

ALMA MATER STUDIORUM – UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

Science, Cognition, and Technology

Ciclo 27

Settore Concorsuale di afferenza: 11/C1 Filosofia teoretica/Theoretical Philosophy

Settore Scientifico disciplinare: M-FIL/01 Filosofia teoretica/Theoretical Philosophy

**KORSGAARD'S CONSTITUTIVISM REFRAMED:
ELUCIDATING AGENCY'S PRACTICAL NORMATIVITY**

Presentata da: Federica Berdini

Coordinatore Dottorato

Prof. Marco Beretta

Relatori

**Prof. Giuliano Pancaldi
Prof.ssa Carla Bagnoli**

Esame finale anno 2017

Declaration

I declare that this dissertation is the result of my individual and independent research. The content within it is original and all of the sources consulted have been properly cited in the footnotes and bibliography.

*I dedicate this work
to the memory of my teacher,
Aldo Giorgio Gargani*

Acknowledgements

As any dissertation, this work is the result of a long process, which I have been lucky enough to share, at different levels and stages, with many people. I'd like to thank Doug Lavin, Dick Moran, Doug Kremm, Byron Davies, Chiara-Camilla Derchi, Gloria Mähringer, Xiaoxi Wu, Matt Rachar, and Marianna Ginocchietti for discussing with me the ideas (and helping scratch the philosophical itches) that led to it. Thanks to Hans Bernhard Schmid and Marina Sbisà for hosting presentations of inchoate versions of it, and providing helpful feedback. Thanks also to Caroline Arruda, for her generous encouragement and insightful feedback.

I want to express my deepest gratitude to my supervisors: Carla Bagnoli, whose guidance and caring support throughout the years have been crucial, and Giuliano Pancaldi, who has supported and advised me in the most delicate phase of this journey.

Thanks to the University of Bologna (three-year PhD scholarship, and a Marco Polo mobility grant), and the OeAD - Austrian Agency for International Cooperation in Education and Research (Ernst Mach grant worldwide) for funding my doctoral studies and the completion of this dissertation, respectively.

Thanks to all the amazing friends that have accompanied, helped, and cheered me up on both sides of the pond: Marianna, Bibi, Lia, Mariangela, Erica, Ilde, Elena, Francesca, Enrica, Adaliss, Byron, Doug, Rita, Filippo, Betty, Ina Ricarda, Bernhard, Sebastian, Sofia, Gloria, Xiaoxi, Michelle and Steve, Carys and Shannon, Sean, SJ, and Riley. I am especially grateful to Steffi Probst and Jess Christian; your friendship and love are precious to me. And thanks to Betsy for having held me with kindness.

Thanks also to my mentors, Barbara Agnese, Claudia Bianchi, and Bernadette Brooten for playing and having played an active role in my life and for providing your loving support. Thanks to my family, and especially to parents, Rosanna and Franco, for the unconditional love and material support throughout these years, and to my brothers Riccardo and Marco. I love you both very much.

And thanks to Michael Baumtrog, who has witnessed and contributed to the development and concretization of this work with patience, curiosity, and trust. Having you in my life, in the last year, has changed perspectives, empowered, and inspired—I am filled with gratitude and love, and excitement for what's next.

Abstract

KORSGAARD'S CONSTITUTIVISM REFRAMED: ELUCIDATING AGENCY'S PRACTICAL NORMATIVITY

Federica Berdini

This dissertation aims to critique and extend upon Christine Korsgaard's theory of agency. To do so, **chapter one** frames the discussion in terms of her larger constitutivist project, and provides definitions and descriptions of key terms in the philosophy of action which prove important for the rest of the dissertation. Unlike much of the burgeoning literature engaging with Korsgaard's work that addresses the metanormative ambitions of her argument, the scope of the present analysis is her first-order metaphysical account of agency. It is at the first-order level, especially with her thesis about the *double nature* of constitutive norms, that the crucial normative step in the argument is made.

Chapter two is dedicated to an extensive reconstruction of Korsgaard's constitutivist account. The attention is focused on Korsgaard's version of the *authorship view* of agency—which provides an articulation of what the agent's psychological structure and organization must be in order for an action to be *owned*, and thus be genuinely *her* action—, on the metaphysics of agency and normativity underpinning Korsgaard's thesis that self-constitution is the function of action, and on her account of responsibility as a practical relation.

Chapter three explicates some problematic cases for Korsgaard's view, by considering instances of so-called “disorders of agency,” and “socially displaced agency.” Evidence for considering these cases as genuine instantiations of agency is presented and discussed, reaching the conclusion that Korsgaard's constitutive norms of agency seem to be neither necessary nor sufficient to properly account for them, which points to the need for a clarification of the metaphysics of agency and of constitutive normativity. In particular, in light of the *Disorders* scenario, arguments are brought against Korsgaard's assimilation of the practical normativity of agency to moral normativity, and the notion of *agential responsibility* predicated upon Hanna Pickard's articulation of *responsibility without blame* is advanced. The analysis of both cases, however, corroborates Korsgaard's relational account of responsibility as answerability, on which the dissertation aims to build a reframing and extension of her theory of agency.

Chapter four sets the stage for the positive contribution of this work by pursuing a conceptual elucidation of the notion of constitutive normativity, which is central to constitutivism as a first-order theory of agency. Three possible interpretations of Korsgaard's characterization of the notion of constitutive norms are presented and one of them—the Practical Interpretation—is endorsed as the best way to make sense of the notion while accounting for the conclusions drawn in chapter three. After a close examination of the Practical Interpretation, it is argued that the constitutive normativity pertaining to human agency is not homogeneous in kind, and the distinction between *socially-generated* constitutive normativity and *non-practice-based* constitutive normativity is introduced. This work is meant to produce a first step in the broader constitutivist project of elucidating how different kinds of normativity hold together.

In **chapter five** the view of an interactionist approach to agency is laid out, by indicating three central components of a first-order constitutivist theory of agency: i) the necessity of a scalar approach, ii) relationality, and iii) context sensitivity and situatedness. Inspired by

Korsgaard's practical account of responsibility, these aspects get considerably strengthened in the relational definition of what it is to be an agent (*Agential Responsibility Claim*) and the proposal of considering agency as inherently socially constituted through interpersonal and social practices (*Interaction Claim*). The reframing of Korsgaard's account advanced here is unique in two respects. First, it suggests incorporating aspects of G. E. M. Anscombe's seminal work on intention—interpreted as an outward-looking and interactional/dialogical approach. Second, and congruently, the notion of *intersubjective recognition* is introduced as a necessary component for the characterization of the social constitution of agency.

After recapitulating the trajectory of the argument, **chapter six** addresses both the theoretical and practical implications of the proposal presented here, as well as some of its limitations and prospects for future research.

KEYWORDS: philosophy of agency; agency theory; theory of agency; authorship views of agency; Korsgaard; constitutivism; constitutive norms; practical normativity; agential responsibility; accountability; disorders of agency, responsibility without blame; intentional action; Anscombe; intersubjective recognition; social authoring; socially displaced agency; social interaction; social constitution of agency

Table of Contents

Title.....	1
Declaration.....	2
Dedication	3
Acknowledgements	4
Abstract.....	5
Table of Contents	7
Abbreviations	9
Chapter One – Introduction	10
1.1 Agency, Agents, and Action Theory	10
1.1.1 The Standard Conception and the Traditional View of Agency and Action	10
1.1.2 Non-Reductive Theories of Agency and Action	11
1.2 Constitutivism	13
1.2.1 Constitutivism and Metaethical Constructivism	14
1.2.2 Constitutivist Strategies and Action Theory	15
1.3 Korsgaard’s Constitutivism.....	17
1.4 Plan of the Work.....	18
Chapter Two – Korsgaard’s Constitutivist Account of Agency	20
2.1 The Distinctive Character of Human Agency	20
2.2 Practical Identity and the Psychology of Action.....	22
2.3 From the Ordinary Concept of Agency to the Normative Conception of Agency	23
2.4 Responsibility	25
2.5 The Metaphysics (and Epistemology) of Agency and Normativity.....	27
2.5.1 Agency as the Activity of Self-Constitution	27
2.5.2 Constitutive Principles and their Double Nature	28
2.5.3 Agency’s Metaphysics and Epistemology	30
Chapter Three – Defective Agency?.....	33
3.1 Constitutive Norms and the Problem of Defective Actions	33
3.2 Disorders of Agency	35
3.3 The Clinical Model of “Responsibility without Blame”	38
3.4 Relevance of the Clinical Model for the Present Work.....	41
3.5 Socially Displaced and Distorted Agency	44

3.6 Relevance of the Phenomenon of Social Authoring for the Present Work	47
3.7 Conclusion	48
Chapter Four – The Metaphysics of Normativity and Agency.....	50
4.1 Disambiguating ‘Normativity’ in Korsgaard’s Constitutivist Argument	50
4.2 Elucidating Constitutive Normativity.....	52
4.2.1 The Metaphysical Interpretation	54
4.2.2 The Individual Psychology Interpretation.....	54
4.2.3 The Practical Interpretation.....	56
4.3 On the Metaphysics of Agency and Normativity.....	57
4.3.1 Scope and Methodology.....	58
4.3.2 The Metaphysics of Agency.....	59
4.4 The Autonomous Systems Account.....	61
4.5 Varieties and Sources of Normativity.....	62
4.6 Constitutive Normativity, regained	64
4.6.1 Non-Practice-Based Constitutive Normativity	65
4.6.2 Socially-Generated Constitutive Normativity.....	66
4.7 Conclusion	69
Chapter Five – Reframing Agency	71
5.1 Central Components of My Reframing.....	71
5.1.1 The Necessity of a Scalar Approach	72
5.1.2 Relationality	73
5.1.3 Context-Sensitivity and Situatedness.....	74
5.2 Anscombe on Intentional Action.....	77
5.3 Relevance of the Anscombean Approach to Intentional Action for the Present Work	79
5.4 Dimensions of Agency	81
5.5 Towards an Interactionist View	83
5.6 Conclusion	85
Chapter Six – Conclusions	86
6.1 Summary	86
6.2 Theoretical and Practical Implications	89
6.3 Limitations and Prospects for Future Work.....	90
6.4 Conclusion	92
Bibliography	93

Abbreviations Used for Korsgaard's Works

SN *The Sources of Normativity*

CA *The Constitution of Agency*

SC *Self-Constitution*

NCA "The Normative Constitution of Agency"

AKC "How to Be an Aristotelian Kantian Constitutivist"

Chapter One – Introduction

This dissertation will provide a critique of, and an extension upon Christine Korsgaard's constitutivist theory of agency. In this chapter I frame the discussion by introducing some of the key notions and projects in philosophy of action and agency (section 1.1), and define the origins and aims of constitutivism as a general strategy (section 1.2). I then provide a preliminary characterization of Korsgaard's version of constitutivism, specify and motivate the scope of my analysis (section 1.3), and end with an overview of the path ahead (section 1.4).

1.1 Agency, Agents, and Action Theory

The rising and flourishing of the philosophy of action over the past sixty years has crucially revolved around conceptual, metaphysical, and epistemological questions about the nature of action and its explanation.¹ Roughly speaking, an agent is a creature capable of acting, and agency is the capacity exercised by the agent when she does something.

1.1.1 The Standard Conception and the Traditional View of Agency and Action

Different *conceptions* and different *theories* of action and agency have been developed,² but the 'standard conception' of action underlying debates in philosophy of action has intentional action at its core. The *standard conception* consists of two central claims: (1) the notion of action is to be explained in terms of intentional action, and (2) there is a close relation between acting intentionally and acting for a reason. Consequently, the *standard conception* characterizes agency as the capacity to act intentionally and for reasons, and someone counts as an agent if she manifests that capacity.

While there is consensus on the standard conception,³ different *theories* of action and agency may result from it, depending on the account of intentional action and reason explanation endorsed. The '*traditional view*' (also referred to as the 'standard view') of action and agency

¹ See Aguilar, Buckareff & Frankish (2011) for the distinction between philosophy of action as a narrow and a broad sub-discipline of philosophy.

² I rely on Schlosser (2015) for this distinction between 'conceptions' and 'theories' of agency.

³ Alternative conceptions differ in that they emphasize agency as 'a power to initiate' irreducible to the capacity to act intentionally and for reasons. For a list of the proponents of such alternative (e.g., C. Ginet, T. O'Connor, E. J. Lowe), I refer to Schlosser (2015, section 2.2).

in contemporary debates is an event-causal theory of intentional action and reason explanation. According to the traditional view, an action is intentional and done for reasons just in case it is *caused* by the right mental states and events (the agent’s beliefs, desires, and intentions) in the right way—that is, by rationalizing the action from the agent’s point of view.⁴ The main theoretical concern guiding the traditional view is that of providing an account of human agency and action which emphasizes a relation of continuity between ‘agential phenomena’ on the one hand, and natural events and the (scientific) explanation thereof, on the other.

Although widespread, the traditional view has been largely criticized for failing to capture the distinctiveness of human agency. In the event-causal picture, the criticism goes, the agent ‘disappears,’ in that she seems to serve just as a ‘passive arena’ for the mental states and events that cause her behavior.⁵ The agent is therefore *reduced* to “the causal roles of agent-involving states and events.” (Schlosser, 2015, 3.1. Cf. also Velleman, 1992/2000, p. 125)⁶ As a result, the traditional view loses grip on the understanding of ordinary practices of responsibility attribution, and on the basic idea that actions are ‘up to’ their agents. Genuine actions—detractors of the traditional view contend—differ from mere behavior and other events that happen to occur to the agent precisely because of the *role* played by the agent in bringing them about.

1.1.2 Non-Reductive Theories of Agency and Action

Accordingly, the task for *non-reductive* theories of agency and action (be they compatible with a causal framework or not)⁷ is that of spelling out an account of the relevant functional organization of the agent’s psychological structure in virtue of which intentional agency⁸ is a form of conscious, self-directed, and guided activity. In other words, non-reductive theories aim at providing a *personal level* (vs. an event-causal) analysis of agency. This kind of analysis

⁴ The leading proponent of the traditional view is Donald Davidson.

⁵ For this criticism, known in the literature as the problem of the ‘disappearance of the agent,’ see Velleman (1992/2000), Ferrero (2009a), Steward (2012). The same contrast emerges also from Korsgaard’s characterization of what being an agent amounts to: “I am not the mere location of a causally effective desire but rather am the *agent* who acts *on* the desire,” SN (p. 228).

⁶ Schlosser (2015, section 3.1). Cf. Velleman, (1992/2000, p. 125).

⁷ Among non-reductive accounts of agency there are agent-causal approaches, and volitionist approaches. For an overview see Schlosser (2015, section 3.1).

⁸ Unless otherwise specified, throughout the discussion I will be using the unqualified term action for ‘intentional action,’ thereby speaking from the standpoint of the *standard conception* of action and agency.

involves accounting not just for the distinctive *psychological* dimension of human agency (e.g., its first-personal phenomenology—the fact that agency comes with a characteristic sense of ownership; the nature of motivation), but also for its *normative* character. This amounts to investigating *agential notions* (choice, autonomy, reason-responsiveness, responsibility, to name a few), and the practices and contexts within which these notions find application, by focusing on the ‘manifest image’—that is, the conceptual framework in terms of which human beings conceive of themselves *as agent*.

An alternate way to delineate the distinction between the traditional (causally reductionist) view and non-reductive theories of agency and action, is to observe that (at least some) non-reductive theories conceive of the philosophy of action and agency as part of practical philosophy, broadly understood to encompass ethical theory, moral psychology, philosophy of practical rationality, and theories of normativity.

Within these more expansive non-reductive theories, so-called ‘dual standpoint theories’ have emerged as leading contenders in the philosophy of action.⁹ According to these theories, “agency can only be understood from a practical and normative standpoint.”¹⁰ While in principle compatible with a broadly naturalistic metaphysical framework, the present work explores therefore a non-traditional *theory* of action and agency, in the sense that the understanding of intentional action and reason explanation endorsed here is distinct from the causal one provided by the traditional view.

The spirit that informs the present work sides with the practical construal of the action-theoretical landscape.¹¹ This dissertation is an attempt to provide an elucidation of the *practical normativity* proper to agency through the examination of one prominent example of a dual standpoint theory: Christine Korsgaard’s constitutivist account of agency—which embraces both a non-reductive view, and the ethical (or broadly normative) implications of such a view.

Generally speaking, constitutivist arguments are such that “they promise to explain the normative force of certain rational requirements in a broadly naturalistic way, without helping themselves to irreducible normative properties *external to the agent*” (Tubert, 2011, p. 344,

⁹ The traditional partitioning is into event-causal, agent-causal, and volitionist approaches to agency; see Schlosser (2015, section 3.3).

¹⁰ Ibid. According to Schlosser dual standpoint theorist are T. Nagel, C. Korsgaard, and A. Bilgrami. I would also add Moran (2001, 2004, 2012), and Bagnoli (2007, 2012).

¹¹ Under this respect, my way of conceiving of action theory is akin to Aguilar et al.’s (2011, pp. 1-2) endorsement of a characterization of the philosophy of action in terms of a broad sub-discipline of philosophy.

italics is mine). Especially in her most recent work on constitutivism, Korsgaard provides a unique account of the metaphysics of agency and normativity. But before getting there, an illustration of constitutivism and a preliminary framing of the aim of the present work is needed.

1.2 Constitutivism

Constitutivism is a philosophical position according to which certain normative claims of practical reason and morality—that is, about which actions an agent has reason (or *ought*) to pursue/perform—can be derived from the features that are essential to being a rational agent:¹² a creature capable of acting on the basis of reasons. Here are a few definitions:

A constitutive theory is a metanormative theory that purports to establish the objective validity and content of practical norms on the basis of the constitutive features of agency. (Ferrero, 2015, p. 883)

Constitutive arguments attempt to establish the normativity of rational requirements by pointing out that we are already committed to them insofar as we are believers or agents [...] to justify requirements anyone would have reason to accept [normative requirements] by explaining what it is to engage in a certain activity or be a certain kind of being. (Tubert, 2010, pp. 656-657)

The insight, in a nutshell, is to ground content and authority of non-hypothetical practical standards in features viewed as distinctive to either agency or action. (Berteau, 2013, p. 81)

The above definitions are an expression of the prevailing line of interpretation of the constitutivist project informing current debates on constitutivism, which emphasizes the distinctively metanormative character of the project. In other words, constitutivism aims to show that the normativity of requirements of practical reason—namely their *objectivity* (non-contingency and universality), and *practicality* (action-guiding character and authority)—is grounded in principles proper to intentional human agency. The constitutivist project, however, originates from somewhat separate debates and theoretical concerns.

¹² Unless otherwise specified, my use of the term ‘agent’ stands for ‘rational agent,’ as opposed, for example, to non-human, animal agents. In other words, the discussion ranges within the scope of the standard conception of agency, i.e., intentional and rational agency.

1.2.1 Constitutivism and Metaethical Constructivism

On the one hand, within moral and metaethical debates, constitutivism should be interpreted as a metanormative project akin to certain versions of *constructivism*.¹³ Metaethical constructivism represents a distinctive alternative to realist and anti-realist approaches to questions about the existence and nature of normative truths, and can be said to be advancing both a metaphysical and a normative thesis. The metaphysical thesis consists in a positive reply to the question about the existence of normative truths, and holds that “these truths are not fixed by facts that are independent of the practical standpoint, [...] rather, they are constituted by what agents would agree to under specified conditions of choice” (Bagnoli, 2011/2015, preamble). As a first-order (moral) normative theory, constructivism holds that the normative principles we ought to accept/follow “are the ones that agents would endorse were they to engage in a hypothetical or idealized process of rational deliberation” (ibid)—call this the normative thesis. In other words, the human capacities for self-reflection, evaluation, practical reasoning, are the legitimate and authoritative processes to appeal to in our quest for normative validation and guidance.

The distinctiveness of constructivism *qua* metaethical position lies therefore in a vindication of the objectivity of normative judgments via the appeal to the self-authenticating character of the evaluative procedure through which these judgements are arrived at, or constructed. The relation between constitutivism and certain varieties of constructivism consists therefore in the appeal, on the part of the latter, to constitutive norms of reasoning/evaluation/agency (ibid, section 7.3).

Emphasizing the tight connection between constitutivism and constructivism, however, should not lead to underestimate the fact that they do not necessarily overlap. First of all, constructivism extends beyond the metaethical debate and as a matter of fact it originally came out as a position in political philosophy. Second, there are various accounts of constructivism and constitutivism, and—as it has been noticed¹⁴—certain versions of the latter might not be compatible with some versions of the former. In particular, the constitutivist strategy as we can find it articulated in recent debates is *not* part of early constructivist theories, notably the one

¹³ Here I refer here mainly to Kantian constructivism. For an exhaustive reconstruction and mapping of constructivism in metaethics see Bagnoli (2011/2015). See also Hanisch & Baiasu (2016), for a characterization of constitutivism that brings to the fore its relation with metaethical constructivism.

¹⁴ Baiasu (2016, n. 1). While this is the case for *A Theory of Justice*, it is less straightforward with respect to the Dewey Lectures.

exemplified by Rawls' *A Theory of Justice*, which is a form of *political* constructivism, more modest than ethical constructivism (ibid, section 1).

1.2.2 Constitutivist Strategies and Action Theory

On the other hand, constitutivism arguably represents one among many recent philosophical approaches which take action theory to have an explanatory priority over theories of practical reasoning.¹⁵ In particular, such approaches focus on establishing what it takes for something to be an action, and in delineating a distinction between full-fledged action and “mere activity” or “mere behavior.” With these distinctions in place, they make a case for the claim that in order for something to be an action at all, reasons for action must have a certain logical form. As a project in philosophy of action distinct from strictly metaethical and metanormative concerns, we can thus say that constitutivism has, first and foremost, a distinctive metaphysical ambition in that it provides an account of what intentional human agency is, and what its essential, or constitutive features are.

The crucial step in the overall constitutivist strategy, however, is made at the very level of the first-order account of what, metaphysically, human agency is. It is here that the essentially normative character of constitutivist *strategies* comes into play¹⁶—that is, with the claim that it is the nature of human agency itself that determines the standards for good or successful agency. In light of the distinctive metaphysical position advanced, I will therefore refer to, and focus on, constitutivism as a kind of *metaphysical-cum-normative* strategy in action theory.

To be sure, there are differences in the scope and strength of the conclusions of extant variants of constitutivism.¹⁷ Christine M. Korsgaard's Kantian account, for example, identifies the constitutive standards of agency with moral principles—the hypothetical and categorical imperatives—qualifying it as a strong form of constitutivism. David Velleman, on the other hand, maintains that moral norms are supported by, but not categorically rationally required

¹⁵ I refer here to Millgram's (2005/2012) way of framing the relation between recent developments in theories of practical reasoning and action theory, and to his characterization of the action-theoretical turn in theories of practical reasoning.

¹⁶ See Enoch (2006) for the distinction between the normative and the metanormative aspects of constitutivism.

¹⁷ As a matter of fact, a plethora of variants can be distinguished within the constitutivist project—i.e., Kantian (Korsgaard, 1996a, 2008, 2009, Engstrom, 2009), Humean (Velleman, 2000, 2009), Aristotelian (Lavin, 2017, Korsgaard, in progress), and Nietzschean (Katsafanas, 2013) approaches.

by, the constitutive aim of action, situating it as a more flexible account (Velleman, 2009).¹⁸ Both, however, define agency in terms of a *functional kind* by claiming that action has a constitutive function, or aim, and both track down the nature and sources of ethical normativity by looking to the constitutive function of action. The very idea of a constitutive aim or function of action, in fact, carries normative consequences, since it sets standards of success with respect to the fulfillment of the aim/function itself.

The capacity to act for reasons at the core of the standard conception of agency is crucial for all the varieties of constitutivism on offer. For example, Korsgaard and Velleman share a similar approach to understanding where agency lies—that is, in the exercise of our distinctive agential capacities—even though they develop independent views of what these distinctive capacities are (the capacity for *self-constitution*, according to Korsgaard, and for *self-understanding*, to Velleman).

Despite their differences, all varieties of the constitutivist strategy can be said to display the following structural features: They provide (i) an account/*theory* of human action—i.e., propounding certain features as the constitutive aim/standards/principles of action—which, in conjunction with (ii) the claim about the inescapability of acting, is aimed at (iii) vindicating the normative authority of normative claims of practical reason from the theory of action/agency endorsed.¹⁹

Unlike most of the extant literature on constitutivism (Enoch, 2006, 2011, Ferrero, 2009b, Millgram, 2011, Bertera, 2013, Arruda, 2016a), in the present work I will not engage in an evaluation of the prospects of constitutivism as a metanormative strategy (step (iii) of the argument). Rather, my focus will be on how *constitutive norms* work in the elaboration of a *theory* of agency and action, and, most importantly, on clarifying the status of the *practical normativity* pertaining to agency, as opposed to moral normativity. This focus will be discussed using Christine Korsgaard’s theory, according to which the normativity of rational agency and of morality are one and the same, and falling short of morality implies falling short of agency altogether, and vice versa.

¹⁸ See Arruda (2016a) for a clear discussion of the distinction between Korsgaard’s “Robustly Normatively Oriented” and Velleman’s “Weakly Normatively Oriented” varieties of constitutivism.

¹⁹ Cf. Katsafanas, forthcoming, and the special issue on constitutivism and Kantian constructivism of *Philosophia* (2016) for a similar characterization of the argumentative strategy common to the different varieties of constitutivism on offer. According to this characterization, constitutivist arguments might endorse a traditional, causal, view of action (e.g., Velleman) or not (e.g., Korsgaard). As stated in the previous section, the focus of the present work is on one example of the latter kind of theory.

1.3 Korsgaard's Constitutivism

The scope of my analysis in the present work is restricted to Korsgaard's theory of agency, which is arguably the most prominent and developed, as well as the most contentious, version of constitutivism on offer. Korsgaard's constitutivist argument can be reconstructed as follows:

- (1) Human beings must act
- (2) In order to act, human beings must unify themselves into agents
- (3) Human beings can only be unified into agents when they cohere with the principles of practical rationality
- (C) So, human beings as unified agents must act in accordance with the principles of practical rationality

The premises of the argument relate to the “inescapability” condition on the one hand (premise 1), and to Korsgaard's account of what, metaphysically, intentional human agency is, and what its essential, or constitutive features are, on the other (premise 2 and 3). As mentioned in the previous section, the core of the constitutivist strategy lies at the level of first-order *metaphysical-cum-normative* account of agency. This, accordingly, will be the level of concern of my analysis throughout this dissertation.

Among the reasons for focusing on the first-order account rather than on the metanormative strategy is the fact that, as it has been pointed out (Katsafanas, 2013, chapters 3-4, Katsafanas, forthcoming), a preliminary clarification of the first-order characterization of agency will prove to be crucial to the viability of the metanormative ambitions of the constitutivist project itself.

The main reason for addressing Korsgaard's account is that it provides an explicit articulation of the notion of *constitutive norms*. Constitutive norms are not just principles regulating a pre-existing activity—they constitute, or define, while at the same time providing guidance for, the very activity one is engaged in by conforming to them. This is Korsgaard's claim about the double nature of constitutive standards:

They are normative, because in performing the activities of which they are the principles, we are guided by them, and yet we can fail to conform to them. But they are also descriptive, because they describe the activities we perform when we are guided by them. (CA, p. 9)

Korsgaard's account of agency has been accused of conflating the concept of action with that of a good action, and of not being compatible with the intuitive claim that in order for something to be subject to a norm, it should be possible for it to violate the norm (Katsafanas,

forthcoming, sections 5.1 and 5.5, and Lavin, 2004, respectively). If these criticisms are correct, then on Korsgaard's account any action must be a good action, and any defective action entails a denial of agency. This raises a number of issues for the messy, actual world within which we ordinarily act, and is especially problematic because the double nature of constitutive norms seems to involve a commitment to provide an account which is both normative *and* descriptive. In the real world, instances of 'defective' and 'bad' actions and agents abound, thereby making Korsgaard's construal of the double nature thesis problematic in the face of the "error constraint" to which agency intuitively must be subject.

This dissertation addresses the challenge from defective actions by analyzing two varieties of alleged breakdowns of agency: cases of 1) disorders of agency and 2) socially displaced agency. The interpretations of the cases that I rely on and endorse, make a strong case for considering them as instances of genuine agency, thereby calling for a clarification of the notion of constitutive normativity relevant to agency.

Overall, my aim in this work is to point to and resolve some of the tensions within Korsgaard's constitutivist theory, by advancing an expansion which strengthens its most valuable insights.

1.4 Plan of the Work

In light of the articulation of the state of the art in philosophy of action provided in the first section of this chapter, the specificity and merit of Korsgaard's account lies in its attempt at an elucidation of the special perspective we occupy, *qua* agents, with regard to our intentional actions, as well as the authoritative character peculiar to the agential standpoint. Accordingly, Korsgaard's theory provides an especially articulated version of the so-called 'authorship views' of agency. These belong to a specific strand in action theory that identifies the distinguishing feature of human actions in their being *authored*—as opposed, for example, to action's stepwise structure, its evaluative character, or its essentially being part of a practice ("calculative," "evaluation," and "practice" views, respectively) (Millgram, 2005/2012).²⁰ Authorship views provide an account of what the agent's *psychological* structure and organization must be in order for her actions to be *owned*.

²⁰ Whether these different approaches count as forms constitutivism themselves and, if so, how do they relate to each other would be a worthwhile research direction to be pursued by the constitutivist project, broadly construed.

In the next chapter I provide a close reconstruction of Korsgaard's authorship view, her relational conception of responsibility, and her constitutivist theory, by paying special attention to spelling out the distinctive metaphysics of agency and normativity underlying them.

In chapter three I present two case studies exemplifying the problem of defective actions and agency. The discussion of these cases brings to the fore the importance of accounting for the agential perspective as connected with the context of interpersonal and social practices within which agency is exercised, and illuminates a tension in Korsgaard's account—the metaphysical turn taken by her constitutivist theory of agency seems to be at odds with the practical character of the overarching constructivist enterprise she takes herself to be engaged in. I provide a critique of Korsgaard's assimilation of the notion of responsibility to moral responsibility, and propose some adjustments to her account that are meant to emphasize the practical (vs. metaphysical) character of her approach in non-moral terms.

Chapter four zeroes in on Korsgaard's characterization of constitutive norms by providing three possible interpretations of the notion, and favoring one of them as the best way to make sense of the conclusions drawn from the discussion of the case studies. I pursue an elucidation of the constitutive normativity pertaining to agency by distinguishing between *practice-based*, or *socially-generated*, and *non-practice-based* constitutive normativity.

In chapter five I string together the claims I advanced throughout the critical part of the dissertation, and outline my proposal of expansion upon Korsgaard's account towards a constitutivist-interactionist view of agency.

Chapter six concludes the dissertation by recapping the argument and the proposal it led to, identifying its theoretical and practical implications, and discussing some of its limitations and directions for further research.

Chapter Two – Korsgaard’s Constitutivist Account of Agency

In this chapter I provide a reconstruction of Korsgaard’s version of constitutivism, with the aim of showing how the notions of rational reflection, choice, authorship, personal identity, self-constitution, and responsibility coalesce into her proposed *normative* theory of agency.

2.1 The Distinctive Character of Human Agency

As I mentioned in the previous chapter, the capacity to act intentionally and for reasons is at the core of all the varieties of constitutivism on offer. For Korsgaard, in particular, human agency *is* rational agency. She defines the faculty of reason in Kantian terms as “the active aspect of the mind,” that is, a faculty or power that stems from the specific kind of self-consciousness proper to human agents. This specific type of self-consciousness is “the source of a psychic complexity not experienced by the other animals,” one that “transforms *psychic unity* from a natural state into something that has to be achieved, into *a task and an activity*” (SC, section 6.4.1, emphasis is mine). In her words:

[...] we human beings are aware, not only that we perceive or desire or fear certain things, but also that we are inclined to believe and to act in certain ways on the basis of these perceptions or desires or fears. We are aware not only of our representations and desires as such but also of the way in which they tend to operate on us. That is what I mean by saying that we are aware of the potential grounds of our beliefs and actions *as potential grounds*. And this awareness is the source of Reason. For once we are aware that we are inclined to believe on the ground of a certain perception, or to act on the ground of a certain desire, we find ourselves faced with a decision, namely, whether we should do that—whether we should draw the conclusion, or perform the action, on the ground in question, or not. Once the space of awareness—of reflective distance, as I like to call it—opens up between the potential ground of an action and the action itself, we must step across that distance with some awareness that we are doing so, and so must be able to endorse the operation of that ground as the basis for what we believe or do. And a ground of belief or action whose operation on us as a ground is one that we can endorse is a reason. This means that the space of reflective distance presents us with both the possibility and the necessity of exerting a kind of control over our beliefs and actions that the other animals probably do not have. We are active, self-directing, with respect to our beliefs and actions to a greater extent than they are. And it is the same fact that we now

both can have, and absolutely require, *reasons* to believe and act as we do. (CA, pp. 4-5. See also SN, section 3.2.1)

It is within the space of reflective distance that the question whether our incentives give us reasons arises. In order to answer that question, we need principles, which determine what we are to count as reasons. Our rational principles then replace our instincts—they will tell us what is an appropriate response to what, what makes what worth doing, what the situation calls for. (SC, section 6.1.7)

According to Korsgaard, action is “one among several forms of ‘rational activity’,” and “acting in the sense relevant to practical reason is that activity that is directed to producing some state of affairs in the world” (CA, p. 10). Thanks to the reflective structure of self-consciousness, however, the relation between an agent and her action is not one of mere production, but rather of *authoring*. Korsgaard’s theory is a version of the so-called authorship views: the distinguishing or essential feature of human action is that it is *attributable* to the person who performs it, to its *author*—an action is “a movement that is attributable to an agent as its author” (SC, p. xi).

The “special form of attribution” that is at stake when identifying someone’s actions²¹ is meant to capture not just the efficacious role of the agent’s relevant mental states in causing her movements, but the fact that *she* is “plac[ing] [her]self fully behind the movement [she is] about to make or the change [she is] about to effect, to endorse it wholeheartedly” (NCA, section 1.1). We are entitled to attribute an action to an agent, Korsgaard continues, when the agent has *contributed* to it in a specific way. In other words, what distinguishes someone’s actions from mere movements or less-than-full-fledged intentional actions, is that the actions are *owned* by the person, that they are attributable to their author *as a whole*, rather than to some desire or other forces working within her. The criteria for whole-person attribution, in turn, rely on the person’s psychic regulation and organization, which requires rational agents to play an *active* role in regulating.

The scope of authorship, in Korsgaard’s Kantian view, is the scope of rational agency, both in its practical and doxastic aspect (see also Moran, 2001). And where rational agency is, there is also a distinctive kind of responsibility, which stands for the *practical* relation between an

²¹ The use of the term ‘action’ here stands for intentional action. I will say more on the distinctions proper to intentional action later (see *infra*, sections 5.1, 5.3).

agent and her own mental states, attitudes, and actions, as she engages in the rational activity of determining what to believe and what to do.

2.2 Practical Identity and the Psychology of Action

The kind of psychic complexity stemming from human, rational self-consciousness, Korsgaard claims, makes it necessary to achieve an agential unity, in virtue of which it is possible to *attribute* the action to the agent as a whole, as its *author*. In order to act, the agent must unify herself by adopting a principle of choice, which enables her to endorse a certain inclination of hers *as a reason*, and to act upon it. The principle of choice confers a normative status on the relevant *incentive*, and by acting according to that principle, by identifying herself with that principle, the agent acts in such a way that her actions are expressive of herself (see SN, section 3.3.1, SC, sections 6.1.7-6.2.5).

When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*. (SN, p. 100)

The necessity of, and capacity for, *normative self-government* proper to human agents brings with it a “distinct form of identity, a norm-governed or practical form of identity, for which we are ourselves responsible” (SC, p. xii): as human beings,

we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical identity—that the other animals lack. It is this sort of identity that makes sense of our practice of holding people responsible, and of the kinds of personal relationships that depend on that practice. [...] It is as the possessor of personal or practical identity that you are the author of your actions, and responsible for them. And yet at the same time it is in choosing your actions that you create that identity. What this means is that you constitute yourself *as* the author of your actions in the very act of choosing them. (SC, section 1.4.3)

It is therefore in virtue of expressing someone’s reflected-upon commitments and values—that is, a practical identity—that actions are attributable to their author as a whole, while at the same time constituting her as an agent. The notion of *practical identity* is central to Korsgaard’s project. A practical identity is “a description under which you value yourself and find your life worth living and your actions to be worth undertaking” (ivi, section 1.4.4, and SN, section

3.3.1). Practical identities represent thus both sources of reasons for us, in that they provide the incentives and principles to govern the choice of our actions, and the output of our actions, when these are properly undertaken.

As concrete, embodied agents, we have many practical identities, corresponding to the contingent roles and relationships we hold within the network of our interpersonal relationships. However, while we acquire or drop different identities in the course of our lives through deliberation and reflection, we cannot drop out of the *practical* necessity of choosing and acting itself. As rational beings, reflection and choice are “the simple, inexorable fact of the human condition,” “our *plight*.”

We must act, and we need reasons in order to act. And unless there are *some* principles with which we identify we will have no reasons to act. Every human being must make [her]self into someone in particular, in order to have reasons to act and to live. Carving out a personal identity for which we are responsible is one of the inescapable tasks of human life. [...]

We owe it to ourselves, to our own humanity, to find some roles that we can fill with integrity and dedication. But in acknowledging that, we commit ourselves to the value of our humanity just as such. (SC, section 1.4.7)²²

Being an agent amounts then to be engaged in the task of unifying oneself into a coherent whole by choosing and acting on reasons, by taking responsibility for one’s own practical identity, and thereby becoming, a specific individual. By providing the specific content on which the formal principle of agency—of self-determined, autonomous efficacy—is imposed (see SC, sections 1.4.7, 2.4.2, and NCA, section 1.3), contingent practical identities determine a functioning self. An action, on the other hand, is authored and attributable to her agent if and only if it is autonomous—i.e., if it is the expression of the process of self-reflection and deliberation.

2.3 From the Ordinary Concept of Agency to the Normative Conception of Agency

Korsgaard’s characterization of agency in terms of authorship is motivated by what she takes to be two features proper to our ordinary concept of agency:

²² On the contingency of our practical identities and on the difference between actively embracing it or passively enduring it, see SN (pp. 239-242).

[i] when we attribute an action to a person, we are suggesting that [s]he has been active, and [ii] when we respond to someone's actions, we take ourselves to be responding to *h[er]*, to features of *h[er]* identity. I am going to call those two features of our ordinary concept of agency *the activity implication* [i] and *the identity implication* [ii]. (NCA, section 1.2)

According to Korsgaard, there is a primarily metaphysical aspect to the *activity implication*, that is the idea that “an agent's *activity* is supposed to be implicated in her agency. The activity implication is meant to capture the intuition that in attributing an action we assume a special kind of relation between it and the person who performed it. Intentional actions are owned, or authored, by the agent as a result of a certain activity on *her* part. This basically amounts to the intuition that overt agency is the ongoing manifestation and exercise of the ability of self-movement and of producing a change, which are initiated and controlled by the agent herself on the basis of rational reflection. The metaphysical aspect of the activity implication, Korsgaard adds, carries therefore an additional, distinctively normative implication, namely the fact that “an agent, just as such, is the kind of thing that can succeed or fail,” which “seems to imply that [she is] under some kind of norm or principle which [she] may live up to or not” (ivi, section 1.3). These norms, the principles governing agency are, according to her, the principles of practical reason.

The *identity implication*, on the other hand, has a distinctively normative import in that

to say that someone did something is to mention what is, in general, a proper ground for responses to her that are normative or at least personal. Characteristically, we hold adult human beings responsible for their actions; more generally, we take people's actions to be legitimate or at least appropriate grounds for responses like love and hate, liking and disliking, gratitude and resentment. In all of these ways, we take a person's actions to reflect something really essential about *her*, to represent her in some way. As I will put it, a person's *identity*, her essential self, seems to be evinced in some special way in her actions, so that when we respond to her actions, we are responding to *her*. (Ibid)

The activity and the identity implications carried by the ordinary attributions of agency are, according to Korsgaard, difficult to explain “on a purely naturalistic conception” of agency. By such a naturalistic conception Korsgaard means a view according to which the sense in which an agent is *active* is when her own mental states or representations play a *causal* role in initiating and guiding her movements. The endorsement of a normative conception such as her self-constitucionalist view, Korsgaard claims, seems to be the “most obvious way” to explain

the two implications “from the first-person point of view”—that is, to secure the distinctive character of the first-person perspective on action, or *agential perspective*. This is the perspective occupied by an agent that, in order to act, needs *reasons* to settle the issue about what she ought to do. In other words, the first-personal, agential point of view is an irreducibly practical, or normative, standpoint; it is the stance of self-reflection and autonomous deliberation from which a rational being comes to act as a unified, coherent whole.

By drawing on Plato’s analogy between the human soul and the constitution of the polis on the one hand, and on the tradition of modern political philosophy, which regards the political state as a kind of agent (notably, Rousseau), on the other, Korsgaard proposes a “constitutional model” of agency. The idea is that of conceiving of individual agency as bearing an inherent symmetry with the way in which a political state comes to act as a single entity by unifying its parts under a “general will”—that is, by acting in accordance with its constitution (SC, ch. 7). According to these views, the capacity to act as a single unified agent consists in “the establishment of normative relationships among the people who make it up, in particular their relationships of authority and mutual obligation,” on the one hand, and in the “success in actually *performing* an action, in doing something that *counts* as an action,” which “depends on conformity to the norms in question” (NCA, section 1.1), on the other.

Korsgaard’s proposal is thus that in the case of individual agency, the relevant normative relations are those through which the person achieves the psychic unity necessary to act, and by so doing, constituting her agency. It is in the very act of choosing her action, Korsgaard claims, that an agent constitutes herself as its author. As a result of the processes of practical reflection and deliberation she actively engages in, the agent authorizes those actions as *her own*, she authors them, and comes to a position of *authority* over her own intentional actions. It is by acting according to the principles of practical reason that the agent comes to normatively relate to herself.

2.4 Responsibility

The *activity* and the *identity* implications correspond to the first-personal, agential perspective, and the second- and third-personal perspectives on agency, respectively. The linchpin between these two aspects of agency, Korsgaard explains, is the concept of *responsibility* (AKC, section 3.1, p. 33). Following Strawson, Korsgaard understands responsibility in terms of a *practical* relation, rather than a metaphysical property: “Responsibility is in the first instance something

taken rather than something assigned” (Korsgaard, 1992/1996b, p. 189).²³ Accordingly, holding someone responsible means to treat her as a person responsive to reasons and capable of acting on them, and to expect to be treated by her in the same way. Holding someone responsible, Korsgaard claims, amounts to entertaining a relation of *reciprocity* with her, acknowledging and being prepared to equal demands for justification of one’s actions, attitudes, and goals. Responsibility is then, in a fundamental sense, the relation of mutual reciprocity we enter by participating in the normative practice of asking for, and giving, *reasons*.

Ask yourself, what is a reason? It is not just a consideration on which you in fact act, but one in which you are supposed to act; it is not just a motive, but rather a normative claim, exerting authority over other people and yourself at other times. To say that you have a reason is to say something *relational*, something which implies the existence of another, at least another self. It announces that you have a claim on another, or acknowledges her claim on you. For normative claims are not the claims of a metaphysical world of values upon us: they are claims we make on ourselves and each other. (Korsgaard, 1993/1996b, p. 301)

Throughout her production, Korsgaard has developed an account of responsibility as *answerability* based on argument from the publicity of reasons. “The possibility of personal interaction,” the argument goes, “depends on the possibility of shared deliberation,” which “in turn depends on a certain conception of reasons. Our reasons must be [...] public reasons, reasons whose normative force can extend across the boundaries between people” (SC, section 9.4.5). According to the Kantian constructivist conception of reasons Korsgaard endorses, public reasons are “things that emerge in the interaction between people,” and are thus essentially intersubjective (*ibid.*).²⁴

²³ Watson (1996) articulates the distinction between ‘two faces of responsibility’ (i.e., attributability and accountability): Strawson-inspired accounts (accountability views) regard responsibility as a matter of interpersonal relations—i.e., the agent is held accountable within a network of social and interpersonal relationships—whereas according to attributability views an agent responsible just in case her action ‘disclose’ her commitments and evaluative judgments.

²⁴ Korsgaard’s proposal bears some similarities with other Strawsonian relational approaches to moral responsibility—notably, Stephen Darwall’s. Unlike Darwall, however, Korsgaard does not embrace a thorough second-personal approach, as she anchors the starting point for morality in the deliberative standpoint of the individual agent (on Darwall, cf. Lavin (2008)). R. J. Wallace develops a “normativist” account of moral responsibility, i.e., conditions for being responsible amount to *norms* for appropriately holding someone responsible (treating her as such). On the normativist approach to responsibility, see also Maher (2010).

To sum up, the normative conception of agency is one that accounts for the special kind of authority of the agential perspective, which carries responsibilities and privileges both to ourselves and others. By acting on (essentially *shareable*) reasons, one establishes a relation of authority over herself, making herself accountable to herself and others, and thereby constitutes herself as an agent according to the principles of practical reason.

2.5 The Metaphysics (and Epistemology) of Agency and Normativity

Korsgaard crucially characterizes agency and responsibility in practical terms, thereby opposing her own view to those taking a ‘metaphysical route’ to defining these concepts.²⁵ However, her most recent work importantly aims at making sense of both the metaphysical and normative aspects involved in the ordinary concept of agency (the activity and identity implications). To this end, a cornerstone of her proposal is the characterization of agency in terms of an *activity*—the activity of self-constitution. And it is in order to clarify the distinctive metaphysics of agency and of the practical normativity proper to it²⁶ that Korsgaard introduces the notion of constitutive norms, or normative principles. The reconstruction that follows will illustrate both issues, in turn.

2.5.1 Agency as the Activity of Self-Constitution

First and foremost, the conception of normativity endorsed by Korsgaard is one that she believes to be shared by Plato, Aristotle, and Kant, according to which “normative principles are in general principles of the unification of manifolds, multiplicities, or [...] *mere heaps*, into objects of particular kinds” (SC, section 2.1.1). More specifically, Korsgaard appeals to

²⁵ I borrow the distinction between ‘metaphysical’ and ‘practical’ routes to responsibility from Zheng (2016), who takes it to correspond to the responsibility-as-attributability vs. responsibility-as-accountability distinction. See also Wolf (1990, pp. 16-17) for the distinction between the ‘metaphysical’ and the ‘pragmatic’ approach to responsibility and free will. I agree with Zheng in considering Korsgaard’s most recent work as an example of the metaphysical route to responsibility, rather than the practical one. This interpretation is in contrast with Korsgaard’s own way of conceiving of her account of responsibility in earlier works, where she takes sides with Strawson in emphasizing the relational/interpersonal aspect of the notion of responsibility (cf. previous section). For the sake of a clearer reconstruction of Korsgaard’s position here, I postpone the articulation of my suggestion that the ‘Korsgaardian constitutivism’ go back on the ‘practical’ route to the next chapters.

²⁶ Chapter 2 of SC is entirely devoted to “The Metaphysics of Normativity.” However, the appeal to the idea of “constitutive norms” is far from being exclusive to the most recent work, as we can find it at least as early as in SN, as well as in Korsgaard (1999/2008). See Bagnoli (2009) for a discussion of the transition from Korsgaard’s early constructivist proposal to her late appeal to constitutivism, which “amounts to importing a significant piece of metaphysics into the picture.”

Aristotle's extension of his account of artifactual identity to the characterization of a living thing in terms of a self-maintaining form—that is, “a functionally related set of powers” engaged into self-constituting processes. The specific difference of human agency—in relation to agency as a genus encompassing self-determined kind of efficacy—consists in the capacity of being motivated and guided by reason.

As we have seen, for Korsgaard being an agent means to be faced with, and being engaged in, the task of acting and choosing—and it is by acting on the basis of reasons that a person comes to constitute herself into an agent, bringing about the kind of unity that makes it possible to ascribe actions to her as their author. In an important sense, then, there is no agent prior to her choosing and acting—action *is* therefore self-constitution (ivi, section 1.4.8), and agency *is* the activity of self-constitution. Korsgaard anticipates the objection that his characterization of agency seems to lead to a paradox, the *paradox of self-constitution*: “How can you constitute yourself, create yourself, unless you are already there?” (ivi, 1.4.3). The paradox, however, stems from erroneously conceiving of self-constitution as “a state that we achieve and from which action then issues.” On the contrary, Korsgaard insists, self-constitution should be conceptualized in terms of an activity: “what it is to be a person, or a rational agent, is just to be engaged in the activity of constantly making yourself into a person” (ivi, section 2.4.1; cf. supra section 2.2).

Korsgaard's characterization of agency in terms of activity builds upon Aristotle's teleological conceptions of objects and activities, and works, in conjunction with her distinctive interpretation of the “function argument,”²⁷ to the endorsement of the claim that the function of action is to constitute the agent (ivi, section 5.1.3). The organization in virtue of which a thing is the kind of thing it is—its functional organization—carries with it normative implications, in that it supports normative judgments about the thing itself.

2.5.2 Constitutive Principles and their Double Nature

Korsgaard calls *constitutive standards*, or internal standards, the standards that “apply to a thing simply in virtue of its being the kind of thing that it is” (ivi, section 2.1.2). Constitutive

²⁷ I am aware that the talk about ‘function’ and ‘teleological structure’ carries some difficulties. Here, however, I am not introducing the Function Argument nor the set of objections and counter-objections to it in the literature, since referring to it here is just functional to the reconstruction of Korsgaard's account. For Korsgaard's defense and discussion of the argument, see her “Aristotle's Function Argument,” CA (essay 4).

standards have a double nature, in that they are descriptive and normative at once with respect to the object they are standard for:

They are descriptive because an object must meet them, or at least aspire to meet them, in order to be what it is. And they are normative because an object to which they apply can fail to meet them, at least to some extent, and is subject to criticism if it does not. This double nature finds expression in the fact that we can criticize such object either by saying that they are poor object of their kind (“That’s a poor encyclopedia, it isn’t up to date.”), or by saying that they are not such objects at all (“That’s not an encyclopedia: it’s just a compendium of nineteenth-century opinion!”). (CA, p. 8)

Analogously, constitutive principles—i.e., constitutive standards applying to activities—relate the performer/agent to the corresponding activity in such a way that “we can say that unless you are following the principle in question, you are not performing the activity at all” (ivi, p. 9).

Although Korsgaard does not herself trace the connection, reference to the notion of *constitutive rule* and the relevant debates might be helpful here to further clarify the point.²⁸ Constitutive rules are such that they constitute, rather than merely regulate, the relevant practice. Metaphysically speaking, they create, or bring into existence, the practice they are rules for (see Glüer & Wikforss, 2009/2015, section 1.2). In their constitutive role, they are therefore logically and conceptually prior to the practice in that they define it and set the very conditions for its possibility and meaning. Constitutive rules have at the same time a normative function, in that they establish what is allowed or prohibited within the activity/practice, and normative force, inasmuch as from the point of view of the participants they provide guidance about to what to do (cf. Rawls, 1955).

One potential problem with the double nature of constitutive rules is that a violation would automatically imply exclusion from the practice, as a move not compliant with the rules would cease the act from counting as an instance of the practice altogether. I have referred to this as

²⁸ Berteau (2013, pp. 84-89) carefully spells out the connection between the debate on constitutive rules and the one about the normative status of constitutive standards of agency. Even though the notions of constitutive rules (as articulated by G.H. von Wright and J. Searle), secondary rules (by H. L. A. Hart), and practice rules (by J. Rawls) are traceable back to different sorts of debates, they present several similarities and are to some extent relatable, deepening our understanding of different aspects of what I will call, in chapter 4, *socially-generated* constitutive normativity.

the problem of defective action, of which this work is meant to provide clarification (chapters three and four), and a solution (chapter five).

2.5.3 Agency's Metaphysics and Epistemology

According to Korsgaard, a constitutive principle is therefore an 'internal standard' arising "from the very nature of the object or activity to which it applies." Activities are distinguished from mere mechanical processes or sequences of events in that they are "by their nature, directed, self-guided, by those who engage in them, even if they are not directed or guided with reference to external goals" (CA, pp. 7-9).²⁹ For an activity to be self-guided means that it is an *essentially conscious* activity, without necessarily requiring that the one who engages in it be conscious *that* she is doing it (cf. SN, pp. 234-237, especially n. 23). "In the case of essentially goal-directed activities," Korsgaard continues, "constitutive principles arise from the constitutive standards of the goals to which they are directed" (CA, p. 8). The internal goal/standard of agency being self-constitution, successful actions are those through which an agent comes to constitute herself as an internally coherent and unified whole.

The kind of metaphysics Korsgaard endorses for the concept of activity is a Platonic one:

the "precise sense" or perfect version of an activity stands in a complex relation to the activity, because it is at once both normative and constitutive. Although it is not true that you are not performing an activity at all unless you do it precisely, it is true that you have to be *guided by* the precise version of the activity in order to be performing the activity at all. And at the same time the precise sense sets normative standards for the activity. (SC, section 2.1.5)

The concept of action, Korsgaard acknowledges, comes in degrees, and an agent may act in ways that unify her to a greater or lesser degree (ivi, sections 1.4.8, 8.5; CA, essay 3). Those actions are still actions, but are *defective*, in that "they fail to constitute their agents as the unified authors of their actions" (SC, section 2.1.6). Bad and defective actions are thus 'tokens' of "the same activity—the activity of self-constitution—badly done" (CA, p. 126; cf. also SC, section 6.4.9).

²⁹ Cf. NCA (section 1.3): "Agency is liable to a standard of success and failure from the inside, that is, by virtue of its own nature: such liability is part of its constitution."

The constitutive principles of agency—that is, the standards applying to the activity of self-constitution—are, according to Korsgaard, the principles of practical reason (specifically, the Hypothetical and Categorical Imperative). This comes from the very nature of action, as Korsgaard follows Aristotle and Kant in considering an action as “involving both an act and an end, an act done for the sake of an end,” and reasons—the considerations supporting the choice of an action—as having an essentially reflexive structure. The description of an action incorporates therefore a principle (a Kantian maxim, an Aristotelian *logos*), and the categorical imperative in its universal law formulation is both descriptive of and normative for the activity of acting for a reason (CA, p. 219; SC, section 1.2).³⁰

The normative force of the principles of practical reason is analogous to that of the principles of logical inference: “they must be principles *in accordance* with which we operate—either well or badly” (CA, p. 7; cf. SC, sections 5.1.3, 5.2). The parallel runs also at the epistemological level, as

Knowledge of the normativity of practical principles doesn’t transform them into premises which are then applied, any more than knowledge of the normativity of logical principles, like modus ponens, transforms those into premises which are then applied. (SC, section 4.2.3)

Accordingly, the aim of a philosophical theory of agency in spelling out the constitutive principles of agency is to illuminate a distinctively practical kind of knowledge:

What Philosophy leads to when it formulates such principles and reveals the sources of their normativity is not knowledge which is then to be applied, but rather self-knowledge. That is to say, it leads to a self-conscious appreciation of what you are and of how you work, which will make you better at being what you are and at working in that way. (Ibid)

Leaving aside the epistemological aspects of agency, which Korsgaard herself does not engage, the present reconstruction of the metaphysics of agency reveals a source of tension between the ‘problem of defective actions’ potentially arising from Korsgaard’s

³⁰ Both the Hypothetical and the Categorical Imperatives (HI and CI) are constitutive principles of action: the former (the principle of instrumental reason) concerns the agent’s efficacy, the latter her autonomy—together they are the constitutive standards of an agent. The HI, however, is already contained into the CI (SC, sections 4.4, 5.1).

characterization of the double nature of constitutive norms, and her acknowledgement that the notions of agency and action come in degree.

The next two chapters are dedicated to an exploration of the problem of defective actions via the analysis of two case studies, and to an elucidation of the notion of constitutive norms, respectively. Together, they provide a critique of Korsgaard's account—in particular, her conflation of constitutive normativity with moral normativity—, and set the stage for my reframing and expansion of the theory, in chapter five.

Chapter Three – Defective Agency?

In the previous chapter I provided a comprehensive and abstract reconstruction of Korsgaard's constitutivist theory of agency. This chapter brings the theory into practice, through the analysis of two case studies: cases of *disorders of agency* and of *socially displaced agency*.

Since these cases represent two varieties of alleged breakdowns of agency, addressing them is particularly relevant to Korsgaard's theory, in light of what in the literature is known as the problem of defective actions. The interpretations of the cases that I rely on and endorse, however, make a strong case for considering them as instances of genuine agency, thereby occasioning a critical assessment of Korsgaard's account. This will be done with a view to providing a highlight of the theory's most valuable features, upon which I will build my positive proposal in chapter five. Importantly, both of the following cases emphasizes that agency *constitutively* involves an interpersonal, social, and institutional infrastructure, thus corroborating Korsgaard's core claim that responsibility is a practical relation.

3.1 Constitutive Norms and the Problem of Defective Actions

Constitutive norms, Korsgaard claims, are descriptive and normative at once in that they both provide guidance, and constitute, or define, the very activity one is engaged in by being guided by, and conforming to, them:

They are normative, because in performing the activities of which they are the principles, we are guided by them, and yet we can fail to conform to them. But they are also descriptive, because they describe the activities we perform when we are guided by them. (CA, p. 9)

This type of double-edged sword is rare in philosophical theory, where the normative or the ideal is traditionally separated from the descriptive or the real. Most often, philosophers postulate an idea as a recognized and separated standard to aim at to improve practice. Accordingly, recognizing the mix of the normative and descriptive in agency already sets Korsgaard's account off on a unique foot.

There seems, however, to be a problem with the thesis about the double nature of constitutive principles of agency. As Korsgaard says, since constitutive principles are normative, we can

fail to conform to them,³¹ thus suggesting that they are ‘norms for action.’ But given that the principles are also descriptive, and a violation of a description is impossible, it would make one who fails to conform to them inherently defective *as* an agent. Since Korsgaard accepts the possibility of bad action, then, according to her characterization of constitutive norms, a violation would sanction the defective status of the action. In turn, a defective action would fail to constitute its author as an agent, up until the exclusion from the kind-membership individuated by the norms themselves. The descriptive character of constitutive norms seems to make them ‘norms for being.’ Agents, according to Korsgaard, are *evaluative* kinds: “the agent, just as such, is the kind of thing that can succeed or fail” (NCA, section 1.2). By failing to conform to the constitutive principles of the activity of self-constitution, one performs defective actions that make her less of a fully unified agent, and thus less of an agent—“a mere heap” (SC, section 10.2.3; cf. also section 8.2.4):

Since the function of action is self-constitution, [...] bad actions, *defective* actions, are ones that fail to constitute their agents as the unified authors of their actions. (Ivi, section 2.1.6)

To put the problem in a highly schematic form, Korsgaard’s characterization seems to present us with the following situation: constitutive norms of agency/action have a double nature—they are both descriptive and normative with regard to the activity of self-constitution. They describe what an action is (they are norms for being), and at the same time sanction as defective instances of actions that do not conform to them (they provide norms for action). In addition, given that the relation between an agent and her actions is not one of production, but rather of authoring *and* self-constitution, defectiveness of actions *transmits* to the agent, making her less of a fully unified, and therefore less of an agent. In Korsgaard’s words:

[...] if your action is unsuccessful and you do not bring about the state of affairs that you intended, it is not (or not just) the action that is ineffective. It is *you* that is ineffective. It is not as if you were effective in producing the action, but then the action, once out there on its own, failed, like a defective machine you have invented and then let loose on the world. The action is not your product: it was *you* that failed. An unsuccessful action renders you ineffective. Therefore a successful action is one that renders you effective. A similar point holds for autonomy [...] Therefore a successful action, an

³¹ The ‘error constraint’ is usually thought to be a requirement for a norm to be prescriptive, or action guiding; see Lavin (2004) for a critical discussion of this assumption.

action that is good as an action, is one that renders its agent both efficacious and autonomous. (SC, section 5.1.4)

It is the essential nature of action that it has a certain metaphysical property [...] But in order to have that metaphysical property it must have a certain normative property [...] This explains why the action must meet the normative standard: *it just isn't action* if it doesn't. (Ivi, section 8.1.3)

According to Korsgaard, consistency, self-reflection, and self-unification are necessary conditions for unified agency, which are also requirements for intentional action (cf. Arruda, 2016a). Autonomy and efficacy are the “essential characteristics of an agent,” and actions that fail to constitute the agent as autonomous and efficacious fail to constitute the agent altogether:

If you fail to follow the Kantian imperatives you will not be efficacious and autonomous, and then you will not be an agent. Your action constitutes you as an agent by being chosen in a way that renders you, the agent, efficacious and autonomous. (Ivi, section 5.1.3 ff.)

For the moment, this rough outline of the problem of defective actions should suffice, as it will become clearer through the discussion in the next sections. A thorough examination of Korsgaard's characterization of constitutive norms of agency/action will be undertaken in chapter four, and chapter five will be devoted to a reframing of Korsgaard's constitutivist account of agency. In this chapter I am going to discuss cases of ‘disorders of agency,’ which are meant to show that while the agent's actions might ‘misfire’ or fail to display Korsgaard's required pattern of consistency, self-reflection and self-unification (and often, morality), her agency does not fade nor disrupt. Moreover, by relying on evidence emerging from therapeutic treatments of disorders of agency, I argue, *contra* Korsgaard, in favor of a clear distinction of the notion of *agential responsibility* from that of moral responsibility, claiming that only the former is necessary for agency (cf. *supra*, sections 2.2, 2.4).

3.2 Disorders of Agency

In mental health-care practice there is a category of conditions referred to as “personality disorders” (PDs).³² The term ‘personality’ stands for a set of traits that incline one to

³² Although I am keenly aware of the complex conceptual and methodological issues concerning the nature and classification of so-called ‘mental illnesses’ I will not engage in any of the relevant philosophical discussions here. PDs, however, are not categorized as ‘illnesses’ nor ‘diseases,’ but rather as a ‘disorders.’

behave/act, think, and feel in stable ways with respect to certain circumstances, which make her ‘the kind of person she is.’ PDs are defined by the *Diagnostic and Statistical Manual of Mental Disorders* (DSM IV-TR, 1994, p. 689) as:

- A. An enduring pattern of experience and behavior that deviates markedly from the expectations of an individual’s culture. This pattern is manifested in two (or more) of the following areas:
 - 1. Cognition (i.e., ways of perceiving and interpreting self, other people, and events);
 - 2. Affectivity (i.e., the range, intensity, lability, and appropriateness of emotional response);
 - 3. Interpersonal functioning; and
 - 4. Impulse control.
- B. The enduring pattern is inflexible and pervasive across a broad range of personal and social situations.
- C. The enduring pattern leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.
- D. The pattern is stable and of long duration, and its onset can be traced back at least to adolescence or early adulthood.
- E. The enduring pattern is not better accounted for as a manifestation or consequence of another mental disorder.
- F. The enduring pattern is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., head trauma).

There are several features of PDs that make them a relevant case for a philosophical account of agency and for the present discussion, in particular. First, PDs generally, have what Hanna Pickard calls a “Janus-faced nature”—i.e., two distinct but closely connected characteristics. Beside causing pain and distress to the individual, the problematic traits and patterns of behavior involved in PDs also often involve impairment of areas of interpersonal functioning, as well as harm to others. For this reason, PDs are considered “a psychiatric condition that essentially involves a moral component” (Pickard, 2011, pp. 182-183), and whose management poses unique scientific, philosophical, legal, clinical, and practical challenges.³³

³³ *Philosophy, Psychiatry, & Psychology* (2011:18 (2)), edited by Pickard, is “a multidisciplinary effort to further understanding of personality disorders, drawing on scientific, clinical, philosophical, social, and legal perspectives. It also includes a first-person narrative of an ex-service user’s experience of living with personality disorder” (Pickard, 2013, p. 1148). See also <http://personalitydisorder.org.uk/about/> and <http://personalitydisorder.org.uk/wp-content/uploads/2015/04/PD-No-longer-a-diagnosis-of-exclusion.pdf>

Second, Pickard points to the fact that PDs lie on a *continuum* with ‘normal’ human personality—normal traits are “taken to the extreme in PD[s]” (Pickard, 2011, p. 182). The relationship of ‘continuity’ between PDs and normal traits seems to emerge from the close similarities with our ordinary experience and behavior evoked by the first-personal narratives of what it is like to live with PD. It is important to stress, however, that the what counts as ‘normal’ is always articulated in relation to the expectations of a given culture (cf. DSM’s characterization above), implying that a major role in the definition of disorders of agency related to PDs is played by societal and cultural factors.

Third, and most importantly, diagnoses of PDs are prompted by symptoms typically involving patterns of behavior (actions or omissions thereof) and attitudes, which play a crucial role in the maintenance of the condition:

They could be emotionally cruel, or extremely angry and threatening without just cause; they might self-harm or disengage from the Community without explanation, provoking high levels of anxiety in others concerned for their well-being; they might shirk their Community tasks and responsibilities, leaving others to pick up the work. (Pickard, 2013, p. 1135)³⁴

Conversely, with the appropriate therapeutic, interpersonal, and societal support, people with PDs can successfully work to change their problematic behavioral patterns and traits (Pickard, 2011, p. 182). In other words, the possibility for *change* and recovery for people with PDs crucially involves the exercise of their *agency* within an appropriate therapeutic interpersonal environment. For this reason, and as a result of a clash of intuitions and beliefs she herself experienced in approaching the psychiatric practice, Pickard coined the term “disorders of agency,” thereby instigating an important shift in focus from ‘personality’ to ‘agency.’

Finally, unlike other kinds of disorders of agency (e.g., those associated with schizophrenia and Alzheimer), discussion of cases of PDs and their treatment in the context of therapeutic communities can be successfully conducted quite independently from any specific empirical hypothesis concerning the psychological architecture underlying the impairment (i.e., Sense

for an insight into the development of policies by the National Institute for Mental Health in England (NIMHE) supporting a therapeutic framework for the treatment of people with PDs.

³⁴ For the present discussion I draw on Pickard (2013), based on Pickard’s own working experience as a clinician in a Therapeutic Community for service users with PDs.

of Agency/Sense of Ownership, and memory, respectively).³⁵ By considering disorders of agency related to PDs, the focus is therefore prominently on the level of interpersonal/institutional contexts, relations, and practices, thereby providing a privileged insight into the conceptual and normative framework of agency and responsibility.

Along with the interpretation of these cases provided by Pickard and her specific proposal, my claim will be that cases of disorders of agency bear considerable significance for an understanding of the way we conceptualize the notion of responsibility and the relevant normative practices in non-clinical contexts. This represents a minimal level of normative reflection—that is, which contexts should we consider when we model the notion/concept of responsibility, whether those carry normative preferences (if so, which those are), and whether a homogeneous model would fit our practical needs for the concept of responsibility.

3.3 The Clinical Model of “Responsibility without Blame”

Pickard’s research and analysis is of further pertinence to my assessment of Korsgaard’s account of agency for the following reasons. According to Pickard, clinical practice involving service users (i.e., patients) with disorders of agency seems to be at odds with both philosophical and ordinary beliefs about the close relation between responsibility and evaluative reactive attitudes, where reactive attitudes are considered the evaluative emotional responses elicited by the kind of attitude toward us manifested by other people’s behavior, like resentment, blame, praise, gratitude. The idea at the core of both the ordinary and the philosophical perspective, Pickard observes, is that responsibility is an essentially moral notion and that reactive attitudes are integral—either constitutively or, more weakly, via a relation of appropriateness—to the practice of holding people responsible. In other words, holding someone responsible characteristically involves morally evaluative, personal reactive attitudes.

Korsgaard seems to be on the same page with respect to the connection between responsibility and reactive attitudes. If we think about Korsgaard’s characterization of the two implications of the ordinary concept of agency—and of the linchpin between the two, the notion of responsibility—we can recall that there were two claims at play. The first, which I will refer to as the *Identification Thesis*, is about the intimate relationship between agents and their actions via the notion of personal, or practical identity—of which the actions are a

³⁵ Now that this distinction is in place, I henceforth use the term ‘disorders of agency’ narrowly—or in Pickard’s sense, so to speak—to refer just to disorders of agency related to PDs.

manifestation or expression, thereby qualifying as *her* actions—since “the most essential part of the person is *constituted* by her actions” (SC, section 5.5.1).³⁶ The second claim states that the reactions to someone’s actions—the practical attitudes constitutive of treating someone as an agent—are evaluations/assessments, and as such, crucially involve standards of appropriateness. Korsgaard’s interpretation of the kind of appropriateness at stake is moral (as is the one proper to common-sense).³⁷ I shall call this the *Moral Evaluation Claim*.

The two claims work together in Korsgaard’s theory in that agency is characterized both in metaphysical terms (as the activity of self-constitution) and in practical terms (as choosing what to do, thereby taking responsibility for a specific practical identity that enacts a certain pattern of self-reflection and self-deliberation). Since the practical, for Korsgaard, is subsumed into the moral (the CI is *the* constitutive principle of agency), then the fundamental dimension of evaluation for agency is the moral one.³⁸

Paying attention to therapeutic contexts, however, sheds light on the fact that the relationships among ‘failures’ of agency (including those involving morally relevant aspects), competent agency, responsibility ascriptions, and personal identity, is more complex than Korsgaard seems to have it.

Crucial for the treatment of disorders of agency is the undertaking, on the part of the practitioners, of a stance of “responsibility without blame.” In Pickard’s words:

Service users were responsible for their actions and omissions and accountable to the Community for them, but an attitude of compassion and empathy prevailed, and they were not blamed. (Pickard, 2013, p. 1135)

³⁶ Here is the full quote: “The intimate connection between person and action does not rest in the fact that action is caused by the most essential part of the person, but rather in the fact that the most essential part of the person is *constituted* by her actions.” I use the *Identification Thesis* quite vaguely to refer both to the *metaphysical* and the *practical* relation between an agent and *her* actions. In the next chapter (section 4.3), I distinguish between the metaphysical and the practical counterparts of the *Identification Thesis*. Later in this chapter I use it in a more restrictive way, to refer just to the practical counterpart. Cf. *infra*, section 3.5.

³⁷ The reactive attitudes which are supposed to be involved in the identity implication of the ordinary concept of agency are “normative or at least personal.” Cf. *supra*, section 2.3, where I reconstruct the part of Korsgaard’s argument that relies on the implications of the ordinary concept of agency.

³⁸ Cf. SC (section 8.5.2): “[...] integrity in the metaphysical sense—the unity of agency—and in the moral sense—goodness—are one and the same property.” Here Korsgaard refers to Kant’s and Plato’s arguments, highlighting the same pattern of establishing that a normative property (“universalizability in Kant’s argument, justice in Plato”) just is the counterpart of a metaphysical property (“autonomy in Kant’s argument, constitutional unity in Plato’s”). Her account follows the same strategy; in other words, Korsgaard endorses the “ancient metaphysical thesis of the identification of the real with the good;” (*ivi*, section 2.1.1). See also CA (pp. 110-113).

For Pickard, blame is a complex emotion involving a negative affect towards the addressee, which is characteristically accompanied by a sense of entitlement to that negative reactive attitude in relation to the addressee's wrongdoing. The notion of responsibility without blame captures the fact that in their therapeutic relations and interactions with practitioners, service users can be successfully held accountable for their actions without being morally evaluated. In other words, in these cases, usual components of holding someone accountable for her actions, such as ascribing blameworthiness and imposing consequences, remain—with the exception of the expression of moral blame. Still, without blame responsibility is preserved, and successfully goes through, despite or perhaps in virtue of that very omission.

Dissecting the elements at play in Pickard's model of responsibility without blame reveals that Therapeutic Communities are interaction spaces between service users and mental health practitioners that focus on and encourage change in service users' behavior. The therapeutic nature of these spaces lies in both the goal of the interaction (i.e., fostering change of maladaptive patterns of behavior; improvement of interpersonal functioning), and the type of relationship between the service users and practitioners (i.e., between a user and a provider of a health care service, which is a distinctively asymmetrical relationship). The specificity of the situation (in terms of goal and type of relationship) notwithstanding, "it is a presumption of treatment," Pickard explains, "that service users have choice and control over their behavior and can therefore be asked to take responsibility for it, as we naturally say." This presumption pervades and structures the interactions within Therapeutic Communities, in that it is made explicit through "the language of agency and responsibility." What service users usually lack, is "full conscious knowledge of why they are behaving as they do, or what the full effects of their behavior on others may be," as well as a full control over their behavior. Evidence from clinical treatment shows that service users routinely exercise choice and control over their behavior "when they have incentive, motivation, and genuinely want to do so" (ivi, pp. 1137-1138). Control and conscious knowledge are therefore *graded* notions.

The fact that this presumption is operative and constitutes a necessary condition for effective treatment is crucial, as it means that agency exercised in clinical contexts shares the same features, in different degrees, of nonclinical/ordinary contexts. In particular, the parallel with the ordinary practice of holding people responsible is maintained in that judgements and ascriptions of blameworthiness are not suspended, and the demands (i.e., specific requests, and potential imposition of negative consequences) and expectations involved in participating in the practice of regarding oneself and holding each other responsible are not lifted either. What practitioners refrain from doing—thanks to appropriate training and cultivation—is to express

blame towards service users (where the expression of blame might be considered a minimal level of acting on it). By taking a blame-free participatory stance—which is, importantly, different from Strawson’s objective stance—practitioners enable an environment where service users are not absolved from responsibility, and their agency is positively affirmed, rather than denied.

Pickard’s contention is therefore that if the evaluation at stake in reactive attitudes is construed as moral in kind—thus involving moral standards in the ascription of agency and responsibility—too close of a connection between responsibility and morality is established; in particular, one that obscures the fact that responsibility is a distinctively *agential* concept, alongside, but not synonymous with or a part of, morality. The experience stemming from therapeutic treatments of disorders of agency seems to testify exactly to the conceptual and practical viability of the distinction between moral and *agential responsibility*, as the exercise of agential capacities can be enhanced specifically by refraining from the moral evaluation involved in blaming.³⁹

In light of the conceptual possibility and practical viability of the clinical stance of responsibility without blame, Pickard develops a conceptual framework that clearly distinguishes responsibility, blameworthiness, and blame. She distinguishes two varieties of blame: affective and detached. Affective blame involves a characteristic ‘sting’ and a moral connotation that represents a hindrance to the effective exercise and improvement of the service user’s agential capacities. On the other hand, what makes detached blame conducive to effectively securing service users’ accountability and change in behavior is that it is devoid of “a sense of entitlement to any negative reactive attitudes and emotions one might experience, no matter what the service user has done” (ivi, pp. 1141, 1146).⁴⁰

3.4 Relevance of the Clinical Model for the Present Work

What emerges from the clinical practice with sufferers of disorders of agency shows that while their participation in responsibility practices—i.e., being considered as responsible agents and

³⁹ To be sure, the distinction between causal, or explanatory, responsibility and moral responsibility is a customary one in debates on responsibility. For a “structured taxonomy of responsibility concepts” (STRC) see Vincent (2011). See del Corral (2015) for a characterization of ‘agential responsibility’ as explanatory responsibility.

⁴⁰ Pickard’s notion of ‘affective’ blame is similar to, but differs under important respects from, Angela Smith’s notion of ‘active’ blame. See Pickard (2013, p. 1141 n. 13).

held in relations of accountability—is indeed a necessary condition for them to effectively be accountable and take responsibility for their actions, thereby exercising agential capacities, moral evaluation is not essential to it, and could actually be a hindrance. I therefore side with Pickard in maintaining that these concepts should be best understood in morally neutral terms.

Getting back to the assessment of Korsgaard’s proposal, this consideration seems to pose a serious challenge to the *Moral Evaluation Claim*—that is, morally evaluative reactive attitudes are not necessary for the practice of holding someone responsible. My suggestion is therefore that the *Moral Evaluation Claim* be turned into the *Agential Responsibility Claim*. While this move constitutes a seemingly minor adjustment to Korsgaard’s theory, it is one with big repercussions, in that a more nuanced account of agency—in particular, one that clearly distinguishes between moral responsibility and agential responsibility—would thereby broaden the scope of humans we can consider, and who can understandably consider themselves, as agents.

Elaborating on the elements pinpointed by Pickard as central to the presumption of treatment in clinical contexts,⁴¹ I will regard as *agential* those capacities in virtue of which an agent is responsible (without blame), or can take responsibility for, her behavior. Agential capacities include: i) conscious knowledge of what one is doing,⁴² ii) choice, and iii) a degree of control over one’s behavior. Accordingly, I propose the following claim:

Agential Responsibility Claim: treating someone as an agent is to treat her as accountable/answerable for her actions, which presupposes that she has knowledge of what she is doing, can exercise choice and a degree of control over her behavior.

It is this presupposition’s structuring of the interpersonal attitudes and reactions to one’s behavior that proves to be a necessary condition for exercising agency. In other words, it is not just the agent’s individual psychological arrangement that matters for agency, but rather the fact any such capacity of the individual agent is actually acknowledged/recognized, and

⁴¹ Here I draw on Pickard’s characterization of what is presupposed by effective clinical treatment. (Ivi, p. 1141).

⁴² Pickard’s understanding of this notion is admittedly an intuitive one. By ‘conscious knowledge’ of behavior she refers to “the way we normally know what we are doing when we are doing when we do it. [...] Normally, we have some knowledge of why we are acting, some knowledge of how we are acting, some knowledge of what we intend in acting, and some knowledge of what effects our actions have on the world” (ivi, p. 1136 n. 5).

interacted with, within an interpersonal relation. What will be involved in actually treating someone as responsible will vary depending on the context, the nature of the relationship with her, along with a consideration of her past personal history. In any case, it will involve some concrete effects in terms of re-action/response, to the agent's actions.

While corroborating the relevance of Korsgaard's approach to responsibility as a practical relation for the notion of agency, treatment of disorders of agency indicates that those who in Korsgaard's terms would be 'mere heaps,' (and therefore less of fully unified agents) (cf. *supra*, section 3.1) would rather seem to be always already agents, and never cease to be such (see also Millgram, 2014). The process of recovery for people with disorders of agency—whose actions/omissions would, according to Korsgaard, consistently fall short of making them *unified* agents—shows that any degree of realization of one's agential capacities, as manifested in her actions, is sufficient to hold her accountable for her behavior, and for her to effectively take responsibility for those actions. A 'disorder of agency' does not make the person who is diagnosed with it *defective qua* agent. The defect, then, would not be in her agency. Rather, it would be in the dysfunctional dynamics of the agent's proximal relational environment, and in those resulting from her past history (including the complex, overlapping influence of many different structural societal factors).

Overall, disorders of agency and their treatments point to the relational, interactional character of agency and responsibility, as well as their social dimension. The discussion in the present chapter is meant to challenge the very idea of *defectiveness* attached to the term 'disorder,' stemming from the placement of moral blame on an individual.

Alternately, we should consider agency as a complex and interactive capacity, the exercise of which crucially depends on interpersonal relationships and contextual factors—at *any* degree or stage of actualization. A characterization of agency in these terms would imply not just that a lack of appropriate interpersonal relationships and material opportunities might importantly impact on the development and exercise of the capacity of agency—which Korsgaard herself would certainly concede (cf. *supra*, section 2.2)—but eventually lead to the following stronger claim:

Interaction Claim: agency itself is *constitutively* a function of one's location within a network of social and interpersonal relationships,⁴³ as well as of the quality of these relationships, as perceived by the agent.⁴⁴

The hypothesis central to the reframing of constitutivism *as a theory of agency* that I pursue in this work intends therefore to build upon and strengthen Korsgaard's practical approach to agency and responsibility. On the other hand, it carries with it important differences, that will be illustrated in due course.

3.5 Socially Displaced and Distorted Agency

By challenging the categorization of 'disorders' of agency as instances of defective agency, the previous sections were meant to make a case for distinguishing agency as a capacity from its actualization in an agent's actions. The connection between agential capacities and the *practice* of holding someone responsible for their actions has been emphasized by introducing the *Interaction Claim*. Taken together, these two moves are designed to shift the focus from conceiving of the *Identification Thesis* (regarding the relation between an agent and her actions) as a metaphysical relation (of self-constitution), to a practical relation—i.e., taking place within interpersonal and social interaction.

A further instructive case to illustrate the interactive/relational character of agency, and therefore to support the *Interaction Claim*, is that of agency exercised in contexts of oppression and violence. In the *Disorders* scenario discussed above, the role of interpersonal relations and practices of accountability was showed to be a necessary condition for the agent to effectively develop and exercise their agential capacities. In the *Displacement* scenario presented in this section, the *Interaction Claim* gains (indirect) support by looking at the negative impact that societal relations might have on individual agents' ability to actualize their agency. In these cases, agency operates *despite/against/within* structures of power and oppression which systematically distorts the intentions of some agents. As a result, the characteristic practical

⁴³ For example, the cases of feral children: being isolated from other people prevents them from having agency in any usual sense of the word. Another example is that of solitary confinement, that has proven to have disintegrating/destroying effects on sense of self (see Gallagher (2014, section 3) for an overview), which I would expect to extend to agency as well.

⁴⁴ Think, for example, of the importance of trust and trustful relationships for adequately developing and exercising one's agential capacities, or of the (enhancing/diminishing) effects connected with interiorizing certain (positive/negative) social labels. Cf. *infra*, section 5.1.3.

relation of ‘authorship’ between an agent and the actions she performs, is bent to the effect that the actions’ locus of significance is *displaced* from their agents.

The *Displacement* scenario represents a problem for Korsgaard’s theory because a consistent failure in the agent’s efficacy—i.e., the “success in actually *performing* the action, in doing something that *counts* as an action” (NCA, section 1.1)—would eventually fail to make her an agent at all, if anything by disempowering the agents’ potential for agency via the internalization of derogatory labels and other mechanisms of social discrimination and oppression.⁴⁵ The interpretation of the *Displacement* scenario that I favor in the following section, on the contrary, accounts for the ‘displacement effect’ without obliterating the agency of the disenfranchised individuals. The explanation I will provide for how that is the case relies on the *Interaction Claim* and calls for a further investigation of its connection with the *Identification Thesis*. Before getting there, let me introduce the cases of displaced agency.

The process through which agency gets displaced from an agent has been characterized as a form of “social authoring:”

As an example of social authoring of action, we can refer to the 2005 media coverage of Hurricane Katrina, which fixated on reports of looting in flooded New Orleans (Sommers et al. 2006). Two photographs published by different news agencies captured the public’s attention: One photo features white-skinned people traveling through the flooded area carrying food, and includes the caption, “Two residents wade through chest-deep water after finding bread and soda from a local grocery store after Hurricane Katrina came through the area in New Orleans, Louisiana” (Agence France Press 2005). The other image is nearly identical except the subject of the photo is a black man, and this time the caption reads, “A young man walks through chest deep flood water after looting a grocery store in New Orleans” (Associated Press 2005). Defining their actions as “finding” and “looting,” the captions diverged in how they narrated the same apparent intentional action. Although the caption for the white-skinned agents abstained from normative explanations about their actions, the caption for the black agent affirmatively criminalized him and his actions, and institutionalized that characterization by reporting it as news. [...] Some have contended that the divergent captions are a consequence of different standards for the two news agencies that published their respective photos and captions. However, analysts have shown that the captions were consistent with the media’s

⁴⁵ See essays collected in Andersen & Collins (2016). For an overview on internalized oppression see Willett et al. (1999/2015); on epistemic oppression and epistemic injustice see Dotson (2013, 2012), and Fricker (1999, 2007).

overall pattern of racialized characterizations of individuals surviving the aftermath of the hurricane (for example, Sommer et al. 2006). [...]

Scenarios such as the looting/finding example highlight how intention is not just authored by the agent, but is also *socially* authored through others' discernment and translation of that action. The term *social authoring* is meant to convey a relationship of production between "observer" and "act." I distinguish between social authoring and "social reading" of an act: To read an act is to apprehend an existing meaning, but to author an act is to create something new. When facilitated by reasoning designed to reinforce and rationalize systems of domination, social authoring relies on and further entrenches an institutionally sanctioned distortion of the intentions of some agents. (Bierria, 2014, pp. 129-130, 142 n. 1)⁴⁶

Alisa Bierria's example and analysis quoted above provides us with an important theoretical perspective on the *social dimension* of agency. Korsgaard's acknowledgement of this dimension can be found, in a minimal form, in her account of responsibility as a practical relation of mutual reciprocity entered by the agent, inasmuch as she is engaged in the activities of practical reasoning and of exchanging reasons (cf. *supra*, section 2.4). Significantly, Korsgaard's claim that the practical relations entered by participating in practices of mutual accountability seems to imply that those relations are inherently symmetrical, and actualized by default in virtue of the agent's acting in conformity with the principles of practical reasoning.

The phenomenon of socially displaced agency, however, reveals that it is not just the normative relations among the agent's mental states, established through the activity of practical reasoning and the exchange of reasons (i.e., not just her individual process of self-constitution) that matter for agency. What is relevant for the successful performance of an agent's action is also the *uptake* on the part of other agents—that is, a relation of dependence on others for the action to be understood as having a certain meaning, and to take the corresponding intended effect.⁴⁷ The uptake is in turn a function of the normative relations among the agent and her social and institutional environment, and those relations might also be such that the authoring process of the agent's intentional actions is displaced from her and

⁴⁶ See the original paper for the references in the quote.

⁴⁷ Here I am relying on a parallel between agency and linguistic communication, where the notion of *uptake* and *reciprocity* on the part of an audience are considered a condition for the successful performance of an illocutionary act by a speaker. See Austin (1962/1975), Sbisà (2009), Hornsby (1995), Kukla (2014). Cf. *infra*, sections 4.6.2 and 5.1.3, n. 76.

subject to distortion—the meaning of her actions is thus defined away from her. In the finding/looting case, it is a feature of the agent’s identity as a member of a racialized group that triggers a criminalized labelling of the action, thereby revealing the structural character of racism as a form of discrimination and oppression—that is, embedded, interiorized, reproduced, reinforced, and even legitimized, within social structures.

3.6 Relevance of the Phenomenon of Social Authoring for the Present Work

To be sure, the process of social authoring of an action is not inherently bad. If anything, it is key to understanding the role played by collective intentionality for both individual and shared/collective agency. However, the potential for, and persistence of, mismatches between the action as intended by the subject who performs it, and the action as attributed to her (or, as Bierria claims, authored away from her), raises the question of what the relation between the *reading* and the *authoring* of an action is, and calls for a closer examination of the distinction between the *exercise of agency*, *attributions/ascriptions of agency*, and their reciprocal connection.

How do these questions impact on the *Identification Thesis* that I have identified as central to Korsgaard’s account? How do individual and collective agency relate to each other? How does morality—or better, lack thereof, in terms of various forms of structural injustices pervading and structuring agents’ field of possibilities for action—connect to agency, after all? All of these issues are deserving of their own dedicated study in the future.

For present purposes, the most pressing question prompted by cases of socially displaced/distorted agency via social authoring is whether social uptake is necessary for agency. Are cases of displaced agency ‘failures’ of agency? Bierria is very straightforward in claiming that although the phenomenon of social authoring displaces some agents’ agency, it does not ‘disable’ nor ‘erase’ it: While individual agency is sanctioned/validated by social uptake, it does not depend on it. Characterizing the phenomenon in terms of a ‘failure’ of the oppressed subject’s agency, Bierria contends, is inaccurate and misleading:

Disenfranchised agents are ‘doing their part’ to reflect, reason, visualize, anticipate —*intend*— and meaningfully act according to their intention [...] The failure in these cases does not emerge from these actors’ agency, but lives in the distortion of these subjects’ intentions by others as well as the social and political conditions that legitimize those distortions. (Ivi, pp. 136-136)

Bierria's proposal is then to drop "a binary or scaled model of agency that gauges subjects as having more or less, abled or disabled, or successful or failed agency," and adopt instead a "heterogeneous framework" of agency, which acknowledges different *kinds* of agency. The kind of agency exercised by an agent—which, Bierria proposes, can be hegemonic, alien, or insurgent/resistant/subversive—depends on how factors pertaining to the social dimension "position her in relation to others when practicing intentional action in the social sphere" (ibid). Whether hegemonic, insurgent or otherwise, I will explore the hypothesis that the exercise of any kind of agency involves some sort of interpersonal uptake or recognition (be it that of a minority, down to just another person).

Without following Bierria's proposal of adopting a heterogeneous framework of agency, her analysis of the phenomenon of displaced agency is important in its own for the aims at issue. It helps clarify that human agency is an inherently *relational/interactive* phenomenon, and that the reciprocal position of the interacting agents is crucial. Overall, Bierria's analysis corroborates the conclusions drawn from the case of disorders of agency. That is, that *agential power* and *status* are *constitutively* sensitive to the position an agent occupies in the relevant network of societal and institutional normative relations. These relations provide the *architecture* (cf. Haslanger, 2015), or *infrastructure*—so to speak—within which agents come to be and operate. Individual agents, on their part, do, however, also play an active role in creating, maintaining/reinforcing, or challenging and reforming their infrastructure.

3.7 Conclusion

In this chapter I examined two alleged types of defective agency—*disorders of agency*, and *socially displaced agency*—, and presented evidence in favor of considering them as instances of genuine agency instead. This revealed Korsgaard's constitutive norms of agency as neither necessary nor sufficient to account for the alleged defects, which prompted a targeted criticism of her account, followed by the proposal of a few emendations.

My criticism addressed Korsgaard's assimilation of the constitutive normativity of agency to moral normativity, and the problematic consequences of the double nature of constitutive norms. The problem is that the assimilation results in the potential exclusion of defective agents from the category of agents altogether. My positive proposals for adjustment take up and strengthen elements of Korsgaard's own account, such as her relational account of responsibility—supported by the discussion of both the *Disorders* and the *Displacement* scenarios—, and her acknowledgement that the notions of agency and action come in degrees.

I argued in favor of i) a morally neutral notion of *agential responsibility*, which features a reference to the agent's capacities exercised and developed (to different degrees) through participation in interpersonal practices of accountability (without blame) for one's actions (the *Agential Responsibility Claim*), and ii) a relational and context-sensitive definition of agency, which takes agency to be constitutively a function of one's location within interpersonal and social relationships, as well as of the quality of these relationships as perceived by the agent (the *Interaction Claim*).

In the next chapter I undertake a clarification of the metaphysics of agency and of the notion of constitutive normativity, which sets the stage for the specification of my positive proposal in chapter five. It is also required to make sense of Korsgaard's characterization of the notion of constitutive norms and the core of her theory—the *Identification Thesis* identified above (section 3.3)—in light of the conclusions reached through the discussion of the *Disorders* and *Displacement* scenarios.

Chapter Four – The Metaphysics of Normativity and Agency

In this chapter I provide a conceptual elucidation of the notion of constitutive norms/normativity, which is central to constitutivism as a first-order theory of agency. I present three possible interpretations (which are not meant to be exhaustive) of Korsgaard's characterization of the notion of constitutive norms. I endorse one of them—the Practical Interpretation—as the best way to make sense of the notion while accounting for some of the conclusions drawn in the previous chapter. For Korsgaard, the constitutive norms governing agency amount to the norms governing practical reasoning and morality. In this chapter I argue that the constitutive normativity pertaining to human agency is not 'homogeneous' in kind. In other words, the point of this chapter is to show that there are more norms governing agency than just the norms for practical reasoning—and, in particular, that agency necessarily involves social norms distinct from any notion of practical reasoning.

I analyze the core of Korsgaard's constitutivist theory, exemplified by the *Identification Thesis* introduced in the previous chapter, as consisting of two parts, and take up the metaphysical part of the claim (about functional normativity), and the practical one (on attributability) in turn. By drawing on a theoretical framework on the metaphysics of normativity developed within the philosophy of biology, I distinguish two varieties of constitutive normativity (*socially-generated* and *non-practice-based*), and provide a preliminary articulation of the socially-generated constitutive normativity inspired by the speech-act-theoretical account of the performative dimension of language.

Taking this detour on the metaphysics of normativity will clear the way for the proposal of a reframing of agency in the next chapter, which expands on the practice-based aspects of Korsgaard's constitutivist theory and positively takes into account the methodological constraints set by the analysis of the alleged cases of defective actions conducted in chapter three.

4.1 Disambiguating 'Normativity' in Korsgaard's Constitutivist Argument

An intended strength of Korsgaard's constitutivist project is its aiming at bringing together *metaphysical* as well as *psychological* aspects of normativity. Korsgaard's use of the terms 'normative' and 'normativity,' however, oscillates between different meanings. In order to clarify which sense of 'normativity' is at stake at which stage in her argument, let us start by

looking at what kind of role the specific stage or step plays in the overall constitutivist argument.

Korsgaard's constitutivist argument, as reconstructed in chapter one above, is the following:

- (1) Human beings must act
- (2) In order to act, human beings must unify themselves into agents
- (3) Human beings can only be unified into agents when they cohere with the principles of practical rationality
- (C) So, human beings as unified agents must act in accordance with the principles of practical rationality

Having set aside any intention to assess the meta-normative ambition of Korsgaard's project in its foundational aspect—that is, the vindication of normative authority of normative claims of practical reason—the aim of the present work is to clarify the relevant sense of normativity pertaining to the first-order level of the constitutivist theoretical nucleus advanced by Korsgaard. This constitutivist theoretical nucleus amounts to the *metaphysical-cum-normative* claim about action/agency having a constitutive aim/function, which sets normative principles/standards (premises 2 and 3 in the argument), along with the claim about the inescapability of acting (premise 1).

Now, the interpretation of the *Disorders* scenario endorsed in the previous chapter, if sound, shows that the *Moral Evaluation Claim* at play in Korsgaard's account is highly contentious.⁴⁸ I have claimed that a notion of *agential* responsibility as distinct from moral responsibility carries more predictive and explanatory power regarding the exercise of agency. To recall, Korsgaard claims that the constitutive principle of agency is the moral law.⁴⁹ If my conclusion that human beings can and do 'unify' themselves into agents *outside* the moral law is correct, then premise 3 of Korsgaard's constitutivist argument is false. The constitutive normativity proper to agency does not *per se* entail the normativity of morality.

⁴⁸ I.e., a moral failure does not imply a failure of agency; bad/defective actions do not transmit failure to the agent *qua* agent.

⁴⁹ Cf. SC (pp. xii-xiii): "the kind of unity that is necessary for action cannot be achieved without a commitment to morality. [...] the only way in which you can constitute yourself well is by governing yourself in accordance with universal principles which you can will as laws for every rational being. It follows that you can't maintain the integrity you need in order to be an agent with your own identity on any terms short of morality itself. [...] a commitment to the moral law is built right into the activity that, by virtue of being human, we are necessarily engaged in: the activity of making something of ourselves. The moral law is the law of self-constitution, and as such, it is a constitutive principle of human life itself."

It is important to stress once again that the step in the argument that carries the normative weight/import/implications is made in the characterization of what, metaphysically, agency is (roughly, premise 2). Having ruled out moral normativity, what does, then, *constitutive normativity*—i.e., the normativity of constitutive norms/principles governing the constitution of agents—consist in? What is its nature, and what is the kind of necessity attached to it? In what follows, I provide three possible interpretations of Korsgaard’s characterization of constitutive norms of agency in order to assess the plausibility of those interpretations in light of the analysis of cases of supposed defective agency conducted in the previous chapter.

4.2 Elucidating Constitutive Normativity

The constitutivist theoretical nucleus of Korsgaard’s account—as exemplified by premise 2: “In order to act, human beings must unify themselves into agents”—amounts to the following two claims. On the one hand, there is the claim that (i) agency/action is a functional kind, or that the function of action is to constitute the one who exercises it/performs it into a unified agent, and denotes the special (metaphysical) relation between the agent and her actions: a relation of self-constitution. In other words, there is no (unified) agent preexisting her actions. On the other hand, we have the claim that (ii) the essential feature of an action is whole-person attributability, or its being authored; where being authored/whole-person attributability refers to the (practical) relation between an agent and her actions as mediated by the notion of personal, or practical identity, which denotes the agent’s engagement in practical reasoning and deliberation.

According to what I have identified as the *Identification Thesis*, someone’s actions, Korsgaard claims, constitute her most essential part, and are the expression/manifestation of *her* practical identity, or capacity for normative self-government.⁵⁰ With this rough recap of Korsgaard’s constitutivist theory of agency/action in place, let us turn to the task of clarifying the notion of constitutive normativity operating within it.

As we saw in chapter two, Korsgaard puts forwards an elaborated account of the psychology and metaphysics of agency, revolving around the notion of constitutive principles, or norms. Let start, then, with a look into what constitutive norms are. According to Korsgaard’s most

⁵⁰ Cf. *supra*, sections 3.3, 2.2.

recent characterization of constitutive norms, they “arise from the nature of the object.” The idea, she specifies, breaks up in two parts:

Constitution Requirement (C): “unless the object conforms to the standard, it ceases to be the kind of object that it is.”

Self-constitution Requirement (SC): “the object *makes itself* into the kind of object that it is *by conforming* to the standard.” (AKC, section 1.2)

This characterization of constitutive norms is meant to capture what for Korsgaard is a crucial feature of constitutive normativity—*internality*—according to which an entity’s liability to standards is part of its constitution.⁵¹ In other words, the source of constitutive norms of agency lies in the nature of agency itself. The characterization is also meant to encode the double nature of constitutive norms/principles—i.e., the fact that the norms have both descriptive and normative implications with regard to the activity they are norms of. They *constitute* the very activity they govern, thereby describing the conditions for the possibility of agency itself; and they are *in force*, thereby exerting a normative role, for the individual, under pain of kind-membership exclusion.⁵² This double nature is the hallmark of constitutive normativity.

At this point, having exposed the supposed cases of defective agency as instances of genuine agency in chapter three, we are left with the task of determining if Korsgaard’s characterization of the constitutive norms of agency can be interpreted in a way that accounts for the conclusions reached at the end of those cases. If there is an interpretation, then we merely need to take it up, along with the suggested changes to the normative theory of agency suggested in chapter three, to be able to include those with disorders of agency and those whose agency is subject to displacement as agents proper. If not, we will have to suggest alterations to her more foundational characterization of the constitutive norms of agency to cover the appropriate anomalies. In what follows I present three possible interpretations of Korsgaard’s characterization of constitutive norms reconstructed above.

⁵¹ Cf. NCA (section 1.3): “Agency is liable to a standard of success and failure from the inside, that is, by virtue of its own nature: such liability is part of its constitution.”

⁵² The normative character of constitutive norms has been targeted by the ‘Why be an agent?’ objection, exemplified by Enoch (2006, 2011). Undertaking a reconstruction and assessment of the objection is beyond the scope of the present work, but represents an important issue to account for given constitutivist strategies’ claims about the inescapability and non-optionality of agency. For a defense of constitutivism against the objection see Ferrero (2009b).

4.2.1 The Metaphysical Interpretation

According to what I will call the Metaphysical Interpretation, the constitutive principles of agency are metaphysical principles *describing* the formal principle of agency—that is, self-determined efficacy. As a result, kind-membership (i.e., whether someone is an agent and genuinely exercises agency) depends on a certain relation between her and her behavior being in place—that is, by the fact that actions are those pieces of behavior instantiating a certain metaphysical property (autonomy/constitutional unity) that reflects on the agent. The kind of necessity associated with constitutive norms, under this interpretation, would amount to *metaphysical necessity*.

This interpretation is supported by Korsgaard's deliberate pursuit, within her constitutivist project, of a characterization of the metaphysics of normativity, which results in several distinct claims. These are: her endorsement of the 'classical view' that normative principles are principles of the unification of the manifold; the claim about the interdependence of the metaphysical and the normative properties essential to the nature of action (SC, sections 2.1.1, 8.1.3); and, finally, her statement that the appropriate metaphysics for the concept of activity is a Platonic metaphysics. Besides inheriting many potentially problematic issues associated with each of the individual claims supporting it (which I am going to overlook, here), the major problem with the Metaphysical Interpretation is that it is inherently unstable, in that it tends to collapse on the Individual Psychology Interpretation (presented below).

Korsgaard takes agency's metaphysical property (autonomy/constitutional unity) to ultimately amount to the agent's *psychic unity*, issuing from the process of self-reflection, and resulting in her maxims/principles of choice displaying a certain normative property (universalizability/justice) (ivi, sections 8.1.3, 10.2.3). The way the formal principle of agency is given specific content, Korsgaard claims, is through the contingent practical identities, which—by providing the agent with reasons for acting—determine a functioning self. However, as mentioned, on this account the metaphysical necessity associated with constitutive norms, would collapse into a sort of *psychological necessity*. Let us turn, then, to the Individual Psychology Interpretation.

4.2.2 The Individual Psychology Interpretation

According to the Individual Psychology Interpretation, the constitutive norms of agency *describe* the psychological structure that an individual agent must have to be able to act at all,

or to author an action. This was, in a nutshell, the idea behind Korsgaard's version of the authorship view.⁵³ As we have seen, according to Korsgaard, exercising agency amounts to being engaged in a self-guiding activity—that is, an essentially conscious activity—even though the explicit awareness that one is engaged in such an activity does not necessarily have to be available to the agent. And yet, the compliance to agency's constitutive norms is crucially mediated by the agent's own conception of herself as an agent. This is where Korsgaard's notion of practical/personal identity, and the normative role associated with it—i.e., the fact that it provides reasons for acting—come into play.

If we interpret the constitutive norms of agency along the Individual Psychology Interpretation, however, we have to face the following predicament—compliance (to any degree, and at any level of conscious awareness) with the constitutive norms of agency would be not necessary for an individual to be engaged in the activity of self-constitution because less than fully unified selves are always already agents. Compliance with the norms of agency under the Individual Psychology Interpretation, would also be not sufficient for successfully exercising agency, since perfectly unified selves without appropriate external/contextual/societal backup are not able to act as they intend.

To recall, the conclusion reached in discussing the *Disorders* scenario was that what proves crucial to the exercise of someone's agency is her being held in interpersonal relationships, being treated as legitimate terms in relationships of accountability independently of the psychological quality (i.e., the realization of higher order self-reflection capacities, ensuing in a more unified self) their actions might display. If anything, it is through inclusion and participation in practices of accountability that the relevant psychological organization can be implemented. As to the *Displacement* case, the conclusion reached was that even when agents successfully intend and meaningfully act according to their intentions, the normative relations between them and their environment can be such that the characteristic relation of authorship they entertain to their actions is systematically displaced/distorted, and yet, their agency is neither disabled nor erased.

What these scenarios show is that an interpretation of constitutive norms of agency focused on the Individual Psychology is therefore partial, and urges us to look beyond the individual agent's psychological arrangement. In particular, both cases of alleged defective agency point

⁵³ This is also David Velleman's view. For a characterization of the authorship view, see Millgram (2005/2012, section 2).

to the necessity to look at the individual's interpersonal and social relationships. Let us, then, consider a final possible interpretation of constitutive norms—the Practical Interpretation.

4.2.3 The Practical Interpretation

According to the Practical Interpretation, the constitutive norms of agency *describe* the practice of agency. Exercising agency means to participate into a practice—a rule-/norm-governed activity—whose dynamics/functioning can be described through the constitutive norms.

The Practical Interpretation of constitutive normativity is textually well supported, as well. Korsgaard's takes herself to be engaged in providing a *practical* account of responsibility.⁵⁴ She follows the Strawsonian strategy of accounting for the notion of responsibility in terms of an agent's being in a network of interpersonal relations, *qua* member of a certain community.⁵⁵ This practical account was explicitly propounded in opposition to metaphysical approaches to responsibility (Korsgaard, 1992/1996b), and, plausibly, to merely psychological ones. It is a *practical* account because it refers to the normative practice of holding each other responsible for our *actions*, characterized by interpersonal relationships, attitudes, and expectations towards others *qua* persons. Participating in such a practice amounts to inhabiting the perspective of a practically deliberating agent and entering practical relations of reciprocity with other agents—i.e., treating them as responsible persons and committing oneself to be so treated—through the exchange of reasons which makes joint deliberation and commitment possible. Inasmuch as one first-personally engages in and evaluates practical reasoning and deliberation, constitutive norms set standards of evaluation and specify what is permissible, forbidden or prescribed for being acknowledged and being held in the practice, thereby playing a markedly *normative* function.

By focusing on agency *qua* interpersonal activity/practice, the Practical Interpretation seems well suited to positively account for the conclusion drawn from *Disorders* and *Displacement* cases. In particular, it fits in with the relational definition of what it is to be an agent exemplified

⁵⁴ Although Korsgaard's and other Strawsonian accounts concern 'moral' responsibility, I take their 'practical route' to defining the notion to coherently hold also for what I have labelled as 'agential responsibility.'

⁵⁵ As anticipated (cf. *supra*, sections 2.4 n. 24, and 2.5 n. 25), I see the metaphysical turn in Korsgaard's latest work to be a departure from the practical character of the constructivist enterprise she takes herself to be engaged in. For an interpretation of the evolution of Korsgaard along this trajectory—i.e., seeing her latest account as expressing a metaphysical, rather than practical, route to agential notions—see Zheng (2016, p. 65). Zheng's focus is on moral responsibility.

by my *Agential Responsibility Claim*. The issue remains open, however, about the provenance (biological, psychological, cognitive, rational, conventional, social, moral, ...) of constitutive norms of agency, and therefore about whether constitutive norms of agency are ‘homogeneous’ in kind. In light of the discussion of the cases of alleged defective agency, there are *prima facie* reasons for hypothesizing that they are not. Notably, the normativity pertaining to agency is not exhausted by a single kind of normativity (i.e., rational normativity) and features instead a fair amount of conventionality and *practice-based* norms, along with possibly being subject to other sorts of (*non-practice-based*) normative constraints—including milder, *descriptive* forms of normativity.

In other words, even if principles of practical rationality are constitutive—to varying degrees—of human agency, other conditions might equally qualify as ‘constitutive’ of (internal/essential to) intentional human agency, as the following sections will illustrate. If this is the case, the indigenous variety of constitutive norms/normativity, along with the consolidated philosophical tendency to consider different kinds of normativity as irreducible to one another,⁵⁶ calls for further investigation in the notion: Are different kinds of norms all equally constitutive of agency? Are they constitutive of the same *aspects* of agency? The coherence of a constitutivist account of agency hinges on disentangling the different varieties of normativity associated with the notion of constitutive norms.

4.3 On the Metaphysics of Agency and Normativity

In the above section, I considered three possible interpretations of Korsgaard’s characterization of constitutive norms. I did this because clarifying what kind of norms constitutive norms are, or what kind of normativity constitutive normativity is, represents a crucial step for assessing the viability of Korsgaard’s constitutivist theory of agency. The result of the comparison revealed that the Practical Interpretation is the most promising to positively account for the alleged anomalous cases of *Disorders* and *Displacement* discussed in the last chapter, while at the same time being consistent with the spirit of Korsgaard’s overall philosophical project, conceived of in terms of a practical enterprise (CA, pp. 23, 321-326), as well as preserving the

⁵⁶ I am assuming that natural facts display a form of *descriptive normativity*. If this assumption holds, *loci classici* of the irreducibility claim would then be David Hume’s ‘is/ought’ problem and G. E. Moore’s ‘open question’ argument. More on ‘natural normativity’ framework *infra*, section 4.4 ff. For a basic distinction of varieties of normativity between *soft* and *robust* kinds irreducible to one another, see, for example, Zaiber & Smith (2007).

central features of the Individual Psychology Interpretation. In the remainder of this chapter I will undertake a conceptual clarification necessary to take positive steps towards a reframing of Korsgaard's constitutivist theory of agency and action which fully embraces and elaborates on the Practical Interpretation.

4.3.1 Scope and Methodology

First of all, I shall tackle some methodological issues to make clear how I operate in providing such a clarification, following Amie Thomasson's distinction among three different kinds of questions asked in contemporary metaphysics. First, Thomasson identifies *existence questions*: questions about what does and does not exist. Answering such questions produces an ontology. Existence questions might be, in Carnap's sense, *internal*—i.e., questions “asked within (or using) a linguistic framework,” whose “answers may be ‘found either by purely logical methods or by empirical methods, depending upon whether the framework is a logical or a factual one’,”—and *external*—“questions about the existence or reality of *the system of entities as a whole*” (Thomasson, 2012, section 1.1).⁵⁷ Second, there are *relational questions*, targeting how many ‘levels of reality’ are there and in what kind of relation (of reduction, supervenience, truthmaking, grounding) they stand to each other. Finally, *modal questions* ask what “properties an object *must* have to be of a certain type, [...] about *what it would take* for something of a given sort (or for an individual) to exist: about the existence *conditions* for things of various kinds” (ivi, preamble).

Taking up Korsgaard's project of providing a constitutivist theory of human agency amounts to answering the modal question ‘What would it take for someone to be an agent?’ restricted to this/the actual world (or nearby possible ones). In so far as the question asks for identity conditions, it is also an existence question, specifically, of the internal kind. The question is asked within the linguistic and conceptual framework of the ‘manifest image’ of human agency—i.e., the conceptual framework in terms of which human beings conceive of themselves *as* agents (cf. supra, section 1.1.2). As to the relational questions, the constitutivist theory of agency I am after is non-reductive: both the psychological (in a broad sense, to encompass cognitive, conative and volitional elements) and the normative *dimensions* of agency are equally recognized as parts of reality. While no attempt is made to spell out their reciprocal metaphysical relation and argue for it, I endorse a commitment to a naturalistically

⁵⁷ See the original paper for the references to Carnap in the quote.

respectable framework, broadly conceived.⁵⁸ Inasmuch as I am concerned with the metaphysics of agency and normativity in the remainder of this chapter, and in my distinction of different *dimensions* of agency in the next chapter, these distinctions should be borne in mind.

4.3.2 The Metaphysics of Agency

An adequate clarification of the metaphysics of constitutive normativity requires us to take a step back to Korsgaard's characterization of the metaphysics of agency—that is, to the claims that (i) the function of agency/action is self-constitution, and that (ii) the essential feature of action is whole-person attributability/being authored. These two claims can be considered as two sides of the same coin (i.e., the *Identification Thesis*; cf. supra, sections 3.3 n. 36, and 4.2).

The claim that the essential feature of action is whole-person attributability/being authored is, at least *prima facie*, easily relatable to the Practical Interpretation of constitutive normativity sketched above. On one reading, if the essential feature of an action is its attributability, then the very concept of action and the normative standards associated with it inherently/necessarily involve interaction with other agents and the existence of a practice. Beyond the minimal (largely sub-personal and pre-conceptual) sense of agency and self-efficacy associated with our interaction with the physical environment, the existence of some other—even if just mediated by the use of language or other cultural tools—represents the background of meaning against which we can recognize ourselves as agents on the richer conceptual level we are concerned with here.⁵⁹ The constitutive normativity at stake in this case is distinctively social. Spelling out the metaphysics of agency in terms of social practices places us in the irreducibly normative province of social ontology. Recasting the *Identification Thesis* through the lenses of the Practical Interpretation, I will refer to the second claim as its practical counterpart, or *IT_P*.

How to make sense of the claim that the function of agency is self-constitution in strictly metaphysical terms (i.e., the metaphysical counterpart of the *Identification Thesis*, or *IT_M*), without falling into the highly controversial terrain proper to any functional claim invoking

⁵⁸ By naturalistic framework I mean a non-reductive methodological naturalism—i.e., the idea that the phenomenon at issue is liable to an explanation in thoroughly naturalistic terms—and, in a broader sense, the idea that attention should be paid to empirical information about the agents' situatedness within social structures and practices—in other words, the *material circumstances* of agency.

⁵⁹ The claim that the existence of a practice is presupposed—or, in Rawls' (1955) words, “stage-setting”—for agency *qua* social activity is compatible with the possibility of cases of authoring or self-attribution of an action in isolation/when no one else is around.

teleological notions, is however, far from straightforward. This is especially true in relation to a commitment to naturalistic respectability, with which Korsgaard's account is admittedly supposed to be compatible.⁶⁰

The rationale behind the appeal to the notion of function is that it provides a principled way for identifying entities. What an entity is, amounts to its functional organization, and its functional organization is such that it engenders standards of success and failure internal/indigenous to the entity itself. Korsgaard relies on Aristotle's characterization of living things in terms of self-maintaining forms, whose existence depends on their being engaged in the very activity of self-maintenance and continuous reproduction of their specific functional organization (cf. *supra*, section 2.5.1).⁶¹

Are the normative standards engendered by self-constitution as the function of action the same as those involved in the *practice-based* constitutive normativity? A practice is a rule- or norm-governed activity in which certain behaviors count as fitting and others are out of place. In other words, a practice is characterized by normative standards whose violation/deviation from the norm warrants a negative evaluation, disqualification/exclusion, or sanction/penalty. As mentioned, this kind of normativity seems to be distinctively practical (norms for action) (See Glüer & Wikforss (2009/2015, section 1.2) and Rawls (1955) for a characterization of the notion of constitutive norms. Cf. *supra*, section 2.5.2 n. 28).

In contrast, given that the notion of proper function is appealed to in the attempt to pin down the identity conditions of an entity, a violation of those conditions would make it a defective/malfunctioning entity, or another kind of entity altogether (norms for being) (cf. *supra*, section 3.1). In this case, the functional organization of an entity sets the conditions for its existence and the standards for its evaluation, whose normative character is *non-practice-based*, yet seemingly constitutive of agency as well.

To explore how these different kinds of constitutive normativity interweave into the metaphysics of agency, I shall look into the Autonomous Systems Account of functional

⁶⁰ For an overview on the lively debate around teleological notions in biology, see Allen (2003). Korsgaard's declaration of compatibility between her defense of teleology and Darwinian approaches can be found in SC (section 2.3 "In Defense of Teleology," see in particular section 2.3.1). See Moreno & Mossio (2015b) for a historical reconstruction that traces back to Kant the use of the notion of teleology and purposiveness in the "holistic and circular organization of biological systems," or self-organization, and contextualizes the "Kantian-inspired organicist ideas" in the most recent research paradigms in biology.

⁶¹ For Korsgaard's arguments see SC (section 2.3) "In Defense of Teleology," and "Aristotle's Function Argument", CA (essay 4).

normativity, and use the taxonomy of the varieties of normativity it provides as a framework against which some aspects of the constitutive normativity of agency might be tentatively understood.

4.4 The Autonomous Systems Account

Within contemporary debates in the philosophy of biology, the Aristotelian idea of functional normativity has seen a resurgence. This is especially the case as it pertains to the organizational structure of living systems.⁶² One such account is the Autonomous Systems Account (ASA) of functional normativity.⁶³ The ASA looks at normativity as a pervasive phenomenon “inherent in the organization or form of living systems, specifically in the form that generates their unity and hence explains their existence” (Christensen, 2012, p. 104). Within this framework, the normativity proper to human agency and rationality are grounded in a more basic kind of normativity, shared with non-human forms of *autonomous systems* such as other animals and simple biological organisms.

A system is *autonomous* when it actively contributes to generate the conditions of its own unity and ongoing persistence, by building an *infrastructure*—a “persistent, relatively stable structure that shapes more dynamic system-maintaining processes, with the cell membrane of living cells being a paradigm example” (ivi, p. 106). The infrastructure supports and contributes to the system’s self-perpetuation, often by playing a regulative role in the organization of the system. The autonomy of a system is thus defined as relative to the system *as a whole*, and *in relation to* the relevant infrastructure (ibid). The resulting functional organization generates *normative constraints* to which the system as a whole is subject. As a consequence, these normative constraints represent, minimally, but fundamentally, the conditions that must hold in order for the system to exist.

An autonomous system’s ontology, in turn, is specified by the different levels at which the regulative processes and interactions contributing to the system’s existence take place. The ontology of complex autonomous systems (like rational agents) is specified in terms of “hierarchically structured forms of organization,”—the autonomous system, the cognitive

⁶² See Christensen (2012, p. 105 n. 1) for references to the current debate on functional normativity.

⁶³ Christensen (2012) and Christensen & Bickhard (2002); the present discussion relies on the former. See also the ‘organisational approach’ in the study of biological autonomy developed by Moreno & Mossio (2015a).

agent, the social individual, the person—“each of which impose normative constraints” (ivi, p. 110). The system’s ontology generates a *normative cascade* ranging from fundamental persistence norms, up to general agency and cognitive norms, skill domain norms, and individual-specific cognitive norms. For example, to be a competent cognitive agent, an agent needs working memory, reasoning, and higher order emotional regulation, along with other kinds of processes and activities and the associated variety of functional norms.

In light of the theoretical tools provided by the ASA, the case of human agency can be analyzed as follows. Human agents are complex autonomous systems made up of a network of interdependent processes (physical, physiological, cognitive, affective, conceptual, metacognitive, volitional, social, etc.). There are thus many levels at which human agents’ self-maintenance/functioning is carried out, and for each level there are specific relations of dependence on different kinds of processes and on the relevant infrastructure.

The consequence of such a view for the first-order constitutivist theory I am concerned with is that we are provided with a naturalistically oriented way of cashing out the notion of functional organization and its normative character as the source of the unity and existence conditions of a living system. Functional normativity in the sense specified by the ASA might therefore be a legitimate candidate for understanding the notion of constitutive normativity in the *non-practice-based* sense outlined above (i.e., constitutive norms as norms for being).

4.5 Varieties and Sources of Normativity

An important feature of the ASA is that it allows us to distinguish between different *normative perspectives*. Specifically, a case is made for conceptualizing normativity as extending beyond the realm of rational agency, and admitting of the normative perspective of the autonomous systems considered as whole, alongside the normative perspective of individual persons. Different kinds of normative evaluation (/normativity) may be thought of as having a parallel structure, insofar as the following elements are identifiable:

- a) a normative perspective (persons, autonomous systems) for which things matter,
- b) the nature/basis of mattering/significance (relevance to the person, relevance to autonomous systems), and

- c) the mechanism through which entities with normative perspective respond to the normative facts constituted by the respective relevance relations (rationality, regulation). (Ivi, pp. 108-9)⁶⁴

The ASA provides arguments in favor of a naturalist grounding of the normativity of personhood and rationality—that is, it aims at providing an explanation of normativity which is “consistent with the natural emergence of th[is] phenomen[on]” (Bickhard, 2002, p. 3).⁶⁵ It does so by situating it within the more basic kind of normativity pertaining to autonomous systems, addressing therefore the question about the origins of the normativity proper of rationality and personhood. The grounding relation between different kinds of normativity might be interpreted, however, not just in terms of ‘origins,’ but also of ‘constitution,’ according to which,

persons are not just descended from autonomous agents, they *are* autonomous agents: a person is constituted as a certain kind of autonomous agent in the base sense of autonomy, and this makes an important contribution to the normativity of personhood. (Christensen, 2012, p. 108)

The relevance of the ASA for the present purposes is neither in its metanormative implications, especially foundationalist ones, nor in the answers it provides to metaphysical questions of the relational kind described above (section 4.3.1). Rather, it helps us advance our understanding of what the constitutive normativity proper to agency amounts to.

Gathering treatments of normativity across diverse philosophical literature, the ASA distinguishes different *kinds* of normativity: on the one hand there is *descriptive normativity*, a minimal kind of normativity which refers to the possibility for a system to depart from the norm, but does not evaluate nonconformance as ‘bad,’ or ‘wrong.’⁶⁶

On the other hand, *evaluative normativity* involves a comparison between actual and alternative states, and comes in different degrees of strength/robustness, corresponding to:

⁶⁴ The analysis of the structural features of normative evaluation is based on distinctions drawn by Joseph Raz (1999). See the original paper for the references.

⁶⁵ Together with Christensen, Bickhard is a major advocate of a naturalistic approach to functional normativity, or “natural normativity framework.” In the philosophical tradition, see Philippa Foot’s (2001) neo-Aristotelian account of rational and ethical normativity as forms of *natural* normativity.

⁶⁶ Descriptive normativity would be the one pertaining to natural states of affairs. Cf. *supra*, section 4.2.3 n. 56.

- (1) *valuations* (axiology): a certain state being ‘good,’ or ‘better/worse than’ another;
- (2) *prescriptions* (deontic theory): stating how things ‘ought’ to be;
- (3) *constitutive norms*: ‘specify rules which must hold if something is to exist;’ ‘are per se non-evaluative, though they can inform evaluations in conjunction with other information, such as an agreement (perhaps tacit) to play by the rules’—usually involving ‘paradigmatic performance norms.’ (Ivi, p. 105)

The rich taxonomy of normativity delineated by the ASA, along with the generic structural features shared by different kinds of normative evaluation, provide us with a basic set of theoretical tools to solve some of the difficulties met in providing a coherent interpretation of Korsgaard’s attempt to bring together metaphysical and psychological aspects of normativity. After this detour on the metaphysics of normativity, I shall provide a provisional characterization of constitutive normativity and its heterogeneity.

4.6 Constitutive Normativity, regained

Agency has a dual nature, roughly paralleling the nature-nurture debate: it is in part due to our social practices, and in part to our biological, cognitive, and psychological constitution—our being autonomous agents in the terms specified by the ASA. The constitutive norms pertaining to agency can therefore be thought of as having a parallel dual nature: *socially generated*, or *practice-based* constitutive normativity, and *non-practice-based* constitutive normativity.

While a full exploration of the distinction between natural and social roots of (constitutive) normativity exceeds the possibility of adequate exploration here, it is worth pointing to some recent work by Katharina Nieswandt (2016) that might help clarify some of the points I am making. By providing a compelling interpretation of G. E. M. Anscombe’s writings, Nieswandt sets the stage for the development of a rich and coherent Anscombean metaethical framework, grounded on the distinction between two kinds of necessity. Specifically, she draws on Anscombe’s distinction between a “practice-internal or ‘conventional’ necessity” (i.e., “the necessity imposed by a rule, a right or a promise”), and a “practice-external or ‘Aristotelian’ necessity,” according to which human goods, or standards for human flourishing, are grounded in human nature.

Nieswandt’s elaboration of the Anscombean framework is interesting for present purposes because it is yet another approach aimed at elucidating how different kinds of norms—i.e., norms of distinct sources and with different kinds of necessity associated to them (biological,

practical, legal, ethical, etc.)—equally pertain/contribute to the *constitution* of distinctive kinds of entities such as human agents. In particular, it is the individuation of the specificity of the practice-internal, or conventional necessity, that is relevant here, as I arrive at similar conclusions while drawing from a different theoretical approach (section 4.6.2). In line with the Anscombean distinction between two kinds of necessity, my proposal is therefore meant to be noncommittal about metaethical and evaluatively normative issues. My aim, once again, is to provide an elucidation of the notion of constitutive normativity relevant for a first-order constitutivist theory of agency.

4.6.1 Non-Practice-Based Constitutive Normativity

How does the ASA relate to the constitutivist theory of agency I am concerned with here? Insofar as we consider agents as autonomous systems—which represent one possible normative perspective from which we can consider an agent’s ontology—the norms governing the different processes (physiological, psychological, cognitive) contributing to the system’s functional organization are constitutive in that they represent the unity and existence conditions for the system considered as a whole. This is a first, *non-practice-based* way of interpreting constitutive normativity, which amounts to functional normativity as specified by the ASA.

Since the characterization of an autonomous system constitutively incorporates a reference to the infrastructure utilized by the system for its self-maintenance, the wholeness of the system and its identity are possible just against the background of the infrastructure and circumstances sustaining it. Given the processual and dynamic character of a system’s functional organization, a human agent *qua* autonomous system, is an inherently dynamic entity, which actively contributes to its own constitution. Being an agent and exercising agency amounts to being engaged in processes and activities in interaction with a rich infrastructure. I am following here an important conceptual distinction operated by theorists of biological autonomy, who identify

two interrelated, and yet conceptually distinct, dimensions of biological autonomy: the *constitutive* one, which largely determines the identity of the system; and the *interactive* one, which, far from being a mere side effect of the constitutive dimension, deals with the inherent functional interactions that the organisms must maintain with the environment. These two dimensions are intimately related and equally necessary. [...] In particular, the emphasis on the interactive dimension implies [...] that autonomy should not be confused with independence: an autonomous system must interact with its environment in order to maintain its organisation. [...] This is what

grounds the agential dimension of autonomy. (Moreno & Mossio, 2015b, p. xxviii)

The interpretation of the IT_M that I am favoring here understands agency as an activity of self-constitution which is inherently situated and interactive, thereby providing a footing for the *Interaction Claim* advanced in chapter three. A consequence of this interpretation, then, would be that it is a matter of constitutive normativity in a *non-practice-based* sense—i.e., independently from any sort of necessity established *within* the practice itself, or *by* it—, that human beings socially organize themselves, and act in the context of practices they themselves create/establish. The kind of necessity associated with non-practice-based constitutive normativity would be something analogous to the necessity of natural-historical judgments, or ‘Aristotelian categoricals’ (or Aristotelian necessity, in Anscombe’s sense), of the form “The S is/has/does F,” and “S’s do F” (Thompson, 2008, Foot, 2001).

Human agency is thus *constitutively* social, in that social structures and practices are not just metaphysically (/logically) prior to individual agency, but also empirically and practically necessary to it, as they provide the material and interactive conditions for individual self-awareness and selfhood. There is in fact a consistent body of empirical evidence (coming from clinical, cognitive, developmental psychology, as well as neurophysiology) supporting the hypothesis of a close interdependence between human subjectivity, and the embodied and interpersonal dimensions of human cognition and meta-cognition, which informs rich strands of research in philosophy of mind and action.⁶⁷

4.6.2 Socially-Generated Constitutive Normativity

For human beings, kind-membership to agency has both natural and social roots. As autonomous systems and cognizers, human beings are born with agential capacities that are actualized and operated in the context of relationships (material, interpersonal, social), and have effects (physical, psychological, cognitive, normative) on their environment (natural and social), on themselves, and others. As stated in the *Agential Responsibility Claim*, the notion of an agent is inherently relational. Agents constitute themselves *qua* social individuals, within social practices and interactions.

⁶⁷ Among these there are phenomenological approaches to self-consciousness (see Gallagher & Zahavi (2005/2014) for an overview, Pacherie (2013) for the specific case of collective action), and for embodied and social cognition see Wilson & Foglia (2011/2015) and de Bruin et al. (2012), respectively. Cf. Meini (2007, ch 4).

Human agency as the activity of self-constitution (keeping Korsgaard’s central claim, in its practical counterpart *IT_P*) takes therefore the form of concrete, embodied, situated social practices—crucially, through conceptually mediated practices of self- and other-ascriptions,⁶⁸ which carry further, discursively and socially-generated normativity. Socially-generated normativity ranges from the norms governing acts of promising, commanding, consenting, pardoning, etc., to actions requiring more complex and structured social institutions, such as voting, marrying, and baptizing—all of which have *normative* states of affairs (consisting of commitments, rights, obligations, licenses, etc.) associated with their performance.

By *socially-generated constitutive normativity*, I mean the normative dimension associated with human practices, which in the next chapter (section 5.4) I will refer to as the *performative* dimension of agency. The performative dimension of agency is roughly modelled on the performative dimension of language as characterized in Speech Act Theory (SAT) developed within the philosophy of language (Austin, 1962/1975, Searle, 1969, Sbisà, 2002, 2007), and beyond, notably by feminist philosophers in a *social account of language use* (Hornsby, 1995, Kukla, 2014, to name just a few).

A ‘speech act’ (or ‘illocutionary act’) is a term of art that refers to the “performance of an act *in* saying something as opposed to the performance of an act *of* saying something” (Austin 1962/1975, pp. 99-100), thereby pointing out that language is used not just to describe how things are, but to *do* things:

The use of language is a sort of social *action*, consisting of the production and reception of utterances. People do things with words—this is action; and (though this may at first seem a strange way to put it) they do things to one another—this is social action: They tell one another things, or ask them things, or try to persuade them of things, or whatever. (Hornsby, 1995, pp. 129-130)

As such, the term ‘performative’ refers to the production of normative states of affairs—“that is, states consisting of the conjunction or disjunction of an agent with a modal predicate belonging to the deontic kind” (Sbisà, 2013, p. 32)—associated with the performance of speech acts *qua* social actions. By extension, and more broadly, with the term ‘performative’ I refer to

⁶⁸ See Thalos & Andreou (2009), for an evolutionary informed account of the distinctiveness of homo sapiens as a bonding species—i.e., capable of collective agency. Bonding, the authors claim, is a “universal of mammalian life.” They emphasize the hyper-sociality of our species to indicate how cooperation, rather than self-interest and individualism, should be understood as the default theoretical assumption.

the acts of doing or omitting to do something, generally identifiable by means of the specific changes/effects/outcomes in the agent's environment that their performance brings about.⁶⁹

The normativity proper to the performative dimension of agency is socially-generated and conventional in that it is associated with the production of *conventional effects*. These effects are “made possible by the social frame and brought about thanks to the kind of agreement between [interactors] about what is being done, which we may call uptake,” and the mark of their conventionality is their “defeasibility,” i.e., “the liability to being annulled in particular circumstances” (Sbisà, 2014, pp. 621-622).⁷⁰

This kind of socially-generated normativity is *constitutive* in the sense that the activities governed by constitutive norms do not exist apart from those very norms. In Korsgaard's *descriptive-cum-normative* sense of the term, the activities are at once defined, or constituted, and regulated by them (cf. *supra*, sections 4.3.2 n. 59, and 2.5.2). By illuminating their defeasible character—i.e., that their constitutive function “is only exercised against a background of intersubjective agreement” (Sbisà, forthcoming, section 1)—the speech-act-theoretical analysis endorsed here is a good candidate for understanding the *practice-based* (yet non-moral) counterpart of constitutive normativity that emerged from the Practical Interpretation advanced above (section 4.2.3).

In my view, the gist of a constitutivist theory of agency consists in an explanation of the complex organization of processes and practices by means of which human agents can recognize themselves as inherently *autonomous*, yet *social* entities. In other words, as individuals occupying the characteristic first-personal (singular and plural) agential perspective of deliberating and evaluating agents, who are autonomous in a distinct manner, richer than the biological sense of the term, and therefore closer to Korsgaard's use. Such a project is grounded in an elucidation of how different kinds of normativity hold together, and the present work is meant to produce a first step in that direction.

⁶⁹ For the distinction between *behavior* and *performance* (negative and positive) in an Austinian framework, see Ginocchietti (2016). Generally speaking, the ‘problem’ of action individuation necessarily points to broader issues of social ontology, which, however, deserve an investigation of their own.

⁷⁰ For an alternative interpretation of the conventionality of speech acts—in terms of conventionality of the (semantic) *means* by which they are performed, rather than their *effects*—see Searle (1969). The distinction is Sbisà's (2009).

4.7 Conclusion

In this chapter I have analyzed Korsgaard's characterization of constitutive norms by providing three possible ways to interpret it. I have endorsed the Practical Interpretation—based on Korsgaard's own practical account of responsibility—as the best candidate to account for, at least in principle, the conclusions drawn in chapter three. There I made a case for considering cases of alleged defective agents as genuinely exercising agency, and concluded that the constitutive normativity proper to agency does not *per se* entail the normativity of morality. Here, I advanced the claim that Korsgaard's criteria for agency, if we look at the phenomenon of agency in merely individual psychological terms, are neither necessary nor sufficient for capturing why those cases legitimately count as genuine instances of agency. As we will see in the next chapter (section 5.2.1), even highlighting the social aspects of Korsgaard's account will not be enough to overcome those objections. Taken together, this means that there is room for an understanding of constitutive normativity/norms of agency in practical, yet non-moral, terms.

While giving prominence to norms of practical reasoning, the Practical Interpretation also pointed out that constitutive normativity is not exhausted by them. Insofar as exercising agency means participating into social practices, there are other sorts of norms (notably, social norms such as those governing the social meaning of actions) which might qualify as constitutive of agency, just as norms of practical reasoning do.

Under the hypothesis of an indigenous variety of constitutive norms/normativity, I turned to what I identified as the core of Korsgaard's constitutivist theory—the *Identification Thesis*—to understand what notion of constitutive normativity it relies on. I first focused on the metaphysical counterpart of the *Identification Thesis*, or *IT_M* (as opposed to its practical counterpart, or *IT_P*)—that is, Korsgaard's appeal to the functional claim that the function of action is to constitute the one who exercises it/performs it into a unified agent. I introduced the Autonomous Systems Account (ASA) of functional normativity, and suggested that functional normativity—i.e., as the unity and existence conditions for a system's self-maintenance and development in connection with its infrastructure—be considered as a *non-practice-based* variety of constitutive normativity.

As to the *IT_P*, I spelled out a *socially-generated* or *practice-based* variety of constitutive normativity, modelled after the normativity proper to the performance of speech acts *qua* social actions as specified by Speech Act Theory (SAT). This preliminary specification of the performative dimension of agency is functional to some analytical distinctions on which I will

rely for my proposal in the next chapter, where I also provide a reframing of agency which positively accounts for the *Disorders* and *Displacement* scenarios and elaborates on the hypothesis of the social constitution of agency prompted by their analysis.

Chapter Five – Reframing Agency

The moral drawn from the discussion of the *Disorders* and *Displacement* cases is that a conceptualization of human agency that focuses on the individual psychology of agents allows us to capture just one aspect of the phenomenon. The central thesis advanced in this work is that *agency* should be conceived of as a complex and interactive capacity, and that the exercise of this capacity, at any degree and stage of actualization, crucially depends on interpersonal relationships and contextual factors. As such, an account of agency limited to individual psychology is incomplete. This thesis found an early expression in chapter three with my proposal to complement Korsgaard's theory with the *Interaction Claim*, which helps to make sense of the intuitions and reasons provided in favor of considering those scenarios as cases in which agency is, to different extents and despite adverse contextual circumstances, exercised.

In this chapter I clarify the relationship between the practical counterpart of the *Identification Thesis* (*IT_P*), the *Agential Responsibility Claim*, and the *Interaction Claim* introduced in chapter three. My aim is to advance an account of the constitution of agency that improves upon Korsgaard's own authorship view of agency and action. I do so by combining G. E. M. Anscombe's characterization of intentional action with the theoretical notion of *selective intersubjective recognition*, in order to focus on the role played by the practices of holding people responsible (and the distortions thereof), and those of identity-ascriptions, in the exercise of agency. Although the conceptual distinctions operated and the theoretical tools adopted here represent a departure from Korsgaard's framework, they have the potential to elucidate how a strengthening of her *practical* (vs. metaphysical) approach to responsibility and agency might be further developed within a first-order constitutivist theory of agency inspired by her own account; or so I will argue. Accordingly, I will conclude by presenting the central components of my reframing and expansion of Korsgaard's constitutivist theory of agency and authorship view.

5.1 Central Components of My Reframing

With the preliminary clarification of the metaphysics of agency and normativity conducted in the previous chapter in place, I shall now proceed to articulate the building blocks for the reframing of the first-order level constitutivist theory of agency that I am pursuing. First of all, it should be noted that my reframing fits in with the standard conception of agency and action

introduced in chapter one. There, I provided a rough characterization of agency as the capacity exercised/manifested by agents when they *do* something. The sense of *doing something* relevant to capturing the specificity of human agency is customarily contrasted with happenings in the world, and the outward performance of merely (non-actional) bodily movements of an agent (e.g. reflexes, spasms). According to the standard conception of action and agency in the philosophy of action,⁷¹ someone is an agent if she manifests the capacity to act intentionally and for reasons. While fitting in with the standard conception outlined in chapter one, some important differences are worth highlighting.

5.1.1 The Necessity of a Scalar Approach

Korsgaard's construal of the connection between intentional and rational agency is quite strong. Her Kantian account of intentional action takes rational action to be "the only genuine kind of intentional action" (Arruda, 2016a, section 1.1; SC, sections 1.4.8, 3.1.1, 9.7.6). While acknowledging that the notion of agency comes in degrees,⁷² her intellectualist account of intentional action, combined with an interpretation of the *Identification Thesis* along metaphysical lines (cf. *supra*, sections 3.1, 3.3-3.5, 4.3), engendered the problem of defective actions. For Korsgaard, defective actions fail to fully unify their performer; they reflect negatively on, or transmit failure to, the agent's status itself by making her 'ineffective,' and 'less of an agent.' The core of her account is that agency requires unification, which in turn demands quite a lot from the individual agent's psychology in terms of the development of higher-order capacities for rational self-reflection and practical deliberation. In sum, Korsgaard's account sets the standards for *genuine* agency quite high in terms of agential capacities.⁷³

Despite the high standards Korsgaard sets, her inclusion of a scalar approach is necessary lest incorrect lines be drawn demarcating agents and non-agents. Incorrectly identifying someone as a non-agent has obvious moral implications. As such, leaving flexibility in who may be counted among the set of agents by allowing agents to be more or less genuine/full

⁷¹ As we saw, the two claims around which the standard conception of agency and action revolves around are that (1) the notion of action is to be explained in terms of intentional action, and that (2) there is a close relation between acting intentionally and acting for a reason. Cf. *supra*, section 1.1.1.

⁷² SC (section 8.5.1): "the extent to which one is unified, and so is an agent, is a matter of degree."

⁷³ See Arruda (ms.) for a critique of the widespread tendency among theories of agency of different orientations to construe their explananda in ideal terms.

agents, is necessary. Generally, I use the term ‘scalar’ to point to the fluidity between the notions of absent and ideal agency, without intending an ability to derive any precise measurement between the two.

One of the tasks for a full elaboration of such approach might be to provide indications for the individuation of certain threshold levels of agential capacities (e.g., absent, minimal, full, ideal, etc.). Although undertaking this task is beyond the scope of the present work, the reference in the *Agential Responsibility Claim* I introduced in chapter three (cf. supra, section 3.4),⁷⁴ to the agential capacities in terms of graded notions, is meant to start responding to the necessity of a scalar approach.

In light of the argument built up throughout this work—i.e., that being held in interpersonal relationships and participating in practices of accountability for actions is constitutive of what makes an agent an agent, at *any* degree of development and actualization of an agent’s agential capacities—I will argue that the scalar notion of agency extends well beyond the psychic unification of an agent and into other considerations such as recognition and intention (infra, section 5.5).

In my reframed account, the possession, *to some degree*, of these agential capacities and the exercise of such form of control in the appropriate relational contexts will be considered as necessary conditions for agency, but no commitment is made to the realization of such capacities to the highest degree. This marks an important shift of focus in the development of a scalar approach, in that regardless of the degree at which the agential capacities (and possibly agential unification) might be realized, the distinctive feature of intentional actions that such capacities point to is their liability to a characteristic form of *agential control*, such that the agent can legitimately be *intersubjectively recognized* as their *author*.

5.1.2 Relationality

The *Disorders* scenario showed that participation in interpersonal relationships—i.e., being recognized and treated as a participant into practices of accountability—is a necessary condition for an agent to successfully exercise their agential capacities, and plausibly to even develop them in the first place. This consideration lead to the second and most prominent component of my reframing: relationality. I phrased the *Agential Responsibility Claim* to give

⁷⁴ To recall, I proposed to turn what I identified as the *Moral Evaluation Claim* at play in Korsgaard’s characterization of agency and responsibility, into the *Agential Responsibility Claim*.

special prominence to the relational aspect of agency. While the possession of the relevant agential capacities figures as a presumption informing the practice of responsibility ascription, the focal point of the claim is the relational definition of what it is to be an agent. To be an agent, so the claim goes, is to be a participant in a network of interpersonal relationships and practices of accountability for one's actions.

In Korsgaard's account, the dialogical character of normative reasons—i.e., that a properly normative reason must be shareable and essentially public⁷⁵—and the relational notion of responsibility as answerability are the elements that are meant to ensure/secure the social dimension of the activity of self-constitution. For Korsgaard, relationality and the social dimension of agency enter the picture with the relations of reciprocity and mutual accountability that the already extant agent enters in when exchanging reasons and engaging in practical deliberation (in isolation and with others). For me the practice itself is a key element for understanding agency in the first place. An agent is constituted *with* others. In other words, on my account agency itself is constituted in the shared space where the authoring of intentional actions (more or less reasoned) gets recognition by the relevant participants in the practices of exchanging reasons and accountability for actions.

5.1.3 Context-Sensitivity and Situatedness

According to Korsgaard, *attributability* is the distinctive feature of human actions, and the authorship view she proposes is meant to account precisely for this feature. In a nutshell, a certain kind of psychological *constitution*—which, in Korsgaard's view, is the result of the agent's successful engagement in practical reasoning—must be in place in order for her actions to be *authored*. The relation of authorship between an agent and her intentional actions emerging from Korsgaard's account has it that actions are a self-determined and efficacious expression of self. In this way, actions are *attributable* to the agents who have *authored* them, and authorship depends on the psychological structure of the agent allowing for morally and rationally coherent self-unification (cf. Arruda, 2016a, section 1.1).

⁷⁵ Note that Korsgaard's is a Kantian conception of the normativity of reasons, which she defends also by appealing to Wittgenstein's 'private language argument' for the essential publicity of meaning. Even though my concern in this work is not a discussion of the sources of the normativity of practical reasons, it should be noted that the Kantian is just but one position on the issue, which is also related to the question regarding the variety of reasons for action, beyond normative ones. For an overview on the distinction between *normative*, *motivating*, and *explanatory* reasons see Alvarez (2016). For a discussion of the shareability issue and a defense of Korsgaard's position see van Willigenburg (2002).

However, as the discussion of the *Displacement* case has shown, successfully carrying out the process of self-constitution at the reflective level (the one relevant for Korsgaard's characterization of intentional action), is not sufficient for bringing about the action as intended by the agent. We need to look 'outside' the agent, into the features of the actual context in which the action is performed.

According to Bierria's explanation of the *Displacement* cases, it is not just that certain actions get *attributed* to an agent, independently of, or despite, the agent's actual intentions. In the situations described by Bierria, the *social reading* of the actions performed by disenfranchised agents is such that it systematically trumps the agents' ability to act as they intend in the social sphere. In conditions of oppression and discrimination the characteristic first-personal relation of *authoring* between an agent and her intentional actions gets hijacked, resulting in a distortion of the agent's intentions, and the process of authoring of the actions is displaced from the agent actually *performing* them.

The conclusion I draw from Bierria's characterization and discussion of the phenomenon of the 'social authoring' of an action, is that it points to the inherently context-sensitive and situated character of agency. The authoring process of an action is dependent on the features and specific dynamics of the context in which the agential capacities are exercised, as well as to the agent's position/status with respect to the broader social environment in which she acts. Social categories and structures (including systems of oppression and discrimination) are also *internalized*, to the effect that the social perspectives on us as particular historically and socially situated agents are also acted on, producing the distinctively psychological harms of negatively affecting the agents' own sense of self-efficacy and motivation.

Psychological research on internal motivation (Ryan & Deci, 2000) points to three factors involved in the development of an agent's integrated sense of self, volition, and initiative, as well as her wellbeing experience and the quality of her performance. These are the individual's experience of autonomy, competence, and relatedness, corresponding to the psychological needs of feeling agentic, effective, and connected to others.

Put concisely, agency is dependent on *intersubjective recognition*, which might range from being thoroughly affirmative to deeply distorting. I should emphasize that in my reframing the notion of recognition⁷⁶ is not a theoretical tool to help extrapolate the individual's intention in

⁷⁶ The notion of recognition is central to debates in political and moral philosophy. My use of the notion here is independent from these debates, and is broadly inspired by the notion of communicative uptake and reciprocity developed in pragmatics, within Speech Act Theory (SAT) (notably by works in the Austinian

acting, but an empirical necessity in the very process of constitution of our agency. In addition to the *Agential Responsibility Claim*, I therefore suggested that the reframing of the constitutivist theory of agency include the *Interaction Claim*, according to which agency is *constitutively* a function of the agent's location within a network of social and interpersonal relationships, as well as of the quality of those relationships, as perceived by the agent herself.

However, in agreement with Bierria's interpretation of the *Displacement* case, in chapter three I concluded that even if the inherent context-sensitivity and situatedness of agency makes it liable to the phenomenon of social authoring, there is no denial of agency. The intersubjective recognition necessary for genuine agency seems therefore to be *selective*. In other words, you are not an agent because everyone recognizes you as such, you are an agent because at least some people, in the relevant interpersonal relationships and groups, recognize you as such. The *context-sensitive* and *situated* character of agency, then, provides the theoretical resources for accounting for the selectiveness hypothesis I advance here.

Postulating the dependence of agency on *selective intersubjective recognition* is compatible with Bierria's proposal to adopt a heterogeneous framework of agency, which allows us to distinguish between different *kinds* of agency, such as transformative, alien, and insurgent:

“Transformative agency” might cover action intended to fundamentally overturn conditions of systematic oppression, especially (but not exclusively) through collective action, such as through community organizing, movement building, or political advocacy. [...]

Alien agency facilitates action that intentionally creates meaning apart from dominant structures of oppression and the people who endorse them. [...] Alien agentic action does not seek to transform systemic conditions of oppression, but is resistant in that it facilitates action that is preoccupied with cultivating its own universe of meaning and practice that affirms that which is unvalued—in this case, black life—and is therefore ambivalent about or even encourages its illegibility within the dominant public sphere. [...]

Insurgent agency is employed by subjects who intentionally act in unstable and precarious circumstances that are difficult to escape or alter, and who craft provisional and makeshift practices of opposition that subvert, but still

tradition, e.g. Sbisà (2007, 2009), especially by Hornsby (1995), who highlights the role this notion plays in the pragmatic dynamics of silencing and other forms of speech disempowering and conversational injustice (see also Kukla (2014), Ayala & Vasilyeva (2016). Arruda (2016b) develops an account of what mutual recognition consists in from the point of view of the philosophy of action.

remain defined by, conditions of power. [...] This kind of agency, this hustle, is not usually celebrated as an idealized model of liberatory action, but it is a kind of intentional action that requires imagination and strategic thinking, and that is accessible for use in circumstances in which one is isolated and has few options. (Bierria, 2014, pp. 139-141)

The necessity of recognizing different *kinds* of agency comes, in my view, as a corollary of the *Interaction Claim*. By considering context-sensitivity and situatedness as constitutive features of agency we can explain why denial of agency by ‘non-affiliated others’ usually (and luckily) misfires,⁷⁷ and make sense of the variety of ways in which agents can and do act and organize despite adverse social, historical, and political conditions, and eventually subvert them.

At the same time, by looking at agents in broader contexts and conditions in which they operate, an approach centered on the *social constitution* of agency like the one I am proposing forces us to consider macro-phenomena such as power structures—which are usually opaque to approaches that focus on the individual agential standpoint⁷⁸—as the background architecture that both constrains and enables agency.

So far so good, as to an outline of the central components of my reframing of Korsgaard’s constitutivist theory of agency. All of this, however, still leaves the substance of the notions of intentional action and its connection to the *Identity Thesis* indefinite, which I shall address in the following sections. Although we cannot rely on Korsgaard’s account of intentional action for the reasons I have outlined here and in chapter three, the characterization provided by G. E. M. Anscombe can be used to help explicate how constitutive features of agency such as relationality and context-sensitivity relate to the notions of (first-personal) agential standpoint and agential authority. Anscombe’s account also helps us flesh out some of the scalar aspects of first-person and attributed intentionality.

5.2 Anscombe on Intentional Action

According to G. E. M. Anscombe, intentional actions are those “to which a certain sense of the question ‘Why?’ is given application”—the relevant sense being “that in which the answer, if

⁷⁷ A “misfire” is a borrowed Austinian term.

⁷⁸ On the theoretical superiority of (social) structural explanations over explanations that focus on individual agents see Haslanger (2015).

positive, gives a reason for acting” (Anscombe, 1957/2000, p. 9). The term ‘intentional,’ Anscombe claims, does not involve any mental state/event or interior act (ivi, pp. 2, 8-9, 36, 49) nor an extra-feature (ivi, pp. 28-9) accompanying the agent’s behavior. Rather, it “has reference to a *form* of description of events” (ivi, p. 84). More specifically, among the several possible descriptions of a person’s bodily movements, what counts as *her* actions are those individuated by the ‘descriptions under which’ she *knows what she is doing*: “it is the agent’s knowledge of what [s]he is doing that gives the descriptions under which what is going on is the execution of an intention” (ivi, p. 87).

The kinds of answers by means of which the ‘Why?’ question is given or refused application reveal important features of the notion of intentional action. For example, answers mentioning some future state of affairs (‘I am doing A in order to do B’), or wider descriptions of what one is doing (‘I am doing A because I am doing B’), reveal the *progressive*, or *imperfective*, character of action and its *teleological/explanatory* structure (ivi, pp. 45-7),⁷⁹ respectively. By refusing application to the ‘Why?’ question, answers like ‘I was not aware I was doing that,’ and ‘I observed I was doing that’ shed light on the *self-conscious* character of intentional action, and on the *non-observational* and *non-evidential* character of the kind of knowledge it consists in.

On the other hand, Anscombe states that answers like ‘I just thought I would’ or ‘It was an impulse’ or ‘For no particular reason’ or ‘It was an idle action—I was just doodling’ are *not* a rejection of the question: “The question is not refused application because the answer to it says that there is *no* reason, any more than the question of how much money I have in my pocket is refused application by the answer ‘None’” (ivi, p. 25). This flexibility shows that an agent may be more or less aware of the intentions behind their actions. Some may be fully conscious, deliberate, and complexly reasoned. Others, not. All, however, count as intentional actions.

Following Aquinas, Anscombe defines the kind of knowledge an agent has of her intentional actions as *practical knowledge*, which is importantly contrasted with speculative knowledge in that the former is “the cause of what it understands,” while the latter is derived from the object known. Practical knowledge is distinctively “exercised in the action” and is the *formal* cause of the agent’s intentional actions, in the sense that “without it what happens does not come under the description—execution of intentions” (ivi, pp. 84, 87-89). By acting intentionally, “I

⁷⁹ By teleological/explanatory structure Anscombe refers to the means-end relation that action descriptions offered in answer to the Why?-question often take.

do what happens. That is to say, when the description of what happens is the very thing which I should say I was doing, then there is no distinction between my doing and the thing's happening" (ivi, pp. 52-53).

5.3 Relevance of the Anscombean Approach to Intentional Action for the Present Work

The exegesis of Anscombe's characterization of intentional action in terms of practical knowledge, and the elaborations thereon, are complex and thriving (see Ford, Hornsby, & Stoutland, 2011). However, for present purposes a quite general highlighting of some of its central insights will suffice.

Anscombe provides a distinctively non-reductive,⁸⁰ outward-looking⁸¹ (Anscombe, 1957/2000, pp. 9, 49) approach to intentional action. 'Intentional' does not stand for any mental states accompanying the action; rather, it is a formal category,⁸² as it marks a description of events. At the same time, her account of intentional action is agent-centered. Any intentional action is given 'under a description' (ivi, pp. 52-3),⁸³ and indeed there might be several descriptions for an action. The relevant description, however, is fixed by directly addressing the agent for an explanation of her action.⁸⁴ The description under which an action counts as intentional is the one under which the agent knows what she is doing, which brings to the fore "the agent's specific understanding of the structure of an action-in-progress" (Lavin, 2016, p. 624), and hence the self-conscious and distinctively first-personal character of intentional action. In the displacement case, no agent was asked about their intentions before producing

⁸⁰ In the sense that the account of intentional action she provides does not reduce it to mental states of the individual agent. Cf. *supra*, section 1.1.2.

⁸¹ The distinction between inward-looking and outward-looking approaches to agency is Stout's (2005). Cf. also Stoutland (2011, pp. 24, 28).

⁸² See Thompson (2008) and Lavin (2016) for a development of this Anscombean idea within a Fregean logical infrastructure, into the project of an 'analytical Aristotelianism.'

⁸³ Note that Anscombe's 'under-a-description' condition for action individuation is very different from Davidson's, despite having been often associated with it. Emphasizing this difference is important as it reflects deeper differences in methods and approaches to action/agency in their equally pioneering work; so much that they are customarily seen as the initiators of two distinct strands or traditions in the philosophy of action. For Anscombe's own stand on this issue see her (1979).

⁸⁴ The scenarios in Anscombe's examples are always 'personal', see Lavin (2016, pp. 623-624): "we are asked to imagine *the agent herself* being asked 'Why?' about the action underway;" "Anscombe's question 'Why?' is characteristically posed to some 'you,' someone who would answer with a sentence beginning with 'I'." Cf. Thompson (2011).

the opposing headlines. The reporters' inability to address the agents also highlights the fact that only a select few can be in a position to address the agent for an explanation, which coincides with my characterization of agency as *selective* intersubjective recognition.

Importantly, Anscombe describes the kind of understanding an agent has of her own intentional action as *practical*, arguably in (at least)⁸⁵ the following twofold sense. First, in that the agent's conscious awareness of her action plays a crucial role in practically orienting her throughout the execution of the relevant action (Pickard, 2004). The '*agential awareness*'—articulated by the agent in response to the Why?-question, typically in the form of a means-end order such as 'I am doing A because I am doing B,' or 'I am doing A in order to do B'—is "an awareness that does not merely record but determines the order, and thus the progress, it comprehends" (Lavin, 2016, p. 622). Second, in a stronger sense, the agent's knowledge of her intentional actions is practical in that the potential for *rational control* that the agent has over them puts her in a (normatively) practical relation of '*agential authority*' with respect to them (Moran, 2001, 2004; Bagnoli, 2007, 2013).

Both features of intentional action—availability to conscious awareness, and susceptibility to rational control, justification, and practical deliberation—are central to the understanding of agential capacities relied on in the *Agential Responsibility Claim*, and represent a significant point of continuity with Korsgaard's account.⁸⁶ However, beyond Korsgaard's suggestions,⁸⁷ Anscombe's account encourages the hypothesis of understanding intentional actions, along with actions of cognate species such as merely voluntary actions, on the one hand, and actions done for reasons, on the other, as structured on a spectrum of increasingly complex cognitive capacities and degrees of availability to the agent's conscious control (see Levy, 2013).

Anscombe's emphasis on the *immediacy* and *authority* proper to the practical kind of knowledge the agent has of her intentional actions (cf. Moran, 2001) importantly speaks to

⁸⁵ This is not meant to be an investigation into Anscombe's notion of practical knowledge. Such investigation has been conducted, among others, by Moran (2004), Rödl (2011), Schwenkler (2015).

⁸⁶ One important difference between the two is Korsgaard's explicit aversion to the notion of practical knowledge. For a reconstruction and discussion of Korsgaard's arguments against the notion of 'applied knowledge' see Bagnoli (2013, pp. 156-160). Bagnoli proposes a 'return to Anscombe' and defends a form of *practical cognitivism* out of Anscombe's account of practical knowledge within the framework of Kantian metaethical constructivism—i.e., practical knowledge is understood as the "knowledge of oneself as a practical subject."

⁸⁷ SC (sections 5.4.7 and 6.1.3). Korsgaard's focus, however, is on different kinds of action and agency as displayed by human and non-human agents. Therefore, the criticism that her account is unduly committed to consider *human* actions that fall short of rationality as less than full-fledged actions, and that her account captures just one aspect of human agency, still applies.

both the commonsense and the philosophical understanding of the centrality of the agential standpoint in explaining the phenomenon of human agency. As a consequence, one of the main tendencies in the current literature has been to interpret Anscombe's characterization of practical knowledge in terms of a form of self-knowledge.⁸⁸

In partial departure from this tendency, my elaboration on Anscombe's work is meant to emphasize instead her distinctive attempt at providing an outward-looking characterization of intentional action (cf. Stout, 2005) which keeps the agent center stage, without restricting the focus to her individual psychology. Specifically, my aim is to shift the focus from the competences and capacities required for agency, to the broader context of the external and intersubjective conditions involved in the processes of authoring of an action. With the Anscombean picture of intentional action in place, I shall now proceed to lay out my reconceptualization of the constitution of agency.

5.4 Dimensions of Agency

To begin with some analytical clarity, I distinguish three different dimensions of human agency: *psychological*, *cognitive*, and *performative*. In introducing these distinctions, it is important to emphasize that, *qua* theoretical distinctions, they are not supposed to be explanatorily exhaustive categories to understand the phenomenon of human agency, but rather useful tools to analyze some aspects of its complexity in theoretical isolation/abstraction from each other.⁸⁹

By the *psychological* dimension of human agency, I refer to the psychological processes and sub-personal mechanisms underlying the performance and (sub-agentive) control of an action, on the one hand, *and* to the personal-level experience of intentional actions available to the agent's agential awareness, on the other.

Empirical studies on the role of consciousness in the initiation and control of actions investigate the hypothesis of a sub-personal psychological system, the so-called 'comparator

⁸⁸ For an example of how the idea of practical knowledge as a form of self-knowledge has been developed along very different directions in the literature, compare the works by David Velleman (reductive cognitivism about practical knowledge), Richard Moran (constitutivism about practical knowledge), and Carla Bagnoli (practical cognitivism).

⁸⁹ The methodological assumption guiding this distinction is that an investigation into each dimension separately from the overall explanation of their reciprocal connection, and, broadly speaking, from addressing *relational* questions, metaphysically speaking (see *supra*, section 4.3.1), is a worthwhile theoretical task in itself.

model of motor control.’ According to several studies (Frith et al., 2000, Haggard, 2005, Bayne & Pacherie, 2007, Pacherie, 2008, 2011), that system is partially responsible for the distinctive phenomenology of agency, or the sense the agent has that she is the author of her actions. These inquiries are meant to shed light on the complex workings of agency, characterized by the interplay between the agent’s conscious awareness and rational control, and sub-personal processes.

In other words, I take the term ‘psychological’ to apply to the self-conscious dimension distinctive of human subjectivity and phenomenological experience, independently of the *degree* of conscious awareness and cognitive penetrability on the part of the agent’s higher level cognitive and meta-cognitive capacities.

The *cognitive* dimension is distinct from, but functions as a bridge between, the psychological and performative. It encompasses an agent’s non-conceptual (procedural) capacities for regulation and monitoring of her own mental processes and bodily actions (Proust, 2010),⁹⁰ for directed attention, as well as higher order conceptual and reflective capacities for epistemic and practical rationality.

The *performative* dimension (cf. supra, section 4.6.2) pertains to the interpersonal and social practices, which rely on, and gain meaning from, the societal infrastructure within which these practices take place. I use the notion of *societal infrastructure* as an umbrella term, to include the set of cognitive, conceptual, and organizational tools—the paradigm case being language, up to the social constructions/categories, beliefs and attitudes constituting a culture—, various coordination devices (i.e., conventions, rules, institutions), and collective epistemic resources making up the social world.

Distinguishing these three dimensions helps provide a better articulation of the scope of a theory of agency, the focus of which can be on:

1. ***agency as a capacity*** (agential capacities/powers), exercised by concrete agents, to which different constitutive norms (both *non-practice-based* and *socially-generated*, along with the characterization thereof provided in chapter four) apply, depending on whether we (third-personally) consider them as
 - a. biological (autonomous) systems

⁹⁰ Cf. Arango-Munoz (2015) on Proust’s 2013 book, which articulates the metacognitive approach to mental agency in full.

- b. cognizers
 - c. social individuals
 - d. persons
2. ***agency as a social activity*** (a set of procedures, practices, institutions), actualized in the concrete *actions performed* (actual performances and reactions—either positive or negative—executed, what is done)
 3. ***agency from within*** the embodied (first- and second-personal) agential standpoints

My proposal brings the performative dimension of agency into the picture, to emphasize that intentional action is social action. A social action in the minimal sense proper of the performative dimension, is an action whose identification and attribution takes place within the context of a structured practice—even though not necessarily in an institutional context.

My reconceptualization is therefore meant to inherit the Anscombean outward-looking approach to agency in the sense implied by the thesis that any intentional action is given ‘under a description,’ along with her specification of the interactive process of reaching the ‘relevant’ description emerging from the reiterative application of the Why?-question. The agent’s social and interpersonal environments provide not only the interactive conditions for individual self-awareness and self-experience⁹¹ crucially involved in the exercise of agency itself, but also the very conceptual tools to frame one’s and others’ intentional actions ‘under a description.’⁹² Dialogically, we learn the meaning that our actions ‘make’ in the public sphere. The performative dimension I am pointing to represents the *meeting ground* (Dotson, 2014, p. 92) where the authoring of intentional actions takes place.

5.5 Towards an Interactionist View

What constitutes a human agent? Korsgaard is right in pointing out that we are born as creatures with certain agential capacities, who cannot help but act. We are nearly always also born into a social world, a world populated by others. However, not all of those others are or should be

⁹¹ On the relational and social nature of selfhood, agency, and autonomy analyzed from a feminist perspective, see the essays in Mackenzie & Stoljar (2000); for an overview, see Stoljar (2013). See also Gallagher (2014). For a (quite radical) challenge to the distinction between a minimal and an interpersonally constituted sense of self, and the proposal of a reframing of psychotic conditions such as schizophrenia in relational terms, see Ratcliffe (ms.).

⁹² See Thompson, 2008 for a logical treatment of practical concepts that emphasizes the constitutive normative relation between intentional action, and human practices and forms of life. Thompson’s work expands both on Philippa Foot’s work (2001), and on the program put forward by Anscombe (1958).

considered equal from an agential point of view. Only those in our immediate surroundings are witness to our first actions, our first acts of agency and thus start to attribute it to us.

Endowed with this *minimal agency*, we are immersed in interpersonal settings where attributions of actions and agency become ‘thicker.’ As our cognitive, psychological, and social skills gradually develop, we are recognized and begin to recognize ourselves as *authors* of intentional actions through participation in the normative practice and exchange of practical reasons. By mastering and exercising our agential capacities, as simultaneous reasoners and actors, we become able to engage in increasingly complex social practices and relationships, wherein we are more fully recognized for our developing complexity.

Those closest to us, that is, those with a concern for our agency—such as parents, siblings, extended family and close friends—provide the heart of our interactions and thus function as the first primary recognizers of our agency. In this way, the simultaneous and interrelated process of becoming and being recognized as an agent happens primarily through what I have called *selective intersubjective recognition*. It is selective because we are not recognized as agents by all of society, but by a select social group concerned with our agency who recognizes the intentions with which we are authoring our actions.

Identifying the same capacity in our select group allows us to mutually reciprocate our *authored agency*, recognizing their agency as well. As our reciprocated agency flourishes, we are able to then recognize and interact with our agency on a macro level—taking note of our attributed place in broader social groupings beyond the selective group from which we developed. Of course we are always only more or less aware of these macro structures. Sometimes they completely hijack our intentions while other times we have the ability to negotiate them. Nevertheless, being recognized and interacting with our select intersubjective groups, while acknowledging our and their places in a broader social context, could be considered the *paradigm agency* of responsible and free individuals.

Importantly, a paradigm agent is not an ideal agent. Developing and exercising agency is a complex and progressive process, and the trajectory of developing from a minimal into a paradigm agent is neither linear nor fixed. Full adults gain and lose ‘components’ of what makes a paradigm agent all the time, which is why, ontogenetically speaking, human agency is best captured by a model employing scalar notions through and through.

Finally, the criteria for potential agents to be recognized and feel as such should not be thought of as hard-and-fast. This is not just because individual agents are constituted *with* others through *selective intersubjective recognition*, and therefore the criteria might vary across the different kinds of proximal interactional contexts where agency is exercised. It is also

because the criteria themselves are subject to a certain extent of intersubjective negotiation and renegotiation in the broader choreography of the social dialectic of agency.⁹³

On the whole, my aim in this work has been to expand beyond Korsgaard's authorship view, and to highlight human agency's constitutive dependence on *selective intersubjective recognition* on the one hand and *extended social recognition* on the other, providing the preliminary reconceptualization for the development of a constitutivist theory of agency in an interactionist key.

5.6 Conclusion

In this chapter I have spelled out three distinct components of a first-order constitutivist theory of agency: i) the necessity of a scalar approach, ii) relationality, and iii) context-sensitivity and situatedness. These components were inspired by Korsgaard's indication that action and agency are graded notions, and by her practical account of responsibility as answerability. I strengthened these aspects into the suggestions of adopting a thoroughly scalar approach to agency, and of considering agency as inherently socially constituted through interpersonal and social practices.

In line with the argument built throughout the preceding chapters, I proposed a two-pronged path to expand on Korsgaard's account, and outlined a characterization of agency that centralizes interaction rather than individual psychology. First, I suggested that some aspects of G. E. M. Anscombe's work—interpreted as an outward-looking and interactional/dialogical approach to intentional action—be incorporated in the reframed account. Second, I introduced the notion of intersubjective recognition as a necessary component for the characterization of the social constitution of agency.

A reframing of the constitutivist theory along the suggested components is capable of accounting for cases of *Disorders* and *Displacement* as legitimate instances of agency, and provide just a first step towards a principled articulation of what dysfunctional interactional dynamics consist of, and how the burden of agential and moral responsibility distributes when they occur.

⁹³ On the 'social dialectic of agency,' cf. Hegelian theories of action and practical reason—notably, the essays in Laitinen & Sandis (2010).

Chapter Six – Conclusions

In this concluding chapter I recapitulate the trajectory of my argument throughout this work, articulate both its theoretical and practical implications, and discuss some of the limitations of my proposal and prospects for future research.

6.1 Summary

The main goal of this dissertation has been to provide a critique of, and an extension upon Christine Korsgaard's constitutivist theory of agency.

To do so, I started chapter one with some definitions of the key notions in the philosophy of action. There, I also provided an outline of constitutivism's origins and goals in relation to the metaethical and action-theoretical debates, as well as a first characterization of Korsgaard's constitutivism in the context of this broader theoretical landscape. With the preliminaries in place, I stated the scope of my analysis and the aim of this work. Unlike much of the burgeoning literature engaging with Korsgaard's work that addresses the metanormative ambitions of her argument, my interest has been in her first-order *metaphysical-cum-normative* account of agency. It is at this stage, especially with her thesis about the *double nature* of constitutive norms of agency, that the crucial normative step in the argument is made, which carries important consequences for what, metaphysically and practically, we are to consider as legitimate instances of agency and who we are to consider genuine agents. Consequently, throughout the critical part of the dissertation I pursued a clarification of the role played by constitutive norms in Korsgaard's account, and of the status of the *practical normativity* pertaining to agency.

Chapter two was dedicated to an extensive reconstruction of Korsgaard's first-order constitutivist theory of agency. This was conducted by zeroing in on her version of the *authorship view* of action, the metaphysics of agency and normativity underpinning the thesis that self-constitution is the function of action and that agency is an activity of self-constitution, and her relational account of responsibility as answerability. The focus of Korsgaard's authorship view is on the psychological structure necessary for an individual to *own* her action, which grants the characteristic relation of authority making those actions attributable to *her* as a whole, rather than to some desires or forces "working in or on her." In short, according to Korsgaard, genuine actions—i.e., actions that fully constitute the agent as such—are *autonomous* actions, or self-determined expressions of the agent's practical identity, or

capacity for normative self-government. In this way, being an agent amounts to being engaged in the activity of self-constitution, and the constitutive principle defining and governing such activity is the categorical imperative. I also outlined Korsgaard's relational approach to responsibility and the (minimal) social dimension of agency and practical reasoning implied by her account of the normativity of reasons, supported by her arguments from the publicity of reasons and their essentially shareable character.

In chapter three, I addressed what in the literature is known as the problem of defective actions. Despite Korsgaard's explicit acknowledgement that the concepts of action and agency come in degree, one problematic consequence of her account is the tension between the simultaneously descriptive ('norms for being') and normative ('norms for action') character of constitutive norms. Given the relation of self-constitution between an agent and her actions, defective actions—i.e., less than autonomous actions—would somehow 'transmit' a kind of 'failure' to the agent and render her ineffective *qua* agent.

To explicate the problem, I made use of two case studies illustrating instances of 'disorders of agency,' and 'socially displaced agency.' A key point in the *Disorders* scenario, was the presumption for the effective 'treatment' of people with disorders of agency: no matter how developed their agential capacities are, they would be held accountable for their actions (i.e., judged blameworthy for them and asked to deal with their consequences), but not morally blamed for them. By undertaking this stance of *responsibility without blame*—which is, importantly, different from Strawson's objective stance—trained practitioners of mental health services create a therapeutic relation and a space where patients can effectively develop and exercise their agential capacities. In light of the *Disorders* scenario, I argued against Korsgaard's assimilation of the practical normativity of agency to moral normativity, and advanced the morally neutral notion of *agential responsibility* predicated upon Hanna Pickard's articulation of *responsibility without blame*. The discussion of the *Displacement* scenario shed light on the phenomenon of the 'social authoring' of an action—the attribution of intentional action to others based on societal power structures. Acknowledging the power of social authoring led to the observation that while factors external to the agent—such as membership to disenfranchised groups subject to structures of oppression and violence—might have a distorting influence on their capacity to act as they intend, their agency is not thereby erased nor disabled.

Overall, the analysis of both cases corroborated the relational approach of Korsgaard's account of responsibility, which motivated my preliminary proposal to complement her constitutivist theory with the *Interaction Claim*, according to which agency is *constitutively* a

function of the agent's location within a network of social and interpersonal relationships, as well as of the quality of those relationships, as perceived by the agent herself.

Chapter four refined the problem of defective actions by providing a conceptual elucidation of the notion of constitutive normativity, which is central to constitutivism as a first-order theory of agency. I presented three possible interpretations of Korsgaard's characterization of the notion of *constitutive norms*—the Metaphysical Interpretation, the Individual Psychology Interpretation, and the Practical Interpretation. I discarded the Metaphysical Interpretation as unstable, and pointed out that an interpretation of constitutive norms of agency along the Individual Psychology Interpretation is partial. Consequently, I indicated the Practical Interpretation as the best way to make sense of the notion while accounting for the conclusions drawn in chapter three and preserving some of the insights of the Individual Psychology Interpretation. My conclusion was that the constitutive normativity pertaining to human agency is not homogeneous in kind, having both natural and social roots, and that while social conditions are necessary for agency, they are not sufficient. I therefore distinguished between a *non-practice-based* and a *socially-generated* variety of constitutive normativity, inspired by the theoretical framework on the metaphysics of normativity developed by the Autonomous System Account (ASA) of functional normativity, and by the Speech Act Theory (SAT), respectively.

In chapter five, I outlined my view of an interactionist approach to agency by indicating three central components that a first-order constitutivist theory of agency should account for: i) the necessity of a scalar approach, ii) relationality, and iii) context sensitivity and situatedness. Inspired by Korsgaard's acknowledgement that agency comes in degrees and by her practical account of responsibility, these aspects are considerably strengthened in my relational definition of what it is to be an agent (introduced in chapter three as the *Agential Responsibility Claim*), on the one hand, and the proposal of considering agency as inherently socially constituted through interpersonal and social practices (*Interaction Claim*), on the other. Relationality and the necessity of a scalar approach were accounted for in my *Agential Responsibility Claim*, stating that treating someone as an agent is to recognize and treat her as a legitimate participant in practices of accountability, which presupposes that she has, to *some degree*, certain agential capacities (i.e., knowledge of what she is doing, ability to exercise choice and a degree of control over her behavior). The *Interaction Claim* (introduced in chapter three) accounted for the inherent context-sensitivity and situatedness of agency, while receiving further refinement in my claim that (*selective*) *intersubjective recognition* is a necessary component for the successful constitution and exercise of agency.

In the simplest terms possible, my view of agency switches the emphasis from individual psychology to social interaction. Whereas Korsgaard prioritizes individual psychological unity, I consider social interaction necessary for individual psychological unity itself, thereby privileging its place in the constitutivist account. Nevertheless, while necessary, social interaction is not sufficient for ‘paradigm agency,’ which importantly involves an individual’s capacities for practical reasoning and deliberation, a certain amount of self-possession (self-regulation and control), and the ability to discern the sense one’s actions make in the relevant interactive and social environments—all of which, of course, is a matter of degree needs case-by-case contextualization.

The peculiarity of my reframing of Korsgaard’s account lies in my proposal to incorporate aspects of G. E. M. Anscombe’s seminal work on intention—interpreted as an outward-looking and interactional/dialogical approach—which can be used to help explicate how constitutive features of agency, such as relationality and context-sensitivity, relate to the notions of a first-personal agential standpoint and agential authority, expanding the scope of the explanation beyond the agent’s individual psychology.

Anscombe’s account also helps us flesh out some of the scalar aspects of first-person and attributed intentionality, opening the way to understanding intentional actions, along with actions of cognate species such as merely voluntary actions and actions done for reasons, as structured on a spectrum of increasingly complex cognitive capacities and degrees of availability to the agent’s conscious control.

Overall, the proposed reframing and expansion of Korsgaard’s first-order constitutivist account of agency helps us make better sense of cases such as *Disorders* and *Displacement* as legitimate instances of agency, and represents a first step towards a constitutivist-interactionist view which highlights agency’s inherent dependence on *interpersonal* and *social recognition*.

6.2 Theoretical and Practical Implications

Beyond the specific goal of providing an expansion of Korsgaard’s constitutivist account, the present work also contributes to the general debate on constitutivism in a few distinct ways. From the point of view of its metanormative ambitions, this dissertation promotes a shift in focus from constitutivism’s hitherto strictly metaethical agenda to the broader project of an elucidation of how different kinds of normativity—i.e., norms of different provenance, such as biological, cognitive, social—hold together in the complex processes of the *social constitution* of agency.

On the first-order metaphysical account of agency, my aim has been to contribute to the debate by illuminating potential points of connection with theories of collective intentionality, on the one hand, and of social ontology, on the other. To do so, I introduced the notion of *intersubjective recognition* and argued for its pervasive role in the development and exercise of individual agency. While there is recent valuable work in the philosophy of action and agency that takes the notion into account,⁹⁴ this is the first attempt, to my knowledge, to incorporate it into a constitutivist theory of agency.

The proposed relational characterization of what it is to be an agent (via the notions of *agential responsibility* and of *intersubjective recognition*), and the preliminary work done throughout the dissertation in the direction of an interactionist, context-sensitive, and situated view of agency are also first steps towards an exploration of the *systemic* character of responsibility. In doing so, I join others⁹⁵ in investigating a conceptual and methodological reorientation towards less individualistic approaches in the philosophy of agency and action.

Even though this dissertation has explicitly set aside the metanormative and ethical aspects of constitutivism, the conceptual clarifications pursued throughout have practical implications of markedly ethical significance. This work has been written on the understanding that theoretical work and categorization are normative practices with concrete practical effects. These effects manifest in domains other than philosophy, such as psychology, social work, and politics, and, more generally and fundamentally, in the collective narratives, attitudes, and practices (ordinary or otherwise) permeated by the language of agency and responsibility.

6.3 Limitations and Prospects for Future Work

While I have limited the scope of the present work to the notion of *agency*, a reframing towards a full constitutivist theory would require at least a tripartite investigation that also examines and connects the complementary and interrelated notions of *action* and *practical reasoning*.

A second limitation of this work pertains to the selection of the cases in chapter three. There, the two cases I chose for the illustration of my points might be considered ‘extraordinary’ compared to more ordinary cases of failures or defects of agency (typically, cases of weakness

⁹⁴ In particular, see Arruda (2016b) for both a detailed state of the play on the issue of recognition in philosophy of action, and an original proposal on what the notion consists in.

⁹⁵ See Cash (2010), Hurley (2011), and Thalos & Andreou (2009), which draw on the hypothesis of Extended and Embodied Cognition in the philosophy of mind, bounded/situated approaches to rationality and practical reasoning, and evolutionary theory, respectively.

of will, or bad actions), and Korsgaard might object that they are beyond the scope of what her theory avowedly tries to account for—i.e., the rational agency of adult human beings (SC, section 5.3). However, aim of my argument has been to emphasize how the continuity between these and more ‘ordinary’ cases (of successful and defective agency alike) points to the necessity of adopting a thoroughly scalar approach to agency, specifically, one in which the relationality, context-sensitivity, and situatedness are integral to what constitutes agency, besides the agential capacities of the individual agent. Besides helping to make my contributions clear-cut, the inclusion of those cases into my proposed reframing of Korsgaard’s theory allowed me to articulate the centrality of her notion of responsibility as a practical relation also to the constitution of agency, leading to an expanded theory capable of more enriching insights. In short, though Korsgaard might object they are beyond the scope of her theory, I argue her theory would be enriched if she included them. I agree, however, that investigating cases of ‘less extraordinary’ defective agency, considered within the relevant contexts, are also likely to produce further valuable insights leading to a further development of the proposals and insights outlined here.

My approach in this work has also been limited to the constitutive (and causally non-reductionist) perspective, which to be sure, is not the only perspective from which interesting and viable theories of agency are developed. Interesting departure points for comparison with the ideas presented here might come from, among others, David Velleman’s and Michael Bratman’s theories which developed different frameworks that must nonetheless address the same or similar issues.

In terms of expansion, given my commitment to provide an outward-looking account of agency and intentional action, I welcome the opportunity of further exploring the fitting of an Anscombean approach to intentionality into a full-fledged theory of agency. ‘Going Anscombean’ represents an extremely intriguing direction of development for some of the points outlined in this work, which would contribute to a burgeoning scholarship in practical philosophy, with important ramifications in other areas of philosophy as well, notably epistemology. In particular, I see great potential in elaborating on Anscombe’s characterization of intentional action in terms of practical knowledge in the direction of an epistemology of competence, or knowledge-how, and other forms of practical knowledge.

Finally, among the points indicated as practical implications of this work, it would be valuable to a number of communities to bring its ethical aspects to full development—possibly within an interdisciplinary oriented research programme/environment—and put the ensuing

theoretical framework in the service of mental health care research, practice, and policy making, and other forms of social policies.

6.4 Conclusion

The motivation for this work stems from the experience of moments and situations of disempowered agency—online, in the news, and while traversing the public sphere. These situations allowed the observation that agency seldom, if ever, disappears; that complicated moments of agency come and go; that *anyone* (to different extents and degrees of severity) is vulnerable to them. Constituting agency is in our power *together*, for better or worse, and we need to own responsibility for our ability and possibility to empower (or disempower) each other.

Providing theoretical legitimacy to agents and agency distant from the paradigm cases is therefore in itself an act of practical recognition, validation, and advocacy. As such, this work is meant to be one small step in the greater fight against stigmatization of people with mental health conditions, racial discrimination, and other forms of thwarting and disempowering of human agency.

Bibliography

- Aguilar, J. H., Buckareff, A. A., & Frankish, K. (2011). Introduction. In J. H. Aguilar, A. A. Buckareff, & K. Frankish, *New Waves in the Philosophy of Action* (pp. 1-9). London: Palgrave Macmillan.
- Allen, C. (2003). Teleological Notions in Biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2009 ed.). <https://plato.stanford.edu/archives/win2009/entries/teleology-biology/>. Retrieved from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/win2009/entries/teleology-biology/>
- Alvarez, M. (2016). Reasons for Action: Justification, Motivation, Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Retrieved from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/win2016/entries/reasons-just-vs-expl/>
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.
- Andersen, M. L., & Collins, P. H. (Eds.). (2016). *Race, Class, and Gender: An Anthology* (9th ed.). Boston, MA: Cengage Learning.
- Anscombe, G. M. (1957/2000). *Intention* (2nd ed.). Cambridge, MA: Harvard University Press.
- Anscombe, G. M. (1958). Modern Moral Philosophy. *Philosophy*, 33(124), 1-19.
- Anscombe, G. M. (1979). Under a Description. *Noûs*, 13(2), 219-233.
- Arango-Munoz, S. (2015). [Review of] Joëlle Proust: The Philosophy of Metacognition: Mental Agency and Self-Awareness. *Minds & Machines*, 25, 297-300.
- Arruda, C. T. (2016a). Constitutivism and the Self-Reflection Requirement. *Philosophia*, doi 10.1007/s11406-016-9744-5.
- Arruda, C. T. (2016b). What We Can Intend: Recognition and Collective Intentionality. *The Southern Journal of Philosophy*, 54(1), 5-26.
- Arruda, C. T. (ms.). Ideal Agents and Minimal Agents.
- Austin, J. L. (1962/1975). *How to Do Things with Words* (2nd ed.). (J. O. Urmson, & M. Sbisà, Eds.) Cambridge, MA: Harvard University Press.
- Ayala, S., & Vasilyeva, N. (2016). Responsibility for Silence. *Journal of Social Philosophy*, 47(3), 256-272.
- Bagnoli, C. (2007). The Authority of Reflection. *Theoria*, 58, 43-52.

- Bagnoli, C. (2009). Review of Korsgaard's *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. *Notre Dame Philosophical Reviews*. <http://ndpr.nd.edu/news/24047-the-constitution-of-agency-essays-on-practical-reason/>.
- Bagnoli, C. (2011/2015). Constructivism in Metaethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 ed.).
- Bagnoli, C. (2012). Morality as Practical Knowledge. *Analytic Philosophy*, 53(1), 61-70.
- Baiasu, S. (2016). Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit Out of the Hat. *Philosophia*, doi 10.1007/s11406-016-9746-3.
- Bayne, T., & Pacherie, E. (2007). Narrators and Comparators: The Architecture of Agentive Self-Awareness. *Synthese*, 159(3), 475-491.
- Bertea, S. (2013). Constitutivism and Normativity: A Qualified Defence. *Philosophical Explorations*, 16(1), 81-95.
- Bickhard, M. H. (2002). Critical Principles: On the Negative Side of Rationality. *New Ideas in Psychology*, 20(1), 1-34.
- Bierria, A. (2014). Missing in Action: Violence, Power, and Discerning Agency. *Hypatia*, 29(1), 129-145.
- Cash, M. (2010). Extended Cognition, Personal Responsibility, and Relational Autonomy. *Phenomenology and the Cognitive Sciences*, 9(4), 645-671.
- Christensen, W. (2012). Natural Sources of Normativity. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 104-112.
- Christensen, W., & Bickhard, M. H. (2002). The Process Dynamics of Normative Function. *The Monist*, 85(1), 3-28.
- de Bruin, L., van Elk, M., & Newen, A. (2012). Reconceptualizing Second-Person Interaction. *Frontiers in Human Neuroscience*, 6(151), doi 10.3389/fnhum.2012.00151.
- del Corral, M. (2015). Excuses and Exemptions. *Tópicos, Revista de Filosofía*, 49, 231-256.
- Dotson, K. (2012). A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers: A Journal of Women Studies*, 33(1), 24-47.
- Dotson, K. (2013). Conceptualizing Epistemic Oppression. *Social Epistemology*, doi 10.1080/02691728.2013.782585.
- Dotson, K. (2014). Black Feminist Me: Answering the Question 'Who Do I Think I Am'. *Diogenes*, 59(3-4), 82-95.
- Engstrom, S. (2009). *The Form of Practical Knowledge*. Cambridge, MA: Harvard University Press.

- Enoch, D. (2006). Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action. *Philosophical Review*, 115, 169-198.
- Enoch, D. (2011). Shmagency Revisited. In M. Brady (Ed.), *New Waves in Metaethics* (pp. 208-233). London: Palgrave Macmillan.
- Ferrero, L. (2009a). Action. In J. Shand (Ed.), *Central Issues of Philosophy* (pp. 137-151). Oxford: Wiley-Blackwell.
- Ferrero, L. (2009b). Constitutivism and the Inescapability of Agency. (R. Schafer-Landau, Ed.) *Oxford Studies in Metaethics*, 4, 303-333.
- Ferrero, L. (2015). Review of Katsafanas' Agency and the Foundations of Ethics: Nietzschean Constitutivism. *Ethics*, 125(3), 883-888.
- Foot, P. (2001). *Natural Goodness*. Oxford: Oxford University Press.
- Ford, A., Hornsby, J., & Stoutland, F. (Eds.). (2011). *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press.
- Fricker, M. (1999). Epistemic Oppression and Epistemic Privilege. *Canadian Journal of Philosophy*, 29(suppl. 1), 191-210.
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.
- Frith, C. D., Blakemore, S., & Wolpert, D. M. (2000). Abnormalities in the Awareness and Control of Action. *Philosophical Transactions*, 355(1404), pp. 1771-1788.
- Gallagher, S. (2014). The Cruel and Unusual Phenomenology of Solitary Confinement. *Frontiers in Psychology*, 5(585), doi 10.3389/fpsyg.2014.00585.
- Gallagher, S., & Zahavi, D. (2005/2014). Phenomenological Approaches to Self-Consciousness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.).
- Ginocchietti, M. (2016). *Actions: A Plea for an Ordinary Framework*. PhD dissertation: Università degli Studi di Trieste.
- Glüer, K., & Wikforss, Å. (2009/2015). The Normativity of Meaning and Content. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 ed.).
- Haggard, P. (2005). Conscious Intention and Motor Cognition. *Trends in Cognitive Sciences*, 9(6), 290-295.
- Hanisch, C., & Baiasu, S. (2016). Constitutivism and Kantian Constructivism in Ethical Theory: Editorial Introduction. *Philosophia*, doi 10.1007/s11406-016-9743-6.
- Haslanger, S. (2015). What Is a (Social) Structural Explanation? *Philosophical Studies*, doi 10.1007/s11098-014-0434-5.

- Hornsby, J. (1995). Disempowered Speech. *Philosophical Topics*, 23(2), 127-147.
- Hurley, S. (2011). The Public Ecology of Responsibility. In C. Knight, & Z. Stemplowska (Eds.), *Responsibility and Distributive Justice* (pp. 187-215). Oxford: Oxford University Press.
- Katsafanas, P. (2013). *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.
- Katsafanas, P. (forthcoming). Constitutivism about Practical Reason. In D. Star (Ed.), *The Oxford Handbook of Reasons and Normativity* (pp. draft of March 6, 2014 <http://people.bu.edu/pkatsa/constitutivism.pdf>). Oxford: Oxford University Press.
- Korsgaard, C. M. (1992/1996b). Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations. *Philosophical Perspectives*, 6, 305-332, reprinted in Korsgaard 1996b, pp. 188-221.
- Korsgaard, C. M. (1993/1996b). The Reasons We Can Share: An Attack on the Distinction Between Agent-Relative and Agent-Neutral Values. *Social Philosophy and Policy*, 10(1), 24-51, reprinted in Korsgaard 1996b, pp. 275-310.
- Korsgaard, C. M. (1996a). *The Sources of Normativity*. (O. O'Neill, Ed.) Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1996b). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1999/2008). Self-Constitution in the Ethics of Plato and Kant. *Journal of Ethics*, 3, 1-29, reprinted in Korsgaard 2008, essay 3.
- Korsgaard, C. M. (2008). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.
- Korsgaard, C. M. (2009). *Self-Constitution. Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Korsgaard, C. M. (2014). The Normative Constitution of Agency. In M. Vargas, & G. Yaffe (Eds.), *Rational and Social Agency: The Philosophy of Michael Bratman* (pp. 190-215). Oxford-New York: Oxford University Press.
- Korsgaard, C. M. (in progress). How to Be an Aristotelian Kantian Constitutivist. <http://www.people.fas.harvard.edu/~korsgaard/CMK.Aristotelian.Kantian.Constitutivism.pdf>.
- Kukla, R. (2014). Performative Force, Convention, and Discursive Injustice. *Hypatia*, 29(2), 440-457.
- Laitinen, A., & Sandis, C. (Eds.). (2010a). *Hegel on Action*. Basingstoke: Palgrave Macmillan.

- Lavin, D. (2004). Practical Reason and the Possibility of Error. *Ethics*, 114(3), 424-457.
- Lavin, D. (2008). Review of Stephen Darwall's *The Second Person Standpoint: Morality, Respect, and Accountability*. *Notre Dame Philosophical Reviews*. <http://ndpr.nd.edu/news/23292-the-second-person-standpoint-morality-respect-and-accountability/>.
- Lavin, D. (2016). Action as a Form of Temporal Unity: On Anscombe's Intention. *Canadian Journal of Philosophy*, 45(5-6), 609-629.
- Levy, N. (2013). Intentional Action First. *Australasian Journal of Philosophy*, 91(4), 705-718.
- Mackenzie, C., & Stoljar, N. (Eds.). (2000). *Relational Autonomy. Feminist Perspectives on Autonomy, Agency, and the Social Self*. New York-Oxford: Oxford University Press.
- Maher, C. (2010). On Being and Holding Responsible. *Philosophical Explorations*, 13(2), 129-140.
- Meini, C. (2007). *Psicologi per natura. Introduzione ai meccanismi cognitivi della psicologia ingenua*. Roma: Carocci editore.
- Millgram, E. (2005/2012). Practical Reason and the Structure of Actions. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2012 ed.).
- Millgram, E. (2011). Critical notice on Korsgaard's Self-Constitution and The Constitution of Agency. *Australasian Journal of Philosophy*, 89(3), 549-556.
- Millgram, E. (2014). Segmented Agency. In M. Vargas, & G. Yaffe (Eds.), *Rational and Social Agency: The Philosophy of Michael Bratman* (pp. 152-189). Oxford-New York: Oxford University Press.
- Moran, R. (2001). *Authority and Estrangement. An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Moran, R. (2004). Anscombe on Practical Knowledge. In J. Hyman, & H. Steward (Eds.), *Agency and Action. Royal Institute of Philosophy Supplement: 55* (pp. 43-68). Cambridge: Cambridge University Press.
- Moran, R. (2012). Self-Knowledge, 'Transparency,' and the Forms of Activity. In D. Smithies, & D. Stoljar (Eds.), *Introspection and Consciousness* (pp. 211-236). Oxford: Oxford University Press.
- Moreno, A., & Mossio, M. (2015a). *Biological Autonomy. A Philosophical and Theoretical Enquiry*. Dordrecht: Springer.
- Moreno, A., & Mossio, M. (2015b). Introduction. In *Moreno & Mossio 2015a* (pp. xix-xxxiv).
- Nieswandt, K. (2016). Anscombe on the Sources of Normativity. *The Journal of Value Inquiry*, doi 10.1007/s10790-016-9562-9.

- O'Connor, T., & Sandis, C. (Eds.). (2010). *A Companion to the Philosophy of Action*. Oxford: Wiley-Blackwell.
- Pacherie, E. (2008). The Phenomenology of Action: A Conceptual Framework. *Cognition*, *107*(1), 179-217.
- Pacherie, E. (2011). Self-Agency. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (online ed., pp. 440-462). Oxford: Oxford University Press.
- Pacherie, E. (2013). How Does It Feel to Act Together? *Phenomenology and the Cognitive Sciences*, doi 10.1007/s11097-013-9329-8.
- Pickard, H. (2004). Knowledge of Action without Observation. *Proceedings of the Aristotelian Society*, *105*(1), pp. 205-230.
- Pickard, H. (2011). What is Personality Disorder? *Philosophy, Psychiatry, & Psychology*, *18*(3), 181-184.
- Pickard, H. (2013). Responsibility without Blame: Philosophical Reflections on Clinical Practice. In K. W. Fulford, & al. (Eds.), *Oxford Handbook of Philosophy and Psychiatry* (pp. 1134-1152). Oxford: Oxford University Press.
- Proust, J. (2010). Mental Acts. In *O'Connor & Sandis 2010* (pp. 209-217).
- Proust, J. (2013). *The Philosophy of Metacognition: Mental Agency and Self-Awareness*. Oxford: Oxford University Press.
- Rödl, S. (2011). Two Forms of Practical Knowledge and Their Unity. In *Ford et al. 2011* (pp. 211-241).
- Ratcliffe, M. (ms.). Selfhood, Schizophrenia, and the Interpersonal Regulation of Experience. https://www.academia.edu/13585253/Selfhood_Schizophrenia_and_the_Interpersonal_Regulation_of_Experience.
- Rawls, J. (1955). Two Concepts of Rules. *The Philosophical Review*, *64*, 3-32.
- Raz, J. (1999). Explaining Normativity: On Rationality and the Justification of Reason. *Ratio*, *12*, 354-379.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*, 54-57.
- Sbisà, M. (2002). Speech Acts in Context. *Language & Communication*, *22*, 421-436.
- Sbisà, M. (2007). How to Read Austin. *Pragmatics*, *17*, 461-473.
- Sbisà, M. (2009). Uptake and Conventionality in Illocution. *Lodz Papers in Pragmatics*, *5*(1), 33-52.
- Sbisà, M. (2013). Locution, Illocution, Perlocution. In M. Sbisà, & K. Turner (Eds.), *Pragmatics of Speech Actions* (pp. 25-76). Berlin: Mouton de Gruyter.

- Sbisà, M. (2014). The Austinian Conception of Illocution and its Implications for Value Judgments and Social Ontology. *Etica & Politica / Ethics & Politics*, XVI(2), 619-631.
- Sbisà, M. (forthcoming). Varieties of Speech Act Norms. In M. Witek, & I. Witzak-Plisiecka (Eds.), *Dynamics and Varieties of Speech Actions. A Theme Issue of Poznan Studies in the Philosophy of Sciences and the Humanities* (p. https://www.academia.edu/26993277/Varieties_of_speech_act_norms.pdf). Leiden: Brill.
- Schlosser, M. (2015). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 ed.).
- Schwenkler, J. (2015). Understanding 'Practical Knowledge'. *Philosophers' Imprint*, 15(15).
- Searle, J. (1969). *Speech Acts*. Cambridge: Cambridge University Press.
- Sommers, S. R., Apfelbaum, E. P., Dukes, K. N., Toosi, N., & Wang, E. J. (2006). Race and Media Coverage of Hurricane Katrina: Analysis, Implications, and Future Research Questions. *Analyses of Social Issues and Public Policy*, 6(1), 39-55.
- Steward, H. (2012). Actions as Processes. *Philosophical Perspectives*, 26, 373-388.
- Stoljar, N. (2013). Feminist Perspectives on Autonomy. In *The Stanford Encyclopedia of Philosophy* (Fall 2015 ed.).
- Stout, R. (2005). *Action*. Chesham: Acumen.
- Stoutland, F. (2011). Summary of Anscombe's Intention. In *Ford et al. 2011* (pp. 23-32).
- Thalos, M., & Andreou, C. (2009). Of Human Bonding: An Essay on the Natural History of Agency. *Public Reason*, 1(2), 46-73.
- Thomasson, A. L. (2012). Research Problems and Methods in Metaphysics. In N. Manson, & R. Barnard (Eds.), *The Continuum Companion to Metaphysics* (pp. 14-45). New York: Continuum International Publishing.
- Thompson, M. (2008). *Life and Action. Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press.
- Thompson, M. (2011). Anscombe's Intention and Practical Knowledge. In *Ford et al. 2011* (pp. 198-210).
- Tubert, A. (2010). Constitutive Arguments. *Philosophy Compass*, 5(8), 656-666.
- Tubert, A. (2011). Korsgaard's Constitutive Arguments and the Principles of Practical Reason. *The Philosophical Quarterly*, 61(243), 341-362.
- van Willigenburg, T. (2002). Shareability and Actual Sharing: Korsgaard's Position on the Publicity of Reasons. *Philosophical Investigations*, 25(2), 172-189.

- Velleman, J. D. (1992/2000). What Happens When Someone Acts? *Mind*, 101, 461-481
reprinted in Velleman, 2000, pp. 123-143.
- Velleman, J. D. (2000). *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Velleman, J. D. (2009). *How We Get Along*. Cambridge: Cambridge University Press.
- Vincent, N. A. (2011). A Structured Taxonomy of Responsibility Concepts. In N. A. Vincent,
& al. (Eds.), *Moral Responsibility: Beyond Free Will and Determinism* (pp. 15-36).
Dordrecht: Springer.
- Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, 24, 227-248.
- Willett, C., Anderson, E., & Meyers, D. (1999/2015). In E. N. Zalta (Ed.), *The Stanford
Encyclopedia of Philosophy* (Winter 2016 ed.).
- Wilson, R. A., & Foglia, L. (2011/2015). Embodied Cognition. In E. N. Zalta (Ed.), *The
Stanford Encyclopedia of Philosophy* (Spring 2017 ed.).
- Wolf, S. (1990). *Freedom Within Reason*. Oxford: Oxford University Press.
- Zaibert, L., & Smith, B. (2007). Varieties of Normativity: An Essay on Social Ontology. In S.
L. Tsohatzidis (Ed.), *Intentional Acts and Institutional Facts* (pp. 157-173). Dordrecht:
Springer.
- Zheng, R. (2016). Attributability, Accountability, and Implicit Bias. In M. Brownstein, & J.
Saul (Eds.), *Implicit Bias and Philosophy. Structural Injustice, and Ethics* (Vol. vol. 2,
pp. 62-89, uncorrected proof). Oxford: Oxford University Press.