result from the discussion of Wolff's first claim and restrict the focus to non-culpable ignorance.

Both views, i.e. the acceptance and rejection of Wolff's second claim, are on a par with regards to the obligations they can impose because claiming that the recipient *ought* to help the victim requires that he *can* be expected to do it, and if he is non-culpably ignorant of the situation, helping the victim is precisely what he *cannot* be expected to do. Therefore, the recipient cannot be obliged to help the victim beyond cases where he is culpably ignorant. Thus, rejecting Wolff's claim does not lead to greater obligations for the recipient than accepting Wolff's claim. Wolff's second claim and its rejection provide the same protection of the victim's autonomy within the consent transaction.

However, Wolff's second claim avoids a flaw that the rejection of it suffers from. Once we have established that the recipient's behaviour determines the validity of consent, rejecting Wolff's second claim would imply the flawed view that *a recipient's non-culpable failure to discharge obligations vitiates the victim's consent*. Accepting Wolff's second claim on the other does not imply such a view and thereby avoids this flaw.

Therefore, the *Recipient-Focus-View* should reject Wolff's first claim but accept his second claim, embracing the view that non-culpable ignorance preserves the validity of consent.

Hence, to come full circle: consent is invalid if and only if the recipient, when obtaining consent, fails to discharge certain important obligations owed to the victim. The recipient fails to discharge such obligations if and only if it was possible and feasible for the recipient to do so and the recipient was either aware of the situation or culpably ignorant of it.

(ii) Accounting for the Janus Face of Consent

A *Recipient-Focus-View* which adopts my position on the recipient's obligations protects the autonomy of the person giving consent, at least as well as an *Influence-Focus-View* could; and given that its focus on the recipient renders the *Recipient-Focus-View* already well-placed to account for the liability of the person receiving consent, the *Recipient-Focus-View* will then account for the Janus face of consent.

The autonomy of the person giving consent requires protection *within* as well as *outside* the consent transaction and the *Recipient-Focus-View* can provide this.

By imposing obligations on the recipient, the *Recipient-Focus-View* provides norms designed to protect a victim *within* the consent transaction. The obligations I developed can indeed be quite demanding, depending on the context and what a person consents to. There is no reason why a *Recipient-Focus-View* could not be as stringent as an *Influence-Focus-View* with regards to norms governing the permissibility of recipient behaviour. In fact, it seems that the *Recipient-Focus-View* I am defending may even provide greater protection, at least when compared to Millum's *Influence-Focus-View*. Depending on how Millum's view on a *reasonable joint decision* is explained in greater detail, arriving at such a decision, which could just be a compromise, may require considerably less than my view focusing on positive obligations of the recipient. Hence, a *Recipient-Focus-View* view provides sufficient protection of the victim *within* a consent transaction.

In addition, just as my view can effectively protect a victim *within* the consent transaction, it can also still endorse all those norms designed to protect a victim *outside* the consent transaction. The *Recipient-Focus-View* can acknowledge that Coercer's influence is a wrong and the negative consequences resulting from the coerced consent transaction will be harmful for Victim. On the *Recipient-Focus-View,* one can still charge

Coercer not only for the wrong of coercion but also for the harmful consequences resulting from the consent transaction.[107] This is true even though the *Recipient-Focus-View* does not consider Coercer's influence to vitiate consent. Hence, on the *Recipient-Focus-View,* there can still be norms proscribing illegitimate conduct by third parties and thereby protecting the autonomy of the person giving consent.[108]

Therefore, the focus on the recipient's behaviour does not prevent my view from protecting the victim's autonomy. On the contrary, such a focus facilitates strong protection of the victim's autonomy. Hence, the *Recipient-Focus-View* not only accounts for the liability of the person receiving consent but also for the autonomy of the person giving consent, and thereby for the Janus face of consent.

2.3.4 Conclusion

I started this chapter by claiming that any satisfactory account of the validity of consent must do justice to the Janus face of consent, i.e. account for the autonomy of the person giving consent and the liability of the person receiving consent, and I claimed that cases of third-party coercion make it particularly difficult to do so. I then discussed the two main approaches for how to deal with cases of third-party coercion: the *Influence-Focus-View* and the *Recipient-Focus View.*

According to the *Influence-Focus-View,* whether an influence vitiates consent depends only on whether the influence displays certain qualities, e.g. being coercive, but

[107] Feinberg correctly pointed out that Victim has such a claim against Coercer and that Coercer is responsible not only for the wrong of coercion but also for what happens subsequent to the coercion. I agree with Feinberg that it is "not unfair to punish C [third party] for the whole wrongful harm of which A's [the recipient of consent] conduct was an innocent but direct contributor, and his own [third party] conduct a wrongful but indirect contributor." See Feinberg, 1986, 195-196. Wolff also argued that Victim can charge Coercer for the consequences of her coerced consent as well as demand compensation if applicable. See also Wolff, 1740, § 578.

[108] On the distinction between the wrong ('delict' or 'tort' in legal terminology) of coercion and the wrong that vitiates consent ('plea' of invalidity in legal terminology) see Martens, 2007, 150, 247, 376.

not on who exerted the influence or on the behaviour of the recipient of consent. I claimed that the *Influence-Focus-View* view leads to a dilemma that requires the rejection of either the *Coercion Principle* or the *Permissibility Principle*, a dilemma that the *Influence-Focus-View* is ultimately unable to solve. Therefore, I claimed, the *Influence-Focus-View* should be rejected.

I then developed an alternative solution to the dilemma based on what I called the *Modified Coercion Principle* and showed how this solution leads to the *Recipient-Focus-View,* according to which the validity of consent depends on the behaviour of the recipient. I claimed that the *Recipient-Focus-View* can convincingly solve the dilemma and is therefore superior to the *Influence-Focus-View.* I then explained in greater detail which kinds of recipient misbehaviour vitiate consent and how the *Recipient-Focus-View* accounts for the Janus face of consent. Focusing on the process of how consent is obtained enables the *Recipient-Focus-View* to be particularly sensitive to the perspective of the recipient and his liability; and spelling out how to understand consent being 'obtained' so as to vitiate consent enabled the *Recipient-Focus-View* to account for the autonomy of the person giving consent. In particular, imposing positive obligations on the recipient and specifying which kinds of misbehaviour vitiate consent facilitated a nuanced protection of the autonomy of the person giving consent, one which is also superior to Millum's *Influence-Focus-View*. On these grounds, I propose the *Recipient-Focus-View* as the correct approach to cases of third-party coercion.

This result is of fundamental significance to consent more generally. If the validity of consent depends on the behaviour of the recipient not only in two-party but also in three-party cases, the validity of consent is genuinely interpersonal. Validity is determined by what happens *within* a consent transaction and not outside of it. Any satisfactory account of the validity and voluntariness of consent must therefore follow the

Recipient-Focus-View, adopt the *Modified Coercion Principle,* and thereby make the behaviour of the recipient of consent central to its approach.

3. My Proposal: *Interpersonal Consenter-Consentee Justification (ICCJ)*

Chapter 2 claimed that any satisfactory account of voluntary consent must focus on the reasons or motives for which a person consents, instead of mere causal links or counterfactuals (2.1), allow the moral legitimacy of an influence to have a bearing on the voluntariness of consent (2.2.), and make the behaviour of the recipient of consent central to its approach (2.3).

To account for these claims, I will now develop and defend the following view, which I call *Interpersonal Consenter-Consentee Justification (ICCJ)*:

Consent is voluntary unless it is motivated by an influence that cannot be interpersonally justified. Interpersonal justification requires that the consentee (i.e. the person receiving consent) can justify to the consenter (i.e. the person giving consent) that he *exerted* the influences or, when influences had a different source, *tolerated* their presence. The principles fundamental to the consenter-consentee relationship provide the basis for interpersonal justification.

ICCJ employs the concept of interpersonal justification, which is also known from contractualism in moral theory and which provides a particularly attractive framework for the claims from Chapter 2. Interpersonal justification does not imply any view about causation or counterfactuals and can therefore be easily combined with my view on a person's reasons or motivation for consenting (2.1). Interpersonal justification can allow the moral legitimacy of an influence to have a bearing on the voluntariness of consent: certain morally illegitimate influences may fail to be interpersonally justifiable and vitiate consent for that reason (2.2). Finally, it is the person receiving consent who will need to

interpersonally justify how he obtained consent and thereby becomes central to ICCJ's approach (2.3).

In the context of treatment, an interpersonal justification is an adequate reply to the question: "How can *you* (*qua* physician) obtain consent from *me* (*qua* patient) in this way?" When I say "in this way" I mean the physician exerting or tolerating an influence on the patient's decision-making. However, it is not necessary for the patient to ask that actual question or for the physician to give an answer to it. It is instead about the *possibility* of giving an interpersonal justification, i.e. that influences could hypothetically be justified.

A reply to the above question is adequate if it shows that the conduct of the person receiving consent is consistent with the principles fundamental to his relationship with the person giving consent. More precisely, I suggest restricting the focus to those principles which afford the kind of protection that also informed consent purports to afford. This overlap between the principles fundamental to the consenter-consentee relationship and the purported protection of informed consent ensures two things: firstly, that ICCJ explains how the principles fundamental to the consenter-consentee relationship can determine the validity of consent and, secondly, that ICCJ is also sufficiently specific to questions of consent, as opposed to other ethical questions in medical settings. ICCJ can thereby distinguish between influences which vitiate consent and influences which do not, even if the latter violate other relationship norms.

Interpersonal justification so understood is not a trivial consequence of justification simpliciter. It is *not* the case that the physician's conduct will be, firstly, justified on grounds unrelated to the patient and then, only as a consequence, be justifiable to anyone, including the patient. The grounds for justification will always be linked to the

consenter-consentee relationship, e.g. patient-physician relationship, and therefore always be relative to the person consenting.[1]

To elaborate further on ICCJ and the principles relevant to interpersonal justification, I will continue to focus on the context of treatment and the patient-physician relationship (3.1). Later, I will explain that almost the same principles apply to the participant-investigator relationship in the context of clinical research (3.2), and explain in detail how ICCJ satisfies the desiderata from Chapter 2 (3.3).

3.1 Treatment

It is a principle fundamental to the patient-physician relationship that the physician ought to respect the patient's autonomy as well as protect and promote the patient's well-being. Requirements of informed consent align with these concerns and primarily purport to protect the patient's autonomy and his well-being. Therefore, ICCJ takes as a starting point the physician's obligation to respect the patient's autonomy and the obligation to act beneficently. Other principles in the patient-physician relationship, e.g. the principle of justice, will not be among the starting points for ICCJ as they do not overlap with the purpose of informed consent.[2]

[1] My proposal may share some features with Scanlon's contractualist account (Scanlon, 1998) and Nagel's deontological account (Nagel, 2012, 53-74: *War and Massacre*). However, my proposal is only inspired by Scanlon's and Nagel's accounts without being a version or an application of them in any stricter sense.

[2] To clarify: such principles are still very important. However, violating them may not lead to the vitiation of consent but to other forms of wrongdoing, which can be handled in ways other than by considering consent invalid. In fact, there are various sanctions available for physician misconduct that do not affect consent, e.g. regulatory sanctions and employment-related sanctions.

I. The Obligation to Respect Autonomy

ICCJ features a second-personal perspective on practical reasoning because its central question "How can *you* (*qua* physician) obtain consent from *me* (*qua* patient) in this way?" (i) explicitly relates physician and patient *to each other*, (ii) assumes the patient's authority to demand justification *from the physician,* and (iii) emphasises that the physician is *accountable to the patient.* [3] In contrast, neither a first-personal perspective (i.e. "How could *I qua* physician obtain consent in this way?") nor a third-personal perspective (i.e. "How could *she qua* physician obtain consent in this way?") implies a reference to these three aspects. This contrast between different personal perspectives not only explains ICCJ in greater detail but also provides a new take on autonomy. It allows for a categorisation of the numerous conceptions of autonomy in the published literature into two main approaches.

On the one hand, autonomy has been understood as "a state of realizing a set of higher-order capacities of rational thought and agency, whereby [on some accounts] practical options are reflectively endorsed and plans of action formulated or brought to execution (with or without the assistance of others)."[4] Unlike ICCJ, autonomy so understood guides practical reasoning from a *first-personal perspective.* It provides an answer to the question "How am *I* to act in order to satisfy the demands of autonomy and be a good ruler over myself?"[5] Requirements of autonomy are satisfied when people actually exercise their relevant capacities. The most prominent version of this approach in the context of informed consent has roots in Mill's philosophy and emphasises individuality and independence. [6] Millian versions of autonomy ramified into three main

[3] See Darwall, 2006 on a second-personal approach to morality and practical reasoning.
[4] Lillehammer, 2012, 197.
[5] Here, the first-personal perspective relates to the patient's point of view.
[6] To clarify: Mill himself did not talk about autonomy but rather about individuality and liberty. Nevertheless, the views on autonomy described in this section resemble some of Mill's statements. For a discussion, see O'Neill, 2002, chapters 2 and 4.

accounts, varying according to whether they focus on autonomous *action*, the autonomous *person*, or an autonomous *life.*

Faden and Beauchamp focus on autonomous *acts* and *choices*. Acts and choices are autonomous if they are intentional, come with understanding, and are free from influences that impede the exercise of certain capacities. Such choices can be made by either an autonomous or a non-autonomous person. Moreover, having the capacity to make autonomous choices does not guarantee that one will make autonomous choices in every situation. [7]

According to G. Dworkin, autonomy is a "characteristic of persons"[8], not of choices. Although there are different ways of explaining autonomy as a characteristic of persons, the most prominent account, also found in Dworkin, refers to the idea of a multi-layered structure of the human mind. According to Dworkin, autonomy is "a second-order capacity to reflect critically upon one's first-order preferences and desires, and the ability either to identify with these or to change them in light of higher-order preferences and values."[9]

Young widens the perspective to an entire life. According to Young, "[a]n autonomous, or self-determining, person chooses or devises a plan for her life."[10] Like Dworkin, Young also appeals to the idea of bringing into accordance one's first order decisions with one's second-order aims and values. A person is self-determining when "making decisions about her life that accord with her own aims and values."[11]

Finally, even though so-called relational accounts of autonomy criticised what they considered an inappropriate focus on individuality in Millian accounts, they also fall

[7] See Faden and Beauchamp, 1986, Part III.
[8] Dworkin, 1988, p. 108. Emphasis added.
[9] Dworkin, 1988, p. 108. Emphasis added. Notice that such an account is closed to what has been labelled "authenticity".
[10] Young, 2009, 520. Emphasis added.
[11] Young, 2009, 531.

into the approach of autonomy as an exercise concept featuring a first-person perspective (albeit one that may relativise the "I" to a "We").[12]

On the other hand, autonomy has been understood as a more Kantian than Millian concept. O'Neill, who defends a Kantian proposal, explains that

"Kant used the term autonomy to refer not to a characteristic of individuals, but to the formal properties of principles of action that can serve for all, and in particular to the combination of law-like form and universal scope."[13]

Following Kant, O'Neill rejects all the previous Millian accounts according to which autonomy is an attribute of acts, persons, their lives, or about individuality and independence. In contrast, O'Neill claims that autonomy "is a matter of acting on certain sorts of principles, and specifically principles of *obligation*"[14] and that it focuses on "a distinctive constraint or requirement, a test that shows which *principles of action could be chosen by all*."[15]

Giving primacy to obligations, O'Neill adheres to the three aspects constitutive of the second-personal perspective explained above. Principles of obligations (i) relate the person who is under an obligation to at least one other person to whom the obligation is owed, (ii) imply that the person to whom the obligation is owed can demand its fulfilment, and (iii) allow the person to whom the obligation is owed to hold the obliged person accountable.

---

[12] According to accounts of relational autonomy, a person's identity is necessarily *relational*, viz. socially embedded and inter-dependent. Therefore, autonomy designates not the fostering of individuality or independence, but rather the promotion of a person's flourishing within the context of versatile relations based on the "values of love, loyalty, friendship, and care" (Herring, 2016, 213). For a good overview of accounts of relational autonomy see Herring, 2014, chapter 3. See also Mackenzie and Stoljar, 2000. Donchin, 2001. Laufer-Ukeles, 2011.

[13] Manson and O'Neill, 2007, 17.

[14] O'Neill, 2002, 84. Emphasis added.

[15] O'Neill, 2002, 84. Emphasis added.

By referring to principles that "could be chosen by all"[16], O'Neill reinforces this second-personal perspective since such principles will have to be explained in terms of the demands that free and rational agents can make on each other's conduct.[17]

Hence, autonomy so understood, like ICCJ, guides practical reasoning from a *second-personal perspective.* It provides an answer to the question "How am I to treat others, i.e. *you,* so as to respond appropriately to their moral status and the claims they can validly make on me?" Requirements of autonomy are satisfied if a person's entitlement to control his situation is respected. As O'Neill explains: "Kantian autonomy is manifested in a life in which duties are met, in which there is *respect for others* and their rights."[18] Accordingly, respect for a patient's autonomy requires freedom from wrongful interference and his having the potential to exercise, as opposed to actually exercising, the capacities for self-governance.[19]

Hence, (Kantian) *autonomy qua respect* and ICCJ converge on the second-personal approach to practical reasoning, whereas (Millian) *autonomy qua exercising autonomous capacities* and ICCJ diverge. Therefore, only autonomy *qua* respect can be understood as extending the idea of interpersonal justification to the concept of autonomy. For this reason, ICCJ will rely on autonomy in terms of respecting the patient as opposed to the patient realising norms of self-governance.[20]

[16] O'Neill, 2002, 84.
[17] See also Darwall, 2006, 35.
[18] O'Neill, 2002, 83. Emphasis added.
[19] For a more detailed discussion on how Kantian moral philosophy connects to a second-personal approach to morality, see Darwall, 2006, especially chapter 9 and pages 35, 118, 242.
[20] Note that ICCJ thereby sidesteps the criticisms against exercise concepts of autonomy, as mentioned at the end of Chapter 1.1.

II. The Obligation to Act Beneficently

Physicians are also under an obligation to act beneficently, i.e. to protect and promote their patient's well-being. The obligation to act beneficently is necessary in order to do justice to the particular vulnerability of the patient, the asymmetry between patient and physician, and the importance of bodily integrity. In order to explain beneficence as part of ICCJ more clearly, I want to elaborate, firstly, on what I take the patient's well-being to consist of and, secondly, on how 'protecting and promoting' it is to be understood.

ICCJ's principles of beneficence focus on four different elements of well-being. All of them also connect to informed consent. (1) There is health, i.e. the proper functioning of body parts and psyche. It is an integral part of a person's well-being and therefore a paramount good.[21] Informed consent is a tool to authorise medical procedures which, ideally speaking, protect, promote, or at least do not (significantly) harm a person's health. (2) There is the good of being in charge of one's own life and decisions.[22] Informed consent provides the setting for a person to make decisions for himself. (3) There is the good of aligning medical procedures with one's overall values and preferences, which can contribute to a person's well-being. Informed consent enables people to choose the procedures that fit best with their values and preferences. (4) There is the relationship with medical personnel that can contribute to a person's well-being. Good rapport reduces the distance or asymmetry between patient and physician and can have immediate positive effects on a person's well-being. Informed consent is a means of interaction between patient and physician and thereby a means to build a potentially valuable relationship.

[21] For a discussion of health and its relation to a person's good and well-being, see Pellegrino, 1988, chapter 6. Schroeder, 2015.
[22] See Dworkin, 1988, 112. Crisp, 1997, 61-62. Cullity, 2007, 24. Groll, 2015, 506-508.

'Protecting' and 'promoting' these elements of well-being includes the principle of Nonmaleficence, according to which the physician must not harm the patient and diminish these elements of his well-being. Nonmaleficence is a negative obligation of non-interference. However, protecting and promoting the patient's well-being also includes beneficence in the narrow sense, which – understood as a positive obligation – requires actively promoting a person's well-being and removing what could diminish it. Hence, I will use the term Beneficence in a wide sense to include both Nonmaleficence (i.e. Do no harm) and beneficence in the narrow sense (i.e. Do good).[23]

Furthermore, protecting and promoting well-being are part of obligations of bene-*ficence* as opposed to bene-*volence.* Bene-*ficence* is a matter of outcome: an act is beneficent if it actually protects and promotes well-being, irrespective of the agent's intention. Bene-*volence*, by contrast, is a matter of intention: an act is benevolent if the agent *intends* to protect and promote well-being, irrespective of the outcome. Ideally, physicians would act beneficently, as well as benevolently, and benevolence is often the agent's guide to beneficence. However, physicians may sometimes act beneficently but not benevolently (e.g. when a physician contributes to a patient's well-being because of self-interest) and vice versa (e.g. when a physician intends to promote a patient's well-being but fails to do so). ICCJ will focus on bene-*ficence* and thereby make the actual impact on a patient's situation, rather than the physician's intention, central to its approach.

Finally, other norms, the law, and the boundaries of the medical context restrict the scope of a physician's obligation to act beneficently. With regards to other norms, a physician is not under an obligation to benefit a patient if doing so involves treating other

[23] This is a terminological choice that is shared by some but not others, e.g. Beauchamp and Childress, 2013. However, nothing substantial depends on whether one uses Beneficence in a wide or narrow sense. See also Frankena, 1987.

patients unjustly, e.g. by giving all the available resources to one patient, or if he thereby violates norms of confidentiality, e.g. by providing information about other patients. With regards to the law, a physician does not have to benefit a patient if doing so involves acting illegally. Finally, a physician is not under an obligation to benefit a person in ways unrelated to the medical context: e.g. a physician does not have to drive her patients back home and pick them up in the morning to save them from a stressful journey on public transport.[24]

III. Combining the Obligation to Respect Autonomy and the Obligation to Act Beneficently

The previous two sections explained a physician's obligations to respect his patient's autonomy and act beneficently. I elaborated on these obligations with an eye to ICCJ's interpersonal perspective: I opted for a second-personal approach to autonomy and also included aspects of well-being that had an interpersonal dimension, i.e. (2) being respected by others as the person in charge of his situation and (4) the valuable rapport with medical personnel.

However, it remains to specify further how these obligations concerning autonomy and beneficence can inform judgments about the voluntariness of consent. In order to do so, I will base ICCJ on three more specific principles that combine and are derived from the obligations to respect autonomy and act beneficently. I will assume that requirements of consent primarily purport to protect autonomy, but also that a convincing account of informed consent must be shaped by both autonomy *and* beneficence.

[24] See Abrams, 1982.

Assisted Decision-Making

Assisted Decision-Making is the principle that obliges the physician (i) not to control his patient's decision-making and (ii) to provide help when the patient has to make his decision under pressure.

(i) The physician controls a patient's decision-making when he compromises the latter's ability or opportunity to make a free decision, e.g. when he threatens to harm a patient if he does not consent, renders non-consent costly in some significant way, prevents the patient from taking part in some aspects of the decision-making, or weakens a patient's emotional or cognitive capacities for decision-making.[25] Controlling a patient's decision-making becomes more severe when it influences the patient in directions that differ greatly from his otherwise existing preference or when the patient lacks feasible means to protect himself against an influence or its consequences, e.g. to protect himself against the threatened removal of his medical care.

The physician does *not* control a patient's decision-making when he affects the decision-making in other ways. Over and above providing medical information, the physician may actively help the patient to clarify views and options, make recommendations, and offer further help when possible and feasible. How much and which assistance is appropriate depends on what a person consents to and the wider circumstances. More serious medical procedures may require greater assistance than less serious procedures. Moreover, in some cases, assistance may require joint deliberation, whereas in others, e.g. emergency cases with little time for a decision, assistance should take the form of a straightforward recommendation.

[25] Shaming may weaken emotional capacities as it intimidates a person and reduces his feeling of self-worth. Forms of distraction, as will be discussed in chapter 4, may weaken cognitive capacities.

Assisted Decision-Making resembles what is sometimes called *shared decision-making.* Like the latter, *Assisted Decision-Making* allows for the physician to engage in shared deliberation with his patients. Unlike the latter, however, *Assisted Decision-Making* makes it explicit that, in certain cases, assisting a patient may require something other than shared deliberation.

Furthermore, *Assisted Decision-Making* differs from narrow models of shared decision-making as it does not amount to a mere division of labour where physicians only provide technical information and patients contribute to the conversation with their values.[26] Physicians may assist their patients more actively and help them to reach a final decision with regards to their overall values too. Moreover, *Assisted Decision-Making* differs from more demanding models of shared decision-making as it does not require the final decision to be mutually agreed upon.[27] The patient retains the final decision-making authority. Finally, *Assisted Decision-Making* is not, unlike some versions of shared decision-making, an alternative to informed consent but rather a device to specify the requirements of informed consent.

(ii) On the other hand, *Assisted Decision-Making* obliges the physician to provide help when the patient has to make his decision under pressure, as in cases of third-party coercion. Accounting for the results from Chapter 2.3, *Assisted Decision-Making* holds that the physician is obliged to offer an option to remove or mitigate the pressure where doing so is possible and feasible. The physician's obligation to do so will arise when an integral part of a patient's well-being is at stake, when the third-party's influence is strong

[26] See Stark, 2013. Joffe, 2009. Sandman, 2009. Entwistle, 2016.
[27] Whitney, 2004. Joffe, 2009. Sandman, 2009. Tilburt, 2011. Stiggelbout, 2012. However, some these models of shared decision-making concede that the patient should have more control over the final decision than the physician. Further discussion of shared decision-making can be found in: Kon, 2012. Olthuis, 2014. Gillick, 2015. Grande, 2015. With regards to shared decision-making, autonomy, well-being, and mutual obligations, see King and Moulton, 2010, 89. Tilburt, 2011, 148-150. Stark, 2013, 13. Kelley, 2015, 5. Sugarman, 2015, 18.

and illegitimate, and when it is feasible for the physician to help. However, whether the physician ought to help will also still depend on the patient's request. The physician is not allowed to act contrary to the patient's preference. Thus, just like in (i), the patient retains the final decision-making authority.

Leaving the final decision-making authority with the patient enables *Assisted Decision-Making* to distinguish between the voluntariness of a decision and its reasonableness or rationality. If the patient has the final say, it may be possible that his decision remains unreasonable or ill-judged in some way. Yet, the patient ought to be respected in his decision-making and, other things being equal, can still consent voluntarily.[28]

Assisted Decision-Making promotes respect for patient autonomy by granting the patient the final decision-making authority as well as by assisting the patient to exercise his decisional capacities. *Assisted Decision-Making* promotes beneficence because (i) the physician's assistance reportedly increases the patient's adherence to the procedure chosen and can thereby promote his health; (ii) it can make it more likely that the patient's decision aligns with his overall values, (iii) it respects the patient as being in charge of his situation, (iv) it develops valuable rapport between the patient and the physician.

Equal Moral Standing

Equal Moral Standing is the principle that obliges the physician to respect a patient's entitlement to plan, revise, and pursue a conception of the good.

This principle is important because people base their medical decision not only on purely medical considerations but also on aspects related to their values and

[28] Note that ICCJ thereby sidesteps another criticism that was levelled against views like Feinberg's, i.e. the criticism of conflating voluntariness with rational choice, which I mentioned at the end of Chapter 1.1.

preferences more generally. For instance, a pianist may accept greater side effects of a particular treatment option to avoid other procedures that could paralyse his hands; an Orthodox Jew may reject certain blood transfusions in order to avoid violating a religious commitment; and a father may undergo strong chemotherapy in order to live until his daughter's next birthday.

I call this principle *Equal 'Moral' Standing* because, in all these cases, people exercise what Rawls called a *moral power*: the capacity to plan, revise, and pursue a conception of the good.[29] I also term it *'Equal' Moral Standing* because it obliges physicians to acknowledge that patients, despite being in an asymmetrical relationship to physicians in other respects, are still on a par with them regarding their entitlement to pursue a conception of the good.

People can valuably pursue a conception of the good in various circumstances. For example, (1) they do so even when they eventually consent to something with great reluctance, e.g. when someone consents to exposure therapy in order to overcome her phobia; (2) when they promote other-regarding interests, i.e. interests concerning other people's fate and preferences, e.g. when a mother takes into account the welfare of her child; (3) when there are only very few alternatives to choose from. In fact, once the situation has become dire, it may become particularly important for a person to be able to select what aligns best with his values and preferences.

These aspects have important implications. If *Equal Moral Standing* ought to inform judgments about voluntariness, then voluntariness may not require the absence of reluctance (from 1), an influence may not be more problematic simply because it induces a person to do what benefits someone else (from 2), and lacking a variety of great options may not automatically negate the voluntariness of a person's action either (from 3).

[29] See Rawls, 2005.

Equal Moral Standing will be called into question when a person is treated paternalistically. Quong, discussing paternalism in general, claimed that a paternalistic act is always "motivated by a *negative judgment"*[30] about the other person's ability to make the right decision or behave reasonably by herself. The relevant abilities here are "practical reasoning, willpower, and emotion management."[31] "*Physical abilities* and *informational deficits* are (…) excluded categories".[32] Such a negative judgment amounts to "denying that another person or group has the necessary capacity, in a given context, to (…) plan, revise, and rationally pursue their own conception of the good"[33] and, therefore, the "paternalizer fails to accord the paternalizee *equal* status."[34]

However, not every paternalistic influence violates *Equal Moral Standing.* An influence violates *Equal Moral Standing* only if it is sufficiently severe, which depends on how much it restricts rather than expands a person's options, the extent to which it is guided by independent standards and ends, rather than a person's own ones, the extent to which it steers a person into a particular option rather than enabling him to make a better decision on his own, and the extent to which the circumstances cannot justify considering a person's judgment temporarily inferior (e.g. pain in emergency cases may justify such a judgment). A physician's attempts to convince his patient that a certain option is best for him, i.e. what Savulescu called "rational non-interventional paternalism"[35], are not paternalistic so as to violate *Equal Moral Standing.*

Equal Moral Standing promotes respect for autonomy as it respects a person's entitlement to pursue his own conception of the good. *Equal Moral Standing* promotes

[30] Quong, 2011, 80. Emphasis in original.
[31] Quong, 2011, 81.
[32] Quong, 2011, 82. Emphasis in original.
[33] Quong, 2011, 101. Emphasis in original.
[34] Quong, 2011, 101. See also Shiffrin, 2000.
[35] Savulescu, 1995.

beneficence as it acknowledges that people are often the best judges of their own good and exercising this capacity itself promotes well-being.[36]

Equal Moral Standing differs from *Assisted Decision-Making* in three main respects. Firstly, *Equal Moral Standing* does not require the physician to provide assistance when a patient is pressured by others but only focuses on the restriction of influences that the physician himself exerts. Secondly, *Equal Moral Standing* is specifically concerned with influences that aim to *benefit* a patient rather than influences in general. It thereby enables ICCJ to separate *beneficent* influences, i.e. influences that aim to benefit the consenter but could be morally wrong due to being paternalistic, from *adversarial* influences, i.e. influences that do *not* aim to benefit the consenter at all and could be morally wrong due to steering someone in a direction contrary to his interests. This distinction is relevant as the former have greater prospects of being interpersonally justifiable, given that only they aim to promote the interests of the person *to whom* an influence needs to be *justifiable.* Thus, ICCJ can potentially treat beneficent influences more leniently than adversarial influences. And thirdly, *Equal Moral Standing,* unlike *Shared Decision-Making,* explains why certain psychological states, e.g. great reluctance, will at least sometimes fail to vitiate consent: they fail to do so when a person pursues his conception of the good as being an integral part of his autonomy.

Trusted Partner

Finally, *Trusted Partner* is the principle that obliges the physician to honour his patient's trust, i.e. the patient's reliance that he will satisfy certain norms when the patient cannot control or monitor him. Honouring a patient's trust requires adherence to two central norms. The first norm, as I argued elsewhere, is:

[36] Dworkin, 1988, 112.

"the *proviso of hypothetical transparency*. Even though a trusting person refrains from 'the precise spelling out of terms' (Greenspan 2003, 161) which the trusted person may operate on, trust always comes with the proviso that the trusted party's relevant actions as well as 'the motives of the trusted party have to be able to be made transparent to the one who extends trust' (Ibid.). This proviso is respected only if full transparency did not lead to the trusting person being justified in feeling resentment or insulted. As Greenspan claimed, 'appropriate trust is trust that would survive a certain kind of knowledge on the part of the person who extends trust' (Ibid.)."[37]

The norm of transparency will be particularly important in cases of manipulation because some manipulative influences are completely hidden from the patient and such non-transparency is precisely what makes them problematic.

The second central norm of trust is what I call *loyalty in power,* which requires that the physician, independent of aspects concerning transparency, never abuses his position of power. This norm applies to the *professional* power that he has *qua physician*, e.g. the power to use confidential information in different ways or the power to control a patient's access to medical care. The norm also applies to the *factual* power that a physician has over his patients due to his asymmetrical relation to them in terms of medical knowledge and state of health. For instance, a physician would abuse his factual power by shaming an already psychologically vulnerable patient. *Loyalty in power* is an important norm because patients need to be certain that their physicians are trustworthy when acting from their much stronger positions.

The principle *Trusted Partner* is violated when either of these two norms is breached, which is what happens when the physician uses non-transparency or his superior position of power to dominate rather than support a patient's decision-making. The more the physician relies on non-transparency or power to dominate the patient's

[37] Kiener, 2018.

decision-making, the more serious the violation of *Trusted Partner.* Moreover, a violation becomes more serious when the patient is in a particularly weak position.

Trusted Partner promotes respect for autonomy because trust has strong instrumental value for autonomy. It facilitates co-operation with medical personnel as an important option for pursuing one's conception of the good.[38] If a patient could not trust his physician, the complexity of medical matters would make it impossible for those without medical training to make decisions. Moreover, trust has also intrinsic value as honouring the trust of another person is part of what it means to respect that person.[39] *Trusted Partner* promotes beneficence as trust fosters co-operation conducive to the patient's well-being, promotes the patient's adherence to medical procedures, promotes valuable rapport, and produces immediate positive side-effects for health outcomes.

Trusted Partner differs from *Assisted Decision-Making* and *Equal Moral Standing* in two ways. Firstly, *Trusted Partner* introduces the abuse of non-transparency and positions of power as additional considerations that can make an influence particularly wrongful. Secondly, *Trusted Partner,* unlike *Assisted Decision-Making* and *Equal Moral Standing,* specifically links to the asymmetry between patient and physician, the patient's vulnerability, and the patient's lack of control over many aspects of his medical situation.

Thus, to conclude: ICCJ bases interpersonal justification on the principles *Assisted Decision-Making, Equal Moral Standing,* and *Trusted Partner,* as derived from a physician's obligation to respect autonomy and act beneficently. Accordingly, consent is

[38] See Castelfranchi and Falcone, 2010, Ch. 7. McLeod, 2015.
[39] See C. O'Neil, 2012. For a detailed discussion of the relation between autonomy and different aspects of trust in bioethics, see O'Neill, 2002, particularly 15-27.

voluntary if and only if the physician's conduct in obtaining consent is in accordance with *all* of these principles. A violation of *one* them is enough to vitiate consent.

3.2 Clinical Research

Up to now I have explained ICCJ with reference to the context of treatment and focused on physicians and patients. I will now address the application of ICCJ to the context of research and the relationship between investigators and participants in a clinical trial.

Treatment and clinical research are fundamentally different. Treatment offers personalised care, whereas clinical research aims to produce generalizable knowledge. Yet, when it comes to consent, ICCJ holds that the principles fundamental to the consenter-consentee relationship are roughly the same in both contexts.

To explain ICCJ's view, let me begin with the principle of beneficence, where the contexts of treatment and research seem to differ most from each other. Admittedly, investigators do not have the same obligations of beneficence as physicians. Unlike physicians, investigators test new medicines and therapies and may subject people to procedures that are not designed to benefit them at all. However, investigators still remain under certain obligations of beneficence.

Conducting clinical research is a privilege accorded to investigators, firstly, by the social consensus that some clinical research ought to be permitted for the sake of scientific progress and, secondly, by those individuals who serve as participants.[40] This privilege is accorded on the understanding that exposing participants to risks of harm for the benefit of future patients is inherently problematic and must therefore be kept to a minimum. It comes with the restriction that procedures exceeding a certain threshold of harm or risk

[40] See Pellegrino, 1992, 362.

must not be performed under any circumstances and this restriction is an obligation of beneficence: investigators are obliged to protect their participants' health within the scope of clinical research.

Moreover, if it is in the investigator's power to prevent harm or promote well-being without sacrificing scientific objectives of considerable importance (or incurring considerable costs), he remains under a moral obligation to treat individuals beneficently. Only by acting beneficently in such situations can investigators do justice to the aforementioned proviso and appropriately respect those individuals who give up or consider giving up legitimate moral claims to safety for the benefit of others.

Finally, other aspects of beneficence, as explained by ICCJ, hold equally for treatment and research. These include the investigator's obligation to help a person to reach a decision that takes into account that person's well-being, the obligation to respect a person as being in charge of his life and decisions, the obligation to help a person to reach a decision that aligns with his values and preferences, and the obligation to promote a valuable relationship to participants.

Thus, aside from the lowered requirement concerning the protection of a person's health, investigators are still under obligations of beneficence.

Furthermore, the obligation to respect autonomy holds equally in the context of treatment and research. Just like physicians, investigators need to respect a person's right to decide which medical procedures to undergo, if any. And given that clinical trials are not designed to benefit those who participate, the importance of respecting a person's decision whether to enrol seems particularly important.

Thus, the more specific principles of ICCJ also apply to research. *Assisted Decision-Making* also obliges investigators not to control their participants' decision-making and to provide help when people have to make their decision under pressure.

Equal Moral Standing obliges investigators to respect their participants' entitlement to plan, revise, and pursue a conception of the good. However, in the context of clinical research, *Equal Moral Standing* may apply only indirectly. An investigator will not try to persuade people to undergo procedures that benefit them and, therefore, *Equal Moral Standing* will not apply to the investigator's influences. However, *Equal Moral Standing* obliges an investigator to tolerate other influences when they are consistent with the person pursuing his conception of the good.

Finally, *Trusted Partner* obliges investigators to honour their participants' trust. In fact, trust may become particularly important in the context of research where participants' interest potentially conflict with scientific interests. Here, it is vital that participants can trust that investigators will not abuse their position of power and always protect participants' interests as required in clinical research.

Hence, save for a slight modification of beneficence, ICCJ's principles apply equally to the context of research and treatment. But to be clear, ICCJ does not challenge the view that research and treatment may be governed by different norms in many other respects. ICCJ will always remain neutral with regards to the central aspects of an alleged difference between research and treatment, such as the permissibility of placebos, the randomised allocation of procedures, the need to stop clinical trials early in certain cases, the obligation to provide post-trial access to medication, the obligation to inform subjects about incidental findings, the duty to provide ancillary care, and the permissibility of medication wash-out periods, etc.[41] It is noteworthy, however, that even here the distinction between research and treatment remains contentious.[42]

[41] See Resnik, 2009, who gives an overview of the topics discussed.

[42] Admittedly, this distinction and the view that different norms apply in each area find support in the published literature. See F.G. Miller together with various colleagues. Emanuel, 2000. Levine, 1986. However, others also challenge this view: (i) some openly argue that norms of treatment apply equally to settings of research. See Jonas, 1969. Freedman, 1987. Miller and Weijer, 2006 & 2007. (ii) Others argue that neither the complete unification of treatment and research nor its complete separation is defensible and

In fact, ICCJ only focuses on the requirements of informed consent and, in this regard, claims that the norms and principles are roughly the same in research and treatment. With this scope in mind, such a claim is hardly novel and is indeed widely accepted. Even those who argue most vehemently for a distinction between research and treatment accept very similar norms when it comes to informed consent, either implicitly[43] or explicitly, like Nelson-Marten:

"A common feature of informed consent in clinical practice [i.e. treatment] and clinical research is that its ethical foundation is respect for human dignity and individual autonomy. Furthermore, the essential nature of informed consent does not materially differ between the two areas."[44]

Finally, this view still allows ICCJ to acknowledge that informed consent may have developed differently in the contexts of research and treatment[45], that it may face different challenges in each context[46], and that it may have different means of enforcement.[47]

argue for a more nuanced approach, e.g. a contextual approach (Resnik, 2009), a partial entrustment model (Richardson, 2004), or a capacity-based approach (Bright, 2012). (iii) Yet, others call into question whether, as a matter of fact, there are such differences between research and treatment in the first place. Kass (2013) and Wendler (2016) point out how clinical practice and research become more and more intertwined, e.g. in the case of learning health care systems or the training of junior doctors.

[43] See Morreim, 2005. F.G Miller and colleagues. Emanuel, 2000. Grady, 2012.

[44] Nelson-Marten and Rich, 1999, 81. See also Berg, 2001, 279. Mazur, 2012, 1. Boddington, 2012, 178.

[45] Whereas informed consent in the context of treatment developed as an answer to medical paternalism and was initially defended in courts, informed consent in clinical research developed as an answer to abuse and was initially presented in international documents such as the Nuremberg Code.

[46] Whereas some challenges arise only in clinical research (e.g. therapeutic misconception and payments for undergoing a procedure), others arise mainly in the context of treatment (e.g. paternalism).

[47] Whereas courts review informed consent in the context of treatment *retrospectively*, ethics committees also review informed consent in clinical research *prospectively*.

3.3 Satisfying the Desiderata

After explaining ICCJ and its fundamental principles, both in the context of treatment (3.1) and research (3.2), let me now return to the desiderata from Chapter 2 and show how ICCJ satisfies them.

ICCJ's focus on interpersonal justification does not imply any view about causation or counterfactuals and can therefore be easily combined with my views from Chapter 2.1. In fact, ICCJ's claim that consent is involuntary if and only if it was *motivated* by a non-justifiable influence allows ICCJ to align neatly with my view on a person's *reasons or motives* for consenting. It is also worth emphasising that consent is motivated by an influence in this sense when consent is either wholly or partly due to that influence. An influence does not need to be the sole cause to motivate consent.

Furthermore, ICCJ allows the moral legitimacy of an influence to have a bearing on the voluntariness of consent: certain morally illegitimate influences cannot be interpersonally justified and vitiate consent for that reason. Moreover, ICCJ tracks the more specific results reached in 2.2:

- interpersonal justification can be lacking even if an influence had only a low impact on a person's decision-making (i.e. even if consent was only *partly due to* that influence) and, therefore, some such influences may vitiate consent, as described in 2.2.1;
- interpersonal justification focuses on the justification *to the person giving consent* and is thereby well-positioned to capture instances of *wronging* a person rather than *wrongdoing* more generally, as explained in the *Mirror Thesis* in 2.2.2;

- ICCJ's specific principles *Assisted Decision-Making, Equal Moral Standing*, and *Trusted Partner* focus on the pressure that influences exert and thereby link to the right to bodily integrity, as also explained in 2.2.2;
- ICCJ's specific principles describe various degrees of seriousness of influences and thereby explain when influences become non-trivial, as required by the third condition of my *Mirror Thesis* in 2.2.2;
- interpersonal justification may be lacking in cases of threats *and* offers, and is therefore broad enough to account for my conclusion at the end of the section on blackmailing;
- interpersonal justification concerns what is *owed to* another person, is specifically based on the consentee's obligations, and therefore provides an intensionally adequate view of consent-vitiating influences, as explained in 2.2.3.

Finally, ICCJ accounts for my claims in Chapter 2.3 and makes the person receiving consent central to its approach. ICCJ explicitly focuses on the person receiving consent and requires that he be able to provide an interpersonal justification for the influences on the consenter's decision-making.

Hence, unlike any other view in the literature as discussed in Chapter 2, ICCJ neatly aligns with all the desiderata and ought therefore to be adopted.

Conclusion Part One

Chapter 1 highlighted important historical connections and defined the central terms of this inquiry. Chapter 2 developed the desiderata for any satisfactory account of voluntary consent. Chapter 3 developed my own proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* and supported it on the basis that it satisfies all these desiderata. Over and above presenting a novel contribution to the voluntariness of consent, this inquiry has already made important contributions to a post-volitional understanding of voluntariness: it explains in greater detail a negative causal proposal as also employed by *Ethical Models* in the 20th century (2.1), elaborates on how the voluntariness of a person's action, e.g. consent, can depend on the morality of another person's action (2.2), and specifies the relation between voluntariness and coercion in cases of consent (2.3). These results conclude *Part One* and my theoretical or *Top-Down* defence of ICCJ. In *Part Two*, I will employ a *Bottom-Up* approach and test ICCJ against specific practical challenges.

Part Two: The Practice of Voluntary Consent

In Part Two, I will address three practical challenges to the voluntariness of consent and argue that my proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* provides a view that is not only distinct from but also superior to other accounts in the literature.

In Chapter 4, I will consider nudging and manipulation: medical personnel use insights from behavioural psychology to exploit certain psychological mechanisms in people's decision-making.

In Chapter 5, I will focus on payments: investigators pay for participation in a clinical trial and thereby incentivise people to undergo procedures that do not provide any medical benefit for them. Here, I will return to Rambha, who I mentioned at the very beginning of this thesis.

Finally, in Chapter 6, I will discuss the pressure in living organ donation: a person may experience moral, implicit social, or outright third-party pressure when he considers whether to donate an organ to a family member.

These topics represent some of the most prominent challenges to the voluntariness of consent and cover a wide variety of different influences on a person's decision-making. Each chapter will start from a real-life case, and to remind the reader: when an influence vitiates consent in my examples, it will always do so by negating the voluntariness of consent, unless stated otherwise.

4. Non-Persuasive Influences: Nudging and Manipulation

34 million people are currently infected with HIV, but only half are aware of their status.[1] The other half is ignorant and unable to ensure they do not spread the disease, or even just protect their own health. Unfortunately, in some regions of the world, still far too few people consent to HIV-screening and with each passing day people die.[2] So, how could we increase HIV-screening and thereby save lives?

Montoy (2018) recently highlighted a powerful response to this challenge. He no longer asked patients whether they wanted to have HIV-screening on top of their other blood tests. He just made HIV-screening the default option. But this change from an opt-in system ('you can have the test if you like') to an opt-out system ('you will be tested unless you decline', where declining only required a brief 'no') had a considerable effect: it increased testing by almost 24%, more than any of the financial incentives commonly used.[3] Other studies have found an even greater impact of default settings.[4]

On the one hand, Montoy's default setting seems innocuous. Montoy neither changed the set of options nor did he make any option more costly to choose. In addition, *something* had to be the default option and if one of those options could have such positive effects, choosing it seems to be justified on grounds of beneficence.

On the other hand, Montoy's default setting seems problematic. The reason why 24% of people, who would otherwise *not* have consented, eventually consented was not because they had better reasons to do so. Montoy took advantage of a feature of their psyche called the *status quo bias*, i.e. people's inherent tendency to accept the default option rather than deviate from it.[5] Although consent was not irrational, it was certainly

[1] See UNAIDS World Aids Day Report 2012, 6.
[2] See Thornton, 2008.
[3] See Montoy, 2018.
[4] See for instance Johnson and Goldstein, 2003.
[5] See Kahneman and Tversky, 1972. McKenzie, 2006. Dean, 2017.

a-rational in the sense that the reason for it (i.e. that they were simply too lazy to depart from the default) could no longer count as a justification, even from their own point of view. Hence, Montoy targeted people's psychology rather than engaging in an open discussion. He steered them in a particular direction.

Montoy's case illustrates a serious challenge for consent. Often, people make medical decisions that are regrettable for some reason: e.g. they decide not to undergo what is in their interests and conducive to public health (e.g. HIV screening, vaccination), they decide not to undergo optimal medical treatment, they decide not to enrol in clinical trials that are urgently needed for scientific progress, or they decide not to allow junior doctors to examine them, whose training may as a result be impeded.

However, in all these cases, people could be subtly influenced into making a different decision. Fifty years of behavioural psychology have revealed several ways to exploit different mechanisms in people's psyches. Most recently, Thaler and Sunstein developed what they call 'nudging' to explain how people's decision-making can be influenced without resorting to either rational persuasion or coercion.

But this situation creates an urgent question: do people still voluntarily consent to a medical procedure if others harness their psychological mechanisms or flaws?[6] Answering this question is critical. We cannot afford to leave to chance how people's decision-making will be affected, given how effective and ubiquitous those influences are. We need to protect people's autonomy, particularly in a medical context where people are already vulnerable and, vis-à-vis medical professionals, find themselves in relations of great asymmetry. But we also need to explore the great potential which influences like Montoy's provide, i.e. the potential to save lives.

[6] To clarify, many of the psychological mechanisms under consideration are not just flaws but may in fact have a good evolutionary justification: having them could be beneficial overall. However, if they affect people's decision-making in certain circumstances, they render decisions flawed.

In this chapter, I will examine the question of how influences like Montoy's, which I will call *non-persuasive influences*, affect the voluntariness of consent. Defending my proposal of *Interpersonal Consenter-Consentee Justification* (ICCJ), I will explain that such influences vitiate consent if and only if the physician (or investigator) cannot interpersonally justify to the patient (or trial participant) that he exerted them. I will proceed as follows: after explaining non-persuasive influences in greater detail (4.1), I will show that the widely adopted criteria for evaluating them are misguided (4.2). I will then present my proposal ICCJ as an alternative and superior approach (4.3).

4.1 Behavioural Psychology

Human behaviour is characterised by bounded rationality, i.e. the indelible tendency of humans to behave either ir-rationally or a-rationally. People behave *irrationally* when they act contrary to what they think is best, as in cases of weakness of will, or when they are subject to cognitive biases causing them to decide contrary to basic norms of logic. People behave *arationally* when the reason for their action is no longer a justification for their action. To come back to Montoy: although people's consent to Montoy's default setting was not irrational, it was certainly arational in the sense that the reasons why people consented (i.e. due to inertia) could no longer justify their decision.

However, people's behaviour is not carved in stone. By arranging the "context in which people make decisions",[7] Thaler and Sunstein argue, we can influence people's choices in different directions. Sunstein and Thaler call this *choice architecture*.[8] The central item in such a choice architect's toolbox is a 'nudge', a form of a non-persuasive influence, which is defined as:

[7] Thaler and Sunstein, 2009, 3.
[8] See Thaler and Sunstein, 2009, 19-109.

"Any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives."[9]

Therefore, Montoy's default option was a nudge. Without forbidding any options or significantly changing their incentives, he changed the situation so that people were more likely to consent to HIV-screening. But as mentioned earlier, Montoy is only the tip of the iceberg. Here are some further characteristics of human decision-making, alongside a quick guide on how to exploit them.

Loss Aversion. People's psychological response to an option varies depending on whether it is presented as a risk of loss or as a chance for gain. This is because the thought of losing something causes people to feel psychological discomfort, the intensity of which can be considerably higher than the intensity of the pleasure they feel when considering obtaining that same object. *Framing* exploits this tendency. Explaining to a patient that a procedure has a 90% success rate rather than a 10% failure rate makes the patient more likely to consent to it, even though – from a logical point of view – there is no difference between 90% success and 10% failure.[10]

Availability Heuristic. Whether a person thinks an event is likely to occur, or how important he considers an aspect to be, depends on how readily that person can think of examples of such an event or aspect.[11] If a patient can easily think of an example of what could happen without treatment, he will consider the potential consequences of non-treatment more likely to occur and more important for his decision, even if they are not. *Salience* and *Priming* can exploit this tendency. *Salience* means drawing a person's attention to certain aspects, e.g. by using vivid images or videos which have proven to be more effective than simple verbal communication. *Priming* means exposing people to certain stimuli that increase sensitivity to certain aspects in

[9] Thaler and Sunstein, 2009, 6.
[10] See Kahneman and Tversky, 1979. Kahneman et al., 1991. Ericson and Fuster, 2014.
[11] See Tversky and Kahneman, 1974. An, 2008.

subsequent and even unrelated tasks. For instance, a patient reading a family magazine in the waiting room will be more sensitive to the impact of his illness on his family than he would have been without prior exposure to the magazine. As it will be easier for him to think of specific examples of how his family could be affected, such potential consequences will appear to him more likely and important.[12]

Affect Heuristic. People's decisions depend on whether they have good or bad feelings in relation to an influence or decision.[13] Such feelings or emotions can be triggered by using certain terminology, various means of expression, or addressing the patient in a particular way. Consider a nurse flirtatiously asking a patient who is unwilling to take his medication: "You won't make me beg you, will you?"[14]

Anchoring Heuristic. Exposure to a certain reference point, e.g. a number, highly influences people's judgments.[15] Before initiating a discussion on real estate, researchers asked participants to multiply the last three digits of their mobile phone number by 1000 (e.g. 456 x 1000 = 456,000). People's later estimation of housing prices significantly correlated with this arbitrary anchor, even though they were also provided with correct information about the housing market.[16] In a similar vein, a physician may ask a patient to consider a very risky procedure, thereby making this the anchor of their discussion. A patient's subsequent consideration of a less risky procedure may then be highly influenced by this anchor.

Confirmation Bias. People are more likely to accept evidence if it fits with the beliefs, preferences, or intuitions they already have. For this reason, people's appreciation of information can greatly vary and, in some cases, can be deeply flawed.[17] For instance, a patient who is already

[12] See Murphy and Zajonc, 1993.
[13] See Slovic, 2007.
[14] See Barnhill, 2014, 56.
[15] See Furnham and Boo, 2011.
[16] See Tversky and Kahneman, 1974.
[17] See Oswald and Grosjean, 2004.

disinclined to undergo a particular treatment may only look for evidence against such treatment rather than aim for a balanced assessment of his situation.

Hot Cold Empathy Gap. People' decision-making in a state of emotional arousal (e.g. anger, pain, hunger) is very different from their decision-making in a relaxed state of mind. For instance, a patient's decision may vary depending on whether he has to make a treatment choice immediately after learning he has cancer or after a cooling-off period.[18] Asking a patient to make a decision at a specific point can therefore have a great impact on that very decision. Unfortunately, people normally fail to understand how great the differences in their decision-making can be.

Herd Behaviour. People sometimes simply follow others instead of making up their own minds. To exploit this tendency, clinical investigators may organise group sessions for informing potential participants to make sure that the first few people who speak say they will enrol. Others who would otherwise not have consented may then follow.[19]

Spotlight Effect. People usually overestimate how much others notice what they are doing and they are often overly concerned with what others think about them. For this reason, making a person aware of how he appears to others can greatly influence his behaviour. For instance, when people are surrounded by mirrors and see themselves as others see them, they often change their behaviour. Placing mirrors in restaurants makes people eat less; placing them in gyms makes people exercise more. In the clinical context, mirrors can be used to reinforce a person's negative feelings about certain parts of his body. Moreover, a patient's decision could also be influenced by making him aware of what colleagues or friends may think about him.[20]

---

[18] See Loewenstein, 2005. To clarify, a cooling-off period is the period *before* a person actually gives consent. In order to avoid confusion, it is worth pointing out that such an understanding diverges from some other contexts where a cooling-off period means the period *after* a person has consented but still retains the right to cancel a contract without penalty.
[19] See Banerjee, 1992. Mandava and Millum, 2013.
[20] See Gilovich, 2000.

Representative Heuristic. People are more likely to consider something to be 'X' if it resembles or represents what they think is the stereotype of 'X'. For instance, people are more likely to consider someone a criminal if that person looks like the stereotype of a criminal.[21] However, as this example illustrates, resemblance to a stereotype can be a poor indicator in individual cases. In the clinical context, a person may not believe she is ill if her situation does not resemble the one of the stereotypically ill person, or she may be more likely to accept some treatment when someone else, whose situation closely resembles her own, is shown to have benefitted from that treatment.

This list of psychological mechanisms and ways of exploiting them is far from being exhaustive, but it should suffice to illustrate the topic of this chapter. Before I proceed, however, I need to mention that I will not aim for a sharp definition of nudging or those *non-persuasive* influences (as I will call them), which exploit certain psychological mechanisms. Rather, I will rely on a certain family resemblance among influences that are neither persuasion nor coercion and I will also include instances of manipulation more generally. Such a wide approach enables me to capture many different non-persuasive influences that could affect a person's decisions. On this basis, I will discuss non-persuasive influences in the context of individual clinical encounters and address the question of whether or under what conditions they vitiate consent.

[21] See Kahneman and Tversky, 1972. Krawczyk and Rachubik, 2019.

4.2 Easy Resistibility, Transparency, and Rationality

4.2.1 The Literature on Non-Persuasive Influences

The main ethical concern about the use of non-persuasive influences is whether such influences undermine the influencee's autonomy. Whereas some have denied that they do by pointing out how nudges can preserve substantial freedom of choice,[22] others have claimed that the influencee's decision-making will be subverted and dominated by another person.[23] However, most of the accounts in the literature on non-persuasive influences neither fully reject nor fully embrace all non-persuasive influences. Rather, they endorse a more nuanced position and claim that non-persuasive influences can be permissible under certain conditions. Amidst the numerous proposals, the debate has moved towards a certain set of conditions. Very often, non-persuasive influences are considered permissible insofar as they are (I) easily resistible, (II) transparent, and (III) preserve rationality. Applied to Montoy, i.e. the HIV-test example from the beginning of this chapter, one can argue that Montoy's default setting is legitimate insofar as it was (I) easily resistible: people could just have said 'no', (II) transparent: Montoy told patients what the default was and in a modified version he may even reveal his intentions behind it, and (III) it did not trigger bluntly irrational reasoning: if anything, it may just have been a-rational. I will now elaborate on these conditions in greater detail before discussing them in their application to consent.

[22] See for instance Saghai, 2013.
[23] See for instance Ploug and Holm, 2015.

I. Easy Resistibility

Sunstein and Thaler claimed that "[t]o count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates."[24] Sunstein and Thaler's remark was highly influential and structured the debate on the permissible use of non-persuasive influences. However, it also left a lacuna. As Sunstein and Thaler admitted,[25] they could not spell out in greater detail what is required for nudges to be "easy and cheap to avoid."[26]

Saghai filled this gap. He started from the claim that influences on a person's decision-making need to be *substantially noncontrolling*, where "A's influence to get B to φ is substantially noncontrolling when B could easily not φ if she did not want to φ."[27]

Saghai then added that "[r]esistibility is a criterion for testing the degree to which an influence is controlling",[28] and claimed that one can "elaborate on a criterion of easy resistibility to ascertain substantial noncontrol."[29] Nudges are "easily resistible"[30] when the person nudged possesses the

(1) "capacity to become aware of A's pressure to get her to ϕ (attention-bringing capacities)",[31] and the

(2) "capacity to inhibit her triggered propensity to ϕ (inhibitory capacities)",[32] and

(3) "is not subject to an influence, or put in circumstances that would significantly undermine the relatively effortless exercise of attention-bringing and inhibitory capacities."[33]

---

[24] Thaler and Sunstein, 2009, 6.
[25] Thaler and Sunstein, 2009, 249.
[26] Thaler and Sunstein, 2009, 6.
[27] Saghai, 2013, 488.
[28] Saghai, 2013, 489.
[29] Saghai, 2013, 489.
[30] Saghai, 2013, 489.
[31] Saghai, 2013, 489.
[32] Saghai, 2013, 489.
[33] Saghai, 2013, 489.

"To qualify as 'relatively' effortless, the influencee should not be expected to have acquired unusually high skills enhancing her attention-bringing and inhibitory capacities."[34] So, in a nutshell, Saghai's condition of easy resistibility comprises the capacity to become aware of an influence and both the ability and the opportunity to act against it.

Proponents of nudging either follow Saghai explicitly[35] or at least focus on the same aspects.[36] Yet, also the critics of nudging focus on those aspects which Saghai highlighted and argue that nudging can never be easily resistible.[37]

Thus, easy resistibility is widely accepted as a condition for the permissible use of nudging. However, some authors still depart from Saghai's view. For instance, Cohen focuses on the extent to which a nudge changes a person's subsequent behaviour and claims that nudges are less problematic when they cause only slight changes in a person's behaviour.[38] This is different from Saghai's view because a nudge may be irresistible in Saghai's sense, yet only slightly change a person's behaviour compared to what he would otherwise have done. However, Cohen's view is related to Saghai's because Cohen is also concerned with spelling out a condition of substantial non-control.

II. Transparency

In addition to easy resistibility, Thaler and Sunstein also introduced a requirement of transparency. When discussing public policy, Thaler and Sunstein argued that governments must be prepared to defend their nudging practices in public and give

[34] Saghai, 2013, 490.
[35] See Engelen, 2019a. Engelen, 2019b.
[36] See Blumenthal-Barby, 2012b, 353.
[37] See Bovens, 2009, 209-210. Hausman and Welch, 2010, 128. Wilkinson, 2013, 345.
[38] See Cohen S., 2013, 8.

citizens a chance to object.[39] Many followed Thaler and Sunstein in emphasising transparency in one form or another, e.g. transparency about every particular nudge, transparency about nudging being generally practiced, or even transparency about the intention behind the nudging.

Proponents of nudging accept the need for transparency. Saghai talks about people's capacity to become aware of an influence. Hansen explicitly claims that "the distinction between transparent and non-transparent nudges (…) serves as a basis for distinguishing the manipulative use of nudges from other types of uses."[40] Finally, Schmidt claims nudges need to be "sufficiently transparent."[41]

On the other side, critics of nudging also refer to transparency. They object to nudges because they are often not transparent or not transparent enough. Dubov and Greenspan claim that the person being non-persuasively influenced may be totally unaware of being influenced by another person at the time the influence is exerted. In such cases, non-persuasive influences keep, as Greenspan argues, certain aspects "inexplicit – off the bargaining table, where they might be subjected to scrutiny and disagreement."[42] Non-persuasive influences are a form of "'masked' agency",[43] which "precludes the possibility of a relationship."[44] Furthermore, Bovens argues that "every Nudge should be such that it is in principle possible for everyone who is watchful to unmask the manipulation"[45] and that influences are "impermissible because they do not satisfy in principle token interference transparency."[46] *Token transparency* requires that a person be informed that nudging is being practiced *as well as* when and how she is

[39] See Thaler and Sunstein, 2009, 243 ff.
[40] Hansen, 2013, 6.
[41] Schmidt, 2019, 533. See also Blumenthal-Barby, 2012b, 347 & 357. Engelen, 2019a, 52.
[42] Greenspan, 2003, 163.
[43] Greenspan, 2003, 158.
[44] Dubov, 2015, 497.
[45] Bovens, 2009, 217.
[46] Bovens, 2009, 217.

being nudged. *Type transparency*, i.e. bringing to people's attention that nudging is generally practiced without informing people when and how they are being nudged, would not be sufficient.[47]

Hence, transparency is widely accepted as a condition for the permissible use of nudging.

III. Rationality

Finally, Thaler and Sunstein claimed that nudging people into irrational choices is "irresponsible"[48] and thereby introduced a condition of rationality for the permissible use of nudging.

Critics of nudging accept such a condition but remained sceptical as to whether nudges can ever satisfy it. Bovens objects to nudges on the grounds that they make one "irrational (…) insofar as what is driving my action does not constitute a reason for my action – i.e. it is not a feature of the action that I endorse as a feature that makes the action desirable."[49] Conly claims that when we nudge other people, "[r]ather than regarding people as generally capable of making good choices, we outmaneuver them by appealing to their irrationality."[50]

Proponents of nudging, on the other hand, also accept the condition of rationality, but support Thaler and Sunstein on this front. Schmidt claims that "nudging not only is compatible with rational agency but can even support it."[51] Cohen claims that "where

[47] See Bovens, 2009, 216-217. Further critique on the basis of a lack of transparency comes from Hausman and Welch, 2010. Wilkinson, 2013. Calo, 2014, 795. Grüne Yanoff, 2016, 177.
[48] Thaler and Sunstein, 2019, 137.
[49] Bovens, 2009, 210. Note that Boven calls 'irrational' what would be 'a-rational' according to my previous definition.
[50] Conly, 2012, 30. Further critics employing a rationality objection include Hausman and Welch, 2010, 134. Grüne-Yanoff, 2012, 636-637. Waldron, 2014. Ploug and Holm, 2015, 34-35, 37.
[51] Schmidt, 2019, 511.

nudging does not encourage irrationality (…) the force of criticism [against nudging] is neutralized."[52] And Hansen argues that nudges are permissible when they do not induce faulty heuristics or encourage irrationality on the part of the person being nudged.[53]

Hence, people widely accept a condition of rationality for the permissible use of nudging. This condition focuses on process rationality rather than outcome rationality. A person demonstrates *process rationality* if and only if the "decision-making procedure should yield accurate results somewhat reliably and not just accidentally."[54] By contrast, a person demonstrates *outcome rationality* if and only if, as a matter of fact, he ends up doing what a rational person would do. Whether the process leading to that result was rational too is immaterial to outcome rationality.

4.2.2 Critique of the Literature on Non-Persuasive Influences

People in the literature on non-persuasive influences widely agree that the three conditions of easy resistibility, transparency, and rationality determine the *general permissibility* of exerting non-persuasive influences. When they focus on consent more specifically, they also suggest that these exact conditions, together with some minor adjustments concerning certain thresholds, determine when non-persuasive influences *vitiate* consent. In this section, I will oppose this view and claim that the conditions of easy resistibility, transparency, and rationality are neither necessary nor sufficient for the voluntariness and validity of consent, and therefore also fail as conditions for the permissible use of non-persuasive influences more generally.

---

[52] Cohen S., 2013, 5. See also Houk, 2016, 21. Engelen, 2019b. Schmidt, 2019.
[53] See Hansen, 2013, 14-15. See also Hanna, 2015, 622-624. Barton, 2015, 343-344.
[54] Schmidt, 2019, 512. This is a less demanding view. For more stringent requirements and other views see Schmidt, 2019. Engelen, 2019b.

I. Why the Conditions are not Necessary to Guarantee the Voluntariness of Consent

Consider the following case based on a real situation from clinical practice:[55]

Stubborn Betty is diagnosed with early stage cervical cancer. Her physician, who happens to be a person of colour, strongly recommends surgery, which would almost certainly cure her, but Betty refuses. She says, 'anyone knows that people with cancer are sick, feel bad and lose weight' and adds that she feels very well. She continues to believe that she is healthy despite clear evidence to the contrary. Betty has a racial bias which prevents her from trusting her doctor's statements. Betty is competent. There are no signs of any mental impairment.

Betty uses a representative heuristic. Since she does not look like the typical ill patient, she does not believe she is ill. Moreover, she is subject to a severe form of confirmation bias. Let me recall, "'[c]onfirmation bias' means that information is searched for, interpreted, and remembered in such a way that it systematically impedes the possibility that the hypothesis could be rejected."[56] Betty's 'hypothesis' is a deep racial bias, i.e. the view that black people are neither intelligent nor trustworthy on medical matters. To confirm this view, she interprets her black doctor's statement as false and prioritises her own feeling of being healthy in a way that systematically impedes the possibility of correcting her beliefs. Finally, Betty may even have been subject to what has been described as the 'backfire effect':[57] the more her black doctor presents evidence contrary to what she believes, the more she clings on to her own opinion.

So far, the story is true. But now, let us go beyond the facts and stipulate that Betty will only take the evidence seriously if the person talking to her is not only white, but also

[55] See Faden, 1977.
[56] Oswald and Grosjean, 2004, 80.
[57] See Nyhan and Reifler, 2013.

a member of her religious community and speaks with the accent from her hometown. Let us also assume that Betty would believe anything from such a person.

Suppose Betty's physician knows this and therefore asks a male nurse who meets these criteria to tell Betty that she needs to take the evidence very seriously. The physician thereby intentionally exploits Betty's confirmation bias. He makes Betty irrationally select information and confirm her false view that people like the nurse are to be trusted without questioning.

In this case, none of the aforementioned three conditions from the literature on non-persuasive influences would be satisfied.

Transparency. Betty is not at all aware that the medical team arranged the situation to exploit her confirmation bias and make her appreciate the medical evidence.

Easy Resistibility. The lack of transparency prevents Betty from becoming aware of the influence and triggering her propensity to act against it. Moreover, Betty is so indoctrinated and biased that she would believe or do anything that the nurse tells her.

Rationality. Betty lacks process rationality: her deliberation is not a reliable procedure by which to arrive at a rational decision. Even on the least demanding accounts of process-rationality, i.e. the views claiming that the "decision-making procedure should yield accurate results *somewhat reliably* and not just accidentally",[58] Betty's deliberation will be procedurally irrational. Trusting a person because of his race, membership of a community, and accent is not at all a reliable procedure in medical matters. It is only in manipulated circumstances that her bias will lead to an acceptable result. Therefore, even if Betty displays outcome rationality, i.e. appreciating the evidence is what a rational person would do, her deliberation is irrational.

[58] Schmidt, 2019, 521. Emphasis added.

Hence, Bovens will consider the influence on Betty clearly impermissible as it was completely non-transparent. Saghai will reject the influence given its irresistibility. And Cohen will have to reject the influence on Betty on the grounds that it encourages irrationality. Therefore, if one relies on the main criteria in the literature, the influence on Betty will be clearly illegitimate; and if these criteria also determine whether an influence vitiates consent, the influence on Betty may very well vitiate her consent too.

However, this conclusion seems too hasty. Exploiting Betty's confirmation bias enabled her to appreciate critical evidence and to deliberate over how to protect her values and interests. It neither worsened the procedural rationality which she would otherwise have been capable of nor did it compromise or unduly control Betty's decision-making. After all, the physician only aimed to make Betty appreciate the evidence and left her free to decide whether she wanted to have surgery in the end. Therefore, the non-persuasive pressure on Betty through exploiting her confirmation bias was neither illegitimate nor did it vitiate her consent. If Betty eventually consents to surgery, the physician's performance will be permissible and will not violate Betty's right to bodily integrity.

Hence, the three conditions in the literature deliver the wrong result: satisfying them is not necessary to guarantee the voluntariness of consent.

II. Why the Conditions are Not Sufficient to Guarantee the Voluntariness of Consent

To see why the three conditions are also not sufficient to guarantee the voluntariness of consent, consider the following example:

Social Cost Alan is publicly known for advocating the obligation to participate in clinical research. Yet, he is highly averse to undergoing medical procedures and has therefore never participated in a trial himself. Investigator Carl knows that Alan would never consent to enrol. He therefore asks Alan during a TV interview, where Alan will incur public embarrassment if he declines. Carl addresses Alan and adds that he is convinced that Alan is a remarkable person and can make such decisions publicly. Alan declares that he will participate.

Carl exploits the spotlight effect: he asks Alan during a TV interview and harnesses his concern about what others think of him. Carl uses the hot-cold empathy gap: he creates a *hot* situation, i.e. a situation of surprise and potential embarrassment, and asks Alan to make an immediate decision. Carl also plays on the affect heuristic*:* he triggers Alan's tendency to respond in ways that make him feel good, i.e. giving him the opportunity to win the approval of his audience rather than incurring public embarrassment. Finally, saying that he would not have asked if he did not consider Alan a remarkable intellectual implies an implicit emotional threat: if Alan does not agree to enrol, he may not be such a remarkable intellectual after all and should no longer be considered one.

In this case, the aforementioned conditions from the literature on non-persuasive influences are all satisfied.

Transparency. Alan, let us assume, knows exactly what Carl is trying to do and which nudges and forms of manipulation he employs. Hence, Carl's influence is fully token transparent.

Rationality. Alan's decision can certainly be outcome-rational. It may be that avoiding public embarrassment is more important to him than avoiding enrolment. However, Alan's decision can also be procedurally rational. Although the array of Carl's non-persuasive influences will have an a-rational impact on Alan's decision-making,

Alan is aware of them, reflects on whether to follow these influences, and thereby arrives at his decision in a way that is a reliable procedure for a rational outcome.

Easy Resistibility. But is Carl's influence also easily resistible? To recall, Saghai claimed that nudges are easily resistible when the person nudged possesses

(1) the "capacity to become aware of A's pressure to get her to ϕ (attention-bringing capacities)",[59]

(2) the "capacity to inhibit her triggered propensity to ϕ (inhibitory capacities)", [60] and

(3) "[she] is not subject to an influence, or put in circumstances that would significantly undermine the relatively effortless *exercise of attention-bringing and inhibitory capacities*",[61] and "[t]o qualify as 'relatively' effortless, the influencee should not be expected to have acquired *unusually high skills* enhancing her attention-bringing and inhibitory capacities."[62]

Alan satisfies (1): he certainly had the capacity to become aware of Carl's pressure and the intention behind it. In fact, everything was fully transparent to him. Alan also satisfies (2): he had the capacity to inhibit his triggered propensity to consent to enrolment. Alan was able to say no or postpone the discussion to a later date. Finally, condition (3) is also met: the situation did not "significantly undermine the relatively effortless exercise of attention-bringing and inhibitory capacities"[63] because, firstly, Alan's "attention-bringing and inhibitory capacities"[64] were not affected at all and, secondly, because he did not need "unusually high skills"[65] to refuse Carl. He only needed to say no. Therefore, following Saghai's definition very closely yields the result that Alan could have easily resisted.

[59] Saghai, 2013, 489.
[60] Saghai, 2013, 489.
[61] Saghai, 2013, 489. Emphasis added.
[62] Saghai, 2013, 490. Emphasis added.
[63] Saghai, 2013, 490.
[64] Saghai, 2013, 490.
[65] Saghai, 2013, 490.

I envisage Saghai objecting to an overly narrow reading of his definition of easy resistibility. Although Saghai emphasises that influences fail to be easily resistible when a person lacks certain skills,[66] he explains that they *also* fail to be easily resistible under

> "circumstances that undermine people's attention-bringing and inhibitory capacities (…) [namely] time pressure, stress (and stress-inducing poverty), submission to perceived authority, fatigue, anxiety, cognitive load and distraction."[67]

Thus, an influence also fails to be easily resistible if, however skilled people are, they could not withstand the influence relatively effortlessly, where "effortlessness is relative to what we can expect of normal individuals."[68] Hence, easy resistibility also depends on the degree of fortitude, resilience, psychological effort, or even kind of insensitivity required to act contrary to an influence.

On this reading, it may be that Alan could *not* have easily resisted Carl's influence as doing so might have required great psychological effort and resilience and, given how Carl put Alan on the spot, it was likely that Alan would succumb to Carl's influence. Hence, Saghai can object to my evaluation of *Social Cost.*

However, if Saghai opts for this response, he will commit himself to a false view. To recall, Saghai claims that "[r]esistibility is a criterion for testing the degree to which an influence is controlling"[69] and holds that an influence vitiates consent if it is sufficiently controlling. He therefore concludes that an influence vitiates consent if it is hard to resist. However, if resistibility is to be understood in terms of resilience and the

---

[66] To clarify: a person may be unable to detect certain non-persuasive influences, let alone easily resist them, unless he has systematically studied what these influences are and how they work, i.e. unless he has certain high skills.
[67] Saghai, 2013, 490.
[68] Saghai, 2013, 490.
[69] Saghai, 2013, 489.

psychological effort it takes to withstand it, resistibility cannot inform judgments about the voluntariness of the influenced person's action. Consider the following examples.

Suppose I offer you a bribe of one million GBP if you steal my neighbour's sofa. This offer may be hard to resist as it requires great resilience or insensitivity to financial gain. Although you have the capacity to become aware of my influence and to resist my offer, the circumstances may be such that you cannot, as Saghai puts it, turn me down relatively effortlessly. Yet, my bribe will *not* be substantially controlling or negate the voluntariness of your theft, let alone your liability to blame and punishment. However, if I threaten to beat you if you decline, things may stand differently, even if the degree of resistibility did not change at all and you happen to be a person who can resist the threat more easily than the bribe.

Similar examples apply to consent. Suppose I offer you one million GBP for a sofa and you really need the money to pay your mortgage. Such an offer is very hard to resist as it requires great resilience or insensitivity to financial gain. But your consenting to sell the sofa is neither substantially controlled nor involuntary. Contrast this case with another scenario where I threaten to beat you unless you sell the sofa to me. It may be easier for you to resist this threat and sue me for damages afterwards than to decline my previous offer. Yet, in this case, your consent to selling was forced, involuntary, and can be voided.[70]

Therefore, the degree of resistibility, understood in terms of resilience, cannot determine the voluntariness of an influenced person's acts. Hence, Saghai faces a dilemma: he must either accept that Alan could have easily resisted, on the skill-based reading, or reject that Alan could have easily resisted, on the resilience-based reading, but

[70] On resistibility in cases of threats and offers, see Wertheimer, 2013.

then commit himself to an account that is unable to inform judgments about the voluntariness of consent.

Having established this, let me now return to those authors proposing the three conditions outlined earlier and apply their statements to *Social Cost.* Those focusing on transparency may approve of Carl's influence. As Hansen pointed out: "the distinction between transparent and non-transparent nudges (…) serves as a basis for distinguishing the manipulative use of nudges from other types of uses."[71] Cohen, on the other hand, seems to approve of Carl's influence to the extent that Alan's response is fully rational, given his claim that "where nudging does not encourage irrationality (…) the force of criticism [against nudging] is neutralized."[72] And on the skill-based reading of resistibility, i.e. the only reading that could be relevant to the validity of consent, Carl's influence was easily resistible.

Hence, these three conditions suggest that Carl's influence is clearly legitimate and therefore also fails to vitiate Carl's consent. Considering Carl's influence to be clearly legitimate is certainly false. The fact that Carl's influence is transparent does not make it any less manipulative since some techniques of manipulation actually work very transparently.[73] The fact that Alan's response was rational does not seem to make the influence less problematic either. After all, coercion also often elicits rational responses. Finally, despite being easily resistible on the skill-based reading, Carl's influence made it costly for Alan to decline to participate. Carl deprived Alan of the opportunity to make a decision about trial participation privately, linked non-participation to the negative consequence of public embarrassment, and thereby reduced the options otherwise available to Alan. As a result, Carl controlled Alan's situation and, if such control can be

[71] Hansen, 2013, 6.
[72] Cohen, 2013, 5.
[73] See Gorin, 2014a and 2014b. Barnhill, 2014.

considered sufficiently serious, Carl's influence may not only be illegitimate but also vitiate Alan's consent. This is true independent of whether, on the resilience-based reading, Alan could relatively effortlessly withstand Carl's influence.[74] Thus, in order to evaluate cases like *Social Cost* appropriately, we need to distinguish between control and resistibility. Control corresponds to the reduction of a person's options and the harm that awaits him when he declines to consent, whereas resistibility describes a person's psychological state when he finds himself in such a situation. *Pace* Saghai, the former but not the latter determines the validity of consent.

Hence, the three conditions deliver the wrong result: easy resistibility, transparency, and rationality are not sufficient to guarantee the voluntariness of consent.

4.3 ICCJ on Non-Persuasive Influences

Applying the conditions of easy resistibility, transparency, and rationality yielded the wrong results in *Stubborn* and *Social Cost.* By contrast, ICCJ yields the correct result for both cases and offers a novel view on non-persuasive influences. In this section, I will explain and defend ICCJ as an alternative and more convincing approach.

In a nutshell, ICCJ holds that consent obtained by a non-persuasive influence is voluntary if and only if the physician can interpersonally justify to the consenter that he exerted the influence. Such justification requires the influence to be in accordance with certain principles fundamental to the consenter-consentee relationship. These principles are respect for autonomy and beneficence, and more specifically *Assisted Decision Making, Equal Moral Standing,* and *Trusted Partner.* An influence is in accordance with

[74] Alan's case could therefore be similar to the one I described in 2.2.1, where a physician's threat to abandon his patient vitiated the latter's consent even though it only had a low impact on that patient's decision-making. It vitiated consent because, like Carl's influence, it illegitimately constrained the situation of the person consenting.

Assisted Decision-Making if it assists a person in his deliberation without taking control of the process of decision-making. It is in accordance with *Equal Moral Standing* if the influence does not compromise a person's power to pursue and revise his own conception of the good. And it is in accordance with *Trusted Partner* if it satisfies the central norms of trust, i.e. the *proviso of hypothetical transparency* and *loyalty in power*.

Generally, an influence is in accordance with ICCJ's principles when it enhances a person's ability to make a reasoned decision and violates these principles when it undermines this ability, as the following explanations show.

4.3.1 Enhancing Decision-Making

A person's ability to make a reasoned decision can be enhanced by means of de-biasing, re-biasing and, in exceptional circumstances, by means of exploiting a particular bias. These means of influencing strengthen people's ability to make a deliberate choice for themselves rather than unduly control their process of deliberation. But they differ from each other in the way that they do this.

De-biasing enhances decision-making by reducing bias without triggering another bias. One can debias people by telling them that they are subject to a certain decision-making flaw such as a status-quo bias or loss aversion,[75] by employing a 'consider the opposite strategy', i.e. asking people to consider pros and cons of a certain decision,[76] by asking people to explain their decision,[77] or by employing a *Cooling Off Period* so as to ensure that current emotional or cognitive overload will not affect a person's decision. If Montoy had used such debiasing strategies, the effect of his default on the patient's decisions about whether to have HIV-screening may have been reduced considerably.

[75] Gelfand, 2016, 605-608.
[76] Trout, 2005, 418-419. See also Gigerenzer, 2015.
[77] Miller and Fagley, 1991. Ploug and Holm, 2015, 35.

Gelfand claimed that debiasing strategies could count as a 'meta-nudge'[78] as they constitute nudges that work on other nudges already in place (e.g. a default option). However, since at least some such debiasing techniques seem closer to an open discussion than to non-persuasive influences, they may be considered an alternative to nudging rather than an instance of it.

Sometimes debiasing is not very effective.[79] In such cases, *re-biasing* may be another option, i.e. the reduction of one bias by triggering another bias. For instance, suppose Walter comes to hospital with an acute leg injury. As he is in pain and very anxious about what will happen next, his decision-making is negatively affected. Due to his anxiety, he is subject to considerably greater loss aversion than usual and any hint about risks or negative consequences would make him even more anxious. Moreover, Walter is also in a state of cognitive and emotional overload. The sudden change of circumstances and arrival at the hospital are hard for Walter to cope with. He has become unable to deliberate calmly about which treatment to pick.[80]

How should the physician deal with such a situation? Debiasing strategies such as a consider-the-opposite strategy may not be optimal. They may only confuse Walter more and use up the scarce mental resources he has left to make a decision. By contrast, re-biasing may fare better. In order to balance Walter's exaggerated loss aversion, the physician could use framing. He could avoid terminology that exacerbates Walter's already heightened sensitivity to considerations of risk and present the information in terms of benefits instead. Furthermore, in order to reduce Walter's cognitive and emotional overload and to calm him down, the physician could trigger Walter's affect heuristic, i.e. people's tendency to be influenced by their emotions in response to an

[78] Gelfand, 2016, 601.
[79] See Ubel, 2005. Scherer, 2015. Halpern, 2016.
[80] Cognitive or emotional overload my lead to people avoiding decision-making altogether. See Iyengar, 2010.

interaction. The physician could deliberately speak in a slow and calm voice to make Walter feel reassured that he will be fine.[81] Such rebiasing strategies will enhance Walter's decision-making and enable him to make as good a decision as possible for him in this situation.

Or consider another re-biasing example: Laura is 65 years old and was just diagnosed with early cancer. She is advised to undergo chemotherapy, which would leave her with very good prospects of recovery. However, Laura remembers her grandmother undergoing chemotherapy many years ago and how debilitating it was at the time. She assumes her experience will be the same.[82] Taking her grandmother's experience as a model for her own experience shows that Laura is subject to availability heuristics: as it is very easy for her to remember her grandmother's illness, she cannot help but believe that it will be similar for her, despite her physician's accurate assurance to the contrary.

In order to correct Laura's distorted picture about current chemotherapy, her physician could show her pictures of present patients, thereby triggering the availability heuristic in a different way. In fact, once Laura can easily think of examples different from her grandmother's experience, the initial bias may be reduced. In addition, Laura's physician may also show her around the treatment centre where Laura can meet current patients. He may introduce her to a woman of her own age, thereby triggering the representative heuristic. If Laura meets a woman similar to her, it is more likely that she will appreciate what current chemotherapy is like. In such a case, the physician can successfully re-bias Laura. By triggering some bias, he corrects Laura's misconceived idea and enables her to think more clearly and accurately about her options.

Finally, in certain exceptional cases, a person's decision-making can be enhanced by exploiting some bias, as was the case in my example *Stubborn*. Betty was subject to a

[81] See Hanna, 2015, who makes a similar point.
[82] Aggarwal (2014) discusses a similar example.

severe confirmation bias: she would not believe any evidence coming from a black doctor. Moreover, she was also subject to a representative heuristic. As she did not look and feel like someone who is ill, she did not believe she was.

Her representation bias could be overcome, but her confirmation bias was so firm that it could not have been reduced in any way. Therefore, the only chance to enhance her decision-making was to exploit that very bias, but to do so in a way that would lead her to the appreciation of her diagnosis. As explained earlier, physicians invited a person who fitted the description of the person she would have no trouble believing. By doing so, the physicians exploited her confirmation bias and turned the situation around so that Betty would consider the evidence and make a decision based upon her values, which would otherwise have been impossible for her.

Debiasing, rebiasing, and exploiting a bias can be consistent with voluntary consent insofar as they enhance a person's decision-making. Consent in Walter's (emergency leg injury), Laura's (past and current chemotherapy), and Betty's (mistrust in the black doctor's diagnosis) cases is voluntary and valid. However, the conditions outlined earlier (i.e. easy resistibility, transparency, and rationality) cannot explain why. As already mentioned, *Stubborn* was impossible to account for. But also in Laura's and Walter's cases, the influences were not easily resistible, i.e. Laura and Walter lacked the required skills to detect the influences in the first place, and the influences were therefore not transparent either.

By contrast, ICCJ can explain why consent is still voluntary. Consent is voluntary because the physician can interpersonally justify to the people consenting that he exerted those influences, based on the principles fundamental to the relationship between him and the person consenting.

In all my examples, *Assisted Decision-Making* is respected because nudges that enhance decision-making neither compromise a person's ability or opportunity to make a free decision nor render non-consent costly in some significant way. In fact, they enable a person to exercise her capacities in a more fruitful way and provide active assistance for the decision-making. Therefore, nudges that enhance decision-making also escape a prominent autonomy-based critique of non-persuasive influences. When trying to pin down what is problematic about nudging, Hausman and Welch claimed that nudging amounts to a "calculated shaping of choices (…) [and therefore] imposes the will of one agent on another."[83] Many other writers followed this diagnosis.[84] Imposition of one's will is problematic as it not only *affects* a person's self-determination but also *disrespects*[85] or *insults*[86] her. But nudges that enhance decision-making cannot be considered a form of "calculated shaping of choices."[87] Cooling off periods, 'consider the opposite strategy', as well as the nudges on Walter, Laura, and Betty do not aim to shape their choices or control the decision-making. They aim to empower Walter, Laura, and Betty so that they can make a good decision for themselves. Walter needs to calm down to make up his mind. Laura needs to appreciate the difference between past and current treatment. And Betty needs to consider the evidence. Hence, by using nudges that enhance decision-making, there is no attempt to impose one's will on another person.

Equal Moral Standing is respected because nudges that enhance decision-making do not compromise a person's entitlement to pursue his conception of the good. The physician who uses these nudges neither picks a particular option for the person influenced nor restricts options in any way. In my examples, the physicians do not try to

[83] Hausman and Welch, 2010, 133.
[84] See Grüne-Yanoff, 2012, 636. Goodwin, 2012, 89. White, 2013, 92. Wilkinson, 2013, 345-347. Baldwin, 2014, 846. Holm and Ploug, 2015, 35. Glod, 2015, 609-610.
[85] See Hausman and Welch, 2010, 134. White, 2013, 118. Glod, 2015, 602.
[86] See Wilkinson, 2013, 345.
[87] Hausman and Welch, 2010, 133.

substitute their own judgment for the judgment of either Walter (emergency leg injury), Laura (past and current chemotherapy), or Betty (mistrust in the black doctor's diagnosis). Rather, the physicians allow Walter, Laura, and Betty to decide for themselves what is best. Thus, Walter, Laura, and even Betty can still make their own decisions and pursue and revise their own conceptions of the good.

Finally, *Trusted Partner* is respected because the physician neither employs non-transparency nor his superior position of power to dominate the decision-making of Walter, Laura, or Betty.

When the physician calms Walter down in the emergency setting, he neither uses non-transparency and nor his stronger position to dominate Walter's decision-making. In fact, he only reduces Walter's cognitive and emotional overload and thereby enables him to make as good a decision as possible for him in this situation.

With regards to Laura, there is no concern over transparency either and when the physician shows her around the treatment centre where Laura can meet current patients, he does not abuse any position of power but rather helps Laura to evaluate her situation accurately and appreciate the differences between past and current chemotherapy.

Finally, in Betty's case, the physician does not tell Betty that he arranged the situation in order to make her appreciate the evidence about her cancer. Thus, in this case, there *is* a clear aspect of non-transparency and the physician uses his position of power to affect Betty's decision-making. However, the physician does not violate *Trusted Partner* as he only employs non-transparency to *support* rather than *dominate* Betty's person's decision-making. Non-transparency is only employed to help Betty make as good a decision as she can. Therefore, the physician's conduct in Betty's case is consistent with *Trusted Partner.* If Betty still feels resentment after finding out about the physician's

conduct, such resentment would be unjustified according to the standards of *Trusted Partner.*

Thus, ICCJ can explain why consent is voluntary in many of the examples discussed and thereby differs from other views in several ways.

Firstly and most importantly, ICCJ accounts for three ways of enhancing a person's decision-making. Other accounts in the literature can account only for debiasing and rebiasing, if at all,[88] but – due to their commitment to resistibility, transparency, and rationality – they fail to appreciate exceptional cases like *Stubborn.*

Secondly, ICCJ denies that nudges can be legitimate only if all other purely rational means of influencing have been exhausted.[89] If a nudge in fact enhances a person's decision-making, such influence will be legitimate and consistent with voluntary consent, even if there had been another non-nudging influence that could have done the same job. For instance, if the physician in Laura's case chose to use nudges instead of rational persuasion, where both would do the job equally well, ICCJ considers nudging to preserve the voluntariness of consent.

Finally, ICCJ rejects views like Engelen's, according to which nudging becomes more problematic when the stakes are high. Consider Walter's case again (i.e. the case of the acute leg injury): he may be in a high-stakes situation where he needs to make an important treatment decision. Yet, the nudges enabling him to make a reasoned decision are not more problematic because the stakes are high. On the contrary, high stakes may even call for nudging in the first place.[90]

[88] See for instance Agarwal, 2014. Buss, 2005. Hanna, 2015. Noggle, 2017. Lemmon, 2019.
[89] See for instance Blumenthal-Barby, 2012a. Bell, 2013, Engelen, 2019a. They all require exhausting other means of persuasion and open discussion before using non-persuasive influences.
[90] Lemmon (2019) also argued that nudging can be legitimate in high-stake situations.

4.3.2 Undermining Decision-Making

Just as a person's ability to make a reasoned decision can be enhanced in different ways, it can also be undermined in different ways, namely by means of distraction, exclusion, and pressure. These means of influencing steer a person's decision-making towards a particular option and, if they exceed a certain threshold, they vitiate consent. They differ from each other in terms of *how* they steer towards particular options.

Distraction undermines decision-making by making certain aspects salient in a way that does not reflect their actual importance. Consider the following example from Noggle:

"Phoebe is a physician whose patient - call him Smoking Joe - is a recalcitrant smoker. Suppose also that Phoebe knows that Joe has irrational fears about contamination. Having tried reasoning with Joe to no avail, Phoebe focuses Joe's attention on the fact that tobacco plants grow outdoors, where they will sometimes come into contact with bird droppings. Of course, Phoebe knows that any bird droppings will be washed away during processing—and she may even acknowledge this to Joe. Yet, suppose that focusing Joe's attention on the possibility of tobacco leaves being touched by bird droppings engages his irrational contamination fears and motivates him to quit smoking. (…) Phoebe achieves her commendable goal by inducing Joe to make a decision based on a true but irrelevant fact, which she has made far more salient than it ought to be, given its irrelevance to the question of whether to quit smoking."[91]

Phoebe uses the availability heuristic to make certain aspects of tobacco production, i.e. bird droppings, very salient to Joe: she may use pictures and videos and she uses the affect heuristic to trigger a particular emotional response in Joe when presenting him with evidence of bird droppings. However, bird droppings are not important enough to deserve

[91] Noggle, 2017, 169.

such a level of salience. Hence, Phoebe distracts Joe. She diverts his attention to a fact that is irrelevant and exploits Joe's irrational phobia. Due to Phoebe's influence, Joe subsequently consents to some treatment for his smoking addiction.

Exclusion undermines a person's decision-making by taking away possibilities of participating in the decision-making process. Suppose Jane has cancer and faces two options: risky surgery that could potentially cure but also kill her, or palliative care that will provide comfort but offers no prospect of recovery. She currently favours surgery and will decide to undergo it if left to her own devices.

However, her physician Sarah thinks this is the wrong decision for Jane, e.g. because she thinks Jane is too old, and wants Jane to consent to palliative care instead. But Sarah knows that an open discussion cannot change Jane's mind. To change Jane's mind, Sarah needs to take advantage of Jane's fear about the risks as well as her concern about the financial burden for her family.

So, Sarah decides to use *priming*, i.e. exposing someone to a certain stimulus in order to change the response to a later stimulus. Before meeting Jane for a final discussion, she makes sure that Jane will be exposed to a range of words that empirical research has shown to increase a person's risk aversion significantly.[92] Sarah also makes sure that Jane, when being in the waiting room, watches a short video which reports how burdensome debts can be for a family. However, Sarah hides the fact that she deliberately arranged the situation so that Jane would be exposed to risk terminology and the video. As a result, Jane's sensitivity to aspects concerning risks and her family is increased considerably.

Such methods of *priming* trigger availability and affect heuristics, but in a way that is completely hidden from the person influenced. Jane has no idea that she is

[92] See Erb, 2002.

influenced in that way. In addition, to avoid Jane's suspicion that she is being influenced, Sarah deliberately separates the exposure to risk terminology and the short video from Jane's medical situation. Finally, Sarah's methods of *priming* trigger automatic responses rather than explicit reasoning on Jane's part. In the end, Sarah's influences are effective: Jane consents to palliative care.

As Sarah harnessed Jane's susceptibility to priming in order to steer Jane in a particular direction, she controlled parts of the decision-making that could have been under Jane's control instead. Hence, Sarah substantially weakened Jane's ability to make an autonomous rational choice.

Finally, one can undermine a person's decision-making by *pressuring* a person into a decision. Recall *Social Cost.* Carl arranged the situation to make it costly and socially embarrassing for Alan to decline. Carl used the spotlight effect and triggered the affect heuristic. He put pressure on Alan by moving him into a situation of decisional weakness and took advantage of it.[93]

The conditions outlined earlier (i.e. easy resistibility, transparency, and rationality) provide a poor guide for such cases. Admittedly, such conditions may be able to explain part of why Jane's (chemotherapy versus palliative care) and Phoebe's (smoking cessation) cases are problematic, e.g. by pointing to aspects of non-transparency or irrationality. However, they are unable to explain *Social Cost,* where all the conditions are met without making the case less problematic. ICCJ provides a better guide. To repeat, ICCJ holds that consent is voluntary if the physician or investigator can interpersonally justify to the people consenting that he exerted those influences, based on the principles fundamental to the consenter-consentee relationship. According to ICCJ, consent is involuntary in all my examples of distraction, exclusion, and pressure.

---

[93] Another way of pressuring someone into a decision is shaming. See Loewenstein and O'Donoghue, 2006. Eyal, 2014, 54. Barton, 2015, 347. Caloia, 2015, 247.

Assisted Decision Making is violated because the influences in my examples of distraction, exclusion, and pressure take excessive control of the decision-making process. In the smoking cessation case, Phoebe controls the decision-making by making clearly irrelevant aspects disproportionately salient for Joe and she thereby weakens his capacities for decision-making. In the case of treatment options for cancer, Sarah controls the decision-making by playing on Jane's psychological mechanisms and failing to discuss risks and financial burdens openly. Sarah thereby prevents Jane from taking part in important aspects of the decision-making. In *Social Cost*, Carl controls the decision-making by increasing the asymmetry of power and taking more control over Alan's decision than he is entitled to have. In fact, Carl makes non-consent costly in a significant way. Therefore, all these influences in no way assist the people in their decision-making but rather weaken their ability or opportunity to make a free decision. Thus, they cannot be interpersonally justified to them and vitiate consent for that reason.

Equal Moral Standing is violated insofar as the nudges in these examples aim to make a person choose what is considered best for him, and where the person nudging does so because he considers the person nudged incapable of making the right decision if left to his own devices. However, this was not the case in all the examples and ICCJ therefore introduces an important distinction.

As explained in Part One, ICCJ distinguishes between *adversarial* and *beneficent* influences. Adversarial influences are those that do *not* aim to benefit the influenced person but rather aim to promote some other interests, e.g. the influences of Carl's trial recruitment in *Social Cost*. By contrast, beneficent influences are those that *do* aim to benefit someone and steer him into what may seem to be in his best interests, e.g. Phoebe nudging Joe into giving up smoking and some form of treatment.

Only beneficent nudges challenge *Equal Moral Standing* because only these nudges can amount to paternalist interference. In adversarial circumstances, there will not be any *paternalistic* interference and, therefore, the relevant principle for assessing them will be *Assisted Decision-Making* rather than *Equal Moral Standing.* Researchers like Carl may very well believe that a person would make a good decision by himself. However, for some reasons, they want him to make a different decision.

ICCJ provides distinctive guidance for each case and also holds that the concern about steering nudges will be reinforced if the nudge is adversarial rather than beneficent, for the following reason: in cases of beneficent influences, there may be some aspects that could be justified to the consenter as the influence is exerted for the sake of him and his interests. In cases of adversarial influence, this is not the case and considerations of beneficence cannot justify the influence. Thus, depending on the end of an influence, nudging can vitiate consent if, in addition to violating *Assisted Decision-Making,* it amounts to serious paternalistic interference and violates *Equal Moral Standing.*

Phoebe (smoking cessation) and Sarah (treatment versus palliative care) consider their patients incapable of making a decision that is best for them, and the way in which they influence their patients displays a serious form of paternalism. Phoebe and Sarah act clearly paternalistically. Phoebe targets Joe's phobia and distracts him so that he will consent to what *she* thinks is best, and Sarah simply decides on her own what is best for Jane. Phoebe's and Sarah's influences are also serious forms of paternalism: Phoebe and Sarah aim to overcome their patients' choices, they sabotage the quality of their patients' decision-making, and they cannot justify their influences on the grounds that their patients' decision-making was temporarily impaired. Moreover, Sarah uses powerful mechanisms to sidestep Jane's attention altogether, where it is unclear what would be best

for Jane. Therefore, Phoebe's and Sarah's influences violate *Equal Moral Standing* and consent is thus rendered involuntary.

Finally, *Trusted Partner* is violated because the nudges in distraction, exclusion, and pressure employ either non-transparency or a position of power to dominate rather than support a person's decision-making. Sarah deliberately hides her influences in order to make Jane consent to palliative care. As Sarah crucially relies on non-transparency in a high-stakes situation for Jane without thereby strengthening Jane's ability to make a decision on her own, Sarah's influence violates *Trusted Partner*.

Phoebe uses her position of power to take advantage of Joe's weakness, i.e. his phobic tendencies. Her influence relies on Joe being in a much weaker position and susceptible to her in ways that he cannot control. Therefore, Phoebe abuses her position of power and thereby also violates *Trusted Partner.*

Finally, investigator Carl takes advantage of the moment when he could control Alan's situation and ignores Alan's interests. Carl simply imposes his own goals on Alan and therefore betrays Alan's trust when acting from his much stronger position.

Hence, the influences in all my examples of distraction, exclusion, and pressure conflicted with ICCJ's principles and consent was therefore involuntary.

Embracing these results, ICCJ provides a novel position in several respects. To begin with, ICCJ accommodates the three conditions of easy resistibility, transparency, and rationality as part of a wider framework, rather than considering them as independent criteria as other authors do. ICCJ acknowledges that the three conditions are relevant when they affect the principles of *Assisted Decision-Making, Equal Moral Standing,* and *Trusted Partner*: non-transparency, not being able to resist an influence (e.g. Jane's case), or inducing irrationality (e.g. Joe's case) can undermine a person's standing in the decision-making process and thereby trigger a response from ICCJ's principles.

However, accommodating the three conditions in this way is still consistent with ICCJ's rejection of the three conditions as independent criteria: taken by themselves, easy resistibility, transparency, and rationality are neither necessary nor sufficient to guarantee the voluntariness of consent.

Secondly, ICCJ is sensitive to different objectives in the use of nudging, i.e. adversarial versus beneficence nudges. According to ICCJ, it makes a difference whether a person is nudged into something that aims to benefit him or into something that aims to serve the interests of others. The principles *Assisted Decision-Making* and *Equal Moral Standing* provide different guidance for each scenario. Other views, such as Saghai's, could not account for the difference between adversarial and beneficent nudges, and the conditions of easy resistibility, transparency, and rationality did not include a distinction between these types of influences either.

Finally, ICCJ embraces a different condition of non-triviality. ICCJ still holds that some forms of distraction, exclusion, and pressure, or steering nudges more generally, may not be serious enough to vitiate consent. However, *pace* most other views, ICCJ does *not* hold that nudges are trivial or non-trivial depending on the extent to which they change a person's behaviour (e.g. Cohen's view) or depending on how easily resistible they are (e.g. Saghai's view). By contrast, ICCJ holds that nudges are trivial or non-trivial depending on whether they create a sufficiently large gap between the salience of some aspects and their actual importance (e.g. in Joe's smoking cessation case), depending on the extent to which they exclude a person from the decision-making (e.g. in Jane's palliative care case), and depending on how costly they make it to refuse to consent (e.g. in Carl's clinical research case).

Conclusion

This chapter examined the question of which non-persuasive influences vitiate consent. Defending ICCJ, I argued that non-persuasive influences vitiate consent if and only if the physician (or investigator) cannot interpersonally justify to the patient (or trial participant) that he exerted them, where such justification rests on the principles fundamental to the consenter-consentee relationship.

ICCJ provides a distinct position by (i) explicitly focusing on individual clinical encounters rather than public policy, (ii) opposing widely accepted conditions for the evaluation of such influences (i.e. easy resistibility, transparency, rationality), and (iii) offering a unified account based on the idea of interpersonal justifiability rather than a set of independent criteria.[94] ICCJ's position is attractive because it not only clearly restricts the use of nudges, but also leaves room for nudging to enhance people's decision-making.[95]

Before I conclude, I wish to return to Montoy, i.e. the HIV-test example from the beginning of this chapter. ICCJ will likely consider Montoy's nudge consistent with voluntariness. If Montoy sets the default as he did and refrains from making it more costly or difficult for people to refuse screening, and if he refrains from using other reinforcing nudges (e.g. priming, shaming), he can interpersonally justify the default nudge to his patients. Announcing the default openly makes it a point for discussion, does not compromise the decision-making process, and thereby aligns with *Assisted Decision-Making*. Actively asking people to make a decision does not compromise their ability or

[94] See Blumenthal-Barby, 2012a. Bell, 2013. Engelen, 2019a. These authors present sets of 3, 6, 9, or even 20 loosely arranged criteria.

[95] Other accounts struggle to find this balance. For instance, Wertheimer plausibly claimed that, on Saghai's view, it remains puzzling how nudges could be both effective in the way psychological research has shown them to be *and* easily resistible. As Wertheimer pointed out: "Can an influence be both easily resistible and yet rarely resisted? (…) Could people easily resist an influence but (virtually) always choose not to do so?" See Wertheimer, 2013, 498.

opportunity to pursue a conception of the good and thereby aligns with *Equal Moral Standing.* Finally, Montoy did not use non-transparency or a position of power to *dominate* his patients' decision-making, and thereby aligns with *Trusted Partner.* Hence, even if Montoy's default was a steering nudge and did not aim to enhance people's decision-making, it failed to violate any of ICCJ's principles and was therefore consistent with voluntary consent.

5. Payment for Research Participation

Let me now return to the example from the beginning of this thesis: Rambha Gajre is a young mother from India who is worried about her son. He is seriously ill and needs life-saving treatment, but Rambha cannot pay for it. She has so little money that she struggles to pay her rent and has been threatened with eviction. To earn enough money for the rent and her son's treatment, Rambha consents to paid participation in a clinical trial. Such participation may be highly intrusive, potentially involve exploitation, and be the consequence of unjust circumstances. Yet, Rambha enrols, and explains her situation as follows:

"Many people commit suicide and I didn't want to become one of those and I didn't want people to think I did it to avoid repaying [i.e. paying the rent]. I have two young kids, 10 and 12 years old. What would become of them? (…) I am helpless, I have to do this [i.e. paid trial participation]. (...) They don't really force us, but I don't have a choice."[1]

In this chapter, I will focus on the question of whether and, if so, under what conditions the offer of payment vitiates a person's consent to participation in a clinical trial. Concerns about consent and voluntariness arise when payments not only reimburse participants for their expenses (e.g. travelling costs to the research facility, parking costs, etc.) but also compensate for time and inconvenience and thereby exert an incentivising influence.[2] Such concerns become reinforced once the background conditions of a participant are as dire as in Rambha's case.

[1] Reference from Wertheimer, 2014, 280-281. See also footnote 1 in the *Introduction* of this thesis.
[2] Hence, I will focus on whether *high* payments are problematic and exclude concerns about too low payments from the scope of this chapter, which are discussed in Phillips (2011) and Belfrage (2016).

To address the question of whether offering payment vitiates consent, I will start from the presumption that *it does not* because consent to financial transactions is normally voluntary, and I will then test whether this presumption can be defeated in the context of paid trial participation.[3] This presumption is only a minimal Archimedean point. It does not rely on any theoretical or libertarian commitments but rather, if anything, only on what seems true in everyday life. We voluntarily consent to work in exchange for money, even if the situation is less than perfect, e.g. when we consent to work in order to avoid becoming homeless.

After providing some background information about clinical trials (5.1), I will survey the three most widely defended challenges to the voluntariness of consent in the context of paid trial participation and claim that payments vitiate consent only in certain cases of increased risks (5.2). I will then explain my own proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* as best able to account for these results (5.3).

Although there is already a very comprehensive debate on paid trial participation in the literature, this chapter will offer a novel contribution with regards to three aspects.

(i) This chapter offers a discussion of payment freed from the question whether payments qualify as 'coercion', 'manipulation', or 'an undue influence', which have so often obstructed rather than assisted the debate.[4] Instead of these questions, I will focus more directly on the question of when payments vitiate consent.[5]

[3] This method is inspired by Radcliffe Richards. See Radcliffe Richards, 2012.
[4] See for instance the disagreement between scholars claiming that offers can never be coercive for conceptual reasons and Research Ethics Committees that regularly call high payments coercive. See Millum and Garnett, 2019. As an example, see the focus on concepts like coercion in the very recent debate in *The American Journal of Bioethics in 2019, 19 (9).*
[5] McMillan succinctly explains how conceptual or theoretical baggage can blind an ethical inquiry and detract from what is ethically important. See McMillan, 2018, 53-62, 81, 122-125.

(ii) This chapter switches the debate from its focus on a person's mental states and circumstances to a focus on the interaction between participant and investigator. I will employ my proposal ICCJ to elaborate on such a perspective.

(iii) Finally, this chapter will highlight how investigators can illegitimately reduce a person's option before offering payment and also draw attention to *modes of payment*, thereby highlighting two challenges to the voluntariness of consent that have not yet received attention in the literature.

5.1 Clinical Trials and Payment for Participation

A *clinical trial* is any investigation involving human subjects that is carried out to advance scientific knowledge about the effects of medicines or therapies.[6] The role of a *human subject* can involve exposure to certain medication or treatment during a period of time when one is also being monitored. The medication or treatment will be initiated at the beginning of the investigation and will be one the participant has not received before ('interventional' trials); or it can involve mere monitoring, where medication or treatment, if people receive it at all, has been prescribed prior to and independently of the investigation ('observational' trials).

Among interventional trials, Phase I trials "are conducted to examine the biological and pharmacological actions of a new treatment (usually a new drug), and its side-effects."[7] Often, participants need to stay overnight in research units and undergo several diagnostic procedures; sometimes, participants will also be infected with a bacterial illness (e.g. typhus or malaria) to test a new therapeutic strategic or vaccine; and sometimes, participants will have to take drugs in increasing doses. By contrast, Phase II

[6] See Hackshaw, 2009, 1-3.
[7] Hackshaw, 2009, 31.

and III trials no longer aim to find new forms of treatment but rather "provide enough evidence to warrant a change in practice"[8], i.e. a change from current routine to a new form of treatment already tested in Phase I trials. Whereas Phase I trials usually involve only healthy participants, Phase II-III trials mostly involve patients whose illnesses investigators want to focus on.

In most countries, a clinical trial has to receive approval from a research ethics committee before it can be initiated. In the UK, the law states:

"A trial shall be initiated only if an ethics committee and the licensing authority comes to the conclusion that the anticipated therapeutic and public health benefits justify the risks."[9]

"Risks" means the likelihood that participants will experience adverse effects on their health as a result of trial participation. On the basis that the "rights, safety, and well-being of the trial subjects are the most important considerations and shall prevail over interests of science and society"[10], risks exceeding a certain degree or inflicting serious harm are absolutely excluded. Other risks can be justified only if they are not "disproportionate to [their] (…) potential benefits."[11] "Benefits" are either therapeutic advantages for the participant or scientific progress as a result of the trial.[12] Payment for participation, however much the participant welcomes it, is *not* considered a benefit.[13] The research

[8] Hackshaw, 2009, 57.
[9] *The Medicines for Human Use (Clinical Trials) Regulation 2004*, Schedule 1, Part 2. See also ibid., Arrangement of Regulations, Part 3, 12. EU Directive 2001/20/EC, Article 3, 2(a). *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine,* Article 16 (iii). *Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research*, Article 7.
[10] Medicines, Schedule 1, Part 2. See also EU Directive 2001/20/EC, Article 4(h), Article 5(h); Convention Article 2; Additional Protocol Chapter II Article 3
[11] Additional Protocol. Article 6, 1.
[12] The EU Directive 2001/20/EC, Article 3, 2(a). Medicines, Schedule 1, Part 2. Convention, Article 17. Additional Protocol, Article 6 (2), 15 (2).
[13] See *The Belmont Report*, Part C, C.2. *International Ethical Guidelines for Biomedical Research Involving Human Subjects* (henceforth "CIOMS-Guide"), Guideline 8.

ethics committee must come to the conclusion that a trial has a favourable risk-benefit ratio independent of payments.[14]

In this chapter, I will only focus on healthy participants in Phase I trials, where problems of payment arise most clearly and forcefully.

5.2 Aversion, Alternatives, and Risk

Sometimes people struggle to satisfy their own and their loved one's basic needs, i.e. needs concerning the minimal provision of health care and the economic resources for food and shelter. In such situations, paid trial participation may become their best option to satisfy these needs and it will normally be very difficult to resist, i.e. it may take great psychological effort or resilience to decline participation. As Vanderpool explained, the offer of paid trial participation may even be "virtually irresistible."[15]

Rambha is in such a situation. She has great difficulty paying for her son's medical treatment and securing shelter. Moreover, paid trial participation has become by far her most feasible option to satisfy these needs. Finally, the offer of paid participation is very hard to resist for Rambha and exerts a strong influence on her decision-making.

Situations where people struggle to satisfy basic needs have often been considered a necessary condition for payments to vitiate consent. Yet, there is also agreement that such situations are not sufficient to render payments problematic, let alone vitiate consent. For instance, if Rambha could earn enough money simply by watering flowers, instead of participating in a trial, and also enjoys gardening work, the moral unease about what

[14] For further details about the legal frameworks, see Mason, 2011, Ch. 19. Jackson, 2006, Ch. 8, with regards to UK law in particular, see pages 477-484.
[15] Vanderpool, 2005, 18. Note that this formulation does not mean anything like a psychological inability to resist. It is just an exaggerated expression for the psychological effort it would take to turn down the offer of paid trial participation.

she is being paid to do would vanish. Hence, a further condition has to be added if payments are to vitiate consent. Three proposals deserve particular attention.

In the rest of this chapter, I will rely on the empirical evidence that monetary offers do not distort a person's judgment about trial participation in any considerable way[16] and only focus on the pressure they may exert on a person's decision-making.

Defeater Candidate 1: Aversion

Vanderpool and Grant highlighted that payments become particularly problematic and potentially vitiate consent if, in addition to the difficulties in satisfying basic needs[17]:

Aversion Participants are offered money to do something they are averse to.

'Being averse' goes beyond people's refusal to do something without payment, viz. "what they otherwise would refuse"[18] to do. It concerns what people "would *strongly* prefer not to do."[19] Assuming this line can be drawn, Vanderpool presents the example of an impecunious college student who is averse to painful muscle biopsies but still consents to undergo them in exchange for money. Grant adds examples of situations where "the aversion is a principled one or a matter of moral scruple."[20] For instance, a religious person may be 'averse' to donating sperm.[21]

---

[16] See Bentley, 2004. Halpern, 2004. Yet, some people in the debate make the contrary assumption. See: Appelbaum, 2009, 34. Phillips, 2011, 213. Largent, 2012, 2. Wertheimer, 2008, 391. Wertheimer, 2012, 248.
[17] Vanderpool talks about "needy college students" and the "poor" for which payments may be "difficult to resist" or even "virtually irresistible" (Vanderpool, 2005, 18). See also Grant, 2004, 727-728.
[18] Vanderpool, 2005, 18.
[19] Vanderpool, 2005, 18. See also Grant, 2004, 728.
[20] Grant, 2004, 728.
[21] See Grant, 2004, 736 footnote 20.

For certain procedures, aversion is a matter of individual disposition: some consider muscle biopsies a minor inconvenience while others panic at the very thought of it. For other procedures, aversion is very widespread, e.g. when procedures become "physically invasive, upset the customary boundaries of privacy, or involve areas of the body generally associated with feelings of shame."[22] However, in any case, aversion to a procedure does not already mean that undergoing that procedure is risky or harmful.

To clarify Vanderpool's and Grant's views further, they do *not* claim that mere aversion to a medical procedure vitiates consent. Peter may be highly averse to dental surgery because it will be painful. Yet, he consents to it to free himself from some illness. James may be highly averse to receiving a blood transfusion because of religious beliefs yet consent to it as part of his leukaemia treatment. Peter and James overcame their aversions after reflecting on their situation and doing so was a genuine exercise of their autonomy.

The situation becomes morally problematic and potentially vitiates consent, Grant argues, when one attempts to "*overcome*"[23] another person's aversion by means of a monetary offer. Furthermore, problems arise when (i) a person's resistance to undergo certain procedures is "reasonable"[24], (ii) the monetary offer "ought to be resisted in some sense"[25], and (iii) a person's acceptance of the offer would therefore go "against his or her better judgment."[26]

(i): A person's resistance is *reasonable* when it is necessary to protect his interests and justified on the basis of the overall situation. So understood, Peter's resistance is *unreasonable*: although declining dental surgery allows him to avoid pain, foregoing

[22] Grant, 2004, 729-730.
[23] Grant, 2004, 729. Emphasis added.
[24] Grant, 2004, 729. Emphasis added.
[25] Grant, 2004, 729. See also Vanderpool, 2005, 18
[26] Grant, 2004, 727. See also pages 728, 729, 734.

urgently needed treatment lacks a plausible justification. By contrast, the impecunious college student's resistance is *reasonable*: not enrolling in the trial protects his interest in avoiding biopsies and, other things being equal, prioritising such an interest over financial gain is justifiable. Similarly, the religious person's resistance is reasonable because prioritising religious commitments over financial gain can be justified.

(ii): Sometimes, however, a person may not only *reasonably* resist certain medical procedures, he also *ought* to resist and decline an offer to the contrary. For instance, the impecunious college student ought to resist the offer of payment when the trial involves "undignified procedures, embarrassing situations, and degrading experiences"[27] but fails to pay a sufficient amount of money. Moreover, the religious person ought to resist an offer if his important religious commitments are of such paramount importance to him that no financial gain could ever justify their violation. Grant is particularly concerned with cases similar to that of the religious person and argues:

"the ethical problem is multiplied where the aversion is a principled one or a matter of moral scruple. To deliberately induce religious people to work on the sabbath by offering large incentives would involve an attempt to get them to act against what they see as their duties; it would be a form of bribery."[28]

And Grant adds that:

"Respect for persons (…) requires respect for certain of their values, beliefs, and preferences. It requires refraining from making seductive offers, offers that ought to be resisted."[29]

[27] Grant, 2004, 729.
[28] Grant, 2004, 728.
[29] Grant, 2004, 729.

(iii) If a person's resistance is "reasonable"[30] and the monetary offer "ought to be resisted in some sense"[31], then accepting such an offer conflicts with at least *a* better judgment, i.e. the justified judgment that one ought to resist such an offer. Such a judgment could be better in an objective sense, independent of the person's endorsed values and preferences, or in a subjective sense, based on a person's own values and preferences. In the former case, the person accepting the offer disagrees with the objective standard and thereby acts against an objectively better judgment. In the latter case, the person agrees with the subjective standard but still acts against it, i.e. acts akratically: despite knowing what is best, he acts against *his* better judgment.

When these three conditions are met, Vanderpool and Grant consider payments to be instances of 'manipulation' or 'undue influence.' As indicated earlier, I will leave aside the conceptual question as to whether payments can be appropriately described with those terms and focus more directly on the question of whether they vitiate consent.

The emphasis on a person's resistance being reasonable or unreasonable is noteworthy because payments can sometimes turn previously *reasonable* resistance into *unreasonable* resistance. Recall the impecunious college student. Suppose he is offered very little money to conquer his aversion. Here, his resistance seems reasonable and he ought not to accept the offer. But if he is offered so much money that he can pay the tuition fees for his entire degree, his continued resistance may become *unreasonable* and there are strong reasons for him to enrol. The same is true of Rambha: it may be quite reasonable for her to overcome a physical aversion to muscle biopsies if she could thereby earn enough money to secure shelter and pay for her son's treatment.

If this line of reasoning is correct, two conclusions follow. Firstly, *raising payments* can remove the problem of incentivising people to overcome a *reasonable*

[30] Grant, 2004, 729. Emphasis added.
[31] Grant, 2004, 729. See also Vanderpool, 2005, 18.

aversion. As a result, one should sometimes only be concerned about *low* rather than *high* payments. Secondly, if offering more money removes the problematic aspect of such payments, then there was no problem with the *voluntariness* of consent in the first place. This is because these payments would not pass the test I developed as part of the *Mirror Thesis* (see 2.2.2 II). There, I argued that only those influences vitiate consent that become more problematic once they are strengthened. However, if strengthening the offer of payment, i.e. by increasing the amount of money, does not make the situation more but less problematic, such an offer of payment cannot vitiate consent in the first place.

But even if the subject's resistance remains reasonable on *moral or religious* grounds, even if any financial incentive ought to be resisted, and even if enrolling conflicts with the subject's better judgment, the offer of payment still fails to vitiate consent. To see this, suppose an investigator offers money to the potential participant Jack but enrolling in the trial would violate duties Jack has towards work: he works in a kitchen and, given the risk of infection, enrolling in the trial would transgress rules of hygiene. Suppose Jack has a reasonable and strong aversion to enrolment - for example, it could be a moral aversion to harming others - the monetary offer ought to be resisted because Jack would otherwise depart from his values and preferences, and enrolling would therefore be against his better judgment, objectively as well as subjectively. But if he does enrol, his consent can still be voluntary. If the investigator performs the trial procedure, he will not assault Jack.[32]

[32] The conclusion that bribing does not negate voluntariness also applies more generally. Suppose a policeman is highly averse to violating his duties. Such aversion is reasonable and any offer to overcome it would be inappropriate, ought to be resisted, and would go against a better judgment. Yet, bribing the policeman will not negate the voluntariness of his acts when he accepts the offer. Other things being equal, he will be liable to blame and, insofar as liability to blame presupposes voluntariness, he must have acted voluntarily. The only scenario in which he may be considered to have acted involuntarily and therefore be exempted from guilt, could be a situation where he lacked other options to satisfy basic needs. In this case, however, his aversion would be immaterial.

This is a case of principled or moral aversion to violating duties rather than aversion to physical pain. In this respect, it is similar to the case of the religious person who is averse to donating sperm. Thus, can we conclude that if financial incentives do not vitiate Jack's consent, then they will also not vitiate the consent of the religious person?

Grant may reject this comparison on the grounds that, in the case of aversion on religious grounds, the person's aversion is more fundamentally connected to his core values and identity than in the case of Jack transgressing rules of hygiene. Thus, even if Jack's consent is voluntary, the religious person's consent may not be.

However, this response is misleading. Suppose that the religious person decides to donate sperm and contacts the investigators after seeing the advertisement of the trial. In this case, it is disrespectful to exclude him on the grounds that his aversion to donating sperm is so fundamental to his core values and identity. Doing so would deny him the authority to decide whether to surmount his aversion in exchange for money or not. In fact, once core values are at stake, the disrespect seems even greater than before. Therefore, it is not *offering money* that is disrespectful (as Grant would argue), but rather *excluding people* and only allowing those to enrol who can show that they do not harbour any aversion. Thus, the difference between Jack and the religious person does not justify treating them differently with regard to the voluntariness of their consent.

Hence, *Aversion* cannot explain when payments vitiate consent.

Defeater Candidate 2: No Acceptable Alternative

Joan McGregor argues that payments vitiate consent if:

No Acceptable Alternative There is no acceptable alternative to paid trial participation.

This means that trial participation is a person's *only* and not just her best way to satisfy basic needs and, therefore, not participating would leave her in an unacceptably bad position. McGregor speaks of "desperate situations"[33] where a "lack of options"[34] makes even the direst proposal "the *only* eligible choice."[35] McGregor's proposal is stricter than Vanderpool's and Grant's, where a person may still have other acceptable options to earn money.

If payment from research participation is the only acceptable alternative, McGregor argues, then offers become "*coercive*"[36] and vitiate consent. Let us assume that Rambha now is in such a situation: trial participation is her only way to earn the money needed to avoid eviction and cover her son's treatment.

McGregor's view is a specification of Serena Olsaretti's account and, therefore, looking at Olsaretti can help to find the strongest version of this view. According to Olsaretti,

"a choice is voluntary if and only if it is not forced, and it is forced if and only if it is made only or primarily because the alternative to it is unacceptable, where the standard for the acceptability of options is an objective standard of well-being, and unacceptable options are those which, by that standard, fall below a certain threshold."[37]

---

[33] McGregor, 2005, 25.
[34] McGregor, 2005, 25.
[35] McGregor, 2005, 25. Emphasis added.
[36] McGregor, 2005, 25. Original emphasis.
[37] Olsaretti, 2016, 444.

Olsaretti claims that people act involuntarily if they do something *only or primarily because* the alternative to it is unacceptable. The italicised phrase is important because it shows that people may sometimes act voluntarily even if they lack an acceptable alternative: they act voluntarily as long as they decide to do something for reasons that are independent of the fact that there is no acceptable alternative to it. On the other hand, people may *not* act voluntarily even if there *are* acceptable alternatives. This is the case if the person reasonably thinks there is no acceptable alternative, e.g. when a person hands over money to a robber after being threatened with a realistic toy gun. Thus, on Olsaretti's view, a person's motivation is critical to the voluntariness of her action.

Olsaretti then claims that the "standard for the acceptability of options is an objective standard of well-being."[38] An option is acceptable if and only if it satisfies core aspects of a person's well-being. An example of an *unacceptable* option is dying a painful death, but also being evicted from one's home or seeing one's son suffer or die, as in Rambha's case.

Finally, Olsaretti claims that voluntariness is necessary for the validity of "consent to (…) surgery"[39], i.e. "whether or not it is permissible (…) for him [i.e. the physician] to carry out the operation."[40] And she adds that the significance of voluntariness for consent is part of a wider "justificatory role"[41] that voluntariness plays: on any satisfactory account, voluntariness "must be able to undergird attributions of responsibility, as people's voluntary choices - for example, the choice of entering a contract, that of making a promise, or *that of consenting to an interference* - can change their and others' entitlements and obligations."[42]

[38] Olsaretti, 2016, 444.
[39] Olsaretti, 2016, 443.
[40] Olsaretti, 2016, 445.
[41] Olsaretti, 2016, 443.
[42] Olsaretti, 2016, 443. Emphasis added.

So understood, *No Acceptable Alternatives* faces an obvious challenge. In medical contexts, the lack of acceptable alternatives does not automatically vitiate consent. To see this, consider the fictitious case of Peter who has been diagnosed with a serious illness. If he does not undergo major surgery, he will die a painful death. Although Peter is afraid of the surgery, he eventually consents, but only for want of an acceptable alternative.

Clinical practice considers consent from people like Peter voluntary and valid, and such a view is indeed justified. This is because consenting to a medical procedure only because one lacks an acceptable alternative can be a genuine exercise of a person's autonomy as well as a promotion of his well-being. Peter may have wished never to be in the situation he finds himself in now. But *once* he is in that situation, selecting the option he prefers most allows him to take control of his life and decision. Moreover, selecting the option he prefers most also allows him to promote his well-being and, assuming that people are usually better judges than others of what is good for them, Peter's choice will likely advance his well-being more than others' choices would.

However, on Olsaretti's and McGregor's views, Peter's consent is involuntary. Death is clearly an unacceptable option. So, how could they respond? Olsaretti identifies the problem and claims:

"the fact that the patient's consent is intentional and informed may suffice to authorize the surgeon's operation, and perhaps also to justify holding the patient to compensate the surgeon for some of the costs the surgeon has to incur."[43]

This is an unsatisfying response. If consent is intentional and well-informed, it does not follow that it is also voluntary, and Olsaretti does not claim otherwise. Hence, Olsaretti

[43] Olsaretti, 2016, 445-446.

is committed to saying that Peter's consent is involuntary but valid, thereby abandoning voluntariness as a necessary condition for the validity of consent. This implication is not only contrary to the orthodox link between validity and voluntariness. It is also inconsistent with her own claims about voluntariness playing a "justificatory role"[44] with regards to the question of "whether or not it is permissible (…) for him [i.e. the physician] to carry out the operation."[45] Hence, I suggest rejecting Olsaretti's and McGregor's views in the context of treatment.

Therefore, if *No Acceptable Alternative* is to explain when payments vitiate consent, there has to be a significant difference between patients like Peter and trial participants like Rambha. The obvious candidate for such a difference is the combination of monetary considerations figuring in Rambha's but not Peter's decision, and those monetary considerations incentivising something that fails to benefit the consenting person's health, i.e. Peter medically benefits from the surgery whereas Rambha does not medically benefit from the trial procedures. With these points in mind, let me reconsider the arguments from well-being and autonomy that justified considering Peter's consent to be voluntary and valid.

Well-Being

The well-being of a person depends on her health and the satisfaction of basic needs. Trial participation often involves a trade-off. The procedures may negatively affect a person's health but the money she earns can help to satisfy other interests. Depending on how health-related and economic interests are balanced, trial participation may or may not be detrimental to a person's overall well-being. Therefore, considerations relating to the

[44] Olsaretti, 2016, 443.
[45] Olsaretti, 2016, 445.

well-being of research participants can differ from those relating to the well-being of the terminally-ill patient.

Considerations of well-being only make a significant difference in some trials, but not in others. Suppose Rambha enrols in a trial testing the effect of certain physical exercises on the quality of sleep. Such a risk-free trial does not compromise any of her health-related interests. Yet, it may help to satisfy other important interests, e.g. by allowing her to pay the rent and cover her son's medical treatment. Hence, enrolment seems conducive to Rambha's overall well-being whereas being denied access to the trial would make Rambha worse off. In the latter case, she would be deprived of what is her only option to pay for her son's treatment and secure shelter.[46] Therefore, in this scenario, undergoing a procedure only for want of an acceptable alternative can still promote a person's well-being.

The situation may be different when clinical trials become more intrusive or risky. However, risk and intrusiveness are additional considerations and not part of Olsaretti's and McGregor's views.[47] In fact, Olsaretti and McGregor exclusively focus on the *unacceptability of foregoing the opportunity of paid trial participation,* rather than the *level of risk or harm involved in participation*. Therefore, they lack the means to reject the argument from well-being on the grounds of risk and intrusiveness.

Hence, both Peter *and* Rambha can promote their well-being by undergoing procedures to which they lack an acceptable alternative.

[46] On removing an option when everything else is unacceptable, see Radcliffe-Richards, 2010, 291. Flanigan, 2018.
[47] See McGregor, 2005, 24.

Autonomy

But what about the argument from autonomy? Is there a significant difference in this regard between Peter and Rambha?

Payments for research participation broaden people's earning potential and thereby promote 'positive autonomy': money enables people to do things they would otherwise be unable to do. The proscription of payments, on the other hand, excludes financial influences and thereby promotes 'negative autonomy': it keeps decisions free from the interference of other people's offers. Since positive and negative autonomy cannot be simultaneously maximised, a sound balance is required.

On this basis, McGregor and Olsaretti could claim that considering Rambha's consent to be valid unduly prioritises positive autonomy over negative autonomy. Therefore, the argument from autonomy only supports the validity of Peter's but not Rambha's consent.

To evaluate this response, let us consider a paradigm case of a wrong prioritisation of positive and negative autonomy. It is uncontroversial that a person ought not to have the power to transfer the title of his property to a burglar while being threatened with death. Allowing transfer of the title only when one's decision is kept free from the interference of the burglar (negative autonomy) is clearly required, whereas enabling the person to organise his life by entering a binding agreement under such circumstances (positive autonomy) lacks justification. This is because the burglar's threat severely wrongs a person and dominates his decision-making.

Rambha's situation is very unlike the burglar example. The offer of payment does not wrong Rambha but rather provides the means to improve an otherwise unacceptable situation. Moreover, Rambha is not dominated as the person in the burglar example is. Participation allows her to regain control over her situation. Finally, investigators did not

cause Rambha's situation. Thus, allowing Rambha to consent to paid trial participation is very different from the burglar case and consistent with a sound balance of positive and negative autonomy.

Hence, there is no difference between Peter's and Rambha's situation that could explain why only Peter but not Rambha consents voluntarily. Thus, *No Acceptable Alternative* fails to explain when payments vitiate consent.

Defeater Candidate 3: Risk

Finally, Nelson argues that payments vitiate consent if:

Risk Participants are offered money to do something where there is "a risk of harm of sufficient seriousness"[48] and the payment is "irresistibly attractive."[49]

The 'seriousness' of risk is a function of the probability that harm will occur and the kind of harm at stake, e.g. harm of a bruise or harm of renal failure. 'Irresistibility' refers to the great psychological effort it would take to decline research participation, but it does not mean a psychological inability to decline the offer.

Nelson claims that payments to undergo risky procedures vitiate consent because they are "manipulative."[50] I take Nelson's core thought to be that payments are 'manipulative', not because they side-step a person's rational deliberation, but because they incentivise a person to act against some of her important interests. But why exactly would such a 'manipulative' incentive vitiate consent? Unfortunately, Nelson does not

[48] Nelson, 2011, 9.
[49] Nelson, 2011, 9.
[50] Nelson, 2011, 9. Emphasis added.

provide an answer. Building on his claims but leaving behind conceptual questions about 'manipulation', the following line of thought seems to present the strongest version of Nelson's position.

High risks affect the situation in three ways. Firstly, as Nelson explains, when risks are high, payments need to exert a strong influence to make a person consent. Secondly, although Nelson does not say this, when risks are high, payments are more likely to incentivise what is contrary to important interests of a participant and become anti-beneficent and illegitimate. Finally, as Nelson implies, when risks are high, the test for valid consent is more exigent, for example, informed consent may require more information and less severe influences compromise voluntariness. Hence, payments for participation in risky trials may vitiate consent because they are strong and illegitimate influences which fail to meet the higher standards required for consent to serious medical procedures.[51]

However, in order to have this effect on the situation, risks must be possible to detect, as the following case illustrates:

Unknown Risk Investigator Sarah wants to develop a new vaccine but the envisaged trial bears such high risks of toxicity that exposing people to them cannot be justified.[52] Unfortunately, neither Sarah nor the Research Ethics

[51] Note that this challenge is different from other challenges discussed in the literature, where risks and harms are sometimes considered to be *indirectly* relevant to voluntariness. Consider the debate on consent to something *harmful:* in the legal cases *R v Carriere* [1987] and *R v Brown* [1994], people consented to a violent fist fight and very painful sadomasochistic sex respectively. The judges deliberated whether there was something else in the background that pressured them into consenting to these things which, they thought, an 'ordinary' person would never consent to. See also Bergelson, 2007, 2010, 2018. Hart, 1963, 33. A similar rationale could apply to consent to *risky* procedures. If the risk is one which, to use Nelson's words, "the person would not ordinarily assume" (Nelson, 2011, 9), there may be the suspicion that something else pressures her into consenting to it. Assumptions about background pressure are reasonable but do not show that risk itself could ever negate the voluntariness of a person's action. I will therefore not discuss this approach in detail.

[52] See Straus, 2012, 111-119, who presents a real example of unanticipated high risks in a clinical trial.

Committee know about these risks and could not possibly have detected them. Rambha enrols in exchange for money.

Due to the unjustified risks of the trial, payments incentivise undergoing procedures that conflict with the required protection of participants' health and are therefore anti-beneficent. Yet, no matter how high such risks are, payments do not vitiate consent. This is because the mere presence of high risks neither increases the pressure on Rambha's decision-making, as Rambha did not know about them, nor leads to Sarah violating her specific obligations of beneficence. Investigators violate their obligations of beneficence only if they knowingly or negligently act against a person's important health interests, but not when they act with due diligence. Moreover, clinical research will inevitably bear the danger of unknown risks and, if such risks materialise, they are better understood as creating additional claims for compensation rather than as vitiating consent to procedures already performed.[53] Hence, in order to vitiate consent, investigators must have at least had a chance to find out about unjustified risks.

However, even when investigators know about them, monetary offers do not automatically vitiate consent. Consider the following case:

Deviant Scientist Investigator Sarah wants to test a new drug but the envisaged trial would involve unjustified risks, which Sarah is aware of. Sarah fakes the documents submitted to the Research Ethics Committee, but tells Rambha everything about the risks. Moreover, she bribes Rambha to remain silent and offers further substantial payment if she participates. Rambha enrols.

[53] Particularly in Phase I trials, it is impossible to anticipate all the risks because it is only through such trials that science can detect and assess risks of certain new forms of treatment. See Straus, 2012, 110.

Performing the procedures in *Deviant Scientist* is impermissible because it violates the restrictions on clinical research that aim to protect the participants' health. These restrictions are paternalistic, i.e. they aim to prevent people from harming themselves, and the deliberate violation of them *alone* makes conducting the trial illegitimate.

But their violation alone does not vitiate participants' consent. For instance, Sarah's colleague George may be so intrigued by the prospect of contributing to scientific progress that he also enrols. Even though the procedures remain impermissible, George can give his voluntary consent and thereby prevent Sarah from committing the kind of wrong that would occur without his voluntary consent.

Therefore, to evaluate *Deviant Scientist* appropriately, we need to distinguish between impermissibility due to the deliberate violation of paternalistic restrictions and impermissibility due to the consent of participants being involuntary. If Rambha's consent is involuntary, it will be due to the aspects that distinguish her situation from George's, i.e. that she does not share the scientific objective of Sarah's trial and, unlike George, experiences high pressure to enrol, given her dire circumstances.

In what follows, I will accept that performing the procedures in *Deviant Scientist* is impermissible for the former reason and only focus on the voluntariness of Rambha's consent. Thus, can Sarah's offer vitiate Rambha's consent, given that it is now a strong and illegitimate influence?

Rambha's consent will be involuntary in only one variation of *Deviant Scientist*, i.e. when Sarah illegitimately reduces Rambha's options before offering payment. Suppose Sarah could either conduct the one risky trial or two safe trials to promote her scientific objective. But Sarah eventually opts for the one risky trial in order to accelerate her research and save money, thereby subordinating Rambha's interest in protecting her health to the comparatively minor scientific interests in saving time and money. In such

a situation, Sarah acts illegitimately and Rambha is entitled to the safe trials because investigators are obliged to act beneficently and respect participants' interests where this is possible and feasible.[54] Hence, if Rambha consents to the risky trial only because the option of safe trials is withheld from her, then Sarah illegitimately constrained Rambha's situation and the pressure on Rambha's decision-making therefore vitiates her consent.

But now suppose that, in order to protect participants' health appropriately, Sarah needs to engage in animal testing first or is required to conduct the safer trials in places where Rambha could not participate; or suppose a safer trial would require greater resources not available at the moment. In these situations, offering the risky trial does not withhold another option from Rambha, and if Sarah and Rambha agree that it is better for them both to select the risky trial, payments do not vitiate consent. In such a situation, Sarah would only make another option available to Rambha and not control Rambha's situation in the way necessary to vitiate consent. After all, Sarah cannot enhance Rambha's control over her situation in any way other than allowing her to enrol in exchange for money. Furthermore, rejecting Rambha as a participant and only enrolling George may amount to greater disrespect than allowing her to enrol alongside George.

Hence, the voluntariness of Rambha's consent does not depend on whether payments are strong and illegitimate influences, when illegitimacy is due to the deliberate violation of a paternalistic restriction. Instead, it depends on whether Sarah illegitimately reduced Rambha's options.

Sometimes, however, risks can create yet another situation where consent is invalid, as the following case exemplifies:

[54] See my argument for obligations of beneficence in clinical research in Chapter 3.2.

Adverse Effect Rambha is halfway through a trial when the procedures cause severe side-effects and reach a level of risk that makes continued participation inconsistent with the protection of her health. Sarah informs Rambha about the situation and tells her that she will not receive any compensation if she withdraws. Rambha remains enrolled.

Adverse Effect shares with *Deviant Scientist* the fact that Sarah knows about unjustified risks, that she deliberately violates a paternalistic restriction, and that she incentivises Rambha to forego the required protection of her health. *Adverse Effect* differs from *Deviant Scientist* in that Rambha is now already halfway through the trial and the trial procedures caused severe side-effects as well as further risks for Rambha. This difference is significant because now (i) Rambha ought to receive full compensation without further participation, whereas she was not entitled to any compensation without participation in *Deviant Scientist*; (ii) withholding compensation violates Rambha's right to withdraw from a trial, and (iii) it therefore also vitiates her consent to continued participation.

(i) Rambha ought to receive full compensation because investigators would otherwise violate an obligation they owe her, i.e. the obligation not to incentivise a course of action that imposes serious risks on her health. As any difference in payment between Rambha's continued participation and her withdrawal would in fact incentivise Rambha to pursue such a course of action, paying the full amount is the only way in which investigators can avoid violating their obligation.

(ii) When Rambha ought to receive full compensation, withholding payment penalises her withdrawal: it withholds from Rambha what she is entitled to. Penalising withdrawal also violates Rambha's right to withdraw from a trial. As a widely accepted official document on clinical conduct, the *CIOMS Guide* explains: the right to withdraw

requires that someone be able to withdraw consent "without penalty or loss of benefits to which he or she would otherwise be entitled."[55]

(iii) Influences which violate the right to withdraw consent also vitiate consent to continued participation. As I argued in 2.2.1, this is because the protection afforded by the right to withdraw consent is not only implied by the general principle of self-determination, but also by the requirement that consent be voluntary. For if an influence violates a person's right to withdraw consent, it qualifies as a controlling influence and is therefore incompatible with consent continuing to be voluntary. Hence, if withholding compensation incentivises Rambha to consent to staying enrolled, her consent will be involuntary for that reason.

Therefore, increased risks *can* lead to a situation where certain modes of payment, e.g. payment upon completion only, vitiate consent to continued participation in a clinical trial.

This argument requires further elaboration in two respects. Firstly, the same line of reasoning may apply when participants become ill during a trial, without the trial procedures having caused the illness. Secondly, people can be entitled to *some* compensation even if they do not have to withdraw on health grounds. They may be entitled to pro rata compensation due to a prior agreement, yet lack the legal means to enforce it. In such cases, withholding payment may also significantly penalise withdrawal and thereby vitiate consent to continued participation.

To conclude: unjustified risks do not vitiate consent if investigators could not have detected them, as shown in *Unknown Risk.* If investigators know about these risks, consent is vitiated only if investigators illegitimately reduce a person's options or penalise

[55] CIOMS Guide, Guideline 9. Other documents echo this position. See section 2.2.1 in Part One of this thesis.

withdrawal.[56] Hence, *Defeater Candidate 3: Risk* led my discussion to certain scenarios where Rambha's consent was indeed involuntary, even if risks alone could not vitiate consent.

5.3 ICCJ on Paid Trial Participation

The debate on paid research participation is divided. On the one hand, people present the three challenges to the voluntariness of consent which I have discussed so far and focus on a person's mental state (aversion), her background conditions (no acceptable alternative), and some aspect of the procedure she consents to (risk). On the other hand (an issue which I have not discussed), people claim that payments *never* vitiate consent because they can never really be an instance of coercion or undue inducement.[57]

Both sides of the debate fail. The former fails because none of the three challenges, taken by themselves, can explain when payments vitiate consent. The latter fails because proponents focus too much on conceptual requirements of coercion or undue inducement to see that some monetary arrangements penalise withdrawal or that investigators sometimes illegitimately withhold options from participants.

In this section, I will show that my own proposal *Interpersonal Consenter-Consentee Justification*, unlike these other views, accounts for all the results reached so

[56] Unfortunately, neither the ethics literature nor policy documents have discussed how investigators can penalise withdrawal. Neither the relevant EU Directive, nor the *Convention*, nor the *Additional Protocol*, nor *Medicines* says anything about modes of payment. The documents referred to are: (1) *EU Directive* 2001/20/EC. (2) *Convention* for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. (3) *Additional Protocol* to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. (4) The *Medicines* for Human Use (Clinical Trials) Regulation 2004. An exception is the CIOMS guide, which says that participants who withdraw because they became ill during the trial should receive as much money as they would have received upon completion. See CIOMS-Guide, Guideline 7. However, the CIOMS Guide does not offer any justification for this view and only presents is as a policy recommendation.

[57] See for instance Wertheimer and Miller, 2008. Emanuel, 2004, 2005a, and 2005b.

far and thereby presents a novel and more convincing approach to consent in the context of clinical research.

With respect to *Defeater Candidate 1*, Vanderpool and Grant claimed that payments vitiate consent if participants are offered money to do something they are averse to, and they presented the cases of an impecunious college student and a religious person as examples. ICCJ follows the result that this position is false and explains the voluntariness of consent with reference to the principles fundamental to the consenter-consentee relationship.

In situations of aversion, payments are in accordance with *Assisted Decision-Making* as they do not compromise a person's decision-making, leave the final decision-making authority to participants, and still allow the investigator to assist participants in their deliberation about whether to enrol. Moreover, *Assisted Decision-Making* can account for cases where people *unreasonably* but still *voluntarily* consent to trial participation (e.g. Jack, who would violate duties at work, the religious person, or the impecunious college student may all unreasonably yet still voluntarily consent to enrolment.). *Assisted Decision-Making* can account for these cases as it explicitly distinguishes between the voluntariness of a decision and its reasonableness. To recall: as *Assisted Decision-Making* emphasises that the person consenting has the final say, it is possible that his decision remains unreasonable or ill-judged in some way. Yet, a person ought to be respected in his decision-making and, other things being equal, can still consent voluntarily.

Furthermore, payments are in accordance with *Equal Moral Standing* as they respect people's entitlement to pursue a conception of the good and decide whether it is worth overcoming an aversion in exchange for money. In fact, *Equal Moral Standing* explicitly claims that people can valuably pursue a conception of the good even when

they eventually consent to something with great reluctance or aversion. Investigators offering payments in these circumstances do not impose their own values on the person consenting and they do not restrict but rather expand that person's options to pursue a conception of the good. By contrast, excluding someone from clinical trials on the basis of his aversion would impose values on a person that he currently does not endorse and restrict his options to pursue a conception of the good. As a result, such exclusion would express a negative judgment about a person's ability to make a reasoned decision and compromise his equal moral standing.

Finally, payments are in accordance with *Trusted Partner* as they satisfy both central norms of trust, i.e. the *proviso of hypothetical transparency* and *loyalty in power*. In situations of aversion, there is no concern over transparency and investigators do not abuse their position of power when offering payment in exchange for trial participation.

With respect to *Defeater Candidate 2,* ICCJ explains why McGregor and Olsaretti are mistaken when they claim that payments vitiate consent if participants enrol only for want of an acceptable alternative.

ICCJ's commitment to respecting a person's autonomy requires that people be allowed to enrol in trials when they find themselves in dire circumstances. Respect for autonomy requires allowing Rambha to make her own decision as long as she is not under any wrongful or dominating influence.

Moreover, ICCJ's commitment to beneficence requires acknowledging that people's well-being depends not solely on their physical well-being but also on the satisfaction of other important needs, to which paid trial participation can contribute.

On this basis, payment to people in situations without acceptable alternatives is consistent with ICCJ's more specific principles. It is consistent with *Assisted Decision-Making* because the investigator can still actively assist but not control a person's

decision-making. Investigators would in no way restrict but only expand a person's option and refrain from controlling a person's otherwise available options.

Moreover, obtaining consent in situations without acceptable alternatives is also consistent with *Equal Moral Standing.* In fact, *Equal Moral Standing* explicitly holds that people can valuably pursue a conception of the good when there are only very few options to choose from. Once the situation has become dire, it may even become particularly important for a person to be able to select what aligns best with her values and preferences. Thus, the offer of paid trial participation may be a significant contribution to a person's ability to pursue her conception of the good.

In addition, obtaining consent from a person in situations without an acceptable alternative is also consistent with *Trusted Partner.* Investigators would neither use non-transparency nor their position of power to dominate a person's decision-making and thereby avoid betraying the trust that prospective participants can expect them to honour.

With respect to *Defeater Candidate 3,* ICCJ explains why *unknown* risks cannot vitiate consent. ICCJ follows my results in Chapter 2.3.3 and focuses on the actual interaction between consenter and consentee, and on what the latter could be expected to have known. Accordingly, it holds that consent can be vitiated only if the person receiving consent violated a norm she could be expected to abide by.

Furthermore, ICCJ also explains why *known* risks vitiate consent if investigators thereby also illegitimately withhold a further option from a prospective participant. In the respective version of *Deviant Scientist,* Sarah completely ignores Rambha's standpoint when she subordinates Rambha's critical interest in the protection of her health to the comparatively minor scientific interests in saving time and money. Therefore, neither on the basis of the obligation to respect Rambha's autonomy nor on the basis of the

obligation to act beneficently can Sarah justify to Rambha how she obtains her consent to paid trial participation.

As a result, Sarah violates *Assisted Decision-Making* because reducing Rambha's option illegitimately constrains Rambha's situation and thereby compromises her opportunity to make a free decision. Sarah's conduct is also particularly serious because Sarah aims to make Rambha do what greatly differs from her otherwise existing preference, because Rambha is in fact entitled to another option, and because Rambha lacks any feasible means to protect herself against Sarah's dominance.

Moreover, Sarah also violates *Trusted Partner* because she abuses her professional power over which trials to conduct and dominates Rambha's decision-making. This is also a particularly serious instance as Sarah greatly relies on her position of power and Rambha is in a very weak position.

However, ICCJ holds that Rambha's consent is voluntary in the first version of *Deviant Scientist,* where Sarah does not withhold any options from Rambha. Here, Sarah does not illegitimately constrain Rambha's situation, ignore Rambha's standpoint, or subordinate Rambha's critical interest in the protection of her health to comparatively minor scientific interests. Sarah allows Rambha to make a decision on her own and expands rather than restricts Rambha's available options. Thus, Sarah can interpersonally justify to Rambha how she obtains her consent to paid trial participation and Rambha's consent is voluntary for that reason.

Finally, ICCJ's focus on the interpersonal interaction between participant and investigator accounts for cases where modes of payment penalise withdrawal and thereby vitiate consent. In *Adverse Effect*, Sarah cannot interpersonally justify to Rambha how she maintains Rambha's consent to continued participation because interpersonal

justification requires taking seriously the interests of the person consenting and thereby opposes the subordination of the legitimate participants' interests to scientific interests.

Sarah violates *Assisted Decision-Making* because penalising withdrawal illegitimately threatens to harm Rambha if she does not consent, renders non-consent costly in a significant way, weakens Rambha's standing in the decision-making, and thereby unduly controls rather than supports Rambha in her decision-making. This is also a particularly serious incident because Rambha is entitled to compensation, there are serious risks for her health, and she is in very weak position compared to Sarah.

Sarah also violates *Trusted Partner* because she abuses her position of power to dominate Rambha's decision-making and thereby betrays the trust participants must be able to bestow upon her.

Hence, ICCJ maintains that Sarah vitiates Rambha's consent in *Adverse Effect* and, on this basis, ICCJ opposes certain modes of payment, e.g. payment upon completion of a trial only.

Conclusion

This chapter focused on the question of whether and, if so, under what conditions paying research subjects vitiates their consent to participation. I answered that paying participants vitiates consent only in certain cases of increased risks when investigators illegitimately withhold an option of safer trials or penalise withdrawal. In all other cases, payments do no vitiate consent. ICCJ's interpersonal focus, unlike other views in the literature, accounted for these results and explained them in greater detail. I therefore hope to have supported the three novel aspects I wanted to bring to the debate: (i) freeing the discussion from conceptual queries about coercion, manipulation, or undue inducement, (ii) switching the debate from its focus on a person's mental states and

circumstances to a focus on the interaction between participant and investigator, using ICCJ to explain this interaction in greater detail, and (iii) highlighting how investigators can illegitimately withhold options as well as how they can penalise withdrawal by means of certain modes of payment.

6. Living Organ Donation

Nashville, USA, Vanderbilt Transplantation Centre. Sarah is fifteen years old and suffers from cystic fibrosis, a genetic disorder that damages a person's lungs and reduces the quality and length of her life considerably. Currently, there is no known cure. As Sarah's condition is worsening, her parents James and Debby consider donating parts of their lungs to her, which would greatly improve her condition, and approach the medical team about this. James and Debby undergo tests and will soon receive the results.[1]

Whereas some people feel very gratified if they can donate and help their child, others are extremely frightened and secretly hope not to be compatible. In interviews, the latter admit to be "scared to death"[2] and "terrified all the way down."[3] Some even say that "[i]t feels like having a gun put to your head, it's like being taken hostage, so what can you do, there is absolutely nothing (...) this is a decision about life and death, it's not about getting a pay rise or buying a car. This is real life drama."[4]

Yet, many of the most frightened eventually feel compelled to donate because they see themselves as under a moral obligation, experience implicit social pressure, or are subject to outright third-party coercion. Empirical evidence shows that, in one form or another, "[p]ressure to donate is a reality."[5]

These situations pose a serious challenge to the voluntariness of consent: is a person's consent to surgery still voluntary if given under any of these conditions of pressure? After all, the standards for consent must be very high: someone would be

[1] This is a case from actual clinical practice, reported by McGee, 1996, 104. The parents' names are fictitious.
[2] See Forsberg, 2004.
[3] See Forsberg, 2004. Further interview statements from frightened people can be found in Knibbe, 2007, 437.
[4] Forsberg, 2004, 374.
[5] Eggeling, 1999, 20. See also Lennerling, 2003. Crowley-Matoka, 2004. Forsberg, 2004. Al-Khader, 2005.

consenting to a procedure that is risky, highly intrusive, and lacking in any medical benefit for him.

Given the need for organs and the ubiquity of pressure, this is a very urgent question. Yet, it has not received sufficient attention in the literature on living organ donation. The current literature offers a fair amount of qualitative interview data, revealing donors' motivations and experiences, as well as several policy recommendations on how to assist people in their decision-making.[6] However, there is very little ethical analysis of the voluntariness of consent.

In this chapter, I will defend my proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* as the correct approach to voluntary consent in this context of organ donation. According to ICCJ, a person's consent is voluntary if and only if the surgeon can interpersonally justify tolerating the pressure on the prospective donor's decision-making when obtaining his consent.[7] ICCJ provides a distinct view by focusing on the interaction between the surgeon and the donor rather than, as all other views do, directly on the psychological distress or social pressure that the donor is subject to. But as I shall argue, this feature not only sets ICCJ apart from other views, it also makes it superior to them.

I will proceed as follows: after explaining various types of pressure in greater detail (6.1), I will outline how the few accounts in the current literature propose to evaluate them and reject these accounts on the basis of their adherence to three misguided assumptions (6.2). Finally, I propose ICCJ as a distinct and more convincing approach to the voluntariness of consent in the context of organ donation (6.3).

[6] See Al-Khader, 2005. Biller Adorno, 2011.
[7] I will use the terms "surgeon" and "physician" interchangeably in this chapter.

6.1 The Prospective Donor's Predicaments

I claimed that people may feel compelled to donate because of moral, implicit social, or outright third-party pressure. In this section, I will give an exemplary scenario for each type of pressure and explain three different challenges to the voluntariness of a donor's consent.

Scenario 1: Moral Pressure

Suppose the tests reveal that James is the only person in his family eligible to donate. If he refuses, Sarah will continue to suffer severely or even die. James is extremely frightened and would strongly prefer not to undergo transplantation surgery. Yet, he recognises a strong moral obligation to donate. The following statements from donor interviews could be representative of him:

> "I felt I was obliged to do it. I wanted to do it. There was really no choice. It was all about trying to save my daughter. They could have whatever they needed."[8]

> "It's my daughter so... there is really nothing to discuss."[9]

> "As my responsibility as a family member (…) I have to do it."[10]

James eventually consents under high psychological distress and only for want of another morally acceptable alternative.

---

[8] Forsberg, 2004, 374.
[9] Lennerling, 2003, 1244.
[10] Majeske, 1996, 94. Further similar statements can be found in: Fujita, 2004, 1426. Knibbe, 2007, 437-438.

Scenario 2: Implicit Social Pressure

Suppose Sarah needs *two* living donors, each giving part of their lungs, and both James and Debby are suitable donors. Debby instantly decided to donate and the family is now waiting for James to make a decision too. James believes that waiting for a cadaveric organ would be an acceptable option and that he is therefore no longer morally obliged to donate.[11] Yet, he feels unable to provide a justification for non-donation that his family would find convincing,[12] and the prospect of failing to provide such justification, to be perceived as a coward or failing father, or even to be blamed for non-donation exerts considerable pressure. It makes it seem impossible for James to act on his actual preference against surgery. As Crouch and Caplan explained:

"Some parents may not want to donate, in fact, may not even feel morally obligated to donate, but would be ashamed to have this attitude revealed to anyone else, especially to family. The opinions of others can be powerful motivating forces."[13]

"Family members will ordinarily feel extraordinary pressure to 'volunteer'. The realization that one could be blamed for the failure to help a spouse, a sibling, or a child may be so frightening that potential donors see themselves as having no choice."[14]

Such implicit social pressure can be problematic and incompatible with the voluntariness of James' consent for three reasons.

[11] 'Cadaveric organ' means an organ from a dead donor.
[12] These situations seem well-documented in clinical practice. For instance, in her interview research, Brown documented a case where a donor understood that their "reservations about (…) donation were valid" (Brown, 2014, 278), but that she "did not perceive that others in the family would find them as a justification for not proceeding with donation" (Brown, 2014, 278).
[13] Crouch, 1999, 277.
[14] Caplan, 1997, 118. See also Mazaris, 2006, 488.

(1) The situation can be problematic because giving James the choice of donation could already harm him as it allows others to blame James if he declines donation.[15] By way of explanation: before Sarah's condition worsened and James became a suitable donor, donation was not an option. Thus, James could not *choose* but only *accept* his situation as a non-donor and others could not ask James to justify any aspect of it. But after Sarah's condition worsened and James became a suitable donor, not undergoing surgery ceased to be the *non-chosen* default. Whatever James does now will be seen as his *choice* and as something he can be called on to justify. In Velleman's words: "He can now choose the status quo [i.e. no surgery] or choose the alternative [i.e. surgery], but he can no longer *have* the status quo without *choosing* it."[16] Thus, James is harmed because he lost non-donation as his *non-chosen* default situation and because he may now be blamed for anything he decides to do.

(2) In addition, James' situation can be problematic because he is under great psychological distress, i.e. distress arising from the combination of his fear of failing to justify non-donation to others and his great aversion to undergoing surgery. The concern about psychological distress may be reinforced once James no longer experiences it because of a moral obligation, but only because of his fear about failing to meet social expectations.

(3) Finally, James's situation can be problematic because he may be unable to continue the relationship with his family as before if he does not donate. His wife and daughter may be greatly disappointed and change their attitudes towards James, even if they do not explicitly say so or threaten James.

---

[15] This section draws on some of Velleman's claims. In his discussion of a right to euthanasia, Velleman claims that giving people an additional option (e.g. whether to consent to a procedure such as euthanasia) can sometimes be harmful to them, even if they are infallible decision-makers. Therefore, we may have good reasons not to allow people to consent to certain procedures.

[16] Velleman, 2015, 10.

Scenario 3: Third Party Pressure

Finally, suppose James is the only person in the family eligible to donate but extremely frightened about the surgery. Moreover, as in scenario 2, he considers waiting for a cadaveric organ an acceptable option and therefore no longer feels morally obliged to donate. However, James's family members disagree and pressure him. His wife Debby tells James that she will divorce him and fire him from the family company unless he donates. Divorce and unemployment would leave James in a very poor financial situation and even make him homeless. James eventually consents but only to prevent Debby from carrying out her threat.

Is James's consent voluntary in these different scenarios?

6.2 Willingness, Relationships, and Moral Agency

6.2.1 The Views in the Literature

The literature on organ donation is divided over the question of when moral or implicit social pressure render consent involuntary. On the one hand, authors defend the

Simple Willingness View Consent is involuntary if a person's unwillingness to donate, i.e. his reluctance, aversion, or fear, exceeds a certain threshold.

Valapour defends a version of this view when she claims that moral or social pressure render consent involuntary if such pressure "is experienced in such a way that a donor feels he/she cannot refuse donation even though he/she does not want to donate."[17]

[17] Valapour, 2008, 198.

When Valapour says "that a donor feels he/she cannot refuse donation", she refers to the fact that the alternative to donation is unacceptable, e.g. because it would lead to the death or severe suffering of one's child, violate an important moral commitment, or damage one's relationships with others.

When Valapour refers to a donor who "does not want to donate", she describes the extent to which a person would prefer not to donate or, in other words, *how unwillingly* a person donates. Here, Valapour explicitly links her account to Al-Khader, who proposed that we "determine the degree of willingness or unwillingness"[18] in order to "screen out donors who are acting under duress"[19] and therefore fail to give their consent voluntarily. On Al-Khader's view, donors are unwilling to the extent that they consent with reluctance, aversion, or fear.

Valapour and Al-Khader make unwillingness central to their view because they consider unwillingness to indicate a conflict with important interests of a person and they claim that, if such conflict is great enough, a person's decision ceases to be truly his own and becomes involuntary.

The *Simple Willingness View* is very influential in the debate on living organ donation. Caplan, Otte, and others have endorsed versions of this view.[20] I called it the *Simple* Willingness View because it does not specify further when an influence renders a person's decision no longer truly his own and because it assumes that a high degree of unwillingness normally has this effect.

The *Simple Willingness View* aligns with common parlance, where the distinction between voluntary and involuntary is often explained in terms of a distinction between willing and reluctant. For instance, *The Oxford English Dictionary* presents 'reluctant' as

[18] Al-Khader, 2005, 338.
[19] Al-Khader, 2005, 338.
[20] See Adams, 1987. Caplan, 1997. Otte, 2004. Biller Adorno, 2011.

synonymous with 'unwilling' and explains the latter as meaning 'involuntary.'[21] And the *Simple Willingness View* also seems to have allies in other philosophical debates. For instance, in the debate on legal responsibility, Hart and Honoré claimed that "[w]hen we take a step reluctantly, as the lesser of two evils, this is one way in which, for common sense and the law, our act is something less than fully voluntary."[22] Moreover, in 20th-century philosophy of action, both Anscombe and Kenny contrasted voluntariness with reluctance.[23]

However, the *Simple Willingness View* has not gone unchallenged. Knibbe opposes the *Simple Willingness View.* He denies that a person's decision ceases to be truly his own or becomes involuntary if he consents highly reluctantly or fearfully and does so only for want of another acceptable alternative. In fact, Knibbe claims that if a mother "wants to become a donor in spite of her fear [this] is an expression of who she is as a mother and of what is most important to her",[24] and therefore consistent with donation being truly her own decision. Moreover, "parents' perception of having no choice should be understood as a result of commitment to their child"[25] and thus also consistent with donation being truly their own decision. And the same seems to apply to James: if he consents due to seeing himself under a moral obligation, his decision may be truly his own, despite his fear, reluctance, and lack of acceptable alternatives.

It is most noteworthy that Knibbe's critique still accepts the link between voluntariness and willingness: he claims that a situation of no choice and great fear or reluctance "does not always limit willingness to donate"[26] and is therefore consistent with voluntariness. Knibbe's critique only challenges the view that moral pressure reduces

[21] See *Oxford English Dictionary:* "voluntary" 7a; "involuntary" 1a; "willing" 3; "unwilling" 2a and 5.
[22] Hart and Honoré, 1985, 156.
[23] See Anscombe, 2000, 89. Kenny, 1985, 31-32.
[24] Knibbe, 2007, 438.
[25] Knibbe, 2007, 436.
[26] Knibbe, 2007, 438.

willingness. In order to reduce willingness and preclude a decision to be a person's own decision, pressure must be sufficiently *external* to a person's agency, and moral pressure, potentially unlike social pressure, is not external in that way.

Knibbe argues that the mistake of the *Simple Willingness View* was the adherence to a picture of the "'moral agent' as being self-interested, independent and free from emotional bonds, commitment and duties."[27] Such an account, Knibbe argues, must have necessarily led to the false view that any pressure opposing self-interests must be detrimental to a person's agency, autonomy, and the voluntariness of his consent. This is a false view because, after all, people are not atomistic units. Their lives and identities are inseparably connected to others.

In order to overcome a false individualistic view of agency and understand voluntariness in the context of living organ donation, Knibbe suggests that "the interpretation of living donor voluntary consent must be based on a *relational view of the moral agent.*"[28] This relational view holds that "relations are seen as constitutive of moral agency",[29] which means that the capacity to act from moral reasons, and being autonomous for that matter, requires that individuals actually have valuable relations with others. As Stoljar seconded, "the presence of certain external conditions is necessary for an agent to be fully or globally autonomous."[30] On this view, agency and autonomy are not merely the non-social result of another social process. Agency and autonomy can only exist and be exercised in the presence of actual relations and if such relations meet certain standards, e.g. not being abusive.

---

[27] Knibbe, 2007, 435.
[28] Knibbe, 2007, 436-437. Emphasis added. Majeske seconds Knibbe's view by suggesting "a theoretical ethical framework that recognises interrelatedness. An ethic of care may seem to be the most compatible approach." Majeske, 1996, 95.
[29] Knibbe, 2007, 436.
[30] See Stoljar, 2011, 377. Stoljar talks about autonomy here, whereas Knibbe focuses on agency. The difference between these two concepts does not matter at this point. The crucial assumption here is that people need *actual* relations to exercise their relational agency as well as their relational autonomy.

Crouch, Spital, and Majeske defend similar views and emphasise the relational aspects of agency and autonomy.[31] They all defend the following second major view in the debate on living organ donation:

Relational Willingness View Consent is involuntary if the pressure on a person's decision-making exceeds a certain threshold *and* fails to promote valuable relationships.

Hence, James's consent is voluntary because the pressure he experiences in fact influences him to act in a way that promotes a valuable relationship. In such a case, James is fully willing to donate and his decision is truly his own.

Let me now move on to third-party pressure. Whereas the literature is divided on how to evaluate voluntariness in situations of moral and implicit social pressure, it is united in how to evaluate voluntariness in cases of third-party pressure. All the views[32] align with the majority in the general debate on consent and focus on the quality of the influence only, e.g. how controlling this influence is and whether it is illegitimate, not on how the recipient of consent responds to the third-party's influence.[33] Thus, they all accept the following view:

Influence-Focus View Consent is involuntary if a third-party's influence on the consenter's decision-making is illegitimate and sufficiently controlling.

[31] See Crouch, 1999. Spital, 2005. Majeske, 1996.
[32] See for instance Adams, 1987, 568. McGee, 1996, 106. Caplan, 1997, 123. Otte, 2004, 318. Valapour, 2008, 198. Chessa, 2014. Hays, 2014.
[33] See for instance Feinberg, 1986. Appelbaum, 2009. Nelson, 2011.

6.2.2 Critique of the Views in the Literature

The literature on organ donation suggests that evaluating the voluntariness of consent requires focusing on willingness, relational moral agency, how strong third-party influences are, and whether or not they are illegitimate. In this section, I will argue that these aspects do not determine the voluntariness of consent. More specifically, I will make three claims: (I) voluntariness cannot be explained in terms of willingness; (II) the standards of relational moral agency cannot inform judgments about the voluntariness of consent; and (III) in cases of third-party pressure, the voluntariness of consent depends on the physician's control over the prospective donor's decision-making rather than on the third-party's influence. As these three claims challenge the fundamental assumptions in the debate on living organ donation, they ask for a new approach to the voluntariness of consent.

I. Separating Voluntariness from Willingness

Voluntariness is not the same as willingness and involuntariness is not the same as unwillingness or reluctance. A person may very reluctantly consent to a major operation in order to save his life or may very reluctantly consent to exposure therapy to treat his phobia. Yet, he can still consent voluntarily. Similar examples also arise when it comes to overcoming one's reluctance for moral reasons instead of prudential interests in one's own health. As Knibbe correctly pointed out, a mother may reluctantly and fearfully consent to donation but can still do so voluntarily. The same is true for James's decision to donate for moral reasons.

In all of these cases, people are strong-willed and self-determined. They do not let themselves be carried away by an immediate preference for comfort, but distance themselves from such preferences and act on a firm decision. To this extent, they are

autonomous, and the reluctance that accompanies their acts does not diminish their autonomy. Quite the contrary, it is a sign of autonomous agency as it shows that they are able to manage their different preferences.

But if reluctance does not undermine autonomy, it cannot undermine the voluntariness of consent either. This is because, in the context of consent, autonomy and voluntariness are closely linked: voluntariness is a condition of the validity of consent, and such conditions are supposed to protect the autonomy of the person giving consent. Hence, conditions of valid consent can only be violated if there is a flaw in the protection of a person's autonomy, but not otherwise. Therefore, for consent to be involuntary there has to be an infringement of a person's autonomy. Thus, as reluctance is consistent with autonomy, it is also consistent with the voluntariness of consent.

Therefore, the distinction between reluctance and willingness does not map onto the distinction between involuntariness and voluntariness. Hyman was correct when he claimed:

"Whether someone does an act willingly and whether she does it voluntarily are different questions."[34]

"Being reluctant contrasts with being willing, and 'voluntary' does not mean the same as 'willing'. (…) An act is done willingly if it is done with a certain attitude or in a certain frame of mind, if the person welcomes doing it or at least does it without reluctance, whereas an act is done voluntarily if the agent does not do it out of ignorance or compulsion (…) 'Voluntary' does not describe a person's attitude or state of mind, whereas 'reluctant' and 'willing' do exactly that. That is why an individual can be willing or reluctant but not voluntary or compulsory, whereas

[34] Hyman, 2015, 96.

an option, such as second-year logic, can be voluntary or compulsory but not willing or reluctant."[35]

Hence, willing and reluctant, unlike voluntary, are psychological terms describing a person's state of mind. Therefore, the debate on organ donation needs to reject the view that voluntariness and involuntariness can be explained in terms of willingness and reluctance. For this reason, the *Simple Willingness View* is false and the *Relational Willingness View* needs to cut its ties to claims about willingness.

II. Relational Moral Agency versus the Voluntariness of Consent

Knibbe's account could be reformulated and separated from the reference to willingness. But even if a modified version aligns with the distinction between voluntariness and willingness, the standards of relational moral agency still cannot inform judgments about the voluntariness of consent. Consider the following case:

Estranged Father Bob has a sixteen-year old daughter, Carry, whom he has not had contact with since she was born. Upon hearing that she needs a kidney, he feels great moral pressure to donate and consents to surgery. Donating a kidney will not promote the relationship to his daughter, as her family made clear that there would be no contact afterwards, and Bob's current family strongly opposes donation. Moreover, Bob's decision to donate would be ill-judged because there is in fact no moral obligation for him to donate to Carry, other donors who are in better health have already agreed to donate, and Bob's current family urgently needs him to stay

[35] Hyman, 2015, 84.

> healthy during the next few weeks so that he can support another sick child in their family.[36]

In some respects, Bob is like James: his daughter needs an organ, he is frightened about surgery, and he feels morally compelled to donate. In other respects, Bob is unlike James: Bob does not have any actual contact or relation to his daughter and donation will not change this; Bob's current family opposes donation and their relation to Bob will become worse if he donates; and finally, Bob's decision to donate – unlike James's – would be ill-judged.

In this situation, Bob's donation would not promote his relationship with his sick daughter Carry because he lacks any actual relations to her and will not be able to establish any following the donation. Views of relational moral agency make clear that they consider the existence of actual relations necessary for relational moral agency and emphasise that "neither biological nor institutionalized family ties fully determine (…) relatedness"[37] as pertinent to relational moral agency.

In addition, donation would harm Bob's relationship with his current family because donation would disrupt his current family and violate Bob's commitment to supporting his family in difficult times. Bob's current family may even ostracise him if he decides to put his health at risk and fails to support them.

Therefore, on Knibbe's view, high moral pressure vitiates Bob's consent, as Bob's decision to donate is ill-judged and fails to promote his relationships to others.

---

[36] Alternatively, there may also be the case of a young college student who, after studying the ethics of altruism, feels great moral pressure to become an anonymous donor. In this case, donation would not promote the relationship to the person receiving the organ but could still disrupt the student's relationship to others, e.g. his family or partner. But even if this is so and he makes an ill-judged decision, he can still consent voluntarily to donation. Thanks to Dominic Wilkinson for suggesting this example.

[37] Majeske, 1996, 100.

However, considering consent involuntary on these grounds seems to be the wrong result. Making an ill-judged decision and failing to promote relationships does not render one's consent involuntary. If Bob, like James, consents because he considers himself under a moral obligation to do so, such consent can be voluntary even if given with great reluctance and distress. The cases of James and Bob are on a par with regards to the voluntariness of consent. Hence, Knibbe's relational view is too strict in cases of moral pressure.

But even beyond cases of *moral* pressure, relational views still seem to arrive at the wrong result, for example, in cases where people consent due to implicit social pressure. Recall the second scenario of James's situation:

Implicit Social Pressure Sarah needs *two* living donors, each giving parts of his lungs, and both James and Debby were shown to be suitable. Debby instantly decided to donate and the family is now waiting for James to make a decision too. James no longer thinks that he is morally obliged to donate but feels unable to provide a justification for non-donation. The prospect of failing to justify non-donation to his family, to be perceived as a coward or failing father, or even to be shamed or blamed for non-donation exerts considerable pressure and makes James consent despite his actual preference against surgery.

Suppose James consents because he wants to avoid damaging his family relationships, which he values much more than avoiding surgery. After talking to the surgeon about his situation, James requests to proceed with donation.

In such a situation, James will have made a judgment about what is most important for him; and the implicit social pressure he is subject to is not created by a particular person. In fact, it is more like a background condition that no one explicitly created or is able to remove; and background conditions do not normally vitiate consent.

In addition, rejecting James as a donor would reduce his control over the situation, whereas allowing James to make a decision maintains James's control and thereby respects his autonomy. Therefore, even if James only consents to avoid damaging his family relationships, his consent can be given autonomously and therefore voluntarily.

However, Knibbe's relational view seems unable to account for this result. Knibbe could support the judgment that James's consent is voluntary only if James's decision is not ill-judged, if donation is conducive to his relational moral agency, and if the implicit social pressure is significantly similar to the moral pressure Knibbe's argument focuses on. However, none of these conditions is necessary for the voluntariness of James's consent. Following the results from *Estranged Father,* the voluntariness of James's consent cannot depend on whether his decision is ill-judged or promotes relational moral agency; and even if social expectations are external to James in a way in which his moral commitment in the previous example way was not, such expectations do not vitiate consent for that reason. Deciding whether to consent remains a valuable exercise of autonomy under implicit social pressure and James's consent can therefore be voluntary.

Given these arguments concerning *Estranged Father* and *Implicit Social Pressure,* we need to develop an account for the connection between the voluntariness of consent and a person's moral agency that is different from Knibbe's. The following line of reasoning provides an alternative.

When a person consents to a medical procedure, he chooses what aligns with his values and preferences, and such values and preferences go beyond general medical

considerations. For instance, a pianist reluctantly accepts greater side effects of some treatment option to avoid another treatment that could paralyse his hands; a person fearfully rejects blood transfusion to avoid violating a religious commitment; and a father reluctantly consents to strong chemotherapy to increase the chances of supporting his family for longer. In all these cases, people purse their conception of the good and thereby exercise what Rawls called a *moral power*.

Exercising this moral power can be understood as exercising one's *moral agency,* broadly construed. Moral agency has to be broadly construed because exercising the Rawlsian moral power does not always involve acting from moral reasons, narrowly understood, but can sometimes only mean acting from reasons related to one's conception of the good. On this reading, moral agency allows people to choose an individualistic life or a life full of social bonds and devotion, a life of religious commitments, or one without them. Thus, this account of moral agency is much wider than Knibbe's. It is not narrowed down to a particular view of relational agency or autonomy, but is rather based on the respect for people's entitlement to choose whatever conception of the good or life plan they think is best for them. Any of these life plans are to be respected and informed consent is an instrument that purports to give people the opportunity to pursue whatever they consider best when choosing a medical procedure.

Therefore, the requirements of informed consent (competence, information, voluntariness) must not be brought into conflict with moral agency so defined. Rather, they must be understood so that a person's consent can become both an expression of his self-determination in general, and also an expression of his individual moral agency in particular.

Experiencing psychological pressure in cases of organ donation, seeing no morally acceptable alternative, and consenting only to fulfil social expectations can all

fall within a person's pursuit of their conception of the good. Hence, considering the donor's consent invalid on these grounds would be misguided: such a view would imply that the exercise of one's moral agency (e.g. seeing no morally acceptable alternative and experiencing pressure) can render void what is supposed to facilitate one's moral agency (e.g. informed consent).

To be clear, I make a negative claim here: exercising one's moral agency cannot vitiate consent if what is consented to is otherwise acceptable. I do not make the positive claim that exercising one's moral agency ensures the validity of consent. For instance, a person may offer to donate his heart to a family member and thereby exercise his moral agency. However, donating one's heart lies outside the scope of what people can consent to and the fact that donation would be an exercise of one's moral agency does not change this. With this qualification, however, the conclusion holds: it is inconsistent to consider consent invalid solely on the grounds that a person experiences the sort of psychological or implicit social pressure at hand or sees no morally acceptable alternative.

Therefore, as long as no other person explicitly pressures someone into consenting to a medical procedure, consent can still be voluntary.

Before I conclude, it is worth highlighting that this conclusion is not defeated by Velleman's concern mentioned earlier, which was about how giving people a choice can harm them. To address Velleman's concern, I first need to clarify its precise focus.

Velleman addresses a question of policy, not a question concerning the validity of consent. He asks whether we should give people a choice to consent to something and not whether they can still consent validly, *if* they have that choice.

Moreover, Velleman presents his argument as being based on the "policy of nonmalfeasance rather than on the policy of paternalism."[38] Velleman's claim is that "we

[38] Velleman, 2015, 15.

must not harm others by giving them choices, not that we must withhold the choices from them lest they harm themselves."[39] After all, James could be an infallible decision-maker and choose what is best for him. But even in such cases, Velleman argues, we will have reasons not to give him a choice about whether to undergo a certain procedure if the harm of giving him that choice is too high. It seems that this is also the position of the *Council of Europe* when it states that, in cases of extreme emergency, living organ donation "should not be recommended."[40]

Therefore, even if Velleman is right and people should not be given the choice of donation in certain situations, for reasons based on a policy of nonmalfeasance, this view does not imply that people lose the capacity to give voluntary and valid consent *if* they have that choice. One may accept that people may strongly prefer not to face a certain choice. One may even accept that it is not valuable for them to have that choice and that, as a matter of policy, people should not have that choice. However, this does not affect the claim that, *if* they are given the choice, they still have the *ability* to consent validly and voluntarily. Hence, even if James is harmed, he can retain his ability to give voluntary consent.

III. Against the Influence-Focus View

Finally, recall the third scenario of James's situation: Debby tells James that she will divorce him and fire him from the family company unless he donates. Divorce and unemployment would leave James in a very poor financial situation and even make him homeless. James eventually consents but only to prevent Debby from carrying out her threat.

---

[39] Velleman, 2015, 15.
[40] *The Council of Europe*, 1997, Recommendation No. R (97) 16, (i). Otte (2004) echoes this position.

The literature on organ donation unanimously claims that the voluntariness of consent depends on whether a third-party's influence is illegitimate and sufficiently controlling. Thus, Debby's threat vitiates James's consent if and only if it is illegitimate and sufficiently controlling. *Pace* these views, I will defend the claim that in cases of explicit third-party pressure the voluntariness of James's consent depends on how the surgeon reacts to the situation, and not directly on Debby's threat.

To defend this claim, I will firstly focus on a different question, namely whether the surgeon could ever permissibly proceed with surgery in such a case. Consider the following scenario:

Final Say James's family knows about his physiological compatibility as a donor and keeps pressuring him. After James informed the surgeon about his situation, the latter agrees with James to contact his family. They meet for a family conversation where the surgeon explains the viability of cadaveric donation and emphasises the importance of making the decision to donate free of pressure. There is also a donor advocate present, i.e. a psychologist supporting James's position. Yet, the family cannot settle the issue and medical personnel are unable to alleviate James's situation. After some deliberation, James requests that the surgeon proceed with the donation. On the basis of the overall situation, he says donating is best for him. He insists that the surgeon no longer offer any help, contact his family, or interfere with his private decision.

Refusing James as a donor would violate James's entitlement to make the final decision. It would also put him in a situation that is, according to his own evaluation, worse for him. On the other hand, accepting James's decision respects him as having the final authority over what to do, at least in the interaction between him and the surgeon, and

allows him to pick the option he considers best for him under the circumstances. Furthermore, having offered any possible help for James, the surgeon did not fail to discharge obligations owed to James. For these reasons, the surgeon can permissibly proceed with surgery.

Contrast this scenario with another case:

No Help According to the hospital's policy, a person will be identified as an *incompatible* donor due to physiological, psychological, or other reasons. 'Other reasons' are broadly understood and include people's requests to be identified as incompatible when pressured by others. The exact reasons for incompatibility will be kept confidential within the medical team and not released to anyone except the tested person. At the moment, only James knows about his physiological compatibility as a donor. His family has not been informed yet and James does not want them to know. He asks for a general statement of incompatibility. Issuing this statement would be in line with the hospital's policy and James's family would subsequently withdraw their pressure. Yet, the surgeon deviates from the usual policy and only issues a specific statement on physiological donor compatibility. James's family keeps pressuring him and he subsequently consents to donating a part of his lungs.

In this case, the surgeon could have helped James very easily. It was possible and feasible for him to provide James with a general statement of incompatibility. Yet, the surgeon decided to deny James this opportunity to escape the threats from his family. The surgeon thereby controlled the circumstances in which James had to make his decision and made him worse off, compared to an otherwise possible scenario. For these reasons, it would be impermissible for the surgeon to perform the surgery.

No Help and *Final Say* do not differ in terms of third-party pressure. In both scenarios, James consents to surgery only because Debby pressured him. However, they differ in terms of the surgeon's reaction: whereas the surgeon did everything he could in *Final Say*, the surgeon in *No Help* failed to help James where doing so was possible and feasible. Therefore, if my evaluation is correct and performing surgery is impermissible in *No Help* but permissible in *Final Say*, permissibility cannot always depend on third-party pressure. The permissibility of surgery can depend on the surgeon's reaction to the situation.

This result can inform judgments about the voluntariness of James's consent. This is because the debate on living organ donation unanimously accepts the Permissibility Principle, discussed in Chapter 2.3, which implies that performing donation surgery can only be permissible if consent is voluntary.[41] Thus, if permissibility requires voluntariness, and if performing surgery is found permissible, then James's consent must have been voluntary. This line of reasoning can be described a simple modus ponens:

1) X is a necessary condition of Y (e.g. voluntary consent is a necessary condition for permissible surgery).
2) Y is given (e.g. surgery is permissible).
3) Therefore, X (e.g. James's consent is voluntary).

To clarify, this argument does *not* claim that the permissibility of surgery explains the voluntariness of consent and thereby reverses the correct explanatory order. It is still the case that the voluntariness of consent explains the permissibility of surgery. This

[41] See for instance Rhodes (2014, 307) who explicitly mentions and defends this view.

argument only makes explicit what follows once we accept the Permissibility Principle and consider surgery permissible.

In order to account for the conclusion in (3), we need to adopt a certain account of voluntary consent. Such an account can start with the premise that voluntary consent requires that the person's decision-making is not controlled by others. But then, one needs to claim that control over a person's consent can take different forms: *global control* is the control coming from outside the interaction between the person giving consent and the person receiving consent, e.g. pressure from third parties like Debby. By contrast, *transactional control* is the control coming from within the interaction between the person receiving consent and the person giving consent, e.g. if the recipient of consent himself threatens the consenter, or if he keeps him in his predicament where it would be possible and feasible to help, as was the case in *No Help*.[42]

In order to accept the Permissibility Principle and the evaluation of the different scenarios of James's situation, we need to adopt the view that the voluntariness of consent depends on *transactional control* rather than *global control*. In *Final Say*, James's consent is voluntary because it was not transactionally controlled, whereas James's consent is involuntary in *No Help* because it was in fact transactionally controlled.

Hence, the voluntariness of consent depends on how the person receiving consent, e.g. the surgeon, reacts to the situation and not just on the influence exerted by third parties. The influence of third parties, and their strength and illegitimacy, may inform what the person receiving consent is obliged to do, but cannot directly vitiate consent.

[42] To clarify: I will consider "exerting the transactional control required to vitiate consent" to be extensionally equivalent to "obtaining consent *by means* of coercion or another seriously illegitimate influence" as explained in Chapter 2.3, and continue to assume that a recipient can vitiate consent in these ways either by exerting an influence himself or by failing to discharge his obligations owed to the person consenting.

Thus, the views in the organ donation literature fail to acknowledge how the surgeon's response can change the permissibility of surgery and that such response determines the voluntariness of consent. Therefore, they cannot explain the difference between *No Help* and *Final Say* and ought to be dismissed.

6.3 ICCJ on Living Organ Donation

The last section rejected three assumptions from the debate on living organ donation: it rejected the views that (I) voluntariness can be explained in terms of willingness, (II) standards of relational moral agency inform the voluntariness of consent, (III) in cases of third-party pressure, the voluntariness of consent depends on whether a third-party's influence is illegitimate and sufficiently controlling. Having rejected these assumptions, we now need a different approach to the voluntariness of consent in the context of living organ donation. In this section, I will defend my own proposal ICCJ as a distinct and superior approach.

According to ICCJ, a person's consent is voluntary if and only if the surgeon can interpersonally justify that he tolerated the pressure on a person's decision-making when obtaining this person's consent. ICCJ provides a distinct view by focusing on the interaction between the surgeon and the prospective donor rather than, as all other views do, directly on psychological distress or social pressure. But this feature not only sets ICCJ apart from other views, it also makes it superior to them.

ICCJ can explain why moral and implicit social pressure do not vitiate consent. ICCJ's principle *Assisted Decision-Making* is satisfied in cases of moral and implicit social pressure because the surgeon, when obtaining consent, would not compromise a person's ability or opportunity to make a free decision. He would neither threaten to harm a patient if he did not consent, render non-consent costly in some way, prevent the donor

from taking part in some aspects of the decision-making, nor weaken a donor's emotional or cognitive capacities for decision-making. In fact, the surgeon could still actively assist the prospective donor by means of a shared process of deliberation. Moreover, simply allowing moral and implicit social pressure to be consistent with voluntary consent already enhances the process of shared decision-making: it refrains from encouraging people to hide their feelings out of fear of being rejected as a donor, and it creates a safe space for an open discussion. Finally, *Assisted Decision-Making* explicitly explains that consent can sometimes be ill-judged but still voluntary, and thereby accounts for cases like *Estranged Father.*

ICCJ's principle *Equal Moral Standing* explicitly requires that the surgeon allow the donor to exercise his Rawlsian moral power and not interfere when the donor pursues his conception of the good. Thus, even if people experience distress or consent with great reluctance, they may still engage in what *Equal Moral Standing* highlights as worthy of protection. As already mentioned, people can valuably pursue a conception even when they eventually consent to something with great reluctance or when they see themselves without acceptable alternatives. In fact, once the situation has become like that, it may be particularly important for a person to be able to select what aligns best with his values and preferences. Therefore, tolerating moral and implicit social pressure can be interpersonally justified.

In addition, ICCJ's principle of *Trusted Partner* is also satisfied. A surgeon would rely neither on non-transparency nor his superior position of power to dominate rather than assist a person's decision-making. In fact, there is no concern over transparency at all and the surgeon is not in any relevant position of power with regards to moral or implicit social pressure.

Hence, accepting consent given under moral pressure can be interpersonally justified, and therefore both James's and Bob's consent is voluntary in the examples described earlier. Thus, ICCJ is superior to other views by accounting not only for the voluntariness of James's consent but also for the voluntariness of Bob's consent in *Estranged Father.*

In addition, ICCJ also presents a more convincing proposal regarding third-party pressure because it can explain the difference between the scenarios mentioned earlier. In *No Help,* the surgeon *cannot* interpersonally justify his failure to help James. He deliberately deviates from clinical routine and makes the statement more specific so that James will run into the threats of his family. Providing the general documents would have required neither considerable effort nor the violation of any general medical obligation. By not providing the statement, the surgeon compromises the decision-making process: he deliberately prevents James from making a decision free from pressure and thereby violates *Assisted Decision-Making.* Moreover, the surgeon violates *Trusted Partner*: he fails to use his position of professional power to help James and alleviate the pressure on his decision-making. Hence, ICCJ explains why performing the surgery in *No Help* is impermissible and bases this explanation on the claim that consent is involuntary due to the illegitimate transactional control exerted by the surgeon.

By contrast, in *Final Say,* the surgeon *can* interpersonally justify that he tolerates the influence and obtains consent. He talks to James about this situation, offers advice, acknowledges James as the final decision-maker, and thereby satisfies *Assisted Decision-Making.* Moreover, he accepts James's request, avoids imposing his own opinion on him, and allows James to pursue his conception of the good, thereby satisfying *Equal Moral Standing.* Finally, he relies on neither non-transparency nor his superior position of power to dominate James's decision in any way, and thereby satisfies *Trusted Partner.* Hence,

ICCJ explains why performing the surgery in *Final Say* is permissible and bases this explanation on the claim that consent is voluntary because the surgeon did not exert undue transactional control over James's decision.

Thus, ICCJ provides a distinct and novel position: unlike other views, it focuses on how the recipient of consent deals with third-party pressure, rather than on the features of the third-party influence. This position is also superior to other accounts because, unlike them, ICCJ aligns with the Permissibility Principle, the evaluation of *Final Say* and *No Help*, and the view that links voluntariness to transactional rather than global control.

Before I conclude, I want to explain that *Assisted Decision-Making* sometimes obliges a physician to provide *false* documentation on donor incompatibility; and if a physician fails to discharge such an obligation when it arises, consent will be involuntary for that reason.[43]

Suppose James's brother-in-law threatens him with severe physical injury if he does not donate and the only way for James to escape this threat is to provide false documentation on his incompatibility. ICCJ implies that, in such cases, it is not only *permissible* for the physician to fake documentation but it is *obligatory* for him to do so.

According to ICCJ, there is a presumption in favour of providing false documentation in these cases, and if such a presumption cannot be defeated, it will be obligatory for the physician to provide false documentation. The presumption arises because providing false documentation serves the two fundamental principles of ICCJ. It serves the principle of beneficence by saving James from severe harm, i.e. either physical injury or undergoing an intrusive operation that he does not want to undergo. And it serves

[43] To remind the reader: I use the terms "surgeon" and "physician" interchangeably. I will refer to the "physician" in the rest of this chapter as it is the more general term and better suited to the claims I will make. However, nothing substantial depends on this terminological choice.

the principle of respect for autonomy as it allows James to regain control over his situation and diminishes the influence of his brother-in-law.

This presumption can be defeated if the physician, *qua physician*, is under an obligation *not* to provide false documentation; and in fact, truthfulness is an important value in the medical context and among professional obligations. Such professional obligations limit the scope to which physicians are obliged to act beneficently and promote their patients' autonomy. However, to see whether this obligation of truthfulness defeats the presumption in favour of providing fake documents in the scenario under consideration, we need to put this obligation into context.

The obligation of truthfulness primarily purports to enable people's self-determination when they decide whether to undergo a medical procedure. A physician needs to tell the truth so that a person can base his decision about medical procedures on correct information. In addition, the norm of truthfulness also purports to maintain trust in the medical profession more generally. It is important that the public can rely on expert statements from physicians.

In the case at hand, telling the truth about donor compatibility would weaken James's self-determination, i.e. it would weaken the self-determination of the person who is about to decide whether to undergo a medical procedure. By contrast, providing false documents would enhance James's self-determination. Hence, following the obligation of truthfulness would not fulfil its primary purpose, while deviating from it would promote such purpose.

In addition, if made public afterwards, having provided false documentation is unlikely to undermine public trust in the medical profession. After all, the physician will only have produced fake documentation to prevent great harm to a prospective donor.

Furthermore, deceiving the coercing party does not seem to be of great concern because the physician is not obliged to facilitate his control over the situation and such control does not seem worthy of protection in the first place. James's brother-in-law may even have forfeited his right to receive correct information.

But there may be some concern about the deception of the potential recipient of the organ. Even though this person may not be entitled to receive information about other people's compatibility, she might make the claim that *if* she is given information, it must be true information so that she can base her deliberation on facts.

Hence, the question arises as to whether the stakes for the prospective donor can be so high that a physician becomes obliged to provide false documents and override the potential entitlement of the prospective recipient of an organ to receive correct information.

According to ICCJ, the physician's decision has to be interpersonally justifiable *to the person giving consent* and such interpersonal justification requires providing fake documentation under the following conditions: if the obligation of truthfulness does not serve its primary purpose, if deviating from such an obligation greatly enhances the control and self-determination of the prospective donor, if there is no other way of helping him, and if there is an immediate threat of severe harm that greatly outweighs the concern for the harm done to the potential recipient of the organ by not telling her the truth. Not providing fake documents would be unjustifiable under these conditions because it would unduly prioritise the norm of truthfulness and the perspective of the potential recipient of the organ over the severe third-party coercion that the prospective donor is subject to and cannot escape in any other way.

Although this implication does not present a precise formula for when physicians become obliged to provide false documentation, it provides guidance on how to evaluate

such a situation: ICCJ identifies a presumption in favour of faking documents, relates to the purpose of the obligation of truthfulness, and requires the balancing of harms on the side of the prospective donor and the harms on the side of the patient waiting for an organ.

This implication of ICCJ opposes the position of McGee and Mazaris who object to such a proposal:

"Offering medical reasons when none exist is problematic, however, because it requires deliberately misleading the patient and this undermines the trust essential to the physician-patient relationship. Perhaps even more important, deception undermines the integrity of the family unit."[44]

"Physicians (…) must avoid lying or falsifying medical documents, both of which can result in catastrophic consequences."[45]

However, ICCJ's implication is a more stringent version of Caplan's position, who argued:

"Transplant teams must even be prepared to lie to protect the ability of prospective donors to make voluntary choices. (…) transplant centers ought to be willing to offer a prospective donor the option of disqualification on medical grounds, even when none exists, in order to ensure that donors understand that they can choose not to participate without being stigmatized, ostracized, or penalized."[46]

[44] McGee, 1996, 108.
[45] Mazaris, 2006, 488.
[46] Caplan, 1993, 1999. See also Caplan, 1997, 123-124. Kallich, 1994, 152.

It is a more stringent version, as being socially stigmatized would not be enough for ICCJ to oblige physicians to provide false medical documentation. According to ICCJ, it has to be a severe and immediate threat against which there is no other means of protection. Hence, the pressure people experience when facing implicit social expectations, as discussed earlier, would not justify providing false documents.

Conclusion

This chapter discussed three types of pressure which people like James can be subject to when considering donating an organ to a family member: moral pressure, implicit social pressure, and third-party pressure. I defended my proposal *Interpersonal Consenter-Consentee Justification (ICCJ)* as the correct approach to voluntary consent. ICCJ provided a distinct view by focusing on the interaction between the surgeon and the prospective donor rather than, as all other views do, directly on psychological distress or social pressure. This feature not only sets ICCJ apart from other views but also makes it superior to them: it allows ICCJ to account for the distinction between voluntariness and willingness, the need to respect whatever conception of the good people choose to pursue, and to accommodate the *Permissibility Principle* in cases of third-party pressure.

Therefore, ICCJ's most distinct feature, i.e. the focus on the interaction between the person giving and the person receiving consent, proved to be appropriate even in cases where strong influences come from psychological distress or from the influence of third parties, and not from the person receiving consent.

These results are particularly important for two reasons: firstly, they explain voluntariness in situations where a person consents to something that offers no medical or financial benefit to him at all, but only aims to benefit another person. Secondly, they explain voluntariness in an area that is, from the perspective of the donor, neither

treatment nor research, but rather a third class of medical procedures which I have not considered before. Hence, understanding voluntariness in this area was required in order to complete a comprehensive picture of the practical challenges to the voluntariness of consent.

Conclusion

Consent protects self-determination: it empowers people to decide how others can permissibly interact with them. But consent can protect self-determination only if people consent *voluntarily*. Being competent and informed does not suffice.

In this thesis, I focused on consent to medical procedures and, developing my proposal of *Interpersonal Consenter-Consentee Justification (ICCJ),* I claimed that consent is voluntary unless it is motivated by an influence that cannot be interpersonally justified. I explained that interpersonal justification requires that the consentee (i.e. the person receiving consent) can justify to the consenter (i.e. the person giving consent) that he exerted these influences or, when influences had a different source, tolerated their presence. The principles fundamental to the consenter-consentee relationship provided the basis for such a justification.

ICCJ introduced the concept of interpersonal justification to the debate on consent and most clearly differs from other views by (i) focusing on the reasons or motivation for which a person consents, instead of mere causal links or counterfactuals, (ii) allowing the moral legitimacy and not just the strength of an influence to have a bearing on the voluntariness of consent, and (iii) making the behaviour of the person receiving consent (the consentee) central to its approach.

On this basis, ICCJ is lenient in some respects: it considers consent voluntary in situations of great reluctance, psychological distress, lack of acceptable alternatives, and even some cases of third-party pressure. ICCJ is strict in other respects, however: it considers consent involuntary when the consentee fails to discharge certain obligations owed to the consenter, e.g. in the case of third-party pressure, even if the consentee does not collaborate with the pressuring third party at all. Moreover, ICCJ acknowledges that

certain influences vitiate consent, e.g. the threat to abandon a patient, even if the extent to which it affects a particular patient is rather low.

I have demonstrated that ICCJ can be supported on the basis of a *Top-Down* approach and general desiderata, as well as on the basis of a *Bottom-Up* approach and the way it deals with particular practical challenges.

However, over and above contributing to a topic in practical ethics, ICCJ also contributes to the more general debate on voluntariness that started in 20th-century philosophy of action and cut ties with theories postulating volitions. In fact, many of my results contribute to the modern post-volitional understanding of voluntariness, e.g. the rejection of counterfactual models of causation and the emphasis on the reasons behind people's consent. Likewise, my view on the relation between intentional action and the voluntariness of consent, my account of how the voluntariness of consent can depend on the morality of another person's influence, and the disambiguation of what I termed the *Coercion Principle* have also contributed to the understanding of voluntariness more generally.

Finally, ICCJ identifies three areas for further important and promising research. First, ICCJ encourages the transfer of the idea of interpersonal justification from the context of medical procedures to other areas of consent, e.g. consent to sexual relations. Second, ICCJ opens up new paths for investigating liability and responsibility. This is because, as Hart explained, there is an

"analogy between conditions that are treated by criminal law as excusing conditions and certain similar conditions that are treated in another branch of the law as invalidating certain civil transactions such as wills, gifts, contracts, and marriages."[1]

[1] Hart 2008, 29.

Thus, just as certain forms of pressure avert liability for wrongdoing, they also vitiate consent. Hence, some of the more general claims in this thesis, e.g. about counterfactuals, causation, and the morality of influences, could be transferred to the debate on wrongdoing and inform judgments about responsibility and liability.

Third and finally, ICCJ connects and contrasts voluntariness with many other concepts (e.g. autonomy, reluctance, willingness, intentions, causation, moral legitimacy, and coercion) and thereby creates the possibility for what Strawson would call a *connective conceptual analysis*, as opposed to a *reductive conceptual analysis*.[2] The former, unlike the latter, does not aim to reduce a concept into simpler parts but rather describes the links to other concepts. Employing the results from this thesis in a connective conceptual analysis may explain the ways in which the concept of voluntariness is significant in moral philosophy more generally.

Thus, I conclude: ICCJ presents a novel view on the voluntariness of consent to medical procedures, it contributes to the larger project of understanding voluntariness independent of theories postulating volitions, and it also identifies three further areas of important and promising research.

[2] See Strawson, 1992, Chapter 2.

References

- Abrams, N. (1982). Scope of Beneficence in Health Care. In E. Shelp (Ed.), *Beneficence and Health Care* (pp. 183–198). Reidel Publishing.
- Adams, R. (1987). Live Organ Donors and Informed Consent. A Difficult Minuet. *The Journal of Legal Medicine*, *8*, 555–586.
- Aggarwal, A., Davies, J., & Sullivan, R. (2014). "Nudge" in the Clinical Consultation – an Acceptable Form of Medical Paternalism? *BMC Medical Ethics*, *15*, 1–6.
- Al-Khader, A. (2005). A Model for Scoring and Grading Willingness of a Potential Living Related Donor. *The Journal of Medical Ethics*, *31*(6), 338–340.
- Alexander, L. (1996). The Moral Magic of Consent (II). *Legal Theory*, *2*(3), 165–174.
- Alexander, L. (2008). Scalar Properties, Binary Judgments. *Journal of Applied Philosophy*, *25*, 85–104.
- Alexander, L. (2014). The Ontology of Consent. *Analytic Philosophy*, *55*, 102–113.
- An, S. (2008). Antidepressant Direct-to-Consumer Advertising and Social Perception of the Prevalence of Depression: Application of the Availability Heuristic. *Health Communication*, *23*(6), 499–505.
- Anscombe, G. E. M. (1979). Under a Description. *Nous*, *13*, 219–233.
- Anscombe, G. E. M. (2000). Intention. Harvard University Press.
- Anscombe, G. E. M. (2008). Sin. In M. Geach & L. Gormally (Eds.), *Faith in a Hard Ground Essays on Religion, Philosophy and Ethics by G.E.M. Anscombe* (pp. 139–188). Imprint Academic.
- Anscombe, G. E. M. (1981). From Parmenides to Wittgenstein, Volume 1. Wiley-Blackwell.
- Anscombe, G. E. M., & Morgenbesser, S. (1963). The Two Kinds of Error in Action. *The Journal of Philosophy*, *60*, 393–401.
- Appelbaum, P., Klitzman, R., & Lidz, C. (2009). Voluntariness of Consent to Research: A Conceptual Model. *The Hastings Center Report*, *39*(1), 30–39.
- Archard, D. (1998). Sexual Consent. Westview Press.
- Aristotle. (2000). Nicomachean Ethics. Translated with Introduction, Notes, and Glossary by Terence Irwin. Second Edition. Hackett Publishing.
- Arneson, R. (1980). Mill versus Paternalism. *Ethics*, *90*, 470–489.
- Bader, R. (2016). Moralized Conceptions of Liberty. In D. Schmidtz & C. E. Pavel (Eds.), *The Oxford Handbook of Freedom* (pp. 59–72). Oxford University Press.
- Bader, R. (2018). Moralizing Liberty. *Oxford Studies in Political Philosophy*, *4*, 141–166.
- Baldwin, R. (2014). From Regulation to Behaviour Change: Giving Nudge the Third Degree. *The Modern Law Reivew, 77*(6), 831–857.

- Banerjee, A. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, *107*, 797–817.
- Barnhill, A. (2014). What is Manipulation? In C. Coons & M. Weber (Eds.), *Manipulation: Theory and Practice* (pp. 51–72). Oxford University Press.
- Barton, A., & Grüne-Yanoff, T. (2015). From Libertarian Paternalism to Nudging - and Beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359.
- Beauchamp, T., & Childress, J. (2013). Principles of Biomedical Ethics (7th edition). Oxford University Press.
- Belfrage, S. (2016). Exploitative, Irresistible, and Coercive Offers: Why Research Participants Should Be Paid Well or Not at All. *Journal of Global Ethics*, *12*(1), 69–86.
- Bell, E., Dubljevic, V., & Racine, E. (2013). Nudging Without Ethical Fudging: Clarifying Physician Obligations to Avoid Ethical Compromise. *The American Journal of Bioethics*, *13*, 18–19.
- Bentley, J. (2004). The Influence of Risk and Monetary Payment on the Research Participation Decision Making Process. *The Journal of Medical Ethics*, *30*(3), 293–298.
- Berg, J., Appelbaum, P., Lidz, C., & Parker, L. (2001). Informed Consent. Legal Theory and Clinical Practice. Oxford University Press.
- Bergelson, V. (2007). The Right to Be Hurt: Testing the Boundaries of Consent. *The George Washington Law Review*, *75*, 165–236.
- Bergelson, V. (2010). Consent to Harm. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent. Theory and Practice* (pp. 163–183). Oxford University Press.
- Bergelson, V. (2018). Consent to Pain. In P. Schaber & A. Müller (Eds.), *The Routledge Handbook of the Ethics of Consent* (pp. 199–210). Routledge.
- Biller-Andorno, N. (2011). Voluntariness in Living-Related Organ Donation. *Transplantation*, *92*(6), 617–619.
- Blumenthal-Barby, J. S. (2012a). Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts. *Kennedy Institute of Ethics Journal*, *22*(4), 345–366.
- Blumenthal-Barby, J. S. (2012b). Seeking Better Health Care Outcomes: The Ethics of Using the "Nudge." *The American Journal of Bioethics*, *12*(2), 1–10.
- Boddington, P. (2012). Ethical Challenges in Genomics Research. Springer.
- Bovens, L. (2009). The Ethics of Nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change Approaches from Philosophy, Economics and Psychology* (pp. 207–219). Springer.
- Bowman, D., Spicer, J., & Iqbal, R. (2012). Informed Consent. Cambridge University Press.
- Brett, N. (1998). Sexual Offenses and Consent. *Canadian Journal of Law and Jurisprudence an International Journal of Legal Thought*, *11*(1), 69–88.

- Bright, P., & Nelson, R. (2012). A Capacity-based Approach for Addressing Ancillary Care Needs: Implications for Research in Resource Limited Settings. *The Journal of Medical Ethics*, *38*(11), 672–676.
- Brown, C. (2014). Pressure and Coercion. In J. Steel (Ed.), *Living Donor Advocacy. An Evolving Role Within Transplantation* (pp. 275–291). Springer.
- Buss, S. (2005). Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints. *Ethics*, *115*(2), 195–235.
- Calo, R. (2014). Code, Nudge, or Notice? *Iowa Law Review*, *99*, 773–802.
- Caloia, B. (2015). Psychic Costs and Broad Learning Effects: Locating Coercion and Threats to Autonomy in the Use of "Nudges" in Public Policy. *Public Affairs Quarterly*, *29*, 235–255.
- Caplan, A. (1993). Must I Be My Brother's Keeper? Ethical Issues in the Use of Living Donors as Sources of Liver and Other Solid Organs. *Transplantation Proceedings*, *25*, 1997–2000.
- Caplan, A. (1997). Am I My Brother's Keeper? Indiana University Press.
- Castelfranchi, C., & Falcone, R. (2010). Trust Theory. A Socio-Cognitive and Computational Model. John Wiley & Sons.
- Chessa, F. (2014). Informed Consent for Living Organ Donation. In J. Steel (Ed.), *Living Donor Advocacy. An Evolving Role Within Transplantation* (pp. 261–274). Springer.
- Chwang, E. (2008). Against the Inalienable Right to Withdraw from Research. *Bioethics*, *22*(7), 370–378.
- Coady, D. (2004). Preempting Preemption. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 1–16). Cambridge, Mass.: MIT Press.
- Cohen, G. A. (1995). Self-Ownership, Freedom, and Equality. Cambridge University Press.
- Cohen, S. (2013). Nudging and Informed Consent. *The American Journal of Bioethics*, *13*(6), 3–11.
- Collins, J. (2004). Preemptive Prevention. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 107–117). Cambridge, Mass.: MIT Press.
- Conly, S. (2012). Against Autonomy. Justifying Coercive Paternalism. Cambridge University Press.
- Crisp, R. (1997). Routledge Philosophy Guidebook to Mill on Utilitarianism. Routledge.
- Crouch, R., & Elliot, C. (1999). Moral Agency and the Family: The Case of Living Related Organ Transplantation. *Cambridge Quarterly of Health Care Ethics*, *8*, 275–287.
- Crowley-Matoka, M., Siegler, M., & Cronin, D. (2004). Long-Term Quality of Life Issues Among Adult-to-Pediatric Living Liver Donors: A Qualitative Exploration. *American Journal of Transplantation*, *4*(5), 744–750.
- Cullity, G. (2007). Beneficence. In R. E. Ashcroft, A. Dawson, H. Draper, & J. R. McMillan (Eds.), *Principles of Health Care Ethics* (pp. 19–26). John Wiley and Sons.

- Darwall, S. (2006). The Second-Person Standpoint. Harvard University Press.
- Davidson, D. (2001). Essays on Actions and Events. Oxford University Press.
- Dean, M., Kibris, Ö., & Masatlioglu, Y. (2017). Limited Attention and Status Quo Bias. *Journal of Economic Theory*, *169*, 93–127.
- Donchin, A. (2001). Understanding Autonomy Relationally: Toward a Reconfiguration of Bioethical Principles. *Journal of Medicine and Philosophy*, *26*(4), 365–386.
- Dougherty, T. (2013). Sex, Lies, and Consent. *Ethics*, *123*(4), 717–744.
- Dubov, A. (2015). Ethical Persuasion: The Rhetoric of Communication in Critical Care. *Journal of Evaluation in Clinical Practice*, *21*(3), 496–502.
- Dworkin, G. (1988). The Theory and Practice of Autonomy. Cambridge University Press.
- Dworkin, G. (2013). Lying and Nudging. *The Journal of Medical Ethics*, *39*(8), 496–497.
- Edwards, S. (2005). Research Participation and the Right to Withdraw. *Bioethics*, *19*(2), 112–130.
- Eggeling, C. (1999). The Psychosocial Implications of Live-related Kidney Donation. *EDTNA ERCA Journal*, *25*(3), 17–20.
- Emanuel, E. (2004). Ending Concerns About Undue Inducement. *The Journal of Law, Medicine Ethics*, *32*, 100–105.
- Emanuel, E. (2005a). Response to Commentators on "Undue Inducement: Nonsense on Stilts?" *The American Journal of Bioethics*, *5*(5), W8–W11.
- Emanuel, E., (2005b). Undue Inducement: Nonsense on Stilts? *The American Journal of Bioethics*, *5*(5), 9–13.
- Emanuel, E., Wendler, D., & Grady, C. (2000). What Makes Clinical Research Ethical? *Jama*, *283*, 2701–2711.
- Engelen, B. (2019a). Ethical Criteria for Health-Promoting Nudges: A Case-by-Case Analysis. *The American Journal of Bioethics*, *19*(5), 48–59.
- Engelen, B. (2019b). Nudging and Rationality: What is there to Worry? *Rationality and Society*, *31*(2), 204–232.
- Entwistle, V., & Watt, I. (2016). Broad versus Narrow Shared Decision Making. In G. Elwyn, A. Edwards, & R. Thompson (Eds.), *Shared Decision Making in Health Care: Achieving Evidence-Based Patient Choice* (pp. 7–12). Oxford University Press.
- Erb, H.-P., Bioy, A., & Hilton, D. (2002). Choice Preferences Without Inferences: Subconscious Priming of Risk Attitudes. *Journal of Behavioral Decision Making*, *15*, 251–262.
- Eyal, N. (2014). Nudging by Shaming, Shaming by Nudging. *International Journal of Health Policy and Management*, *3*(2), 53–56.
- Faden, R., & Beauchamp, T. (1986). A History and Theory of Informed Consent. Oxford University Press.

- Faden, R., & Faden, A. (1977). False Belief and the Refusal of Medical Treatment. *The Journal of Medical Ethics*, *3*(3), 133–136.
- Feinberg, J. (1973). Social Philosophy. Prentice-Hall, Inc. Englewood Cliffs, New Jersey.
- Feinberg, J. (1986). The Moral Limits of the Criminal Law: Volume 3: Harm to Self. Oxford University Press.
- Ferzan, K. (2015). Consent, Culpability, and the Law of Rape. *Ohio State Journal of Criminal Law*, *13*(2), 397–439.
- Flanigan, J. (2018). Sweatshop Regulation and Workers' Choices. *Journal of Business Ethics*, *153*(1), 79–94.
- Forsberg, A., Nilsson, M., Krantz, M., & Olausson, M. (2004). The Essence of Living Parental Liver Donation – Donors' Lived Experiences of Donation to Their Children. *Pediatric Transplantation*, *8*, 372–380.
- Frankena, W. (1987). Beneficence/Benevolence. *Social Philosophy and Policy*, *4*(2), 1–20.
- Freedman, B. (1987). Equipoise and the Ethics of Clinical Research. *The New England Journal of Medicine*, *317*, 141–145.
- Fujita, M., Slingsby, B. T., & Akabayashi, A. (2004). Three Patterns of Voluntary Consent in the Case of Adult-to-Adult Living Related Liver Transplantation in Japan. *Transplantation Proceedings*, *36*(5), 1425–1428.
- Furnham, A., & Boo, H. (2011). A Literature Review of the Anchoring Effect. *The Journal of Socio-Economics*, *40*, 35–42.
- Gardner, J. (2007). Offences and Defences. Oxford University Press.
- Gelfand, S. D. (2016). The Meta-Nudge - A Response to the Claim That the Use of Nudges During the Informed Consent Process is Unavoidable. *Bioethics*, *30*(8), 601–608.
- Gerver, M. (2016). Refugee Repatriation and the Problem of Consent. *British Journal of Political Science*, *48*(4), 855–875.
- Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*, *6*(3), 361–383.
- Gillick, M. (2015). Re-engineering Shared Decision-making. *The Journal of Medical Ethics*, *41*(9), 785–788.
- Gilovich, T., Medvec, V., & Savitsky, K. (2000). The Spotlight Effect in Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance. *Journal of Personality and Social Psychology*, *78*(2), 211–222.
- Glod, W., Department of Philosophy, Florida State University. (2015). How Nudges Often Fail to Treat People According to Their Own Preferences. *Social Theory and Practice*, *41*(4), 599–617.
- Goodwin, T. (2012). Why We Should Reject "Nudge." *Politics*, *32*(2), 85–92.

- Gordon, L. (1966). The Range of Application of 'Voluntary', 'Not Voluntary' and 'Involuntary.' *Analysis*, *26*(5), 149–152.
- Gorin, M. (2014a). Do Manipulators Always Threaten Rationality? *American Philosophical Quarterly*, *51*(1), 51–61.
- Gorin, M. (2014b). Towards a Theory of Interpersonal Manipulation. In C. Coons & M. Weber (Eds.), *Manipulation: Theory and Practice* (pp. 73–97). Oxford University Press.
- Grady, C. (2012). Ethical Principles in Clinical Research. In J. Gallin & F. Ognibene (Eds.), *Principles and Practice of Clinical Research* (pp. 19–30). Elsevier.
- Grande, S. W. (2015). The Reciprocity of Shared Decisions Across the Care Spectrum. *The American Journal of Bioethics*, *15*(9), 25–26.
- Grant, R., & Sugarman, J. (2004). Ethics in Human Subjects Research: Do Incentives Matter? *Journal of Medicine and Philosophy*, *29*(6), 717–738.
- Greenspan, P. (2003). The Problem with Manipulation. *American Philosophical Quarterly*, *40*(2), 155–164.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.
- Groll, D. (2015). Medicine and Well-Being. In G. Fletcher (Ed.), *The Routledge Handbook Of Philosophy Of Well-Being* (pp. 504–516). Routledge.
- Grubb, A., Laing, J., McHale, J., & Kennedy, I. (2010). Principles of Medical Law. Oxford University Press.
- Grüne-Yanoff, T. (2012). Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles. *Social Choice and Welfare*, *38*(4), 635–645.
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge Versus Boost: How Coherent are Policy and Theory? *Minds and Machines*, *26,* 1–35.
- Hackshaw, A. (2009). A Concise Guide to Clinical Trials. Wiley Blackwell.
- Hall, N. (2004). Two Concepts of Causation. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 225–276). Cambridge, Mass.: MIT Press.
- Halpern, S. (2016). Judging Nudges. *The American Journal of Bioethics*, *16*(5), 16–18.
- Halpern, S., Karlawish, J., Casarett, D., Berlin, J., & Asch, D. (2004). Empirical Assessment of Whether Moderate Payments Are Undue or Unjust Inducements for Participation in Clinical Trials. *Archives of Internal Medicine*, *164*, 801–803.
- Hanna, J. (2015). Libertarian Paternalism, Manipulation, and the Shaping of Preferences. *Social Theory and Practice*, *41*(4), 618–643.
- Hansen, P., & Jespersen, A. (2013). Nudge and the Manipulation of Choice. A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation*, *4*(1), 3–28.
- Hart, H. L. A. (1963). Law, Liberty, and Morality. Stanford University Press.
- Hart, H. L. A. (2008). Punishment and Responsibility. Oxford University Press.

- Hart, H. L. A., & Honoré, T. (1985). Causation in the Law. Oxford University Press.
- Hausman, D., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy*, *18*(1), 123–136.
- Hays, R. (2014). Contraindication to Living Donation from an ILDA Perspective. In J. Steel (Ed.), *Living Donor Advocacy. An Evolving Role Within Transplantation.* (pp. 205–219). Springer.
- Healey, R. (2015). The Ontology of Consent: A Reply to Alexander. *Analytic Philosophy*, *56*(4), 354–363.
- Herring, J. (2014). Relational Autonomy and Family Law. Springer.
- Herring, J. (2016). Medical Law and Ethics (6th Edition). Oxford University Press.
- Hitchcock, C. (2007). Prevention, Preemption, and the Principle of Sufficient Reason. *Philosophical Review*, *116*(4), 495–532.
- Hohfeld, W. N. (1913). Some Fundamental Legal Conceptions as Applied in Judicial Reasoning. *The Yale Law Journal*, *23*(1), 16–59.
- Holm, S. (2011). Withdrawing from Research: A Rethink in the Context of Research Biobanks. *Health Care Analysis*, *19*(3), 269–281.
- Houk, T., DiSilvestro, R., & Jensen, M. (2016). Smoke and Mirrors: Subverting Rationality, Positive Freedom, and Their Relevance to Nudging and/or Smoking Policies. *The American Journal of Bioethics*, *16*(7), 20–22.
- Hug, K., Hermerén, G., & Johansson, M. (2012). Withdrawal from Biobank Research: Considerations and the Way Forward. *Stem Cell Reviews and Reports*, *8*(4), 1056–1065.
- Hurd, H. M. (2009). The Moral Magic of Consent. *Legal Theory*, *2*(2), 121–146.
- Hyams, K. (2011). When Consent Doesn't Work: A Rights-Based Case for Limits to Consent's Capacity to Legitimise. *Journal of Moral Philosophy*, *8*(1), 110–138.
- Hyman, J. (2015). Action, Knowledge, and Will. Oxford University Press.
- Iyengar, S. (2010). The Art of Choosing. Hachette UK.
- Jackson, E. (2016). Medical Law. Oxford University Press.
- James, W. (1981). The Works of William James. The Principles of Psychology. Volume 1. (F. Burkhardt, F. Bowers, & I. Skrupskelis, Eds.). Harvard University Press.
- Joffe, S., & Truog, R. (2010). Consent to Medical Care: The Importance of Fiduciary Context. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent. Theory and Practice* (pp. 347–373). Oxford University Press.
- Johnson, E., & Goldstein, D. (2003). Do Defaults Save Lives? *Science, New Series*, *302*(5649), 1338–1339.
- Jonas, H. (1969). Philosophical Reflections on Experimenting with Human Subjects. *Daedalus*, *98*(2), 219–247.

- Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, *3*, 430–454.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292.
- Kahneman, D., Knetsch, J. L., & Thaler, R. (1991). The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193-206.
- Kallich, J., & Merz, J. (1994). The Transplant Imperative: Protecting Living Donors from the Pressure to Donate. *The Journal of Corporation Law*, *20*, 139–154.
- Kass, N. E., Faden, R. R., Goodman, S. N., Provonost, P., Tunis, S., & Beauchamp, T. (2013). The Research-Treatment Distinction: "A Problematic Approach for Determining Which Activities Should Have Ethical Oversight." *The Hastings Center Report*, *43*(1), S4–S15.
- Kazan, P. (1998). Sexual Assault and the Problem of Consent. In S. French, T. Wanda, & M. Purdy (Eds.), *Violence Against Women. Philosophical Perspectives* (pp. 27–42). Cornell University Press.
- Keil, G. (2007). What Do Deviant Causal Chains Deviate From? In C. Lumer & S. Nannini (Eds.), *Intentionality, Deliberation and Autonomy: the Action-theoretic Basis of Practical Philosophy* (pp. 69–90). Ashgate.
- Kelley, M., James, C., Alessi Kraft, S., Korngiebel, D., Wijangco, I., Rosenthal, E., et al. (2015). Patient Perspectives on the Learning Health System: The Importance of Trust and Shared Decision Making. *The American Journal of Bioethics*, *15*(9), 4–17.
- Kenny, A. (1985). The Ivory Tower: Essays in Philosophy and Public Policy. Oxford: Blackwell.
- Kenny, A. (1975). Will, Freedom, and Power. Oxford: Blackwell.
- Kiener, M. (2018). Truth-Telling and Respect for Autonomy. *The American Journal of Bioethics (Neuroscience)*, *9*(3), 193–194.
- Kiener, M. (2019). Consent, Threats, and Offers. *The American Journal of Bioethics*, *19*(9), 66–68.
- Knibbe, M., Maeckelberghe, E., & Verkerk, M. (2007). Confounders in Voluntary Consent about Living Parental Liver Donation: No Choice and Emotions. *Medicine, Health Care and Philosophy*, *10*(4), 433–440.
- Kon, A. (2012). Difficulties in Judging Patient Preferences for Shared Decision-making. *The Journal of Medical Ethics*, *38*(12), 719–720.
- Krawczyk, M., & Rachubik, J. (2019). The Representativeness Heuristic and the Choice of Lottery Tickets: A Field Experiment. *Judgment and Decision Making*, *14*(1), 51–57.
- Kristjánsson, K. (2007). Social Freedom. Cambridge University Press.
- Largent, E., Grady, C., Wertheimer, A., & Miller, F. (2012). Misconceptions about Coercion and Undue Influence: Reflections on the Views of IRB Members. *Bioethics*, *27*(9), 500–507.

- Laufer-Ukeles, P. (2011). Reproductive Choices and Informed Consent: Fetal Interests, Women's Identity, and Relational Autonomy. *American Journal of Law & Medicine*, *37*(4), 567–623.
- Lemmon, M., & Ubel, P.(2019). In Defense of Nudging When the Stakes Are High. *The American Journal of Bioethics*, *19*(5), 62–63.
- Lennerling, A., Forsberg, A., & Nyberg, G. (2003). Becoming a Living Kidney Donor. *Transplantation*, *76*(8), 1243–1247.
- Levine, R. (1988). Ethics and Regulation of Clinical Research. Yale University Press.
- Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567.
- Lewis, D. (2004). Causation as Influence. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 75–106). Cambridge, Mass.: MIT Press.
- Lillehammer, H. (2012). Autonomy, Value, and the First Person. In L. Radoilska (Ed.), *Autonomy and Mental Disorder* (pp. 192–213). Oxford University Press.
- Locke, J. (2009). An Essay Concerning Human Understanding. Oxford University Press.
- Loewenstein, G. (2005). Hot-Cold Empathy Gaps and Medical Decision Making. *Health Psychology*, *24*(4), S49–S56.
- Loewenstein, G., & O'Donoghue, T. (2006). "We Can Do This the Easy or the Hard Way": Negative Emotions, Self-Regulation, and the Law. *The University of Chicago Law Review*, *73*, 183- 206.
- Lowe, J. (2002). Routledge Philosophy Guidebook to Locke on Human Understanding. Routledge.
- Mackenzie, C., & Stoljar, N. (2000). Introduction. In C. Mackenzie & N. Stoljar (Eds.), *Relational Autonomy Feminist Perspectives on Autonomy, Agency, and the Social Self* (pp. 3–31). Oxford University Press.
- Maclean, A. (2009). Autonomy, Informed Consent and Medical Law. Cambridge University Press.
- Majeske, R., Parker, L., & Frader, J. (1996). Search of an Ethical Framework for Consideration of Decisions Regarding Live Donation. In B. Spielman (Ed.), *Organ and Tissue Donation. Ethical, Legal, and Policy Issues* (pp. 89–101). Southern Illinois University Press.
- Mallary, S., Gert, B., & Culver, C. (1986). Family Coercion and Valid Consent. *Theoretical Medicine*, *7*, 123–126.
- Malm, H. (2009). The Ontological Status of Consent and its Implications for the Law on Rape. *Legal Theory*, *2*(2), 147–164.
- Mandava, A., & Millum, J. (2013). Manipulation in the Enrollment of Research Participants. *The Hastings Center Report*, *43*(2), 38–47.

- Manson, N., & O'Neill, O. (2007). Rethinking Informed Consent in Bioethics. Cambridge University Press.
- Martens, S. (2007). Durch Dritte verursachte Willensmängel. Mohr Siebeck.
- Marzilli Ericson, K. M., & Fuster, A. (2014). The Endowment Effect. *Annual Review of Economics*, *6*(1), 555–579.
- Maslen, C. (2004). Causes, Contrasts, and the Nontransitivity of Causation. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 341–357). Cambridge, Mass.: MIT Press.
- Mason, J. K., & Laurie, G. T. (2011). Mason and McCall Smith's Law and Medical Ethics. Oxford University Press.
- Mayr, E. (2011). Understanding Human Agency. Oxford University Press.
- Mazaris, E., & Papalois, V. (2006). Ethical Issues in Living Donor Kidney Transplantation. *Experimental and Clinical Transplantation*, *2*, 485–497.
- Mazur, G. (2012). Informed Consent, Proxy Consent, and Catholic Bioethics. Springer.
- McConnell, T. (2010). The Inalienable Right to Withdraw from Research. *The Journal of Law, Medicine, and Ethics*, *38*(4), 840–846.
- McGee, C., & Fox, M. (1996). Bilateral Lung Transplantation from Living Donors. In B. Spielman (Ed.), *Organ and Tissue Donation. Ethical, Legal, and Policy Issues* (pp. 103–109). Southern Illinois University Press.
- McGregor, J. (2005). "Undue Inducement' as Coercive Offers. *The American Journal of Bioethics*, *5*(5), 24–25.
- McKenzie, C., Liersch, M., & Finkelstein, S. (2006). Recommendations Implicit in Policy Defaults. *Psychological Science*, *17*(5), 414–420.
- McLean, S. A. M. (2010). Autonomy, Consent and the Law. Routledge.
- McLeod, C. (2015). Trust. *Stanford Encyclopedia of Philosophy*.
- McMillan, J. (2018). The Methods of Bioethics. An Essay in Meta-Bioethics. Oxford University Press.
- Miller, D. (1983). Constraints on Freedom. *Ethics*, *94*(1), 66–86.
- Miller, F., & Brody, H. (2001). The Internal Morality of Medicine: An Evolutionary Perspective. *Journal of Medicine and Philosophy*, *26*(6), 581–599.
- Miller, F., & Brody, H. (2003). A Critique of Clinical Equipoise: Therapeutic Misconception in the Ethics of Clinical Trials. *The Hastings Center Report*, *33*(3), 19–28.
- Miller, F., & Wertheimer, A. (2010). Preface to a Theory of Consent Transactions: Beyond Valid Consent. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent. Theory and Practice* (pp. 79–105). Oxford University Press.
- Miller, F., Rosenstein, D., & DeRenzo, E. (2000). Professional Integrity in Clinical Research. *Jama*, *280*(16), 1449–1454.

- Miller, F., & Rosenstein, D. (2003). The Therapeutic Orientation to Clinical Trials. *The New England Journal of Medicine*, *348*(14), 1383–1386.
- Miller, P., & Weijer, C. (2007). Equipoise and the Duty of Care in Clinical Research: A Philosophical Response to Our Critics. *Journal of Medicine and Philosophy*, *32*(2), 117–133.
- Miller, P., & Fagley, N. (1991). The Effects of Framing, Problem Variations, and Providing Rationale on Choice. *Personality and Social Psychology Bulletin*, *17*(5), 517–522.
- Miller, P., & Weijer, C. (2006). Fiduciary Obligation in Clinical Research. *Journal of Law, Medicine, and Ethics*, *34*(2), 424–440.
- Millum, J. (2014). Consent Under Pressure: The Puzzle of Third Party Coercion. *Ethical Theory and Moral Practice*, *17*(1), 113–127.
- Millum, J., & Garnett, M. (2019). How Payment for Research Participation Can be Coercive. *The American Journal of Bioethics*, *19*(9), 21–31.
- Montoy, J., Dow, W., & Kaplan, B. (2018). Cash Incentives versus Defaults for HIV Testing: A Randomized Clinical Trial. *Plos One*, *13*(7), 1–10.
- Morreim, E. H. (2005). The Clinical Investigator as Fiduciary: Discarding a Misguided Idea. *The Journal of Law, Medicine, and Ethics*, *33*(3), 586–598.
- Moulton, B., & King, J. (2010). Aligning Ethics with Medical Decision-Making: The Quest for Informed Patient Choice. *The Journal of Law, Medicine, and Ethics*, *38*, 85–97.
- Mulgan, T. (2004). The Ethics of Killing: Problems at the Margins of Life by Jeff McMahan. *The Canadian Journal of Philosophy*, *34*, 443–459.
- Murphy, S., & Zajonc, R. (1993). Affect, Cognition, and Awareness: Affective Priming With Optimal and Suboptimal Stimulus Exposures. *Journal of Personality and Social Psychology*, *64*(5), 723–739.
- Nagel, T. (2012). Mortal Questions. Cambridge University Press.
- Nelson, R., & Merz, J. (2002). Voluntariness of Consent for Research: An Empirical and Conceptual Review. *Medical Care*, *40*(9), V69–V80.
- Nelson, R., Beauchamp, T., Miller, V. A., Reynolds, W., Ittenbach, R. F., & Luce, M. F. (2011). The Concept of Voluntary Consent. *The American Journal of Bioethics*, *11*(8), 6–16.
- Nelson-Marten, P., & Rich, B. A. (1999). A Historical Perspective of Informed Consent in Clinical Practice and Research. *Seminars in Oncology Nursing*, *15*(2), 81–88.
- Noggle, R. (2017). Manipulation, Salience, and Nudges. *Bioethics*, *32*(3), 164–170.
- Nozick, R. (1974). Anarchy, State, and Utopia. Basic Books.
- Nyhan, B., & Reifler, J. (2013). Which Corrections Work? *New America Foundation. Research Results and Practice Recommendation*, 1–15.
- O'Neil, C. (2012). Lying, Trust, and Gratitude. *Philosophy and Public Affairs*, *40*(4), 301–333.
- O'Neill, O. (2002). Autonomy and Trust in Bioethics. Cambridge University Press.

- Olsaretti, S. (2004). Liberty, Desert and the Market. Cambridge University Press.
- Olsaretti, S. (2016). Voluntariness, Coercion, Self-ownership. In D. Schmidtz & C. E. Pavel (Eds.), *The Oxford Handbook of Freedom* (pp. 439–455). Oxford University Press.
- Olthuis, G., Leget, C., & Grypdonck, M. (2014). Why Shared Decision Making is Not Good Enough: Lessons from Patients. *The Journal of Medical Ethics*, *40*(7), 493–495.
- Oswald, M., & Grosjean, S. (2004). Confirmation Bias. In R. Pohl (Ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgment and Memory* (pp. 79–96). Psychology Press.
- Otte, J.-B., Janssen, M., Rosati, M.-R., & Gonze, D. (2004). Parental Experience with Living-related Donor Liver Transplantation. *Pediatric Transplantation*, *8*, 317–321.
- Parfit, D. (1986). Reasons and Persons. Oxford University Press.
- Paul, L. A. (1998). Problems with Late Preemption. *Analysis*, *58*(1), 48–53.
- Paul, L. A. (2004). Aspect Causation. In J. Collins, E. Hall, & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 205–224). Cambridge, Mass.: MIT Press.
- Paul, L. A. (2013). Varieties of Redundant Causation. In L. A. Paul & N. Hall (Eds.), *Causation: A User's Guide* (pp. 70–172). Oxford University Press.
- Pellegrino, E. D. (1988). For the Patient's Good. Oxford University Press.
- Pellegrino, E. D. (1992). Beneficence, Scientific Autonomy, and Self-Interest: Ethical Dilemmas in Clinical Research. *Cambridge Quarterly of Health Care Ethics*, *1*(4), 361–369.
- Phillips, T. (2011). Exploitation in Payments to Research Subjects. *Bioethics*, *25*(4), 209–219.
- Ploug, T., & Holm, S. (2015). Doctors, Patients, and Nudging in the Clinical Context—Four Views on Nudging and Informed Consent. *The American Journal of Bioethics*, *15*(10), 28–38.
- Pufendorf, S. (1998). Samuel Pufendorf. Gesammelte Werke. (W. Schmidt-Biggemann, Ed.). Akademie Verlag.
- Quigley, M. (2014). Are Health Nudges Coercive? *Monash Bioethics Review*, *32*, 141–158.
- Quong, J. (2011). Liberalism Without Perfection. Oxford University Press.
- Radcliffe-Richards, J. (2012). The Ethics of Transplants. Oxford University Press.
- Rawls, J. (2005). Political Liberalism. Columbia University Press.
- Recanati, F. (1986). On Defining Communicative Infentions. *Mind and Language*, *1*(3), 213–242.
- Resnik, D. (2009). The Clinical Investigator-Subject Relationship: a Contextual Approach. *Philosophy, Ethics, and Humanities in Medicine*, *4*(1), 16–15.
- Rhodes, R. (2014). Autonomy, Agency, and Responsibility: Ethical Concerns for Living Donor Advocates. In J. Steel (Ed.), *Living Donor Advocacy. An Evolving Role Within Transplantation* (pp. 301–310). Springer.

- Richardson, H., & Belsky, L. (2004). The Ancillary-Care Responsibilities of Medical Researchers. *The Hastings Center Report*, *34*(1), 25–33.
- Russell, B. (1995). The Analysis of Mind. Routledge.
- Ryan, C. (1977). Yours, Mine, and Ours: Property Rights and Individual Liberty. *Ethics*, *87*(2), 126–141.
- Ryle, G. (2009). The Concept of Mind. Routledge.
- Saghai, Y. (2013). Salvaging the Concept of Nudge. *The Journal of Medical Ethics*, *39*(8), 487–493.
- Sandman, L., & Munthe, C. (2009). Shared Decision-making and Patient Autonomy. *Theoretical Medicine and Bioethics*, *30*(4), 289–310.
- Savulescu, J. (1995). Rational Non-interventional Paternalism: Why Doctors Ought to Make Judgments of What is Best for their Patients. *The Journal of Medical Ethics*, *21*(6), 327–331.
- Scanlon, T. (1998). What We Owe to Each Other. Harvard University Press.
- Schaefer, O., & Wertheimer, A. (2011). The Right to Withdraw from Research. *Kennedy Institute of Ethics Journal*, *20*(4), 329–352.
- Schaffer, J. (2001). Causation, Influence, and Effluence. *Analysis*, *61*(1), 11–19.
- Scherer, L., de Vries, M., Zikmund-Fisher, B., Witteman, H. O., & Fagerlin, A. (2015). Trust in Deliberation: The Consequences of Deliberative Decision Strategies for Medical Decisions. *Health Psychology*, *34*(11), 1090–1099.
- Schmidt, A. T. (2019). Getting Real on Rationality. Behavioral Science, Nudging, and Public Policy. *Ethics*, *129*(4), 511–543.
- Schroeder, S. A. (2015). Health, Disability, and Well-Being. In G. Fletcher (Ed.), *The Routledge Handbook Of Philosophy Of Well-Being* (pp. 221–232). Routledge.
- Shafer-Landau, R. (2005). Liberalism and Paternalism. *Legal Theory*, *11*, 169–191.
- Sherwin, E. (1996). Infelicitous Sex. *Legal Theory*, *2*(3), 209–231.
- Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs*, *29*(3), 205–250.
- Shnayderman, R. (2013). Social Freedom, Moral Responsibility, Actions and Omissions. *The Philosophical Quarterly*, *63*(253), 716–739.
- Simmons, A. J. (2010). Political Obligation and Consent. In F. Miller & A. Wertheimer (Eds.), *The Ethics of Consent. Theory and Practice* (pp. 305–328). Oxford University Press.
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The Affect Heuristic. *European Journal of Operational Research*, *177*(3), 1333–1352.
- Smith, G. W. (1976). Anarchy, State, and Utopia (Review). *Philosophical Books*, *27*, 87–90.
- Spillman, M., & Sade, R. (2007). Clinical Trials of Xenotransplantation: Waiver of the Right to Withdraw from a Clinical Trial Should Be Required. *The Journal of Law, Medicine, and Ethics*, *35*, 265–272.

- Spital, A. (2005). More on Parental Living Liver Donation for Children with Fulminant Hepatic Failure: Addressing Concerns About Competing Interests, Coercion, Consent and Balancing Acts. *The American Journal of Transplantation*, *5*(11), 2619–2622.
- Stark, M., & Fins, J. J. (2013). What's Not Being Shared in Shared Decision-Making? *The Hastings Center Report*, *43*(4), 13–16.
- Stiggelbout, A. M., Weijden, T. V. D., Wit, M. P. T. D., Frosch, D., Legare, F., Montori, V. M., et al. (2012). Shared Decision Making: Really Putting Patients at the Centre of Healthcare. *British Medical Journal*, *344*, e256–e256.
- Stoljar, N. (2011). Informed Consent and Relational Conceptions of Autonomy. *Journal of Medicine and Philosophy*, *36*(4), 375–384.
- Stout, R. (2010). Deviant Causal Chains. In C. Sandis & T. O'Connor (Eds.), *A Companion to the Philosophy of Action* (pp. 159–165). Wiley-Blackwell.
- Straus, S. E. (2012). Unanticipated Risk in Clinical Research. In J. Gallin & F. Ognibene (Eds.), *Principles and Practice of Clinical Research* (pp. 109–126). Elsevier.
- Strawson, P. (1992). Analysis and Metaphysics: An Introduction to Philosophy. Oxford University Press.
- Sugarman, J. (2015). Is Shared Decision Making an Appropriate Analytic Frame for Research on Medical Practices? *The American Journal of Bioethics*, *15*(9), 18–20.
- Tadros, V. (2016). Wrongs and Crimes. Oxford University Press.
- Thaler, R., & Sunstein, C. (2009). Nudge. Improving Decisions about Health, Wealth and Happiness. Penguin Books.
- Thornton, R. L. (2008). The Demand for, and Impact of, Learning HIV Status. *American Economic Review*, *98*(5), 1829–1863.
- Tilburt, J. (2011). Shared Decision Making After MacIntyre. *Journal of Medicine and Philosophy*, *36*(2), 148–169.
- Trout, J. D. (2005). Paternalism and Cognitive Bias. *Law and Philosophy*, *24*(4), 393–434.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131.
- Ubel, P. A., Loewenstein, G., & Jepson, C. (2005). Disability and Sunshine: Can Hedonic Predictions Be Improved by Drawing Attention to Focusing Illusions or Emotional Adaptation? *Journal of Experimental Psychology: Applied*, *11*(2), 111–123.
- Ulpian. Digest. Retrieved from: http://www.thelatinlibrary.com/justinian/digest4.shtml
- UNAIDS. (2012). UNAIDS World AIDS Day Report.
- Valapour, M. (2008). The Live Organ Donor's Consent: Is It Informed and Voluntary? *Transplantation Reviews*, *22*(3), 196–199.

- Vanderpool, H. Y. (2005). A Quartet of Criticisms. *The American Journal of Bioethics*, *5*(5), 16–19.
- Velleman, J. D. (2015). Beyond Price. Open Book Publishers.
- Waldron, J. (1991). Homelessness and the Issue of Freedom. *UCLA Law Review*, *39*, 295–324.
- Waldron, J. (2014). It's All for Your Own Good. *The New York Review of Books*, 61(15), 1–3. Retrieved from: https://www.nybooks.com/articles/2014/10/09/cass-sunstein-its-all-your-own-good/
- Walker, T. (2018). Consent and Autonomy. In P. Schaber & A. Müller (Eds.), *The Routledge Handbook of the Ethics of Consent* (pp. 131–139). Routledge.
- Wear, S. (1998). Informed Consent. Georgetown University Press.
- Wendler, D., & Johnson, R. (2016). When Clinical Care is Like Research: the Need for Review and Consent. *Theoretical Medicine and Bioethics*, *37*(3), 193–209.
- Wertheimer, A. (1987). Coercion. Princeton University Press.
- Wertheimer, A. (1999). What Is Consent? And Is It Important? *Buffalo Criminal Law Review*, *3*, 557–583.
- Wertheimer, A. (2003). Consent to Sexual Relations. Cambridge University Press.
- Wertheimer, A. (2011). Rethinking the Ethics of Clinical Research. Oxford University Press.
- Wertheimer, A. (2012). Voluntary Consent: Why a Value-Neutral Concept Won't Work. *Journal of Medicine and Philosophy*, *37*(3), 226–254.
- Wertheimer, A. (2013). Should "Nudge" be Salvaged? *The Journal of Medical Ethics*, *39*(8), 498–499.
- Wertheimer, A. (2014). Poverty, Voluntariness, and Consent to Participate in Research. In T. Meyers (Ed.), *Poverty, Agency, and Human Rights* (pp. 272–298). Oxford University Press.
- Wertheimer, A., & Miller, F. G. (2008). Payment for Research Participation: a Coercive Offer? *The Journal of Medical Ethics*, *34*(5), 389–392.
- Westen, P. (2016). The Logic of Consent. Routledge.
- White, M. D. (2013). The Manipulation of Choice. Ethics and Libertarian Paternalism. Palgrave Macmillan.
- Whitney, S. N., McGuire, A. L., & McCullough, L. B. (2004). A Typology of Shared Decision Making, Informed Consent, and Simple Consent. *Annals of Internal Medicine*, *140*(1), 54–9.
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, *61*(2), 341–355.
- Williams, B. (1990). Voluntary Acts and Responsible Agents. *Oxford Journal of Legal Studies*, *10*(1), 1–10.
- Williams, B. (1993). Shame and Necessity. University of California Press.

- Wittgenstein, L. (1958). Preliminary Studies for the "Philosophical Investigations." Basil Blackwell Oxford.
- Wittgenstein, L. (1981). Zettel. Second Edition. (G. E. M. Anscombe & G. Wright, Eds.). Basil Blackwell Oxford.
- Wittgenstein, L. (1988). Remarks on the Philosophy of Psychology. (G. E. M. Anscombe & G. Wright, Eds.). Basil Blackwell Oxford.
- Wittgenstein, L. (2013). Philosophische Untersuchungen. (J. Schulte, Ed.). Suhrkamp.
- Wolff, von, C. F. (1740). Ius Naturae, methodo scientifica pertractatum. Pars tertia, de modo derivativo acquirendi dominium et jus quodcunque praesertim in re alterius, ubi et agitur de officiis circa sermonem, juramentis ac votis nec non usucapione et praescriptione. Francofurti et Lipsiae. Retrieved from Gallica (Bibliothèque Nationale de France): https://gallica.bnf.fr/ark:/12148/bpt6k97383530
- Young, R. (2009). Informed Consent and Patient Autonomy. In P. Singer & H. Kuhse (Eds.), *A Companion to Bioethics* (pp. 530–540). Wiley-Blackwell.

Legal Cases

- *Al Nehayan v Kent* [2018] EWHC 333 (Comm).
- *Barclays Bank v O'Brien* [1994] 1 AC 180.
- *Boulton v Jones* [1857] 2 Hurlstone & Norman 564, 157 ER 232.
- *Ciarlariello v Schacter* [1993] 2 SCR 119 (SCC).
- *Fairbanks v Snow* [1887] 13 NE 596, 599.
- *Freeman v Home Office* [1984] QB 524.
- *Grabowski v Quigley* [1996] 684 A.2d 610 (Pa. Super).
- *Kesarmal v Valliappa Chettiar* [1954] I WLR 380.
- *Mrs U v Centre for Reproductive Medicine* [2002] EWCA Civ 565.
- *R v Brown* [1994] 1 AC 212 HL.
- *R v Carriere* [1987] 35 CCC (3d) 276, 286-87 (Can).
- *R v Elbekkay* [1995] Crim LR 163 CA (Crim Div).
- *Schloendorff v New York Hospital* [1914] 105 N.E. 92 (N.Y.).
- *Talbot v von Boris* [1911] 1 KB 854.
- *Taylor v Albert Einstein Medical Center* [2000] 754 A.2d 650 (Pa.).
- *Vincent v Lake Erie Transportation Co.* [1910] 124 NW 221 Sup Ct (US).
- *Williams v Bayley* [1866] LR 1 HL 200.

International Documents

International Law:

- Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research (2005): https://rm.coe.int/168008371a
- Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (1997): https://rm.coe.int/168007cf98
- The Council of Europe. Committee of Ministers. (1997) Recommendation No. R (97) 16: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b
- EU Directive 2001/20/EC: http://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf

National Law (UK):

- The Medicines for Human Use (Clinical Trials) Regulations (2004): http://www.legislation.gov.uk/uksi/2004/1031/pdfs/uksi_20041031_en.pdf

International Documents (not legally binding but very widely accepted):

- The Belmont Report (1978): https://videocast.nih.gov/pdf/ohrp_belmont_report.pdf
- The Declaration of Helsinki (1964): http://jamanetwork.com/journals/jama/fullarticle/1760318
- International Ethical Guidelines for Biomedical Research Involving Human Subjects ("CIOMS-Guide") (2002): http://www.cioms.ch/publications/layout_guide2002.pdf
- The Nuremberg Code (1947): https://history.nih.gov/research/downloads/nuremberg.pdf
- The Universal Declaration on Bioethics and Human Rights (2005): http://unesdoc.unesco.org/images/0014/001428/142825e.pdf#page=80
- The WMA Medical Ethics Manual (2005): http://www.whcaonline.org/uploads/publications/em_en.pdf

(Verification of online resource addresses: 23rd September 2019)