PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

9

n. 9 - 2015

PHENOMENOLOGY AND MIND

THE ONLINE JOURNAL OF THE FACULTY OF PHILOSOPHY, SAN RAFFAELE UNIVERSITY

JOINT COMMITMENT: COLLECTIVE INTENTIONALITY, NORMS AND JUSTICE

Edited by Francesca De Vecchi and Silvia Tossut

Phenomenology and Mind practices double blind refereeing and publishes in English.

Stefano Canali (Scuola Internazionale Superiore di Studi Avanzati - SISSA)
Ian Carter (Università degli studi di Pavia)
Emanuela Ceva (Università degli studi di Pavia)
Antonio Da Re (Università degli studi di Padova)
Mario De Caro (Università di Roma III)
Corrado Del Bo (Università degli studi di Milano)
Emilio D'Orazio (POLITEIA - Centro per la ricerca e la formazione in politica ed etica, di Milano)
Maurizio Ferrera (Università degli studi Milano)
Luca Fonnesu (Università degli studi di Pavia)
Anna Elisabetta Galeotti (Università del Piemonte Orientale, Vercelli)
Barbara Herman (University of California, Los Angeles - UCLA)
John Horton (Keele University)
Andrea Lavazza (Centro Universitario Internazionale di Arezzo)
Eugenio Lecaldano (Università degli Studi di Roma "La Sapienza")
Neil Levy (University of Melbourne)
Beatrice Magni (Università degli studi di Milano)
Filippo Magni (Università degli studi di Pavia)
Massimo Marassi (Università Cattolica di Milano)
Alberto Martinelli (Università degli Studi di Milano)
Susan Mendus (University of York)
Glyn Morgan (Syracuse University in New York)
Anna Ogliari (Università Vita-Salute San Raffaele)
Valeria Ottonelli (Università degli studi di Genova)
Federico Gustavo Pizzetti (Università degli Studi di Milano)
Mario Ricciardi (Università degli studi di Milano)
Nicola Riva (Università degli Studi di Milano)
Adina Roskies (Dartmouth College)
Giuseppe Sartori (Università degli Studi di Padova)
Karsten R. Stueber (College of the Holy Cross)
Nadia Urbinati (Columbia University)
Corrado Viafora (Università degli studi di Padova)

Cognitive Neurosciences, Philosophy of Mind and Language, Logic (CRESA)
Edoardo Boncinelli (Università Vita-Salute San Raffaele)
Stefano Cappa (Institute for Advanced Study, IUSS, Pavia)
Benedetto de Martino (University College London, UCL)
Claudio de' Sperati (Università Vita-Salute San Raffaele)
Michele Di Francesco (Institute for Advanced Study, IUSS, Pavia)
Massimo Egidi (Libera Università Internazionale degli Studi Sociali Guido Carli di Roma, LUISS Guido Carli, Roma)
Francesco Guala (Università degli studi di Milano)
Vittorio Girotto (Istituto Universitario di Architettura di Venezia, IUAV, Venezia)
Niccolò Guicciardini (Università degli studi di Bergamo)
Diego Marconi (Università degli studi di Torino)
Gianvito Martino (Università Vita-Salute San Raffaele)
Cristina Meini (Università del Piemonte Orientale)
Martin Monti (University of California, Los Angeles, UCLA)
Andrea Moro (Institute for Advanced Study, IUSS, Pavia)
Michael Pauen (Berlin School of Mind and Brain, Humboldt-Universität)
Massimo Piattelli Palmarini (University of Arizona)
Giacomo Rizzolatti (Università degli studi di Parma)
Marco Santambrogio (Università degli studi di Parma)
Achille Varzi (Columbia University)
Nicla Vassallo (Università di Genova)

CONTENTS

CONTENTS

INTRODUCTION

INTRODUCTION

Francesca De Vecchi and Silvia Tossut
Joint Commitment, Human Life and Social Ontology

FRANCESCA DE VECCHI
Università Vita-Salute san Raffaele
francesca.devecchi@unisr.it

SILVIA TOSSUT
Università Vita-Salute san Raffaele
silvia.tossut@gmail.com

JOINT COMMITMENT, HUMAN LIFE AND SOCIAL ONTOLOGY

abstract

In this introductory chapter, we recall some of the crucial aspects of Gilbert's notion of joint commitment. Special attention is devoted to the importance of this notion both for human life in its social aspects (notably, the formation of group beliefs and the constituions of just joint commitments) and for social ontology (in particular, for the understanding of norm and institutions and of the intentionality of groups). Then, we briefly summarize the contents of the contributions collected in the issue.

keywords

joint commitment, social ontology, human collective life

1.
The san Raffaele spring school on joint commitment

The present issue of *Phenomenology and Mind*, "Joint Commitment: Collective Intentionality, Norms and Justice", originates from the Spring School "Joint Commitment: Collective Intentionality, Trust, and Political Obligation" which was organized by the research centres PERSONA (Research centre in phenomenology and sciences of the person) and CeSEP (Research centre in public ethics) and took place at Vita-Salute San Raffaele University in June 15th-17th, 2015.

The keynote speaker was, of course, Margaret Gilbert, the philosopher of the "joint commitment", whose recent book (*Joint Commitment. How We Make the Social World*, 2013) was deeply discussed in the days of the School. Moreover, the school was animated by seventeen speakers – among invited speakers and contributed papers – from ten different countries. The contributed papers were selected in a double-blind review process by international reviewers from four different countries. The days of the Spring School were rich of collaboration and fruitful discussions among young scholars and affirmed philosophers, and this volume collects the great part of the papers presented and discussed at the school.

2.
Joint commitment and human life

Joint commitment is the key-concept labelled by Margaret Gilbert to understand what we are talking about when we speak of what we do, think and feel and of our values, conventions and laws, and therefore to comprehend the structure of our social world. Joint Commitment is a very powerful concept both for our personal and public lives: "How is one to understand the sense of unity, of connection, the sense of the collective 'we'? Given disparate human beings with their own personal beliefs, strivings, and so on, what kind of unity is possible? When we talk about our goals, beliefs, values, and so on–what are we talking about?" (Gilbert 2013, pp. 5-6).

Throughout her writings, Gilbert argues that joint commitment is a fundamental part of human life, since it is at the core of fundamental everyday concepts (some of which, such as institutions, law, walking together, rights and obligations, helping behaviour, collective beliefs, actions and values, etc. are discussed in the contributions collected in this issue).

As she puts it in her last book, her notion of joint commitment covers "a wide range of topics which fall the multifaceted domain of the philosophy of social phenomena" and help addressing "matters of great significance to several philosophical specialties – including ethics, epistemology, political philosophy, philosophy of science, and philosophy of law – and outside philosophy as well" (Gilbert 2013, p. 1).

In the opening contribution of this volume, "Joint commitment: what it is and why it matters", Gilbert argues that the appeal to joint commitment can be justified by way of the "rights

argument", that is, she argues that joint commitment is the most plausible source of the rights inherent in central social phenomena. This approach to the justification of joint commitment is important since it emphasizes that joint commitment has to be understood as the key-concept in the explanation of the normativity of social phenomena. In doing so, Gilbert also suggests that social normativity has not to be confused with moral normativity.

The other contributions collected in the present volume discuss, more or less directly, Gilbert's concept of joint commitment and point out its implications or even, more simply, its connections, both in a positive and negative perspective, with several crucial issues for the social ontology research agenda. Think of the issue of collective beliefs and the variable level of personal commitment and freedom they may imply, to the phenomenon of groups' intentionality and the different forms of collective mental states, acts and actions in it involved, to the problem of the extreme variety of normativity (not just moral vs. social normativity, but also different types of social normativity) and its relation with institutions, norms and laws, and also to the question of the ground of political obligation and political justice and their possible relation with joint commitments. Thus, we decided to organize the articles in the following four sessions: (i) Collective beliefs, (ii) Groups' intentionality, (iii) Shared norms, (iv) Just joint commitments.

3. Joint commitment in the social ontology debate

The contributions collected in Session 1, "Collective beliefs", deal with Gilbert's notion of group belief, with a special focus on the relation among individual and collective doxastic attitudes.

In his contribution, "*Augur augurem videns...* Belief and make-believe in social life", Wojciech Żełaniec focuses on the phenomenon of the "evanescence of the individual", which characterizes our age and which he detect in Gilbert's approach. Gilbert's discourse concerning collective belief is dominated by the question whether the parties to the joint commitment behave in a certain way. On the contrary, "it is precisely that evanescent question – of whether the parties to a joint commitment (have to) believe severally that which they are jointly committed to believe – that is most interesting in the context of the 'evanescence of the individual'". Żełaniec suggests that in the complete lack of individual beliefs (i.e., if people are only educated to internalize the beliefs they are jointly committed to believe) we can have at best "pretended beliefs" leading to something like a "collective schizophrenia".

Silvia Tossut's paper, "On acting *because* of a joint commitment", tackles the issue of the obliteration of the individual too, though she focuses on individual preferences. By using a game theoretical approach, Tossut shows that the Gilbert's holistic interpretation of joint commitment entails the obliteration of individual preferences, and that this makes it irrational to abandon whatever joint commitment.

Leo Townsend, in his paper on "Joint commitment and collective belief: a revisionary proposal", suggests a way out the debate between Gilbert and rejectionism (the thesis that gilbertean collective beliefs amount to acceptances). Townsend argues that Gilbert's account should be revised by saying that when people are jointly committed to believe, "the sort of commitment forged is a commitment to p as true, not a commitment to having the belief that p"; thus, he argues for "a broadening of the notion of joint commitment to include collective doxastic commitments".

Sonja Rinofner-Kreidl's contribution, "Epistemic Authority and Manipulation: Exploring the 'Dark Side' of Social Agency", is not directly concerned with collective beliefs, but tackles the related issue of the way in which people can *manipulate* each other beliefs. The author introduces the notion of "cognitive dissonance", and analyzes an interesting case of manipulation, i.e. that of a beated woman. Rinofner-Kreidl's main concern is "to figure out

3.1 Joint commitment and collective beliefs

how the ideas of reason and (epistemic) authority are connected with the social constitution and self-understanding of the agents involved", and she argues for a distinction between objective and social authority.

3.2
Joint commitment and groups' intentionality

Session 2, "Groups' intentionality", deals with the relation between the groups and the individuals constituting them, in particular with the issue of the relation between genuine groups, which have joint or collective intentions, commitments, feelings, etc., and mere aggregates or sums of individuals.

Adopting (even if implicitly) a very phenomenological attitude, Jacob Heim, in his paper "Commitments in Groups and Commitments of Groups", takes seriously the existence of groups and focuses on the experience of groups' commitments that we can have in everyday life. Heim shows that such experience is not reducible to that of the sum of the commitments of individual group members. First, Heim distinguishes between the individual level of commitments, owed to the group members, and the group level of commitments, which "the group as a single body owes either to itself or to some third party". Second, Heim distinguishes between the *content* and the *holder* of a commitment: "even when individual-level and group-level commitments have the same content, they are understood to have different holders". These distinctions allow him to make sense of the experience of groups' intentionality, and in particular of groups' commitments, as such.

Francesca De Vecchi's contribution, "The Plural Subject Approach to Social Ontology and the Sharing Value issue", inquires the role played by values for the creation of social unity and groups and for their maintenance in existence. This is a topic quite neglected in social ontology and also in Gilbert's social ontology of plural subjects. In referring to Max Schelers' axiology, De Vecchi points out the need of a values account that adequately considers the contribution of shared and collective values for unifying and binding people together in stable groups.

Glenda Satne and Alessandro Salice focus on the phenomenon of "Helping behaviour and joint action in young children" and show that Warneken, Tomasello et alia's (2006) attempt to explain helping behavior in young children assuming the validity of Bratman's theory of shared intentions as the right explanation of joint actions, faces several problems. Satne and Salice suggest that instead of Bratman's idea of "weak" interdependence, Tomasello et alia should better adopt a "robust" idea of interdependence, according to which "individual intentions are not based on individual intentions but are themselves dependent on collective intentions". Such robust interdependence is adopted by those philosophers that describe the phenomenon of collective intentionality not as distributive, but rather as genuinely collective – that is, as the intentionality of a group, or of a *we*. Needless to say, among such authors, beside Raimo Tuomela and John R. Searle, there is, Margaret Gilbert with her account of joint commitment.

Gian Paolo Terravecchia's "A Phenomenology of Social Stances" develops a phenomenology of social stances, that is of intentional acts of taking a position about something, and aims to show that Gilbert's joint commitment account is grounded in just one or two of the several stances which may characterize collective acting. Terravecchia argues that joint commitment needs *accepting* or *assenting*, while there are at least other three stances to be discussed about the creation of collective phenomena. These are: *refusing* (or *rebelling against*), *suffering*, and *making something one's own*. Very interestingly, indeed, Terravecchia points out that Gilbert (as other social ontologists and scientists) tends to reduce the phenomenon of acting together to the phenomenon of joint acting, in the sense of acting in accordance with someone, and to the corresponding joint commitments. Yet, in some cases of collective acting, such as rebellion, people act together with others, without the implication of any joint commitment , and

without the presence of any common goal (at least, among opponents) or social obligation. Thus, Terravecchia concludes that if one aims to account of the phenomenon of collective acting, one should consider the several stances, that is, the complex reality, in which collective acting may be grounded.

The papers collected in the third session, "Shared Norms", discuss, more or less directly, a common, general claim: normativity is said in many ways, and the normativity of social or shared norms is a very multifaceted phenomenon.

3.3 Joint commitment and shared norms

Francesco Guala's paper tackles the problem of "The Normativity of Institutions" and argues against the main trend in social ontology (Gilbert included) to ground the normativity of institutions in collective intentionality. Guala's claim derives from the assumption that there are different kinds of normativity and that in social ontology normativity is rather considered just as a product of collective intentionality. Thus, Guala argues that "many social institutions do not rely on normative commitments engendered by a joint intention", "that there is a viable alternative theory of institutions, and that normativity plays a different role in this theory than the one it plays in the collective intentionality programme". Guala shows that institutions are not necessarily characterized by the normativity of joint commitment, and holds that they are, rather, sets of *rules in equilibrium*, whose main function is to indicate actions promoting coordination and cooperation. "Each rule codifies a behaviour – a set of actions – that solves a problem of coordination".

Joshua Keaton's "The Social Impact Theory of Law" suggests that Gilbert's work on the "normativity of joint commitment can help resolve the intractable debate on legal normativity in philosophy of law", going beyond both Mark Greenberg's recent call to eliminate the problem of legal normativity, by grounding it in moral facts, and Hart's thesis that moral and legal obligations should be held distinct. Keaton proposes what he calls the *Social Impact Theory* (SIT), a substantial variation on the theme on Greensberg's Moral Impact Theory (MIT), according to which the rules that institutions recognize, promulgate, enforce, and practice give rise to "legal obligations, which are just various descriptions and entailments of our pre-existing social commitments – legal normativity is just an expression of rational commitment to group decision-making". Therefore SIT, like MIT, "has no need to posit a unique type of normativity".

Seumas Miller's "Joint Political Rights and Obligations" proposes a very interesting taxonomy of rights and obligations in moral, social and political spheres. Miller distinguishes between *moral rights*, which are *natural rights* because one possesses them by virtue of properties one has *qua* human being (think for instance to the natural moral right not to be tortured that is based on the human beings capacity to suffer physical pain), and *institutional moral rights*, which do not pre-exist to social institutions, but, rather, presuppose institutions, such as the moral right to vote. Moreover, Miller distinguishes between *institutional moral rights and obligations* (e.g. "the right to vote and the right to stand for office embody the human right to autonomy in the institutional setting of the state") and *institutional rights and obligations that are not moral* (e.g. "the right to make the next move in a game of chess": such right is "entirely dependent on the rules of chess and does not entail any moral element in it"). Finally, Miller introduces the category of *joint moral rights and obligations*, argues that "political rights and obligations are in large part joint (moral and institutional) rights and obligations", and distinguishes them by *individual rights and obligations*. Very compellingly indeed, Miller's taxonomy of rights and obligations aims to show that political rights and obligations may be both moral and joint, since they may depend both on properties we have *qua* human beings and on agreements and commitments individuals have freely decided to assume. In other words, Miller' brings the moral normativity together with the social normativity in a very interesting way.

Lorenzo Passerini Glazel's contribution on "Shared Norms and Nomotrophic Behaviour" focuses on the phenomenon of *nomotrophic behaviour* (a variation on the theme of Amedeo G. Conte's *nomotropic* behaviour), described as the "behaviour which aims at the maintenance of a norm in the event of its infringement: it typically consists in a reaction to the (actual or possible) infringement of that norm". "The idea underlying the concept of nomotrophic behaviour is that a social norm that is repeatedly infringed with no reaction may slowly 'atrophy' and vanish (by 'desuetude'); and that its atrophy may be countered through different forms of nomotrophic behaviour". Very originally indeed, Passerini Glazel suggests to construe Gilbert's (2005) (and Devlin's (1959)) idea of "interventions in the lives of others when one thinks that the others' behaviour is wrong", and Niklas Luhmann (1972) concept of "reactions to disappointment of normative expectations" as cases of nomotrophic behaviour. According to Passerini Glazel, both rebukes and demands, in Gilbert's sense, are forms of nomotrophic behaviour. Moreover, he points out the fruitfulness of the concept of nomotrophic behaviour for social ontology in elucidating the epistemological and ontological implications of such concept concerning, respectively, the questions of the *inference of norms* from action and of the *existence of norms.*

3.4 Joint commitment and justice

The fourth session, "Just Joint Commitments", collects contributions dealing with the political consequences of joint commitment (where 'political' should be understood in a broad sense). All the authors of these papers deal with the ways in which joint commitment might be used to explain specific features of our societies and our living in society.

John Horton and Ryan Windeknecht hold that we do have associative political obligations, and argue against the thesis that such obligations are subordinate to principles of global redistributions (the "distributive objection"). The authors show that the two both associative obligation and principles of global redistribution are morally relevant and that only the more extreme claims made by the proponents of the distributive objection should be rejected.

Roberta Sala builds on the realist claim that there are people within liberal-democratic societies whose lives are not inspired by liberal values, and argues that joint commitment is an answer to the question of how those people may not coercively adhere to liberal institutions. Sala suggests that joint commitment can help in explain the notion of 'modus vivendi', if we understand the latter as "a way to be involved (and not coerced) in a social enterprise, more or less willingly, generally aiming at a peaceful coexistence".

In the last paper, Helen Lauer analyzes an important kind of collective endeavour: collusion. The author emphasizes the characteristics of collusion that illuminate some limits of Gilbert's account of joint commitment. For example, the fact that it is not rational for colluding agents to mutually express their readiness to collusion. Lauer suggestion is that there are covert norms and that collusion may be a rational response to prevailing covert norms; if we recognize this feature of the social environment, we have the possibility to understand (and correct?) "socially acceptable hypocrisy".

Thanks to the efforts of all the authors, the volume as whole results in a very intensive and stimulating discussion on the joint commitment concept and on other similar and near concepts which, together with the "joint commitment", constitute a conceptual map of the main interesting problems faced today by social ontology.

JOINT COMMITMENT

JOINT COMMITMENT

Margaret Gilbert
Joint commitment: What It Is and Why It Matters

MARGARET GILBERT
University of California, Irvine
margaret.gilbert@uci.edu

JOINT COMMITMENT: WHAT IT IS AND WHY IT MATTERS*

abstract

There is reason to think that a particular concept of joint commitment informs much human behavior. This paper introduces the concept in question and briefly develops one argument for its centrality in human life. This argument focuses on the inherence in many central social phenomena of a particular kind of right.

keywords

agreements, joint commitment, rights, social phenomena

Introduction   I have long conjectured that the concept of a joint commitment, as I understand it, is a fundamental part of human life.[1] It informs much human behavior, and helps to explain much of that behavior.

This conjecture results from a number of independent investigations of specific everyday concepts.[2] I have argued that central everyday concepts of *acting together*, of *social conventions* and *rules*, and of *an agreement*, among others, have the concept of a joint commitment at their core. I have argued that other good candidates for a joint commitment analysis include central everyday concepts of *collective intention*, *belief*, *valuing*, and *emotion*, and *shared attention*. Thus I have come to believe that the concept of a joint commitment is a fundamental part of the conceptual scheme in terms of which human beings approach one another and, in doing so, construct their social world.

The reference to *construction* is important. I take it that armed with *the concept* of a joint commitment *people jointly commit one another* in various ways, and these joint commitments lie at the heart of our acting together, social conventions, and so on.

Two kinds of concern have been expressed in relation to my previous discussions of the role of joint commitment. The first concern relates to my technical phrase "plural subject". I have used this to refer to any two or more people who are jointly committed with one another in some way. That is: by definition, if and only if certain persons are jointly committed with one another in some way they constitute a *plural subject*. It is important to emphasize that this is a *definition* of the phrase in question. In using it I have not meant to suggest that when a plural subject in my sense is created – *that is*, when a number of people jointly commit one another – *a new center of consciousness* arises – insofar as I understand what it would be for this to happen. Rather, my use of the phrase "plural subject" in its intended meaning reflects my conjecture that an important class of vernacular sentences refer to two or more people who are jointly committed in some way. I have in mind everyday ascriptions of one or more actions –in a broad sense – to a plurality of persons, as in the statement "We want to win", where this is not elliptical for "We both (or we all) want to win", and the statement "We believe that it is going

---

1 I first invoked joint commitment in Gilbert 1989 (p. 198). Gilbert 1987, written later, refers to people being "committed as a body".

2 Gilbert 1996 (pp. 7-15), notes my increasing emphasis on joint commitment from 1989 till 1996. Discussions that further develop and apply the idea of joint commitment include Gilbert 2006 and 2013.

to rain" where this is not elliptical for "We both (or we all) believe that it is going to rain". The label "plural subject" can be used, as defined, without endorsing that conjecture, and that conjecture should be considered on its own merits, regardless of the words in which it is couched.

Turning now to the second concern, some theorists have been loth invoke joint commitment in social theory because in doing so one goes beyond what I have called the *singularist* conceptual scheme: one that encompasses only the personal intentions, beliefs and so on of individual human beings. Part of the idea is this: social theorists already operate with the singularist scheme in describing the actions of individual persons. They should, therefore, try to avoid going beyond it if they can. This concern conjures Ockham's razor. According to that dictum, we are told not to multiply entities *beyond necessity.* So the question is: do we *need* to go beyond the singularist scheme in social theory? Note that if we *do*, then even Ockham would approve our doing so.

A preference for singularism may also sometimes be defended by saying that the notion of a joint commitment is *mysterious* – in contrast with familiar singularist notions. I hope that the discussion that follows will show that this is not so.

I believe that the notion of joint commitment is a philosophically respectable one, and that we need to employ it if we are to give a comprehensive account of the social world. In the rest of this discussion I start by placing joint commitment, as I understand it, within a framework that subsumes some familiar singularist notions. It is hard to see how, if these singularist notions are considered unexceptionable, exception can be taken to the notion of joint commitment. I then justify an appeal to joint commitment in many contexts in light of an argument I call "the rights argument".

**1.
Joint commitment**

According to the broad notion of *commitment* at issue in joint commitment one's commitments as such, are normative constraints on one's behavior that derive from one's own past or present states, actions, and so on.

Among one's commitments thus broadly defined, some have their immediate source in one or more human wills. In that sense, then, they are "commitments of the will". I shall in what follows construe "commitments of the will" in this source-related sense.

How can the human will normatively constrain someone? A good place to start considering this question is by reflecting on the familiar case of an individual person's decision, such as my decision to go to Milan. Personal decisions are typically expressed (if only in private rumination) by such words as "I'll go to Milan!".

What difference does one's decision make to one's situation, normatively speaking? That is a good question and there is now a considerable literature that bears on it. I shall focus on one aspect of that question and discuss it in my own terms. Consider the following case. Sophia decides to have lunch at Cafe Rosso. Later she absent-mindedly walks in the other direction. Realizing this, she presses her hand to her brow and says, critically, "Oh, I meant to go to Cafe Rosso!". I take Sophia to be responding here to the following aspect of the normativity of decision: all else being equal, at least, one acts in error if one acts contrary to a standing decision. I take this to be true even if the decision was made on a whim and has nothing in particular to be said in its favor. It was her decision and she had not changed her mind. There is more to be said about the normativity of decision but here I rest with the minimal point just made, elaborated as follows: I say that one who has made a decision has *sufficient reason* to act in conformity to it, where this means that if he is appropriately responsive to applicable considerations of whatever kind he will so act, all else being equal. Such a person *ought* then so to act – in a generic sense of "ought". Perhaps it can be called the *rational* ought.

Insofar as nothing beyond a personal decision as such is necessary to create the space for

an error, we may say that a personal decision *imposes on the decision-maker* a specific type of *normative constraint* with respect to his future actions. I do not mean to imply that the decision-maker brings it about that, once he has decided to perform some act, he has sufficient reason to do so. Rather, given that normative fact, *he only has to decide* in order to have sufficient reason to act accordingly. In this way the decision-maker *commits himself* to act in a certain way.

A personal decision effects what I refer to as a *personal* commitment of the will. Such a commitment obtains by virtue purely of an act or state of the will of the individual in question. There is no need to make the decision public, to communicate it to others, or anything of that nature, in order to commit oneself. I take it that the broad notion of commitment at issue here is not a specifically *moral* notion – unless one operates with a particularly broad notion of morality. One might talk of someone being *morally committed* in the following kind of context. I have regularly been offering food to a homeless person who sits on the street corner near my house. Though I have made no promise and, for that matter, no decision to continue to do so on a regular basis, I have set up expectations, the person now relies on my help. Perhaps *by virtue of what I have done* I now have sufficient reason to continue, and may then be said to be committed to do so. My commitment here, if there is one, is not a commitment of my will, in the sense introduced above. Rather it is a commitment deriving from moral considerations given what I have done. It is, in other terms, a matter of *moral requirement.* Though my will was involved in these past doings, I did not will *the continuation of my help*, as I might have done.

Returning to personal commitments through personal decisions, an important further point is as follows. I unilaterally *make* and can unilaterally *rescind* my personal decision. Without rescission my decision stands until the time for carrying it out is past, and throughout this time I am personally committed to act accordingly. Evidently, then, the commitment associated with a personal decision has a degree of stability, but it is easy to rid oneself of it by, as we say, *changing one's mind.* As we shall see, from a normative point of view a joint commitment is a better means of pinning oneself – or another – down with respect to one's future actions than is a personal decision.

I turn now to joint commitment. At this point in my discussion *virtually no preliminaries are needed before introducing the basic idea.* This should help to defuse a preference for singularism, in that it shows how little the invocation of joint commitment moves us beyond the singularist conceptual scheme of personal decisions and so on.

A joint commitment is *a commitment of the will.* In a way that will be explained, the wills of *two or more people* impose the commitment on the same two or more people – as one.

Clearly, as with the personal commitments that come from personal decisions, a joint commitment in this sense involves both a particular *process* and a particular *product.* In both cases the process is in broad terms psychological; the product is normative.

In the case of the process of joint commitment, there may be different ways in which two or more people come to be *committed as one.* Thus suppose someone is drowning and can only be rescued if two particular people jointly develop an action plan. It may then be plausible to say that there is now a moral requirement on *them*, jointly to develop such a plan. If so, these people will not be jointly committed *in the sense at issue here*. In order to be jointly committed in that sense, the commitment of the parties *must be imposed by their wills* – in the way shortly to be discussed.

In the case of the product of the relevant process of joint commitment, various alternative possibilities suggest themselves. For instance, it may be that two or more people can jointly impose a commitment on just one of their number, or on some proper subset of their number as one.

I mention these possibilities only to clarify the sense of joint commitment now at issue. Joint commitment in this sense involves *both* a joint process *and* a joint product. The process, in this

case, involves an exercise of will on the part of two or more people; the product is a normative constraint applying to the same two or more people as one.

I focus here on *basic* cases of joint commitment. Here there is no background joint commitment that allows for the creation of a new joint commitment of them all by a single party or proper subset of the parties. I allow, of course, for non-basic cases in which there *is* such a background joint commitment, as when Jane has said to her husband, Tom "You decide where we go" and he responds "Okay!".

In the basic case, on which I focus here, matching expressions of personal readiness to co-create a particular joint commitment are necessary to create that joint commitment.[3] If these expressions are made *openly* and their having been so made is *common knowledge* among the parties, that is sufficient for the creation of the pertinent joint commitment.[4]

This is not a great deal to require for joint commitment. It is, if you like, the *natural social analogue* of a personal intention or, where the expressions are of the right, explicit kind, a personal decision. These can of course be formed within the confines of an individual's mind, and need not be made public. It is natural to suppose, meanwhile, that the social analogue of a personal intention or decision will involve some form of communication between the parties along the lines suggested here.

Suppose, then, that Bob is personally ready to be jointly committed in a particular way with Lily and does his part in the co-creation of such a commitment. That is, he openly expresses his readiness to co-create it in conditions of common knowledge. He understands that the relevant joint commitment will come into being if and only if Lily similarly does her part, and there is common knowledge between them that this has occurred. In other words, he understands that *if and only if a certain condition is satisfied* he and Lily will jointly have committed each other in the way in question. That condition is Lily's corresponding expression of readiness. Their expressions, meanwhile, need not be *conditional in form.*[5]

It is easy to construe familiar types of everyday interaction in terms of this account of joint commitment formation. Suppose that, in conditions of common knowledge between Bob and Lily, Bob says to Lily "Shall we dance?" and Lily responds "Yes! Let's". Bob can be construed as openly expressing his readiness to enter a joint commitment with Lily in favor of their dancing together and Lily's "Yes! Let's" can be construed likewise. As they both understand, nothing remains to be done in order to create the relevant joint commitment. At this point, then, they are jointly committed. I say something about why one *should* construe such interactions in terms of joint commitment in due course.

To fill out the previous discussion somewhat I now make some important further points about joint commitment as I understand it.

A joint commitment is not rescindable by one party unilaterally, but only by the parties together. In some cases there may be special background understandings or explicit preliminaries that allow, in effect, for one person to rescind the commitment. The situation described here is the "default" situation.[6]

People are jointly committed to act as a body in a specified way, where "acting" is taken in a very broad sense. Thus people may jointly commit to *accept as a body a certain goal*, or *plan,* to

---

3 Possibly there are special cases that do not require expressions of readiness jointly to commit in the way in question. For discussion see Gilbert, forthcoming.

4 See Gilbert 1989, Ch. 4 for discussion of the terms and conditions involved here.

5 The phrasing of my initial discussion in Gilbert 1989 has suggested the opposite to some commentators, though I believe that was never my intention. See Gilbert 2013, Ch. 2 for discussion.

6 The *two-person case* has some special features with respect to rescission. See e.g. Gilbert 1996, pp. 14-16; and pp. 381-383.

believe as a body a certain proposition, to *value some particular thing*, and so on. The phrase *as a body* is not sacrosanct, nor is it supposed to be more than a place-holder for a longer description. Take the case of jointly committing to *believe that such-and-such as a body*. One way to describe the content of this joint commitment more expansively is this: to *emulate a single subject of belief by virtue of their several actions and utterances.* Thus, if Betty and Joe are jointly committed to believe as a body that Joe is a brilliant mathematician they must – if they are to fulfill the commitment – talk the appropriate talk, and so on, at least in conversation with each other. Thus Betty must not react with amazement or utter a derogatory comment if Joe tells her he is about to prove an important theorem.

When there is a joint commitment of two or more people, each of the parties is committed, that is, normatively constrained. One may therefore speak of the associated "individual commitments" of the parties. As to the content of these individual commitments, each is presumably committed to promoting the fulfillment of the joint commitment to the best of his or her ability *in conjunction with the actions of the other party or parties.* Depending on the case, this may require careful efforts to coordinate behavior. Given their existence through the joint commitment, these "individual commitments" are not *personal* commitments in that they are not the unilateral creation of the respective persons, and they cannot be unilaterally rescinded.

2.
Why invoke joint commitment?

Why think of joint commitment as a fundamental part of human life? In what follows I focus on one argument for a positive answer. I shall refer to it as *the rights argument.* It starts from the premise that rights of a certain kind inhere in many central social phenomena, and argues that not only is joint commitment a source of rights of the relevant kind, it is the most plausible source of the rights inherent in the phenomena in question.

The premise from which the argument starts requires some initial amplification. It refers to "many central social phenomena". I would include among these, the phenomena mentioned in the *Introduction* – acting together; social conventions and rules, agreements, collective intentions, beliefs, values and emotions; and shared attention.

This premise maintains that rights of a certain kind "inhere" in each of these phenomena. By this I mean that there cannot be a collective belief, say, whose parties lack rights of the relevant kind. A given party can, of course, *waive* his right in the sense of determining not to enforce it, but the right remains as long as the collective belief does. If this is so, a satisfactory account of any one of these phenomena must posit *at its core* a source of rights and correlative obligations.[7] In other terms, relevant rights must come from something in their constitution.

What is the "particular kind" of right in question? It is a particularly important kind of right. In brief: to have such a right to a particular action of a particular agent is to have the standing or authority to demand that action from that agent, and the standing to issue appropriate rebukes to that agent should the action not be performed. In what follows I shall use the label "demand-rights" for rights of this kind.

It may be observed that people often use the term "demand" in a broad sense. Thus a robber with a gun who says "Give me your money!" may be said to have demanded that you hand over your money. Evidently I am using "demand" in a different sense such that, presumably, this robber lacks the standing to demand that you hand over your money.

---

7 For reasons of space, I focus on the rights side of the equation in this discussion, as opposed to the obligation side. In brief, for X to have a right (of the kind in question) to an action of Y's is for Y has a obligation (sometimes called a duty) *to X* (appropriately construed) to perform that action. For detailed discussion of this "equivalence" see Gilbert 2012 and forthcoming.

It is important to distinguish standing from justification. One may have a demand-right to some action yet not be justified, in the circumstances, in demanding the object of one's right. The rights argument goes on to contend that joint commitment is a source of demand-rights. Is this so? Even without an extended discussion a positive answer is compelling. Intuitively, those who are jointly committed in some way do have the standing to demand of any party that he (or she) conform to the joint commitment, should he threaten not to do so, and to rebuke one another for any failure to conform. Thus suppose Anna says "Don't walk away!" to Marco. She could plausibly meet a challenge from Marco to her standing to make this demand, by appeal to a pertinent joint commitment. Spelling things out she could say, for instance, "I joined you in jointly committing us to wait here as a body till Roberta arrived. That puts me in a position to call you to order!".[8]

In the final step towards its conclusion – that joint commitment is a fundamental part of human life – the rights argument maintains that the demand-rights inherent in the central social phenomena in question are most plausibly explained by reference a constitutive joint commitment. There is no space to argue this with respect to each of these phenomena here. I shall argue for a more modest point: at least one of these central social phenomena is most plausibly explained by reference to a constitutive joint commitment. Though more modest, this point alone is enough to get us to the conclusion of the rights argument.

I shall focus on everyday bilateral agreements – whose centrality in human life is clear given that it is hard to imagine a human society that does not involve the making of multiple agreements on a daily basis.[9] I argue, first, that demand-rights inhere in agreements, and, second, that a constitutive joint commitment is the most plausible source of the demand-rights of agreement. Due to space limitations this argument must be briefer than it might be. Arguing the first point is easy, insofar as agreements are generally accepted as canonical sources of demand-rights, whether or not they are so called. Further, it can be argued that the connection between agreements and demand-rights is a tight one. If an agreement was made, demand-rights of the parties to action that conforms to the agreement are there. The nature of the actions agreed-upon, for instance, is not a factor, though it will be a factor in what one ought to do, all things considered. The same goes for the circumstances in which the agreement was made: as long as it *was* made.[10]

An account of agreements that incorporates a constitutive joint commitment suggests itself. An agreement will then be just what it seems to be: a phenomenon from whose mere existence one can immediately infer that the parties have demand-rights against one another to action that conforms to the agreement. Indeed, there is reason to think that the demand-rights of agreements are joint commitment rights whether or not there is another source of such rights. For, as I indicated earlier, it is easy to construe situations in which people make agreements as their making a joint commitment.

What would be the content of an agreement-constituting joint commitment? I have argued elsewhere for an account of agreement as a joint commitment *to endorse as a body a given plan of action*, such as the plan that Anna and Marco will stay where they are till six o'clock.[11] This plausibly aligns interpersonal agreements with personal decisions, which can be seen as involving the endorsement of a particular plan of action for the person in question.

The foregoing makes it clear that there is a good case to be made for a joint commitment

8 For further discussion see Gilbert 2012 and forthcoming.
9 Cf. Feinberg 1970.
10 For extensive discussion of immoral and coerced agreements in relation to this claim, see Gilbert 2006, Ch. 10.
11 Many such plans will require some action of both parties, though some may not. For further discussion relating to this point see Gilbert 2013, Ch. 13.

account of agreements.[12] I now argue briefly that it is doubtful that a better case can be made for another type of account.

First, it is not at all clear that anything other than a joint commitment or something very like it is a source of demand-rights. It is not possible to give this point an extended treatment here.[13] For present purposes, however, it may be helpful to note the following. Perhaps it will be argued that, given a bilateral agreement, it is clear to all of the parties that each of them will expect – or, perhaps, rely on – the other to act in accordance with the agreement, unless he says that acting otherwise is fine with him. Perhaps it will then be said that, in that case, the parties are morally required to act as expected, unless the relevant word has been given. Supposing that all this is so, the existence of a demand-right in either party to the other's compliance with the agreement can be questioned.[14] Even if it were allowed, given that the expectations in question are only likely concomitants of agreements rather than part of their constitution, they will not be associated with demand-rights that inhere in agreements themselves.

In referring, above, to something very like joint commitment I have in mind the situation envisaged earlier in this article which involves the same process but a different product. As a way of interpreting agreement formation I take it that a constitutive joint commitment is a more plausible source of the demand-rights of agreements.

In light of the considerations so far adduced it is reasonable to conclude that joint commitment is the most plausible source of the demand-rights inherent in agreements, one of the central social phenomena adduced by the rights argument. For this reason alone it is reasonable to conclude that joint commitment is a fundamental part of human life. This conclusion can only be strengthened if the point can be made in relation to one or more of the other central social phenomena adduced in the rights argument. Though there is no space to argue that here, I believe it can be made with respect to all of them, and more.[15]

3.
Concluding
remarks

Suppose now that joint commitment is a fundamental part of human life, that people are constantly co-creating joint commitments, whether in the making of agreements, or in a larger range of contexts. Given that they have at least some tendency to act as is appropriate all things considered this is likely to be the explanation of much of their behavior, including their treatment of other people. The social sciences, then, need to pay attention to the concept of joint commitment. So also does the theory of rights. All this suggests that we risk a significant impoverishment of our understanding of the human world if we restrict our accounts of social phenomena to a singularist conceptual scheme.

REFERENCES

Feinberg, J. (1970), "The nature and value of rights", *Journal of Value Inquiry*, 4, pp. 243-260;

Gilbert, M. (1987), "Modeling collective belief", *Synthese*, 73, pp. 185-204;

----. (1989), *On Social Facts*, Routledge and Kegan Paul;

----. (1996), *Living Together*, Rowman and Littlefield, Lanham, MD;

----. (2006), *A Theory of Political Obligation*, Oxford University Press, Oxford;

---

12 For further discussion of the plausibility of a joint commitment account of agreements, in addition to what follows below, see Gilbert 2013, Ch. 13, and 2014. I do not mean to imply, here or elsewhere, that in order to make a joint commitment one must make what amounts to an agreement. I am only suggesting the reverse.

13 There is such a treatment in Gilbert forthcoming.

14 See, in particular, Gilbert 2013, Ch. 12 (orig. 2004) with reference to the Scanlon's principle of fidelity. See also Gilbert forthcoming.

15 For further discussion see Gilbert 2013 and references therein.

----. (2012), "The giving claim-rights their due", in B. Bix and H. Spector (eds.), *Rights*, Ashgate, Farnham, pp. 301-323;

----. (2013), *Joint Commitment*, Oxford University Press, New York;

----. (2014), "The nature of agreements" in *Rational and Social Agency*, M. Vargas and G. Yaffe (eds.), Oxford University Press, New York, pp. 215-256;

----. (forthcoming), *Rights and Demands: A Foundational Inquiry*, Oxford University Press, Oxford.

SESSION 1

SESSION 1

COLLECTIVE BELIEFS

WOJCIECH ŻEŁANIEC
University of Gdańsk
zelaniec@aol.com

AUGUR AUGUREM VIDENS... BELIEF AND MAKE-BELIEVE IN SOCIAL LIFE

abstract

Lévy-Bruhl thought that the primitive mind could not really distinguish itself from the 'collective mind' of the community it was immersed in; the post-modern mind, by contrast, though arguably no less forcibly pressed into the Procrustean bed of various collective beliefs, is very well in the position to dissociate itself from the 'us' whose position it sometimes pretends to represent. Are such beliefs, then, really anyone's beliefs or are they merely 'make-believes'? Though Gilbert in her 'Joint commitment' does not explicitly address this question, I try to reconstruct, from the text of her book, her possible answer to it.

Vetus autem illud Catonis admodum scitum est,
qui mirari se aiebat, quod non rideret haruspex, haruspicem cum vidisset.
Quota enim quaeque res evenit praedicta ab istis?
aut, si evenit quippiam, quid adferri potest, cur non casu id evenerit?
Marcus Tullius Cicero, *De divinatione*, II, 24, 51f.

(But indeed, that was quite a clever remark which Cato made many years ago:
"I wonder," said he, "that a soothsayer doesn't laugh when he sees another soothsayer."
For how many things predicted by them really come true?
If any do come true, then what reason can be advanced
why the agreement of the event with the prophecy was not due to chance?
English by William Armistead Falconer)

Margaret Gilbert's *Joint commitment* (2014) is an impressive piece of contemporary social philosophy. One of its many merits is that it brings to the fore not just perennial truths about social action but, also, certain characteristic traits of the present age. One of such traits is what I should like to call the "evanescence of the individual".
More accurately this should, perhaps, be called the demise, or, better still, the obliteration of the individual, except that it is not always easy to say who the obliterating agent is.
Take, by way of an example, a matter as harmless as class syllabi to be prepared by the instructor for her students at a university. A class syllabus is no doubt a very useful thing for both parties involved, and it makes sense that it should contain such fields as "expected teaching effects" or "social skills acquired during class".[1] In a course on contemporary social philosophy, getting introduced to Gilbert's thought would be one of the most important "teaching effects" to expect. But at Gdańsk, Poland, instructors are discouraged, by the anonymous authors of the respective electronic forms, from giving their own wording to whatever they fill such fields out with, and required, instead, to use standardized, predefined expressions and quasi-algebraic symbols like "J23", or "B52", mysteriously supposed to be

---

1 These titles do not sound very idiomatic in English, as they are translated from Polish, see further on above.

comprehensible to the students.[2] (Thanks Goodness, instructors are still allowed to write their books and papers in their own words, rather than in prefabricated building blocks; but for how long yet?...). It is not known who ever invented such constraints: the University authorities, the Polish government, the European Commission, some other, anonymous and elusive, bureaucrats? Probably not any mere Ortega y Gasset's "masses" since long prevailing over "that single individual" whom Kierkegaard championed a century before Ortega y Gasset and whom he was wont "with joy and gratitude [to] call [his] reader" (Kierkegaard 2009, p. 92). The issue is more redolent of Kafka's *Castle* than of *The revolt of the masses.*
However, the issue Gilbert broaches is far more serious than all that. If I express my beliefs in ready-made phrases, no matter how linguistically objectionable or stylistically maladroit I find them, the beliefs expressed are still *mine*, an *individual's.* Gilbert, by contrast, examines *collective* beliefs and she finds that

> X, Y, and so on, *collectively believe that p*, if and only if X, Y, et al. are jointly committed to believe as a body that p (Gilbert 2014, p. 71).[3]

An old-fashioned epistemologist could take exception[4] to the very idea that believing (an epistemic *state*),[5] collective or not, could be the matter of a *commitment,* normally, he would think, a prerogative of acts, actions, and perhaps dispositions to such. But for Gilbert it is precisely *collectively* (on the *definiendum* side of the above definition) and *jointly* (on the *definiens* side) that make the difference.
What, then, is it for a "population",[6] or simply a group of human individuals, to be "jointly committed to believe as a body that p"? This is, clearly, *the* question to which Gilbert's book as a whole is an answer – and it is not within the purview of this article to deal with *it.* But to start from the basics, right in the Introduction Gilbert tells us that

> [f]or any joint commitment, the parties are jointly committed *to phi as a body*, where "phi" stands for the relevant verb. [That is:] they are jointly committed to emulate, *by virtue of the actions of all,* a single phi-er. For example: they are jointly committed to emulate, by virtue of the actions of each, a single believer of the proposition that justice is the first virtue of institutions (Gilbert 2014, p. 7).

If the above were to be taken as a definition, the objection of circularity would spring to mind, "jointly committed" occurring on both sides of "that is".[7] However, jointly committing oneself

---

2 Such problems, in all fairness to Gdańsk University instructors, appear idyllic as compared with the plight of a growing number of real-life job seekers whose cv's are read first and, in most cases, *only*, by computers running "applicant tracking system" software.
3 Page references to the aforementioned edition. See, too, p. 137 ("The members of a population, P, collectively believe that p if and only if they are jointly committed to believe that p as a body"), and p. 173, where an essentially equivalent definition is given ("*A population, P, believes that p* if and only if the members of P are jointly committed to believe as a body that p"). With respect to belief, "one might say either, 'Population P believes that p,' or, equivalently, 'The members of population P (collectively) believe that p", Gilbert explains (p. 165). On p. 192 Gilbert applies the same idea to axiological beliefs ("The members of a population P *share value V (in the sense they share the belief that item I has certain value)* if and only if the members of P are jointly committed to believe that as a body that item I has that value").
4 Gilbert comes to terms with this on pp. 153f.
5 Or a cognitive one, as Gilbert prefers to call it (p. 163).
6 See p. 165.
7 See p. 141 where Gilbert successfully, to this writer, disposes of another objection of circularity, that pertaining to "believe".

being a human matter, it is impossible to do so without having an ever so ingenuous idea of what one is doing (i.e., precisely, jointly committing oneself) in mind beforehand.[8] As Gilbert carefully explains:

> [If Andrea and Heinrich have a joint commitment, it is because] Heinrich expresses to Andrea his readiness to be jointly committed to espouse the relevant goal as a body, in conditions of common knowledge, and Andrea does likewise. The joint commitment is then in place.
> Note that in order to enter a joint commitment with another person *one must know* [italics W.Ż.] what a joint commitment is. One could not otherwise express one's readiness to enter such a commitment. This does not mean, of course, that the phrase "joint commitment" is part of one's vocabulary (Gilbert 2014, p. 32).

and still less does one need to know and subscribe to a *theory* of joint commitment as sophisticated as Gilbert's. Yet still, it remains that the joint commitment comes into being as soon as, and no sooner than, all parties have expressed their readiness to be jointly committed [and, probably, have taken note of each other's expressions of readiness][9] (*ibidem,* and p. 47f.). This is a hermeneutic, rather than vicious, circle, and the former is – it has been said –[10] victorious.

Concerning the concept of "expressing one's readiness to be jointly committed", Gilbert prudently does not engage in much further analysis:

> It is not clear that there is any very helpful way of breaking down the notion of expressing one's readiness to be jointly committed. It could be said that one makes it clear that *all is in order as far as one's own will is concerned* for the creation of the relevant joint commitment. Importantly, one understands that *a necessary condition of the creation of the joint commitment is corresponding contributions from the other parties* (Gilbert 2014, p. 48).

More problematic, in this writer's view, is another part of the above quasi-definition of "joint commitment", i.e. the part where Gilbert says that the parties involved "are jointly committed to emulate, *by virtue of the actions of all,* a single phi-er. [e.g.]: they are jointly committed to emulate, by virtue of the actions of each, a single believer of the proposition that justice is the first virtue of institutions" (Gilbert 2014, p. 7). What are the relevant *actions* of a believer like that in her capacity of a believer? A *single* believer – and Gilbert is probably right when she says that it is substantial for her theory that the parties to a joint commitment should emulate a *single* performer of whatever they are committed to –[11] first and foremost *believes* (which is not, strictly speaking, an action)[12], for instance that justice is the most valuable virtue of institutions. But parties to a joint commitment, not having quite literally a common mind,[13] cannot emulate that believer *as a body.*[14]Clearly, they can emulate her (the exemplary believer)

---

8 "The concept of [joint commitment] is implicit in everyday discourse" (p. 174) – not just discourse, also everyday practices.

9 "In conditions of common knowledge" (pp. 32, 154).

10 See Maddox (1983)

11 Otherwise her definition of joint commitment would be circular, see note 7.

12 But on pp. 32f. Gilbert tells us that she construes "doing something" – and, presumably, "action" – broadly, so as to include such psychological states as belief; see also p. 193.

13 Gilbert explicitly abjures the "scary monster" of a "group mind" (p. 119), cf. pp. 9f. and 135.

14 Alternatively: *as a unit* or *as one* (p. 33). The scope of "as a body" in Gilbert sometimes gives rise to doubts. Emulate-as-a-body or emulate a phi-er-as-a-body?

severally, each for one's own account, but that is clearly not what Gilbert is aiming at.
At this point (not any definite page in her book but strewn all through her work), Gilbert turns phenomenologist –[15] she starts working with examples, to no small part quite illustrative and convincing ones. This one is particularly clear:

> What is it to be jointly committed to x *as a body* [...]? [It is] to be jointly committed to bring it about as far as is possible that the parties emulate a single x-er. [...] Suppose that John and Doris are jointly committed to believe as a body that Doris, who is ill, will get well. What must each of them now do, in order that the joint commitment is fulfilled? It is not the case that each must personally believe that Doris will get well, something it would be hard if not impossible for either one deliberately to bring about. Rather, in the appropriate circumstances – in particular, when they are together – each is to act in such a way that together they emulate a single believer of the proposition that Doris will get well.

This we know already. But here comes the new part:

> Among other things, whatever he or she personally thinks, neither will in the other's presence baldly declaim that Doris will not recover (p. 348).

This is quite congruent with such explanations as this:

> [T]he joint commitment will be fulfilled [...] if those concerned say that p in appropriate contexts, with an appropriate degree of confidence, and do not call p or obvious corollaries into question. Their behavior generally should be *expressive of the belief that p*[16], in the appropriate contexts (p. 176)

or this one (suppose that the joint commitment is to believe, *as a body*, that an item has a certain value):

> This would be achieved by each of the committed parties doing such things as: confidently stating that item I has value v; refraining from calling this or its obvious corollaries into question; suggesting by actions and emotional expressions that item I has value v; not, therefore, acting contrary to the shared value, nor reporting such contrary actions with bravado. Thus, were each of the parties the mouthpiece or representative of a single person, *one would judge that single person to believe* [italics W.Ż.] that item I had value v (pp. 193f.).

There are more direct ways of creating that impression, for example, letting an actual mouthpiece or representative of a body say such things as "The United States believes that those responsible for these dreadful acts must be punished" (p. 131) or "The campus improvement committee believes that there needs to be a café on campus" (pp. 139, 146), or "The union believes that management is being unreasonable" (p. 164), or "We believe that

---

15 She mentions with irony "analytic philosophers, established purveyors of clarity and rigor" as those who have ignored collective belief (p. 132).
16 This is a critical juncture Gilbert passes by without ceremony: whose belief? An individual's or a collective's? Or generic, perhaps?

the conservation of species is an important goal" (p. 172). But the problem is that while the mouthpiece or representative can be, and more often than not is, actual, "that single person" (the one body as which the parties to the joint commitment try to act) cannot – it is to remain merely fictional, or at best hypothetical (but against this see Christian List's contribution to SRSSP, as well as his and Philip Pettit's publications on the topic). It is for this reason, perhaps, that the question whether the single persons that there actually are, i.e. the parties to the joint commitment, for their own part do or do not believe that item I has value v (or whatever they are jointly committed to believe) must recede and is overshadowed by the question if the parties to the joint commitment *behave* (p. 141) in a certain way, such as "refrain [...] from calling [the proposition they are jointly committed to believe] or its obvious corollaries into question" (p. 193), or from openly, "without preamble",[17] denying it (p. 137). Such behaviour lays one open to rebuke from the other parties to the joint commitment concerned, the "standing to demand conforming [to the joint commitment in question] and to rebuke for non-conformity" (p. 8) being one of the best-explained[18] and most often recurrent elements of Gilbert's theory[19] – and finally even to a kind of ostracism as an "outsider" and no longer "one of us" (p.177, cf. pp. 120, 172).

Yet it is precisely that evanescent question – of whether the parties to a joint commitment (have to) believe severally that which they are jointly committed to believe – that is most interesting in the context of the "evanescence of the individual", mentioned early on in this essay. The way Gilbert addresses this question (and she does so quite often in her book) shows (in Wittgenstein's sense) how less and less important it is getting.

In the Introduction she says clearly: "I do argue that [...] when we[20] [collectively] believe something no one of us needs to believe that thing" and provides a footnote mentioning chapter 7 of her book (p. 9) as the relevant reference; but in chapter 7 itself she is less bold and unambiguous as far as the "no one" part of above clear statement is concerned.

For instance, arguing for a "no" as an answer to the question whether *most*[21] [italics W.Ż.] members of a group must believe that such-and-such in order for it be true that they collectively believe it, Gilbert employs the example of a poetry discussion group thinking

---

17 A "preamble" could be something like "We believe collectively that item I has value v, but speaking personally and for myself, I should say..." see e.g. p. 171.

18 But also, having a considerable explanatory power within Gilbert's theory of collective belief. That is, a substantial part of what (on this theory) collective belief is, is "something that gives the parties to it the standing to rebuke one another for certain types of behaviour challenging that belief," see e.g. p. 172, where this type of behaviour is described as "bluntly [i.e. without a preamble like 'speaking just for myself'?] expressing a view contrary to [what we collectively believe]".

19 The present writer cannot help wondering if this element of Gilbert's theory is not applicable to some cultures more and to some others less, for instance, if it is not "truer" in application to cultures influenced by the diverse versions of Protestantism. Against this seems to militate the circumstance that Gilbert precisely for this element of her doctrine at crucial places invokes the authority of Emile Durkheim (on whom she fathers collective epistemology in general, pp. 166f.) and his theory of *faits sociaux* (e.g. pp. 132, 177). Be it as it may, this hypothetical cultural relativity of the "standing to rebuke" seems to escape Gilbert's notice. What does not, is the important fact that despite the quasi-moral flavour of the rebuke here intended, a joint commitment can be to morally reprehensible acts (pp. 121f.) – and thus there is no (?) *prima facie* moral obligation to act on a joint commitment *qua* such.

20 English works with "we" in such contexts, a pronoun with strong moral overtones (see preceding note) whereas other languages have much less engaging "collective subject" pronouns or constructions: German "*man*", French "*on*", Italian "*si*", Polish "*się*". These are rendered as "you", "one", or "they" in English, but none of these is exact. See Ortega y Gasset (1981, p. 15) on the Spanish "*se*."

21 It is instructive to see how, in Gilbert's text – and this is no "rebuke" directed at her text, for which this author has no standing, but rather a praise for how faithfully she is therein representing certain tendencies of our age – the classical quantifiers "all" and "none" get sometimes, quite frequently in fact, replaced with multal and paucal quantifiers, such as "most". "All" and "none", however, take individuals more seriously than "most".

collectively that a poem is a powerful one. In this group, she says, there is *just one* member who does not find that poem all that good, and even this person nodded her head in approval at the critical moment of the group's making up their minds, thereby expressing her readiness to see the view she personally disagrees with as the established view of the group (pp. 168f.). A group where "all" less one is no longer "most" has two members, by the way.

But then, as if heedless of her cautious quantifying with the multal "most", she goes on to assert: "In this informal type of case [...] there may be a collective belief that p without all or most – *or indeed any* [italics W.Ż.] – members of the population in question believing that p" (p.169). This is, indeed, a "stark", as she herself puts it, conclusion. But if the matter is any different in more formal kinds of cases she does not make quite explicit. Here is a seminar on human rights – a formal entity, presumably – which enters a joint commitment to believe that the notion of group rights is viable. The members of the seminar are under the obligation, derivative of that joint commitment, "to express that belief at least within the confines of the seminar when it is in session" (p.176). But when they are "at large", and a friend approaches one of them with his doubts concerning group rights, the member of the seminar in question is ("presumably", as Gilbert cautiously puts it) free to give him her mind even without a preamble like "personally speaking..." (p.177). Now since there had been no mention of "her" before and "she" is entirely arbitrary, we are entitled, by Universal Generalisation, familiar to logicians, to conclude that *all* parties to the seminar and its concomitant joint commitment may, personally speaking, dissent. Is this what Gilbert means?

A similar impression arises on the occasion of Rose, who, while acknowledging "that she and her friends collectively believe a certain thing [adds] without a sense of fault, that she, *personally,* does not believe it" (p.140). Again, since Rose seems to be an arbitrary selected member of a gang of friends collectively believing in something-or-other, the question arises whether we may apply Universal Generalisation, and conclude that in fact none of these friends personally believes, or needs to believe, that thing.

There are, too, different "populations" comprising the very same individuals, for instance a court of law and a poetry reading group, that may have (collectively) two different opinions on the merits of a certain poem (read for enjoyment by the latter population, and in the context of a legal action by the former one) (pp. 170f.) – again, need it be the case, we may ask, that nearly half of the individuals involved personally hold the one opinion on the poem and another half the other, or is another distribution conceivable, for instance, that one of the opinions is held by no one personally? Something like this is conceivable, to Gilbert, for collective intentions: Olive and Ned may collectively intend to climb a hill, while neither of them has the corresponding individual "contributory intention", as Gilbert calls it, i.e. neither intends "personally" to climb all the way to the top – maybe as a result of the first difficulties (pp. 103f.). "[A]s a *conceptual matter* [italics W.Ż.] when two or more people share an intention, none of them need to have a contributory intention" (p. 103) – does the plural form of the verb "to need" indicate that *all* the individuals involved may *at the same time* not have the contributory intention?[22]

Given the current trend to the obliteration of the individual, the question "Does anyone of us really and sincerely believe what we collectively believe in virtue of a joint commitment?" is not always properly askable, or, indeed, intelligible. Certainly, for anything that "we" collectively believe, there have always been and there will always be quite a few among "us" who do not really share this belief personally, this is no news; but... are they just a few, or are they more – most, or even *all* of those who are party to that collective belief? One can almost

---

22 Something analogous for collective beliefs: Mark and Roz, p. 137.

hear a voice mumbling off-record: "We're too busy peddling various 'corporate identities', so who cares?...". Gilbert does not press the question too hard. She does, it is true, at length discuss, and in this writer's view successfully[23] refute, what she calls "rejectionism" (pp. 133ff.), a position that maintains that the collective belief in the sense of her theory is not at all belief but just a pallid "acceptance".[24] She stresses that collective beliefs are – as are the individual ones – adopted and given up for *epistemic* reasons (pp. 148ff.) or that they cannot be taken on at will (the relevant joint commitments can, pp. 155f.). But she remains mildly non-committal and soft-contoured on various critical issues.

Olive and Ned (pp. 103f.) make one feel sorry, as do Roz and Mark (p. 137): parties to collective intentions or beliefs none of them personally shares. In Poland under Communism (1944–1989) we collectively believed Marxism-Leninism, with (next to) no one believing it personally, with all the practical consequences of that pretended belief, *somewhat* more vexatious than a hill left unclimbed, and with a collective schizophrenia to loom large for generations yet. "But does any one *really* believe what we believe?" "How silly to ask such questions!..." Gilbert observes with Durkheim (the "coercive power of collective beliefs", pp. 131, 157, 160, 172, 177) that people are often "educated" (by officiously asserting what they do not yet believe[25] and by rebukes if they refuse to assert it, or, worse still, assert the opposite, or report such counter-assertions with bravado, p. 193) into internalizing beliefs they are jointly (with someone else) committed to believe.

It would be ironic if such "educated" ones were ultimately to be the only ones to individually believe collective beliefs, the rest just paying lip-service[26] to them or at best "going through the motions" of believing them, Cato's soothsayers ("augurs"), T.S. Eliot's hollow men. Also, one wonders if such "populations" are not doomed to failure, in a rather short run. In Innocent III's times few influential Catholics believed what "the Church" believed, and the edifice was toppling; yet a certain Francis of Assisi did believe – and he saved the Church, according to the pious legend. Maybe every collective, to survive, needs its own St. Francis? This question is as much philosophical as it is empirical.

REFERENCES

Gilbert, M. (2014), *Joint Commitment*, Oxford University Press, New York;

Kierkegaard, S. (2009), *The Point of View,* H.V. Hong & E.H. Hong (eds. and tr.), Princeton University Press, Princeton;

Maddox, R.I. (1983), "Hermeneutic circle – viscious [*sic*] or victorious," *Philosophy Today*, 27 (1), pp. 66-76;

Ortega y Gasset, J. (1981), *El hombre y la gente*, Madrid: Revista de Occidente en Alianza Editorial.

---

23 Too complex to go into here, but see the first paragraph on p. 144.

24 As in assuming something "for the sake of argument" without actually believing it (p. 134, including note 11) though this is probably not the kind of acceptance in terms of which Gilbert's theory of collective belief is most fairly interpreted (see pp. 146f.), even if one believes in/accepts "rejectionism".

25 Pascal's Wager includes a related idea of educating oneself (note well: *oneself*) to believe by behaving like a believer.

26 Isaiah 29:13.

SILVIA TOSSUT
Università Vita-Salute san Raffaele
silvia.tossut@gmail.com

ON ACTING *BECAUSE OF* A JOINT COMMITMENT

abstract

I focus on the motivational role that Gilbert attributes to joint commitment. Using Bacharach's game theoretical analysis of non-summative group agency, I point out that Gilbert's account of social actions conceptually requires the obliteration of individual preferences and individual rationality. Then, I investigate whether acting because of a joint commitment is rational in some sense, focusing primarily on the phenomenon of asking and giving the permission to defect (fair defection) when a joint commitment is in place. I show that the obliteration of individual preferences prevents the possibility of rational fair defection. Finally, I analyze Gilbert's recent suggestion concerning the introduction of personal preferences in her account, and I show that such introduction cannot solve the problem with the (ir)rationality of fair defection, and that personal preferences can hardly be consistently included in Gilbert's account.

1.
Gilbert's classical account of joint commitment

In this paper I assume the general validity of Margaret Gilbert's account, and I focus on a specific problem raising in this framework. The question that I will try to answer is: When I am part of a plural subject, what motivates my actions? What should I *rationally* do?

According to Gilbert, when a joint commitment is in place the people involved must act having this joint commitment as their motivation. Her thesis is that subjection to a joint commitment compels me from a *rational* point of view to act in a way consistent with its content: neither me nor my partners in the activity can remove the joint commitment at will, so I have sufficient reasons to undertake actions consistent with the collective goal expressed by the joint commitment. In Gilbert's words, once people are jointly committed to pursue a goal as a body, the group's goal becomes the unique motivational source for the actions of each of the parties (Gilbert 2006, p. 123), and thus, "a member's action may often be explained without any reference to his or her own personal goals, values, or principles of action" (Gilbert 1996, p. 268). As a consequence, the description of social actions in terms of joint commitment has the advantage of providing a *unified explanation* of the individual contributions to the action, even in the complete absence of reference to personal intentions or desires (Gilbert 2000, pp. 14-36). In other words, when a joint commitment is in place people involved act *because of* their joint commitment. A rational agent acting as a part of a plural subject must behave consistently with the intention to achieve the plural goal, and her behavior must be motivated by the joint commitment.

It is worth noting that the importance of joint commitment consists precisely in its functioning as a unified motivational source, since this unity confers stability to the cooperative outcome. Even though individual rationality can succeed in prescribing cooperation when specific circumstances occur, the existence of a joint commitment allows me to rely on the *assumption* that the others will cooperate. On the contrary, in the individualistic frame I must always consider the possibility that my partner might abandon cooperation.

The explanation of this kind of stability is *psychological*: as List and Pettit (2011, p. 193) put it, when people act in their capacity as group members, they experience a change in the perceived subject of intention and action. The reason is that the pronoun "we" has the power to provoke an immediate psychological response in the subject, inducing one's self-identification as a member of the group. The members of a group are thus the indivisible parts of a single center of agency.

The departure from the singularist approach is also embraced (from a different perspective) by Michael Bacharach (2006), who presents a game theoretical analysis of cooperation quite sympathetic with Gilbert's plural subject theory. Bacharach's analysis is particularly useful to analyze the rationality of acting *because of* a joint commitment.

Bacharach's starting assumptions are (i) that if a group is an agent, then it has its own payoff function (distinct from the payoff functions of its members), and (ii) that, if rational, the group agent will aim at the maximization of *its* expected utility. According to Bacharach – and consistently with Gilbert's holistic perspective – the group's preferences cannot be reduced to the members' preferences, nor can the reasons motivating the group's choice be reduced to the members' reasons. Roughly, cooperative outcomes are explained by the individual agents propensity to *group identify*.

Group identification is a psychological response, primed by some objective features of the game (i.e., the immediate perception of a common interest and of the necessity of the others' contributions).[1] Group identification involves a shift of frame, a change in the agent's conception of herself: a player who group-identifies conceives herself as part of a unified agent (Bacharach 2006, p. 70). The crucial effects of group identification are payoff transformation, cooperation, and agency transformation.

Payoff transformation corresponds to the unification of the motivational sources that is prompted by group identification. Since all the actions of the players are motivated by the group's goal, group identification induces also a *cooperative behavior* within the group, (Bacharach 2006, p. 79). The mechanism that conveys these effects, and which is at the core of Bacharach's account, is the *agency transformation* prompted by group identification. Agency transformation consists in the players' adoption of a novel perspective with respect to the game. In summary, group identification causes the *disappearance* of personal payoffs, and a re-description of the game. (Compare this with Gilbert's idea that being part of a plural subject causes a change in the perceived subject – from the "I" to the "we").

To explain this point, let me consider the game in *figure 1*, which is a "Prisoner's Dilemma" (PD). In PD, each player chooses among "cooperate" (C) or "defect" (D). Given the structure of the game, each player prefers the outcome in which she defects and the other cooperates

2. A game-theoretical explanation of joint commitment

P1

	C	D
C	2,2	-3,3
D	3,-3	1,1

P2 (label for left table)

	C	D
C	4	0
D	0	2

Figure 1 *Figure 2*

---

1 Bacharach takes identification with the group to be basically a non-rational framework effect, so it is not clear where the commitment is supposed to come from. This seems to be an important point to explain the integration of Gilbert's notion of joint commitment within Bacharach's account. I think that this feature is consistent with Gilbert's account to the extent that it is not always clear *why* individuals do enter a joint commitment. In what follows I offer an analysis of the motivations leading to the formation of a joint commitment, at least for those cases in which the framework effect is weak enough to leave space for a voluntary decision. I thank an anonymous referee for suggesting me to clarify this point.

("free-riding" situation); conversely, each player aims at avoiding the situation in which she cooperates and the other defects. The application of individual rationality leads to the solution (D,D). Yet, despite this prescription, both players might obtain higher payoffs if both cooperate.

Bacharach's analysis of cooperation lies on the assumption that group identification causes the players' understanding the PD as the game depicted in *figure 2*: the payoffs of this matrix are the payoffs of a *group agent*, which are achievable through the implementation of a certain profile of action by the original players.[2]

Note that the disappearance of personal payoffs and the new matrix of the game are not merely due to a change in the preferences of the individual agents: it is not simply that a group-identifier wishes spontaneously to promote the collective goal, rather she "thinks of her agential self [...] as a component part of [the group's] agency" (Bacharach 2006, p. 136). Agency transformation entails a change from the payoffs that govern choices for one unit of agency (the individual), in the payoffs that govern choices for another unit of agency (the group). Group identification amounts to feeling a sense of collectivity, which prompts "team-reasoning". Team-reasoning is possible only if the players are in a frame in which first-person plural concepts are activated (Bacharach 2006, p. 135 and p. 141): "if a group-identifier thinks of herself as part of a 'we' [...] it is only for us that she can intelligibly deliberate" (Bacharach 2006, p. 145). The peculiar form of reasoning entailed by group identification, also, supports the holistic interpretation of the "we", which cannot be reduced to the sum of individual motivations and actions: team-reasoning eliminates any reference to the individual, and allows the players to adopt a genuinely collective perspective.

The profile of action (C,C) in *figure 2* – i.e., from the point of view of the collective agent – is *insensitive* to the distribution of gain among the participants.[3] As Gilbert correctly points out, the obliteration of individual preferences due to the adoption of the plural subject perspective is a phenomenon that we can observe more clearly in particular kinds of groups (e.g., in marriage, in which people are likely to arrive to what she calls the "fusion of egos"). But the question is: besides being descriptively accurate, is the obliteration of individual preferences also *rational*?

In particular, Bacharach would say that the cooperative outcome (C,C) in PD *is* rational if we consider the collective agent as the subject of rationality, since (C,C) is the profile of action that maximizes its utility. Group identification transforms the original PD in the game in *figure 2*, which is a game played by the collective agent, and which has a unique "collective solution", the profile of action (C,C).

The peculiarity of the collective solution of the PD is that it is *inconsistent* with the solution prescribed by individual rationality. According to Bacharach, this inconsistency influences the final outcome, to the extent that the players' preferences toward a non-cooperative outcome lower the probability that they will group identify.

Nonetheless, in competitive frameworks, I can still rely on the other's doing her part in the realization of the collective solution. Indeed, Bacharach claims that since group identification is an involuntary psychological response, activated by *objective features* of the situation, in framing the situation as a problem "for us", an individual also gains some sense of *how likely it is* that another

---

2 The discussion is open on the exact mathematical representation of the collective payoff. For example, it might not be a sum, but rather an average of the individual payoffs. What I am claiming here is just that, whatsoever form this measure may taken, it will always be insensitive to individual preferences.

3 A complete account would require the analysis of asymmetric PD games, since such games entail a number of different considerations (e.g., one of the player could justify occasional defection on the basis of egalitarian considerations). Alas, such a complete analysis goes beyond the limits of this paper.

agent, facing the same situation, will we-frame it. Arguably, the more competitive is a game, the less intuitive – and the less probable – will be group identification (Bacharach 2006, p. 75). The obliteration of the individual payoffs entails that the players might be required to act in contrast with the prescriptions of individual rationality. As in the PD, the adoption of the collective perspective results in an efficient solution. An objector might observe that, though the solution (C,C) is *prima facie* efficient for each player since it allows avoiding the worst profile (D,D), it is easy to see that each player can still maximize her expected utility: if I can *assume* that you will do your part in the collective solution, and I am rational, I must defect to obtain the profile (D,C). The answer to this objection is that "collective efficiency" is insensitive to the players' individual payoffs – consistently with Gilbert's non-summative approach to collective notions. The stability provided by the joint commitment is meant to prevent exactly this kind of strategic reasoning: in particular, the obliteration of individual preferences entails that once that I have adopted the collective perspective, and computed the collective solution, I cannot simply turn back to the individualistic point of view and adopt the assumption of your cooperation.

One consequence of the holistic character of joint commitment is that the members of the plural subject feel that none of them can rescind the commitment unilaterally, by simply changing her mind. In detail, Gilbert argues that if one has not been given the permission to defect, she, *being rational*, will not defect, because her motivations for action are not due to her personal preferences but rather dependent on the plural subject's goal (Gilbert 2000, pp. 24-25). I will call this process of asking and giving the permission to abandon a joint commitment "fair defection". Fair defection is meant to be the proper way to abandon a joint commitment, in contrast with simple defection, which allows the "abandoned" members to rebuke the defector. Assuming the descriptive adequacy of Gilbert's claim, I will now examine the rationality of fair defection. For the sake of simplicity, I will refer to the PD illustrated in the previous section, though I think that the conclusions (with minor modifications) are valid for all the cases in which there is a joint commitment.

**3.
Fair defection**

The description of joint commitment in terms of agency transformation suggests that when we consider specific features of collective actions we should distinguish among individual rationality and collective rationality.

Consider the two agents P1 and P2, facing a PD. Imagine that they are jointly committed to implement the collectively efficient outcome (C,C). At a certain time, P1 undertakes the procedure for fair defection, by asking P2 the permission to abandon the joint commitment. Let me focus first on P1 asking P2 the permission to defect, and look at the situation from the individualistic perspective. The first problem is that the adoption of joint commitment prevents P1 from referring to personal preferences as reasons to abandon the collective point of view: if P1's individual preferences have been obliterated, she has no reason to desire defection. For the sake of the argument, let me still assume that P1 can for some reason turn back to individual preferences. As I observed above, is rational for P1 to defect if she *knows* that P2 adopts collective rationality (and thus, plays C); in order to rely on P2 doing her part in the collective solution, however, P1 should better not communicate her decision to stop doing her part in the collective solution.

Also, from the point of view of collective rationality, there are no reasons why a member should prefer defection: the collective solution is the best for the plural subject, regardless the distribution of the gain.

Thus, neither collective nor individual rationality command asking the permission for fair defection: collective rationality prescribes to avoid defection at all, individual rationality prescribes defection without communication for strategic reasons.

Now, turn to P2 reasons to give P1 the permission to defect. If we consider P2 reasons from individual rationality perspective, there is a problem of regression: if P2 adopts individual rationality in deciding whether to give P1 the permission to defect, then P2 has in turn already abandoned collective rationality; if P2 asked P1 the permission to defect, there is a regression problem, while if P2 simply abandoned the joint commitment, then P1 does not need to ask the permission to defect (since the joint commitment is already broken). However, we might admit that P2 can adopt individual rationality, for example, because P1's request signals that the joint commitment has lost the required stability. In this case, individual rationality commands P2 (i) not to give P1 the permission to defect and (ii) to defect: P2 should forbid P1 to defect, so she can rely on P1 cooperation; in this way, P2 can do D and obtain her preferred outcome (C,D).

The other option is that P2 adopts collective rationality in considering whether to give P1 the permission to defect. Yet, from the collective point of view, giving P1 the permission to defect is never rational. As seen in the previous section, cooperation leads to the collective solution of the game (the collectively efficient profile of action), so the plural subject cannot maximize its utility if one of the members ceases to act in accord with the joint commitment. Thus, from the point of view of collective rationality, P2 should not give P1 the permission to defect. As a conclusion, neither individual nor collective rationality allow giving the permission to defect. The above considerations hold only to the extent that personal preferences are obliterated and collective notions are interpreted in a non-summative and non-correlative way. The conclusion is that from the point of view of individual rationality, defection is not only permitted, but required, though it cannot be properly be considered an instance of fair defection in the sense illustrated by Gilbert. From the point of view of collective rationality, defection on the part of one member is never rational – and each member should do everything in her power to avoid the others' defections.

4.
The inclinations plus joint commitment account

So far, I have emphasized that the obliteration of individual preferences is a crucial element of Gilbert's account of joint commitment. Also, I pointed out that this account of joint commitment is inconsistent with the rationality of fair defection.

In her recent book, *Joint Commitment. How We Make the Social World* (2013), Gilbert restates her thesis that joint commitment provides stability to collective actions, by preempting a decision contrary to the collective interest (Gilbert 2013, p. 93).

With specific reference to PD, Gilbert specifies that the adoption of the collective point of view entails that the players accept to do their part in a combination of actions that do not give them what they are most inclined to get. In general, one of the points that Gilbert makes about the motivational force of joint commitment concerns its capacity to "lead to *relatively good outcomes* for all in collective action problems of all kinds" (Gilbert 2013, p. 93; my emphasis). The main problem with this claim is that getting a "relatively good outcome" is not the goal of rationality: by definition, a rational agent does not look at the "relatively good outcomes for all", nor to a "relatively good outcome for herself", but rather at the maximization of her own utility. Note that this holds not only for the individual agents, but also for the collective agent: the plural subject utility (if we persist in a Gilbertean non-summative approach) is insensitive to the distribution of utility among the members; the plural subject, thus, is not interested in the collective solution being "relatively good for all", but rather in its being the best solution for the whole – the collective agent. Despite the problems raised by this specific formulation, however, the point is perfectly consistent with Gilbert's classical account.

I want now to focus on another element, which Gilbert – quite surprisingly – adds to her account in the new formulation. I refer to the introduction of personal inclinations in the explanation of the actions of the parties in a plural subject. As I showed in the previous

sections, the obliteration of individual preferences poses severe limits on Gilbert's account. As a matter of fact, each participant in the activity has personal goals and preferences that differ, even substantially, both from the collective's and from the other participants' goals and preferences.

Perhaps, Gilbert has precisely the intent to avoid these consequences when she says that subjection to a joint commitment does not prevent one's acting according to *her own* best judgment, and that joint commitment does not obliterate one's inclinations. She goes on advancing the proposal of an "inclinations-plus-joint-commitment model of action", on the assumption that such a model might "explain how, though *rationality* requires one to act in a particular way, there may remain a pull in the direction of acting contrary to reason's dictate" (Gilbert 2013, p. 93).

Gilbert tries to combine two elements: the first one is the non-summative notion of joint commitment and the non-correlative apparatus that characterize her account; the other one is the respect of individual rationality.

> Participation in collective agency – in particular, subjection to the underlying joint commitment – does not leave me free to do as I please, *from a rational point of view.* Among other things, it gives me *sufficient reasons to act in a certain way*, reasons I cannot remove at will. Second, it does not – how could it? – deprive me of my capacity to reason and to act according to my own best judgment. I may break away from a collective action in progress at any time – sometimes this may be rationally required, sometimes at least rationally permitted, sometimes not (Gilbert 2013, p. 91).

In the above quotation, Gilbert makes large use of the notion of "rationality". It is unclear whether she is referring to individual or collective rationality, but the meaning of her claims changes drastically depending on the interpretation that we adopt. For example, she says that the subjection to a joint commitment doe not leave me free to do as I please *from a rational point of view.* If I am adopting collective rationality, this claim is trivially true, while if I am adopting individual rationality it is false, for the reasons investigated in the previous sections.

Also, Gilbert argues that one's breaking away from a collective action is sometimes rationally required (or permitted). Yet, the analysis of fair defection provided in *section 3*, showed that individual rationality does always command defection, while collective rationality always preempt defection.

Does the introduction of personal inclinations introduce substantial changes in this analysis? It seems that the possibility of such a change depends on the relations among personal inclinations and joint commitment.

Though Gilbert has merely sketched the inclination-plus-joint-commitment account, without giving much details (indeed, she does not say many thing besides those in the above quotations), I think that with respect to the relation among inclinations and joint commitment there are two main options. The first option is that individual agents involved in the collective activity do not adopt the collective point of view, but rather experiment a change in their individual payoffs due to their perceived relations with the other members. Yet, this explanation is inconsistent with Gilbert's approach, since it removes the role of joint commitment and reduces the problem to one to be solved by individual rationality. The second option holds that individual agents do in fact experiment agency transformation, but – contrary to the holistic interpretation outlined above – the collective payoff is *sensitive* to the distribution of the utility among the players: not only in the sense that a great deal of inequality might prevent the formation of a joint commitment, but also in the sense that the collective payoffs is not a monolithic value, blind to individual gains. This explanation seems

plausible, for it saves both joint commitment and a distinctive role for collective rationality. Yet, it is inconsistent with Gilbert's account of joint commitment, because such a solution rejects non-summativism and non-correlativism.

In conclusion, it seems that there is not a straightforward way to introduce inclinations in Gilbert's explanation of the motivational role of joint commitment, without bringing inconsistencies within the original view.

5.
Conclusion

I showed that one problem with Gilbert's classical account of joint commitment is that it conceptually requires the disappearance of individual preferences. I argued that the obliteration of individual preferences makes unintelligible the phenomenon of fair defection, which is the only way to exit a joint commitment. Then, I argued that the re-introduction of individual preferences recently sketched by Gilbert (2013) cannot respect the holistic spirit of her account. The introduction of an inclinations plus joint commitment account is hardly consistent with Gilbert's general theory. In particular, there are two main risks, for one might exaggerate with the import of joint commitment to the detriment of individual preferences (and rationality), turning back to the classical joint commitment account and its limits; or, on the other hand, the introduction of individual preferences might result in the erasure of the stabilizing role of joint commitment.

I think that the problem examined here is part of a general and unsolved problem for the non-correlative accounts of sociality, concerning the relation among the individual and the collective level of explanation. Arguably, the problem of collective rationality as presented in this paper is likely to find a solution only after a general clarification of such relations.

REFERENCES

Bacharach, M. (2006), *Beyond Individual Choice: Teams and Frames in Game Theory*, N. Gold and R. Sugden (eds.), Princeton University Press, Princeton;

Gilbert, M. (1989), *On Social Facts*, Routledge, London;

----. (1996), *Living Together: Rationality, Sociality, and Obligation*, Rowman and Littlefield, Lanham, MD;

----. (2000), *Sociality and Responsibility: New Essays in Plural Subject Theory*, Rowman and Littlefield, Lanham, MD;

----. (2006), *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*, Oxford University Press, Oxford;

----. (2013) *Joint Commitment. How We Make the Social World*, Oxford University Press, Oxford;

List, Ch. & Pettit, Ph. (2011), *Group Agency*, Oxford University Press, Oxford.

LEO TOWNSEND
*Department of Philosophy, Classics, History
of Art and Ideas, University of Oslo*
l.c.townsend@ifikk.uio.no

JOINT COMMITMENT AND COLLECTIVE BELIEF: A REVISIONARY PROPOSAL*

abstract

According to Margaret Gilbert, two or more people collectively believe that p if and only if they are jointly committed to believe that p as a body. But the way she construes joint commitment in her account – as a commitment of and by the several parties to "doing something as a body" – encourages the thought that the phenomenon accounted for is not that of genuine belief. I explain why this concern arises and explore a different way of construing joint commitment, in order to avoid the concern. This leads me to propose a revised Gilbertian account of collective belief, according to which two or more people collectively believe that p if and only if they are jointly committed to p as true.

Introduction

In a series of publications spanning over 25 years, Margaret Gilbert has developed and defended the following account of "collective belief": "The members of a population, P, collectively believe that *p* if and only if they are jointly committed to believe that *p* as a body" (Gilbert 2002a, p. 137; Gilbert & Pilchman 2014, p. 197).

It is widely accepted that Gilbert's account describes an important and interesting social phenomenon. What is less widely accepted is that the phenomenon in question is that of belief. Many so-called "rejectionist" critics claim that the collective attitude she picks out is better understood as a form of "acceptance" than of belief, since it does not necessarily aim at truth, and is brought about voluntarily.

In this paper, I argue that Gilbert's account, as it stands, is susceptible to something like these criticisms, but I think this calls for revision of the account rather than outright rejection. The paper has two parts. In the first I rehearse the debate between Gilbert and the rejectionists, and suggest that Gilbert's problem arises from her construal of joint commitment in practical terms, as a commitment of the several parties to "doing something as a body". In the second part I explore the possibility of construing joint commitment differently, in order to avoid the rejectionist threat. Drawing on recent work by Pamela Hieronymi I propose a revised account, according to which a set of individuals collectively believes that *p* if and only if they are jointly committed to *p* as true.[1]

1.
Joint commitment to believe that *p* as a body

An example can help to bring Gilbert's account into view:

1.1
Gilbert's account

> Roz personally believes that it is never justified for one country to take up arms against another. When Mark asserts the justifiability of a defensive war, he speaks very forcefully. Rather than argue, Roz decides to agree with him. So she says, "Yes, indeed". It seems that now either of them could properly make the collective belief statement, "We believe that a defensive war is justifiable". And this is true even if Mark was for

---

1 In several places (e.g, Gilbert & Pilchman 2014, p. 198; Gilbert 2002a, pp. 138-140) Gilbert cashes out what is meant by "believe that *p* as a body" in terms of the notion of *emulating a single believer of p*. To be so jointly committed is for several parties each to have incurred duties to speak and act "as would any one of several mouthpieces of the body in question" (Gilbert 2002a, p. 140). Had I focused on this formulation, I might have presented my proposal less as a *revision* of Gilbert's account than as a suggestion for how we should understand her notion of "emulation", in order to forestall the rejectionist threat.

some reason asserting the opposite of *his* personal view when he spoke. In other words, it may be true, for Roz and Mark, that "We believe that *p*", though neither of them personally believes that *p* (Gilbert 2002a, p. 137).

According to Gilbert, what happens in cases like this is that each of the parties expresses his or her readiness to let a particular proposition stand as *their* collective view. Here Mark signals his readiness by making an assertion, while Roz signals hers by assenting. This, it seems, is enough to make the collective belief statement "We believe that a defensive war is justified" true of them, and this is so even if neither personally holds that belief.[2]

Gilbert's primary interest is in the normative fine-structure of such situations. She notes that once each party signals readiness to let some view be established as the collective view, each incurs a distinctive suite of obligations and entitlements in relation to the other(s). Specifically, each owes it to the other(s) to speak and act in ways that accord with them believing that *p* as a body. While no party need pretend to personally believe that *p*, each must declare that *p* when appropriate, and none may express direct disagreement without "preamble" (e.g., "*I personally* do not believe that *p*"). Should any fail to meet these obligations, any other has the standing to rebuke the recalcitrant party.[3]

Gilbert accounts for this normative situation with the notion of *joint commitment*. Like personal commitments, joint commitments are sources of normative constraint and standing. If Jim decides to go to the party then he is committed to going, and so, all else being equal, he *ought* to go. Jim is then *answerable* to himself for going, and might appropriately chide himself should he fail to go (without rescinding his decision). Likewise, if Beth and Sue jointly commit to walking together then, all else being equal, they ought to do so. Each owes it to other to play her part in seeing to it that they walk together, and each has the standing to rebuke the other for failing to do so.

A joint commitment is thus, like a personal commitment, a normatively-significant process of self-binding, except that what is bound is not a single person but two or more people, who thereby constitute a "plural subject". By reciprocally expressing readiness to let "a defensive war is justified" stand as their collective view, Mark and Roz bind themselves to believing as a single body that a defensive war is justified. Since such a commitment seems to explain the distinctive obligations and entitlements Roz and Mark have in relation to one another, Gilbert proposes that this commitment constitutes their collective belief.[4]

1.2 Rejectionism rejected

Many of Gilbert's critics have seized upon examples like the one involving Roz and Mark in order to argue that, regardless of how Gilbert accounts for it, the phenomenon picked out could not be that of belief. This is because the collective view is formed for reasons unrelated to its truth, and is brought about voluntarily.[5] But belief must necessarily aim at truth, and belief cannot be willed,[6] so (they claim) the attitude could not be that of belief.

---

2 It is worth noting that the way cases like this are described reflects a crucial methodological difference between Gilbert and her rejectionist opponents. Gilbert treats the truth- and appropriateness-conditions of collective belief statements as something evident in our practices of making and accepting such statements – i.e., as part of the phenomenon to be explained. Rejectionists, by contrast, tend to treat the truth (or appropriateness) of such statements as beholden to a theory of belief.

3 See Gilbert (1987, p. 199)

4 Note that what the group does (collectively believing that *p*) and what the group is committed to doing (believing that *p* as a single body) are not seen by Gilbert as identical. Rather, the commitment to the latter constitutes the former. Thanks to an anonymous reviewer for urging me to clarify this.

5 See, e.g., Meijers (1999) and Wray (2001).

6 I follow Gilbert in accepting these two requirements on belief. A canonical supporting argument is in Williams (1970).

Let us start with the concern about truth. In the Roz and Mark case we find that neither party personally takes what they collectively believe to be true. Perhaps, then, in letting this view stand as their collective view, the parties are only *going along with* or *accepting* the view, rather than believing it.[7] Moreover, neither party seems to be motivated to let that view stand as their view for reasons pertaining to its truth. We are not told why Mark makes his assertion, but Roz agrees only to avoid an argument. But since Mark's assertion together with Roz's agreement is what establishes the group view, it seems that the group view gets established for non-epistemic reasons. Such reasons would be of the right kind for acceptance, but are of the wrong kind for belief.

Gilbert (2002a, 2014) replies that this objection wrongly assumes that the phenomena picked out at the individual level must constitute the same kind of phenomena at the collective level. Even if the *group members* are simply going along with *p*, it does not follow that the *group itself* is simply going along with *p*. And even if the *group members* participate in the establishment of a group view for non-epistemic reasons it does not follow that the *group itself* is motivated by the same reasons. (Indeed, as Gilbert notes, the members' participatory motives are better viewed as the group view's *causes* than its *reasons*).

What, then, of the rejectionist claim that collective belief is voluntary, so cannot be genuine belief? This can seem plausible because, for Gilbert, in order for a collective belief that *p* to be established, each individual must *act* in a certain way. For example, Mark performs a declarative speech act, while Roz "decides to agree". These two performances are all it takes for their collective belief to be established, and since both performances are produced voluntarily, it seems the collective belief must be too.

Gilbert again replies that the rejectionist is looking in the wrong place – looking to the wills of the group members, rather than to the collective will. Even if the members must voluntarily signal their participation in the establishment of the collective belief, this is no reason to conclude that the group itself voluntarily establishes its belief. So, claims Gilbert (2002a, p. 155), "there is no confrontation with involuntarism" at the collective level.

1.3 Rehabilitating rejectionism

I think that headway in the debate between Gilbert and the rejectionists can be made through a shift of focus. Instead of focusing on the phenomenon picked out in her examples, rejectionists should attend to the joint commitment Gilbert uses to account for that phenomenon, and, in particular, to the *object* of this joint commitment – i.e., *that which it is a commitment to*. The object of the joint commitment in the case of collective belief is "believing that *p* as a body". The question is: how should we understand this object?

Well, according to Gilbert, "joint commitments are always commitments to 'act as a body' [...] where 'acting' is taken in a broad sense" (Gilbert 2002b, p. 41), so as to include the having of psychological states. This suggests that we should interpret the object of the joint commitment in Gilbert's account in an "attitude-centered" way, rather than a "content-centered" way. That is, what the participants in a collective belief jointly commit to is the having of an attitude – the believing that *p* – rather than to the content of that attitude, i.e., *p* itself.[8] This, to my mind, is where the problem in Gilbert's account lies.

---

7 There are various competing accounts of the attitude of acceptance in the received literature, but I abstract away from their differences here. For my purposes the key features are, first, that acceptance need not aim at truth in quite the same way as belief, and, second, that acceptance is under the control of the will to a greater extent than belief.

8 If jointly committing to the having of the belief that *p* is understood – as Gilbert sometimes explains it – as jointly committing to emulating a single believer of *p*, then perhaps the two options presented here are not so starkly opposed. After all, it may be that the way to understand what it is to emulate a single believer of p is in terms of commitment to p as true. Cf. footnote 1 above.

You can begin to see the problem by noting that, in the individual case, a person could fully commit to the having of a belief that *p*, and not yet believe that *p*. Perhaps I recognize that, all things considered, the belief that my work is going along nicely would be a good belief for me to have, and so commit myself to believing – that is, to having the belief – that my work is going along nicely. This might amount to my intending to inculcate the belief that my work is going along nicely, but it does not yet amount to my so believing.

I think the same trouble arises in the collective case. There may be practical or prudential reasons why a set of people should jointly commit to believing some proposition as a body. For instance, it might be best for a tobacco company's business if they could construct themselves as a plural subject of the belief that smoking does not cause cancer, and so, recognizing this, they commit themselves to believing – that is, to having the belief – that smoking does not cause cancer. As rejectionists have urged, it seems strange to say that the company now *genuinely believes* that smoking does not cause cancer. After all, they have only considered whether the belief would be a good belief to have, not whether its content – that smoking does not cause cancer – is true.

The problem, as I see it, is that many of the reasons which would be of the right kind for *jointly committing to believing that p as a single body* are reasons which would be of the wrong kind for *believing that p.* Chief amongst these would be considerations that bear on the question of whether believing that *p* as a single body would be good for us to do, regardless of whether *p* is true. To be sure, *p* being true (or likely true, or well supported by evidence) would sometimes be a good reason to jointly commit to believing that *p* as a single body, but it is far from the only such reason. Another good reason might be that so believing would maximize profit, and indeed this sort of reason could bear decisively on the question of whether to commit (or jointly commit) to believing that *p*. But, as we have seen, to be moved by some such reason is not yet to have formed the belief that *p*. So it is unclear how any such commitment could, as Gilbert suggests it could, constitute any such belief.

2. Joint commitment to *p* as true

If rejectionism can be rehabilitated in this way then Gilbert's account is not, as it stands, an account of something properly called belief. But since the problem seems to be generated by the way she construes joint commitment, perhaps what is called for is not an outright rejection of the account but rather a revision of that construal. In the rest of the paper I draw on recent work by Pamela Hieronymi to explore this possibility.

2.1 Doxastic commitments

Hieronymi observes that certain propositional attitudes can be formed or revised simply by settling for oneself a question. Intentions are like this, as are beliefs: "If you [...] settle for yourself the question of whether p, you have thus, ipso facto, formed a belief that p" (Hieronymi 2005, p. 447).

Hieronymi calls these "commitment-constituted attitudes", since what it is to have them is just to have committed oneself to an answer to some question. To be so committed is also to have incurred certain normative statuses: "If I believe *p*, then I am committed to *p* as true, that is, I am answerable to questions and criticisms that would be answered by the considerations that bear on whether *p*" (Hieronymi 2005, pp. 449-450).

However, not all propositional attitudes embody one's answer to a certain kind of question; some are more like a mental exercise or undertaking. Hieronymi calls these attitudes "action-like", and includes among them, "supposing that *p* for the sake of argument, imagining there's no heaven, or remembering your keys" (2005, p. 451). To this we can add *acceptances*, as least as they are understood in the debate between Gilbert and the rejectionists.[9]

---

9 Gilbert herself (2002a, p. 134) suggests that supposing for the sake of argument is a paradigm case of acceptance.

There are, of course, questions *relevant to* the formation of "action-like" attitudes, such as whether they would be useful, or somehow good attitudes to have. But settling any such question does not yet amount to forming that attitude. Instead it amounts to forming a different, second-order attitude – the intention to inculcate the first-order attitude.

Though I will not here attempt to support Hieronymi's distinction between commitment-constituted and action-like attitudes, I do think the taxonomy it engenders could aid Gilbert in her debate with the rejectionists. This is because, by Hieronymi's lights, acceptance is not a commitment-constituted attitude, but belief is. So if some kind of *Hieronymian* commitment constitutes the phenomenon at issue then that phenomenon could not be acceptance, but might be belief. The threat of rejectionism arises for Gilbert from what I called her "attitude-centered" reading of the joint commitment she sees as constitutive of collective belief. What Hieronymi offers is a "content-centered" understanding of this commitment, which I think could help Gilbert avoid the rejectionist threat.

In light of this, here is the revised version of Gilbert's account I propose:

> The members of a population, P, collectively believe that *p* if and only if they are jointly committed to *p* as true.

For the remainder of the paper I briefly elaborate and assess this proposal.

2.2 Collective doxastic commitments

What is it for the members of a population to be jointly committed to *p* as true? On my proposal, they must have settled for themselves the question of whether *p*. How this can happen is suggested by one of Gilbert's own examples:

> There are three states in a particular alliance [... represented by] Peter, Antoine, and Karl. Previously the alliance had come to believe that the way to achieve its goal G was to bomb country C, or, for short, it had come to believe that *g*. Now each member prefers, for its own reasons, that the alliance's bombing of C be discontinued. It therefore wants to bring it about that the alliance cease to believe that *g*.
> Karl speaks first, in the name of his own country. He is quite likely to say: "Is bombing C really going to achieve G?". In other words, he is likely to question the *truth of g* [...] Peter might appropriately say, "It's not clear that it is [...] I'd say that bombing isn't likely to achieve G. Given the people we are up against, it is just as likely to have the opposite effect!". Karl and Antoine might eagerly approve this, thus establishing for the alliance a new collective belief [...].
> Given this scenario, on what basis did the alliance give up its belief that *g*? It did so for this reason: given the character of the people the alliance is up against, the bombing of C is not – after all – likely to achieve goal G (Gilbert 2002a, p. 149).

The example purports to show, *contra* the claims of rejectionists, that a group can form or revise a group view for properly epistemic reasons, even when the reasons of the individuals for forming or revising the group view are prudential or practical. So even when the participatory motives of group members are of the wrong kind for belief, the group itself might believe for the right kind of reasons.

This seems correct. What seems wrong, however, is what Gilbert's account implies about how we should characterize what is going on in the example. It seems wrong to think that what the delegates are doing is jointly committing to believing not-*g* as a body. For it is not the collective belief that *g* which they are discussing, but rather the matter of *g*. In Hieronymian terms, the delegates are engaged in the joint activity of settling a question. But it is not the

question of whether they should go on believing that *g*, but the question of *g* itself. It seems legitimate to call the outcome of their inquiry the forging of a commitment. But the sort of commitment forged is a commitment to *p* as true, not a commitment to the having of the belief that *p*.

The revised account thus seems adequate to account for at least one of Gilbert's examples. I do not, however, wish to claim it can vindicate all the cases she offers of "acknowledged group belief". Roz and Mark, for example, do not seem to be genuinely settling for themselves the question of whether a defensive war is justified, and so I suspect the Hieronymian approach would side with Gilbert's critics in rejecting theirs as a case of genuine belief. This is so even if (as Gilbert suggests) we are apt, pre-theoretically, to understand a case like theirs as a case of belief. So the revised account I am proposing is also a somewhat *revisionary* account, requiring, in some cases, that we revise our intuitive judgments.

2.3 Assessing the proposal

Where it does find application, the revisionary proposal seems well equipped to account for the normative fine-structure Gilbert observes. A commitment to *p* as true, no less than a commitment to a course of action, is a source of normative constraint and standing. On account of one's doxastic commitment to *p*, one is liable to certain challenges to *p*, obliged to provide reasons of the right sort when challenged, subject to normative appraisal, and a suitable target of reactive attitudes on the basis of such appraisal. Someone committed to *p* as true also arguably has the normative standing to assert that *p*, and to arrange her practical life – including her dealings with others – around *p*'s being the case. And, just as with a joint commitment of the practical sort, the parties to a joint doxastic commitment will incur distinctive *individual* obligations and entitlements: they will need to see to it that they all speak and act in ways that accord with their collective commitment to *p* as true.

However, it might be thought that the revisionary proposal, whatever its independent prospects, does not cohere with Gilbert's broader philosophical enterprise. Gilbert uses the same construal of joint commitment – the construal which sees the object of joint commitment as always some kind of "acting as a body" (in a broad sense of "acting") – to account for many different facets of the social world, but I am proposing she gives up this construal for the case of collective beliefs. Could she accept my proposal?

I would like to think that she could. This because the Hieronymian approach to commitment does not *exclude* but can actually *encompass* Gilbertian "commitments of the will". Thus, when Gilbert stipulates that a commitment is a "fact [...] about what one has reason to do" (Gilbert 2013, p. 899), she can be seen to be referring to what, for Hieronymi, is just one kind of commitment. It is the kind formed by settling for oneself a question of the form, "Would X-ing be good to do?". But there are other sorts of questions, the settling of which amounts to the forming of other sorts of attitudes, amongst which are beliefs.

My revisionary proposal is, at this point, only a sketch – there are a number of important details that must still be filled out. What exactly is it for a group of people to settle for themselves a question of whether or not *p*? Might they settle such a question in a way that differs from how they have settled that question for themselves individually?[10] And to what extent is the will – of the individuals or of the collective itself – involved in such an exercise? Depending on the answer to these and other related questions, the revisionary proposal may end up being a more or less radical departure from Gilbert's original account than first envisaged. Still, I think the proposal is worth pursuing, because, as it stands, Gilbert's account

Conclusion

---

10 Thanks to Jacob Heim and Silvia Tossut for emphasizing to me the importance of this question.

seems susceptible to the grave threat of rejectionism. If collective belief is constituted by a Gilbertian joint commitment – a commitment to doing something as a body – then collective believers will routinely and non-criticisably form their collective views in ways unbecoming of belief. Nevertheless, I think that Gilbert's account can be revised to avoid these problems, and hence maintained as an account of genuine belief. What is needed is a broadening of the notion of joint commitment, to include collective doxastic commitment.

REFERENCES

Gilbert, M. (1987), "Modelling collective belief", *Synthese*,73(1), pp. 185-204;

----. (2002a), "Belief and acceptance as features of groups", in M. Gilbert (2014), *Joint Commitment*, Oxford University Press, New York, pp. 131-162;

----. (2002b), "Considerations on Joint Commitment", in M. Gilbert (2014), *Joint Commitment,* Oxford University Press, New York, pp. 37-57;

----. (2013), "Commitment", in H. LaFollette (ed.), *The International Encyclopedia of Ethics*, Blackwell, London, pp. 899-905;

Gilbert, M. & Pilchman, D. (2014), "Belief, acceptance, and what happens in groups", in J. Lackey (ed.), *Essays in Collective Epistemology*, Oxford University Press, Oxford, pp. 189-212;

Hieronymi, P. (2005), "The wrong kind of reason", *Journal of Philosophy*, CII(9), pp. 437-457;

Meijers, A. (1999), "Believing and accepting as a group", in A. Meijers (ed.), *Belief, Cognition and the Will*, Tilburg University Press, Tilburg, pp. 59-71;

Williams, B. (1970), "Deciding to believe", in B. Williams, *Problems of the Self*, Cambridge University Press, Cambridge, pp. 136-151;

Wray, K. (2001.), "Collective belief and acceptance", *Synthese* 129, pp. 319-333.

SONJA RINOFNER-KREIDL
University of Graz
sonja.rinofner@uni-graz.at

EPISTEMIC AUTHORITY AND MANIPULATION: EXPLORING THE 'DARK SIDE' OF SOCIAL AGENCY

abstract

Part of our social reality results from explicitly acknowledging sharing certain ideas, emotions, and value-commitments. Hence it has been addressed as manifestation of joint commitments, shared intentions, we-reasoning or collective intentionality. Yet relevant parts of our social reality come into existence in a different way. There are various forms of intentionally opaque social agency that are founded upon implicit agreements, thereby conveying uncertainty and blurred roles of acting. Their overall character may, for instance, result from an in-built epistemic asymmetry that tends to blur the distinction between voluntary and non-voluntary modes of acting. Arguably, manipulation is a case in point. My main concern in analyzing manipulative actions is to figure out how the ideas of reason and (epistemic) authority are connected with the social constitution and self-understanding of the agents involved.

Manipulation is widespread (if not to say pervasive) in all areas of human life. It prevails in intimate relationships just as much as in anonymous inter-group interactions. Among the well-known, more or less innocent utilizations of manipulations are techniques to guide our attention, to highlight ideas and arguments according to varying circumstances and audiences, to establish loyalty in larger groups of people and to arouse emotions suited to advance the realization of one's aims whenever these aims cannot be brought about unless others join in. The targets of manipulative techniques in these and other cases are modes of belief formation, processes of deliberation and decision, emotional episodes or attitudes, motives for acting as well as modes of acting.

How does manipulation enter the picture given the actual constitution of human beings? It is a basic and conspicuous fact that although the entire content of an agent's mental life is responsive to socially acquired modes of feeling, thinking, perceiving, and evaluating, it nonetheless evades any direct epistemic access by her fellow men. Our first-personal mode of experience and the ubiquituous need for making ourselves understood by expressing meaning in terms of sounds, signs, bodily behavior, and actions is the most fundamental and powerful lever of manipulation. On these anthropological conditions manipulative actions, moreover, benefit from the limited range and imperfect nature of human reasoning and acting, its interwoveness with desires, emotions and moods, and its susceptibility to weakness of will and other forms of irrationality. It is important to bear in mind this general framework which enables us to understand the possibility of manipulation from the point of view of anthropology (or, for that matter, metaphysics). As soon as we take the stance of everyday life our concern, however, is not to explain the very possibility of manipulation. We rather are interested in its concrete manifestations. We want to come to know how to recognize specific forms of manipulation, how to cope with them and what to think about them with regard to our social relations and our self-understanding as rational and moral agents. For this purpose we are eager to learn more about those concepts and conceptualizations that might be useful for analyzing different brands of manipulation as they often are involved in such diverse activities as advertising, fundraising, coaching soccer teams, child rearing and teaching tasks of all sorts.

In this paper I discuss some specific problems associated with manipulative action. I do so in three steps. First, I present some provisos concerning the term "manipulation", followed by a more precise introduction of the notion of manipulation that I consider proper for my present purpose. Furthermore, I draw attention to several conceptual distinctions that help

to grasp the idea of authority as far as it is involved in manipulative actions. Focusing on the ideas of authority and reason will pave the way for challenging some well-entrenched views with regard to manipulation later on. This especially refers to an important qualification of the idea that manipulator and manipulee are related to each other in an asymmetrical relation (part I). In a second step, it will be argued that there are specific cases, classified as border-line (BOLI) cases, that are interesting owing to some unexpected and irregular forms of social "cooperation" which are realized in a subliminal mode. My suggestion will be to analyze BOLI cases of manipulation by means of Leon Festinger's theory of cognitive dissonance (CODI). Doing so allows for considering manipulative actions as multi-layered phenomena which show a more complicated structure than usually assumed (part II). In the third and final part I summarize the outcome of the former investigation with a view to the following questions: Do BOLI cases of manipulation supplement or supplant the prevailing understanding of manipulation? In particular, does their accurate description go beyond the received view on how manipulator and manipulee are tied together (part III)?

Before we enter a more detailed discussion of manipulation a few remarks are in order with regard to my usage of the term "manipulation" as far as two interrelated and controversial aspects are concerned. I consider both aspects as crucial concerning the interpretation of manipulation, yet I do not have enough space to deal with them in a more extensive manner here. The first aspect touches upon the intention of the manipulator which is either taken to be necessary or not. The second aspect refers to the juxtaposition of a neutral or merely descriptive usage of the term "manipulation", on the one hand, and a so-called moralized usage, on the other. My approach with regard to both issues, ultimately, leads back to the anthropological setting mentioned above.

I.

As is well-known, manipulative actions greatly benefit from certain psychological regularities and biasing effects. Among these is, for example, the so-called *framing effect* according to which the interpretation of information strongly depends on how it is presented to us – whether, for instance, it is presented in a way that invites for a thoroughly pessimistic or optimistic reading. These are well-confirmed results of psychological research. From a philosophical point of view, we are interested in the presuppositions or the basic facts that underlie such specific distortions of our evidential behavior. The most important basic fact is that human agents, as far as their self-perception and self-understanding with regard to their beliefs, emotions, and capabilities in general are concerned, to a large extent depend on (and know themselves as depending on) other-perception. How we feel and what we expect of ourselves in terms of typical or average human accomplishments depends on how we perceive others to perceive us, or expect others to think about us and approach us in the light of certain normative demands. We should therefore recognize that a human agent's self-understanding as an epistemic *and* a moral subject is imbued with "social facts", comprising prevailing ideas, value commitments, common practices concerning emotional expression and the like. If this is true of human existence in general we should not be surprised to see that manipulative actions, which hinder the manipulee to meet certain epistemic ideals, do have an impact on the self-understanding of the manipulee (and, in other ways, potentially on the manipulator, too). In particular, one should expect such an impact on condition that the term "manipulation" indicates the existence of a manipulator's intention – even if he does not need to be fully aware of this intention in all possible circumstances that go hand in hand with manipulative actions. From the manipulee's point of view, it does make a crucial difference, both for one's relation to the manipulator and one's self-understanding, if one retrospectively realizes that one has been deceived. This difference is remarkable even in those cases where the manipulator convincingly argues that the deception has been effectuated *for the other*

person's own good. (Here, I cannot broach the important issue of how to analyze the intimate connection between manipulation and paternalistic action.) The following line of reasoning is based on the "intentional reading" of manipulative actions sketched above. Hence there is no accidental manipulation according to my terminological decision and understanding.

By my lights, the aspect of moral assessment, in a certain sense, enters the stage in the very first step of ascertaining the range of relevant experiences and suitable concepts. However, I do not think that it is appropriate to consider manipulations as *tout court* morally impermissible. If we proceeded like this we had to withdraw the designation "manipulation" in all those cases where closer investigation shows that it is doubtful whether or not the relevant mode of acting could be justified *in the respective context.* Instead of excluding unclear and mixed cases for the benefit of a crystal-clear moral assessment I follow the strategy to start with a broader understanding of manipulation which takes seriously our context-sensible and much more shaded everyday experience of manipulative actions. It is in line with this view not to exclude cases of justified manipulation beforehand. Manipulative actions, of course, should never be treated in a light-minded fashion. They always stand in need of justification. It is only on certain conditions that manipulation can possibly be justified. Plausible examples comprise, above all, instances of paternalistic manipulation, especially where this kind of intervention seems to be unavoidable due to the lack of any rational alternative. This is true, for instance, with regard to the upbringing and education of young children. Among those who elaborated this very briefly sketched general attitude toward a normative assessment of manipulation, is Marcia Baron. She convincingly argues that manipulating people is not always wrong although it is generally wrong. This allows for individual cases of manipulation that, under the given circumstances, appear as the best option. Acknowledging such cases does not, as Baron stresses, change the overall approach to consider manipulative actions as generally objectionable. Neither does it mean to swerve from the general demand for justification (cf. Baron 2014, p. 106 f). Especially with a view to the great variety of relevant cases of manipulation this seems to be a reasonable view. (It also finds support in the idea of an accurate moral phenomenology that is meant to pre-date our conceptual decisions and theoretical work in ethics.) Whether or not single instances of manipulation are morally permissible depends on the precise aims that are meant to be realized in the contexts at issue, the intent of the manipulator, the means used in order to secure success, and the mental condition and overall capabilities of the manipulees. Yet regardless of how many cases of justified manipulation we are able to glean and agree upon, this does not neutralize the generally objectionable nature of manipulative actions.[1]

When I now go ahead with introducing some basic conceptual issues I do so in a minimalist sense. My purpose is to establish a suitable framework for discussion of manipulative actions. As a proper starting-point we may raise the following question: does manipulation and how does manipulation violate epistemic authority? Answering this question, first of all, requires outlining my understanding of "manipulation" and "epistemic authority".

The idea of manipulation I am going to work with can be explained as follows. When x (*manipulator*) performs a manipulative action by communicating with y (*manipulee*) the following conditions must be fulfilled: (i) x has the intention to influence y's beliefs or belief formation, her emotional attitudes, her decisions, her motives for acting or modes of acting

---

1 Arguing along these lines, Baron therefore feels committed to a moralized understanding of the term "manipulation". However, she carefully wards off an understanding according to which we should equate "moralized" with "immoral". "In holding that 'manipulative' is a moralized term I was claiming not that it is part of the meaning of the term that what it refers to is *immoral*, but only that it is morally objectionable – but of course something can be morally objectionable yet be morally justified and in that sense not immoral" (Baron 2014, pp. 98 f).

in a non-argumentative though non-coercive way[2]; (ii) x does so with a view to a certain end he has in mind when he sets going the manipulative action; he wants to meet the respective end by arguing y into believing, feeling or acting in the determined way; (iii) while exerting this intentional influence on y x is eager not to let y know about the real nature of their interaction; this policy of concealment or opaque communication, respectively, includes both the end x aims to realize and the specific ways by means of which he strives to do so.[3]

According to the above, the manipulative action establishes a relation between at least two (and possibly more) persons that can be characterized as epistemically asymmetrical insofar as there is a relevant lack of knowledge on part of the manipulee and a corresponding privilege on part of the manipulator. This asymmetry gives rise to the presumption that the epistemic authority of the manipulee is (more or less seriously) impaired. In this vein, the above definition addresses the intuition that lies beneath the four standard accounts of manipulation presently discussed. Correspondingly, manipulation is understood as founded upon deception (*The Deception View*), as imposing harm on the manipulee (*The Harm-based View*), as impairing the manipulee's autonomy (*The Autonomy-Undermining View*) and as bypassing or subverting her rational capacities (*The Bypass or Subvert View*) (cf. Gorin 2014a).

Now, how does the idea of epistemic authority come into play with regard to the agents involved in manipulative actions? There are two aspects that should be considered. On the one hand, there are requirements of rational and moral agency. On the other hand, one may link up epistemic authority with the idea of expertise. As to the first aspect, "epistemic authority", generally, refers to a properly functioning reason-responsiveness which is a prerequisite of rational behavior. In particular, there are certain rational standards with regard to meeting epistemic ideals that specify our idea of epistemic authority. These standards, for instance, determine how belief formation and belief changes, if ideally rational, should take place,

---

2 Or: in a not straightforwardly and manifestly (strongly) coercive way. In the present context I cannot enter discussion of the notion of coercion although it is important to give a more precise account of how manipulation and coercion are related to each other and nonetheless differ. According to my understanding, the designation "manipulative actions" picks out certain modes of acting that are placed somewhere in between an ideally rational behavior of reasoning and persuasion, on the one side, and various forms of strong coercion and violence, on the other side. Manipulations are clearly distinct from both rational persuasion and strong coercion, especially if the latter is mixed up with physical threat. Yet I take it that manifestations of manipulation allow for gradually approaching the one or the other opposite. There are manipulations that, for instance, come close to *coercive offers* and therefore are farther away from rational and moral acceptance than, for instance, manipulations that merely consist in emphasizing different aspects of one's talk according to different audiences. As weak forms of coercion may be indistinguishable from manipulation, it also seems that there is no overall clear-cut distinction between rhetoric and manipulation: the former can gradually transform into the latter.

3 In this respect, manipulation and coercion largely coincide. Both are distinct from violence that exclusively operates on the body, and they seem to be distinct from violence in the very same manner. "Not all violent acts coerce, and not all coercion uses violence. Some acts of violence aim only at another's body: for example, acts done by those who have run amok, what we call 'mindless violence'. Violence can be mute and brute. It need not demand anything of its victims or of others; there may be no implied conditions that victims or others can meet in order to avert it. Coercion (including coercion that uses violence) is different: it has propositional content. Coercers have to communicate with those whom they coerce, and fail if they merely destroy agents whose compliance they seek. This is sometimes hard to see because victims of coercion may also be victims of violence inflicted by their coercers. For example, if a coercer tortures a child in order to get her to reveal where somebody can be found, the child is a victim both of coercion and of violence: violence is the means to her coercion. However, other examples show that victims of coercion and victims of violence undertaken to coerce may be distinct: if a coercer tortures a child in order to get her father to reveal somebody's whereabouts, of which the child knows nothing, then it is the father who is the victim of coercion, although violence is done to the child. It is the father who can comply or refuse to comply; the child can do neither. In yet other cases coercers inflict no violence. They may rely on threat, menace and gesture that suggest varied harms to achieve their ends. Expert coercers concentrate on securing compliance; violence is important to them only when it produces results more effectively than other approaches" (O'Neill 2000, pp. 82 f).

how one should gather evidences and take seriously counterevidences, whether and on what conditions testimony could be considered a reliable source of knowledge, and so on. The second aspect that comes to mind when talking about epistemic authority refers to an agent's knowledge, especially with regard to certain fields of knowledge. A person's expertise may include a broad range of different fields and types of knowledge. For instance, it may cover technical, mathematical, and ethical knowledge.

What is important in the present context is that both aspects, that is, the requirements and methodological ideals of rational and moral agency, on the one hand, and the idea of expertise, on the other, need to be distinguished from various forms of social authority (cf. De George 1985, pp. 26-61). Especially with a view to manipulative actions, it is crucial to recognize that epistemic authority does not include and must be distinguished from executive authority which is a special form of social authority. "In general, an executive authority has the right or power to act for or on someone else. A nonexecutive authority does not" (De George 1985, pp. 22). Epistemic authority is a non-executive type of authority. Rational agency, for instance with regard to belief changes or deliberative processes that aim at determining the right mode of acting, is a *content-related* form of thinking, judging and communicating that operates in terms of rational persuasion. Contrary to this, if someone acts on behalf of her social authority she does not expect that others follow her views or demands solely on basis of objective evidence and arguments. Rather, she relies upon a certain social structure, including hierarchical relations or relations of subordination. It is on behalf of a certain social position and the relating equipment with power, authorization, and prerogatives that the agent's demands are rendered efficacious and are meant to be rightfully efficacious. The other person who is expected to meet the relating demands is not addressed with a view to her content-related rational abilities. She is addressed with regard to her *attitude* towards her social embeddedness in general and her social roles in particular.[4]

Based on the above definition of "manipulation" and "authority of reason" we can now answer the question whether manipulation does violate and how it does violate epistemic authority. Given that we agree on the above characterization of manipulation as an asymmetric relation we may add a further question: Whose epistemic authority is in danger to be violated owing to manipulative actions?

On the one hand, *the manipulator* acts in the light of his unchallenged epistemic authority insofar as, on usual conditions, he has no evidence to doubt that he really has the intention he has, that he wants to realize the end he is aware of as ultimate purpose of his manipulative interference. Acting as an epistemic authority in this way, of course, does not include complete control on the factual success. However, in order to "sincerely" and determinately push through his deceiving project, that is, maintaining the intentional structure of his manipulative action, he must inevitably (thoug possibly unnoticed) transgress the limits of epistemic authority: for the sake of his success he must act as if having epistemic authority would, by the same token, legitimate the exertion of executive authority.

The manipulee, on the other hand, is engaged in forming beliefs, making up her mind and pondering motives and modes of acting *that unbeknownst to her the manipulator has made*

---

4 For a special interpretation of the above distinction see Hampton 1998, pp. 83-122. Jean Hampton juxtapposes an *authority of reason* whose obligatory force is grounded in necessity (i.e., in necessary reasons to act, choose, or believe in a certain way in certain circumstances), on the one hand, and a *psycho-social authority*, on the other. In case of the latter the obligatory force of "its reason to act, choose, or believe in a certain way in certain circumstances makes reference (only) to certain contingent facts about the society and the psychology of the people who take themselves to be subject to the norm. Error theories, expressivist theories, and cognitive theories of normativity presuppose this conception of normative authority" (Hampton 1998, p. 99).

palatable for her. To the extent that there is an epistemic authority operative on part of the manipulee, we must assume that its normal efficiency (e.g. with regard to gathering and checking evidences and counter-evidences) is more or less seriously hampered. Given that there is at least some amount of epistemic authority left on part of the manipulee, it does not suffice to argue that someone falls prey to manipulation because a manipulator successfully conceals his true intention, thereby deceiving the manipulee. To some extent at least, it also must be possible to ascribe epistemic responsibility to the manipulee who, in some ways or others, fails to (fully) realize what is going on. Given that we encounter real cases of manipulation the manipulee's epistemic authority undoubtedly is impaired. Yet it is not annulled as it were in case of a straightforward loss of intellectual competence as it may occur, for instance, in the wake of brain surgery. In terms of competence the manipulee's epistemic authority is in working order[5] although its actual manifestation is inadequate (or "dysfunctional") due to another person's willful interference. The manipulee, so far as her first-personal experience of the interaction with the manipulator is concerned, acts within the proper limits of epistemic authority: she is sincerely engaged in the relating business of deliberating or acting. She nonetheless is not or only gradually aware of the epistemic role played by her interaction partner. However, as long as we consider her as a competent rational and moral agent at all, we must assume that, *in principle, she could have seen through the attempt of deceiving her* (and actually would do so on proper conditions). The above reference to *the principle of alternative possibilities* (someone could have decided or acted otherwise than she actually did) means that notwithstanding given restrictions on the manipulee's freedom of action, which come in the wake of successful manipulations, we do ascribe freedom of will and, correspondingly, freedom of thought to the agent. Refraining from doing so is neither in the interest of the manipulee whom we then would not consider a rationally and morally respectable agent any more. Paradoxical as this may appear at first sight, it is not in the interest of the manipulator either. In order to realize his manipulative intention the manipulator must be interested in arousing the impression that the manipulee's compliance has been granted voluntarily.[6] It is therefore important for the manipulator, too, that we do not categorically deny the manipulee's ability to see through manipulative encroachments on her mental life.

It is, however, a different question whether a manipulee can *reasonably be expected* to unmask a manipulative action *given the particular circumstances at hand.* Answering this question, among others, required investigating the more or less skillfully practiced deception on part of the manipulator, the overall character of the personal or impersonal relation between manipulator and manipulee as well as the latter's former experiences with similar situations. To the extent that the manipulee given the concrete situation, could unmask the manipulation but abstains from doing so it is legitimate to raise the issue of shared responsibility for manipulative actions notwithstanding their epistemically asymmetrical structure. Considering the possibility of shared responsibility on conditions of manipulative actions does not amount to denying that the responsibility originally and to a considerably larger extent lies with the manipulator. Still, granting that the manipulee's response, on

---

5 Strictly speaking, assuming this to be so presupposes that the manipulee's history of epistemic experiences has not mostly or strongly been determined by exposure to manipulative actions. For simplicity's sake I suppose here that we are talking about a single episode of manipulation. Hence I leave untouched the question what sorts of effects are to be expected from habitualized manipulations.

6 Depending on the concrete circumstances it may be no easy task at all to arouse this impression. It requires making the manipulee feel as if she responded to a legitimate manifestation of epistemic authority while the other person actually wants her to acquiesce in demands that originate from a self-appointed social authority.

certain conditions, *can* give rise to a co-responsibility acknowledges and strenghtens her autonomy as a rational and moral agent. Given appropriate conditions, which offer a real chance to unveil the manipulator's true purpose, we indeed expect the manipulee to regain the stance of voluntary action by either approving or disapproving of the manipulative action. The epistemic asymmetry in typical cases of manipulation characterizes the manipulee's acting as non-voluntary although she may (like to) consider herself to act voluntarily. In a garden-variety of cases of manipulation the overall situation, however, seems to be more adequately described by acknowledging the non-voluntariness of her action. On this condition, it is plausible to argue that the manipulee does not know the true meaning of her own responses and actions though the question why she fails to acquire this knowledge leaves room for the above sketched intricacies of shared responsibility. Yet, first and foremost, being subject to a manipulative action means that a person is lead to approve of certain beliefs, decisions or actions that in some sense are forced upon her on part of a manipulator who does not care about her own desires, beliefs, and ends. The *non-voluntary* character, as seen from the manipulee's point of view, mirrors this recklessness on part of the manipulator. Usually, the manipulator will act *voluntarily* although it is easy to construct situations in which he is acting involuntary. This is the case when, for instance, a kidnapper threatens to kill someone's child if the father does not succeed in manipulating some third person in order to set going a certain action whose beneficiary is the kidnapper. Contrary to the manipulee's non-voluntary acting, a person who is subject to coercion acts *involuntarily*: he is fully aware of the coercer's intention and his own position; both coercer and coerced are clear about the fact that the latter does not want to be treated like that. What then is the benefit of going beyond the usual distinction between voluntary and involuntary actions and considering non-voluntary actions as a third category? As argued above, using this additional category allows for a more accurate description of the manipulee's epistemic stance. Considering the non-voluntary character of the manipulee's compliance we do not need to (tacitly) deny her epistemic authority although we are ready to grant that its workings are (considerably) impaired. Characterizing her response as a non-voluntary mode of acting leaves room for a more active role on part of the "victims" of manipulations: under certain circumstances, she may somehow avoid to become aware of the irregular, that is, rationally insufficient character of the belief formation process she is running through.

The following section takes up the above considerations, including the intuitions grasped by the four standard accounts of manipulation (deceiving, doing harm, undermining autonomy, bypassing or subverting rational capacities), as a useful starting point. Yet I shall argue that they neither comprise the variety of cases nor the full complexity of paradigmatic everyday experiences of manipulation. I shall designate those cases that complicate and partially undermine the standard views of manipulative actions as "borderline (BOLI) cases". I take it that analyzing such cases challenges the idea that the above-mentioned standard views could be applied in an easy and unequivocal way. Part of this is to make explicit certain idealizing assumptions that, for instance, seep into the analysis by restricting one's attention to episodic instances instead of considering manipulation in the light of historical concerns, both with a view to the habitualized of modes of acting and biographical 'depths' of the agents involved.

II. As announced earlier, my proposal is to analyze BOLI cases of manipulation by referring to Leon Festinger's theory of cognitive dissonance (CODI). First, let me shortly explain why I think that it is suitable and promising to work with this theory in the present context. Festinger starts his investigation by pointing out that in human minds unfit relations between different cognitive states regularly occur. These relations either hold between different beliefs

or between beliefs and those actions that one reasonably expects to be motivated by these beliefs. The unfit relations at issue are called "inconsistencies" or "dissonances". Two elements x and y stand in a dissonant relation if, "*considering these two alone, the obverse of one element would follow from the other. [...]* x and y are dissonant if not-x follows from y." (Festinger 1957, p. 13) Based on this definition, Festinger introduces the following theses:

> 1. The existence of dissonance, being psychologically uncomfortable, will motivate the person to try to reduce the dissonance and achieve consonance.
> 2. When dissonance is present, in addition to trying to reduce it, the person will actively avoid situations and information which would likely increase the dissonance (Festinger 1957, p. 3).

The sum total of all those modes of behavior that are suited to reduce dissonance could be called "strategies of dissonance reduction". Festinger himself does not use the term "strategies" in the present context. It is indeed questionable to do so because any talk about strategies assumes an explicit choice to avail oneself of certain patterns of response. Yet, as I will explain shortly, it is not clear whether it is suitable to conjure up the terminology of choices and consciously employed decision procedures when it comes to describe different modes of coping with dissonant cognitive states. Festinger does not do so. He explicitly denies that human beings are engaged in reflection and deliberation when trying to reduce dissonance. One distinctive benefit of the CODI account is that it allows for understanding human behavior within a broad range of different types of situations. In particular, it sheds light on attempts to rationalize inconsistencies. This is remarkable insofar as an often used strategy of manipulation consists in arguing other persons into adhering to exploded views or giving up well-grounded ideas by encouraging a more or less biassed treatment of evidences, introducing additional, yet inconclusive arguments or constructing indirect lines of reasoning. Proceeding like this, the agent, on the surface of things, produces "justifications". He does so by ignoring, downplaying or misconstruing the relevance of counter-evidences.

What is the alleged role of rationalization with regard to our discussion of manipulation? Two issues force themselves upon us which I would like to mention though I cannot dig into them here in a more detailed manner. From the point of view of the manipulee who is engaged in reducing more or less obvious inconsistencies, it is, first, near at hand to think about "self-manipulation". It is, however, controversial whether it is of any help to introduce this term which even may appear as self-refuting. Not surprisingly, the same objection has been brought forth with regard to the term "self-deception" (cf. Rinofner-Kreidl 2012a, 2012b). Secondly, analyzing manipulative actions we should bear in mind that there is a variety of different forms of cooperation human agents enter and a variety of motives for doing so. In some cases forms of cooperation and motives present themselves as mutually dependent and vary accordingly; in other cases they do not. Occasionally, human agents cooperate with regard to jointly realizing ends that are unanimously acknowledged as worthwhile or even vital for living a human life. Yet they also cooperate for a variety of other reasons. They link together in order to commit murder or robbery, to exploit the working power of the poor and uneducated, or to instigate others to join for cyber-mobbing. What is of primary interest in the present context (and this leads us back to the issue of rationalization) is that human agents also cooperate, consciously or not, for the purpose of shaping their own minds, that is, reworking their past experiences and their self-understanding. Things like that need not be done within the horizon of solitary reflection. In a certain sense and intensity, we are constantly occupied with responding to other persons' experiences and self-understandings without even realizing that we are so engaged. Part of this also is that, for the better or worse, we are used to play on other persons'

inconsistencies, as Festinger puts it. This can take a variety of different forms that should interest us if we are about to investigate manipulations. People, for instance, cooperate in order to share certain demanding tasks like upbringing their kids or re-shaping the communal caring institutions for the elderly or for disabled persons. However, on a different level of description it can also be adequate to depict what they are doing as an attempt to render invisible their fears and accumulated scruples or disappointments. Thus, whenever we are absorbed in individual or collective actions we shape our public life and the world as a whole and ourselves, that is, the mindset by means of which we approach the world.

As I will argue in the following, we should take into account a deep level description with regard to certain cases of manipulation. Doing so we should also feel encouraged to specify the precise meaning and range of so-called "cooperations". In any case, it is natural to include tacit and pre-reflective modes of cooperation which seem to be widespread in all human societies. I take it that the relating patterns of behavior differ depending on how more or less remote they are from processes of reflective endorsement and paradigmatic cases of explicitly acknowledged (manifest) forms of cooperation. Take, for example, the tacit and pre-reflective modes of communication and cooperation that are triggered by processes of emotional contagion which crop up on occasions like terror attacks, running amok or public weddings. These processes, of course, are different from the pre-reflective modes of communication and cooperation that are involved in the habitual rationalization of an alcoholic's akratic behavior which is tacitly supported by his wife and family in order to ward off outbursts of aggression. What is important, in the first place, is whether we are able to make sense of and define a suitably broad range of different forms of social cooperation, including unusual or subliminal forms that on first sight may appear paradoxical. Basically, *cooperation* requires that there is some non-incidental mutual relation between the agents that comes into being on occasion of a certain (succession of) action(s) and that the latter is adequately characterized by holding that *the mode of one's acting is in accordance with another's mode of acting*. Thereby, "acting in accordance with" does not necessarily require that the agents involved share an overall *common* aim whose realization is brought about by acting together.

According to a widespread view, the idea of cooperation and the idea of manipulation are mutually exclusive. Acknowledging this view, which brings to bear the above-mentioned standard items (deceiving, doing harm, undermining autonomy, bypassing or subverting rational capacities), one may argue as follows. There is good evidence to consider manipulative actions as falling beyond the scope of cooperative actions. There is not even a minor overlapping between the two. This seems true whatever liberal modes of interpretation we advance with regard to the notion of cooperation. Accordingly, embarking on manipulation amounts to a harsh denial of cooperation. Moreover, it is the manipulee who has to pay the prize for the manipulator's autonomous decision to realize the action in question. Human agents who are exposed to manipulative actions risk losing control over their own processes of thinking, belief formation and acting. In any case, it is the very intention of the manipulator to withdraw the power of control from the manipulee with regard to her own mental states and actions. Realizing this intention is a fundamental encroachment on the autonomy of another person whether or not this other person, in the concrete situation, is fully or only dimly aware of what is going on or even totally ignorant of it. Given this to be a correct description of the received view on manipulation it is near at hand to conclude that for rejecting the manipulator's mode of acting it is sufficient to argue as follows: The other person would deny, and justifiably denied, being treated like that if she were (fully) aware of what is going on (given that there is a real chance for her, to attain this awareness).

For the present purpose, let us put aside those special cases of *weak paternalism* in which there is no other option than taking control over issues that are of vital interest for the other person

who is unable to do so herself. In such cases we are faced with a severely reduced physical condition that excludes reasonable judgment, for instance due to so-called persistent vegetative state[7] or otherwise severely retarded mental state. Having in mind such situations, which require different treatment, we could enter discussion whether cases of weak paternalism should be classified as exceptional cases of a justified manipulation. I hesitate to do so because these cases differ from other instances of justified manipulation with regard to the crucial aspect of the manipulator's intention to deceive the other person and his denial to disclose his true and overall purpose. Leaving aside weak paternalism and similar issues we focus on the majority of manipulative actions which take place on less extreme conditions. With a view to these non-exceptional cases, I wish to argue as follows. Whenever a manipulator intentionally takes control over another person's mental states (beliefs, emotional states, modes of acting), he acts as if the other person need not be taken seriously as a rational and moral agent. In this sense, the manipulator acts recklessly with regard to the other person's warranted claim to think, decide and act on conditions that no other person is allowed to surreptitiously tinker with in order to disguise the true situation (as it presents itself to the manipulator). To be sure, this is *not* to maintain that it were fully in the agent's own power to determine or modify these conditions if the manipulator abstained from his interference. It is part of the contingencies of human life that agents never have complete control on the conditions of their thinking and acting. Neither do they gain (with presumably very few exceptions) complete transparency with regard to these conditions. However, if a person, according to some received standard, is a rationally competent thinker and morally mature agent she may legitimately demand not to be deceived (or mislead or intentionally be left in the dark) about the conditions of her own thinking and acting. I therefore hold that the recklessness of the manipulator consists in his disrespecting the manipulee's rational and moral authority, that is, in his disregarding or actively overruling it. It is the issue of respect that lies beneath when it comes to assess the permissibility or non-permissibility of manipulative actions. According to my understanding, this is equally true of moral and epistemic assessments. What is the upshot of this consideration? *The manipulee, if undeceived and asked, can reasonably be expected to have good reasons to deny being subject to manipulative actions.* Pointing at special cases of justified manipulation does not annul this general assessment.[8] It merely reminds us that the complexity of those situations in which human agents interact with one another can well lead us to an all things considered judgment according to which a concrete manipulation, *notwithstanding its wrongful nature when considered as a general type of action*, should be accepted in order to secure other benefits (value commitments).

Coming back to the CODI theses should help us to see why this is not the whole story. Or why the manipulee, notwithstanding this basic threat of losing respect and self-respect, can find herself in a surprisingly unclear or equivocal state of mind as far as her resistance to a manipulation is concerned that she is dimly or openly aware of. (It will soon become clear why a certain amount of minimal awareness of the situation is a necessary requirement of the following line of reasoning.) I want to argue that, referring to a certain segment of manipulative actions, the CODI theses helps to understand why those who suffer from the effects of manipulation nonetheless do not actively strive for disentangling from the manipulator's grip, but seem to acquiesce to this detrimental treatment.[9] It is vital for understanding the cases at

---

7 Recently, a presumably more suitable designation has been proposed: "Unresponsive Wakefulness Syndrome" (Syndrom reaktionsloser Wachheit).

8 Take, for instance, situations where a life-threatening danger has to be warded off without having a chance to exhaustingly discuss the situation in due time with the person who risks losing her life.

9 The following description does not take up the further thought that denying a person her proper respect as a

issue that the manipulee's acquiescence does not take place in terms of an explicit agreement and approval. The manipulee, who is at least dimly aware of what is going on, suffers from the manipulator's encroachment. She does so in a direct or literal sense concerning the beliefs, decisions, or modes of action that the manipulator wants her to accept. She also suffers in an indirect sense with a view to her self-understanding that is considerably impaired and challenged by the manipulative action. If this is a correct description how then could it happen that manipulees, in certain types of situations, nevertheless do show the above-mentioned unclear or equivocal state of mind, meaning that they *have good reasons to resist the manipulation, but they also have good reasons not to do so*?[10] While the former tendency has been extensively discussed in the literature the second has been largely ignored or at least not explained in a satisfying manner. This task can be met by utilizing the CODI theory.

From the point of view of the person who suffers psychological discomfort due to cognitive dissonances rationalization is a proper means to avoid putting up with dissonances. From various reasons, an individual agent's rationalizations can be either socially supported and reinforced or socially complicated and impeded (cf. Rinofner-Kreidl 2010). Granted that the occurrence of dissonances, among others, vary with cultural backgrounds and the past experiences of the persons involved (cf. Festinger 1957, p. 15), we may assume that the precise amount of psychological distress that manipulative actions bring forth on part of the manipulee vary from case to case. Rendering invisible or rationalizing dissonances will require correspondingly varying endeavors. Without entering into the relating details, which leave ample room for psychological research, let us capture the basic fact: "Only rarely, if ever, are they [the inconsistences, i.e., dissonances that regularly occur in human minds, SR] accepted psychologically *as inconsistencies* by the person involved. Usually more or less successful attempts are made to rationalize them" (Festinger 1957, p. 2).

If this is right we come to see what might, with due caution and proper restriction to special cases of manipulative actions, be recognized as an indirect and tacit or subliminal benefit on part of the manipulee. Accordingly, there is a strong motivation not to explicitly approve of this benefit and even, if possible, not to realize that one is subject to an action (or a recurring pattern of interpersonal behavior) that forces upon oneself the undesired and degrading role of a manipulee. However, this negative or even hardly bearable action has the unintended positive effect that it helps the manipulee to reduce her cognitive dissonance. Let us take an everyday example, which combines coercion, manipulation and physical threat or violence in order to make clear what is at stake here and what kind of processes are meant by "dissonance reduction".[11]

---

rational and moral agent does have an impact on her motivation and actual (real) ability to offer resistance to those who intend to denigrate, coerce or manipulate her.

10 Compare for instance the discussion on whether autonomy, if conceived as a real capability of real persons instead of an ideal capability of entirely reasonable ideal agents, actually is overdemanding, and particularly so in certain fields of applied ethics (medical treatment). Given that autonomy can occasionally be experienced as "threatening" in this sense, one should reckon with a certain proclivity for voluntary subjection to paternalistic intervention.

11 As indicated above, it is difficult to give a clear-cut distinction between coercion and manipulation (see fn. 2 and fn. 3). Even subsuming the notion of manipulation under the notion of coercion, thereby choosing to work with a more encompassing notion of coercion, is not absurd. Among those who follow this path is Alan Rosenbaum who asserts that "coercion and physical force are not the same thing as many are inclined to suppose. Instead, physical force is an instrument, among others, of coercion and not its essence; for coercion can be shown to occur in the absence of physical force. Other techniques for influencing someone's behavior, such as paternalism, manipulation, psychosurgery, bribes, incentives, and offers, will be of considerable interest when we attempt [...] to further distinguish coercion from its various instruments, as well as to demarcate the often blurry boundary line between the coercive and the noncoercive aspects of these techniques" (Rosenbaum 1986, pp. 38 f).

The humiliated woman: Christine is a 46 years old woman who has been married for more than twenty years. On the whole, her marriage is an average one. It could be worse. At least, this is what Christine herself says if she compares her situation with those of other couples she knows. Christine and her husband Lucas have got two children (Kira, 20, and Paul, 19) who recently left their parent's house. Kira went abroad for doing her studies and Paul successfully applied for his first job in another town not far away. In course of the years Lucas has turned into an alcoholic although he stubbornly denies this to be the case. Especially when drunken, he becomes rude and aggressive, shouting at his wife and scolding her for her lazy, meaningless life as a 'luxury woman', as he says, who spends her time with shopping and painting her face. Regularly, such scenes end with heavy arguing and, occasionally, with Lucas' battering his wife. When this happened for the first time, Christine was deeply terrified and enraged. She was determined to leave her husband. When she told him that she would not any more tolerate his violent behavior and increasing addictiveness, he calmly replied that if she really dared to leave him and turn their intimate matters into a neighborhood gossip, thereby ruining his existence as an entrepreneur, he would thoroughly get her down. This totally unexpected kind of response left Christine speechless. A few days later, Lucas comes back to her previous conversation and tells her that meanwhile he had talked with a friend of his, a lawyer, who had assured him that Christine, given that she could not prove any maltreatment, would not have the slightest chance to get through with her wild stories. Moreover, Lucas reminds her that she does not have any income of her own and no considerable savings and that it certainly will be pretty hard for her to find a job since she has left her former job as a secretary two decades ago. On another occasion, Lucas assures her that, should she decide to talk with Kira and Paul, he would deny everything. This should be easy, according to Lucas, because they never witnessed anything suspicious. They simply had no reason to believe her and certainly would not do so given that they always had admired their smart and entertaining father. Suddenly, the week after, Lucas begins to talk about Christine's and his first dates and how delighted and happy they were having met each other. Lucas also mentions the sad story of Christine's childhood and reminds her how often, in course of the past years, she had told him what good luck she had to escape the dull atmosphere at her parent's house and live with Lucas. In course of the next months, Luca's behavior wavers between conciliatory moods, making Christine promises and trying to drink less, and his well-known outbursts of rage and intimation. Christine feels depressed and paralyzed. She is unable to take the initiative and leave her husband. Moreover, the very self-assured manner in which he speaks with her and his unshakably confidence to have control over the situation exert a deeply distracting effect on her. Especially Lucas' story about how he 'rescued' her from the overly conservative and soul-destroying atmosphere at her parent's home, how she then started an entirely new life do not leave her untouched. On the other hand, she is full of fear and sorrow when she thinks about how an independent life of her own could look like. Yet she feels that Lucas could be right as far as her missing experiences and working practice are concerned. She feels trapped.

How should we analyze Christine's situations according to the CODI account? It is obvious that she has good reasons to get out of her violent and degrading marriage with Lucas. On the other hand, from her point of view, she also has good reasons not to leave her husband, given the good times they had and all the things she enjoyed over the years. In this particular situation of conflicting demands it is impossible to relieve the pressure of the resulting dissonance by reducing its overall importance. Both options (leaving her husband and not doing so) are

very important and consequential to Christine (cf. Festinger 1957, p. 22). The magnitude of the dissonance, which is caused by the existential impact of the elements (options) involved, therefore increases the pressure to reduce it.[12] An important part of the dissonance at issue is the disquieting mixture of feelings Christine has to cope with: gratefulness, joy and happiness when she thinks about the early years of her marriage, shame, fear, anger, indignation and depression as far as her present situation is concerned, feelings of uncertainties and doubts when she thinks about the future. Lucas skillfully addresses this bunch of feelings. His well-chosen stories and reminders increasingly take a grip on his wife. By asking Christina to confirm their shared history as couple and a flourishing family Lucas offers a *reward for compliance.* By telling her about his conversation with the lawyer, by threatening to lie to their adult children and by confronting her with her obvious shortcomings with regard to higher education, professional life and career planning Lucas offers a subtle variety of *punishments for non-compliance.* Both manipulative strategies – rewarding compliance and punishing non-compliance – are efficient because of Lucas intimate knowledge about those things that are important or even essential in his wife's life: staying in good relations with her children, having an intact family, living in safety and harmony. Provided with these information Lucas is able to act as an *expert manipulator* who avails himself of his knowledge about his wife's sense of self and integrity.[13]

It should be noted that, according to Festinger, dissonances are both culture-dependent and dependent on past experiences. We therefore should say that "two cognitive elements may be dissonant for a person living in one culture and not for a person living in another, or for a person with one set of experiences and not for a person with another" (Festinger 1957, p. 15). Granting this to be so, it is near at hand to assume that the better a manipulator is acquainted with the relevant backgrounds the more efficient or successful his manipulation will be. Fructifying another person's dissonances greatly benefits from knowing about the "sore points" in the other person's history of experiences. Obviously, considering the manipulee's relevant self-understanding goes beyond its episodic statement. It requires digging into past experiences and their appropriation. "Self-understanding", if informative at all in investigating different sorts of manipulative actions, must be understood as a "historical" notion which might even have moral implications. If we agree that we are stronger obliged not to disappoint the expectations of our dearest that we will behave in a trustworthy manner towards them, manipulating those with whom one is intimately connected and familiar, therefore having the best chance to push one's ideas through (as an expert manipulator), is even worse from a moral point of view than trying to non-argumentatively manipulating strangers.

The CODI account helps to understand how persons like Christine cope with deeply ambiguous motivations and incompatible tendencies to act. Going ahead without constantly suffering from the tension produced by the opposing elements is only possible by either changing her behavior, that is, changing her life, or by changing her beliefs. Given that Christine does not succeed in leaving her unhappy marriage the only way to reduce her cognitively dissonant situation lies in re-interpreting it in a way that it loses its depressing outlook. The memories

---

12 Cf. "*If two elements are dissonant with one another, the magnitude of the dissonance will be a function of the importance of the elements.* The more these elements are important to, or valued by, the person, the greater will be the magnitude of a dissonant relation between them" (Festinger 1957, p. 16).

13 In introducing this designation I follow Onora O'Neill who coined the phrase *expert coercer.* Cf. O'Neill 2000, pp. 82, 90-95. See e.g.: "An unrefusable 'offer' is not, indeed, one where non-compliance is made logically or physically impossible for all victims; it is one that a particular victim cannot refuse without deep damage to sense of self or identity" (O'Neill 2000, p. 91).

and thoughts Lucas' utters, his ingratiating tone as well as his angry reproaches and threats all are meant to manipulatively stimulate a re-interpretation of the entire situation and Christine's part in it, in particular. Lucas' non-argumentative influence, which pushes the right buttons by addressing Christine's individual uncertainties and weaknesses, aims at de-evaluating her considerations about having a divorce. The more Christine suffers from her incompatible motives and related alternatives for acting and self-understanding, the more she will be susceptible to Lucas' manipulative interventions. In the light of Lucas' offers – in terms of both rewards for compliance and punishments for non-compliance – Christine is likely to gain a more positive interpretation of the overall situation and, consequently, will give up her project of starting a new life of her own.[14] If this is the intended outcome on part of the manipulator, Lucas' interventions have succeeded in making her "*complicit in a way in which brute violence does not*" (O'Neill 2000, p. 89).[15] Complicity does not undermine the manipulative character of the act. Yet it renders the situation more complicated in terms of shared responsibility. Hence we may talk about BOLI cases of manipulation.

Let us take stock. What do have we attained so far? I have argued that the CODI theory gives us a proper theoretical tool to understand the deeply ambivalent motivational situation on part of the manipulee in BOLI cases of manipulation. Following this account can lead to the prima facie paradoxical statement that the manipulee, however indignantly denying if explicitly asked, "receives support" from the manipulator in neglecting to adequately see the situation at hand and, therefore, tacitly approves of the manipulator's attempts to argue her into certain beliefs or modes of acting (or forbearance, respectively) *if this presents itself as the only efficient way of reducing her cognitive dissonance*.[16] On certain conditions, it will even be plausible to describe the epistemically deficient grasp on reality on part of the manipulee as a paradoxical striving for self-relief[17] which may go hand in hand with self-deception. That is, the manipulee looks for "receiving support" or gladly accepts support in neglecting to adequately attend to her own self and self-understanding.[18] As stressed above those specific brands of manipulative actions that can be analyzed by means of CODI do not represent the most typical cases of manipulation. Yet if the above analysis is on the right track, we should also take into account cases of irregular manipulations which appear as more complex and multi-layered interactions in the way described above. Their "rationale" or overall nature cannot be understood if we exclusively restrict ourselves to the more obvious aspects. A CODI based analysis of manipulative actions offers a paradigmatic view on how deeply and how tightly the psychological make-up of human beings is connected with various forms of social interacting.

III. Current philosophical literature on manipulation offers analyses with regard to the following aspects: a) the manipulator's intention to non-argumentatively influence another person so

---

14 Cf. "Manipulation can make someone fall short of one ideal while causing her to meet another ideal" (Barnhill 2014, p. 63).

15 Though O'Neill refers to coercive encroachments when uttering the above statement it also perfectly fits the situation of manipulation as described above.

16 Cf. "In general, establishing a social reality by gaining the agreement and support of other people is one of the major ways in which a cognition can be changed when the pressures to change it are present" (Festinger 1957, p. 21).

17 Although this is a markedly Heideggerian idea ("Seinsentlastung") one may benefit from it in other contexts without going into the details of Heidegger's existential-phenomenological ontology.

18 Relating to the example discussed above, an additional complication can occur which I cannot dig into in the present context. It can happen that self-deception or a thoroughly biased access to reality (e. g. in terms of a resentment) also comes into play on part of the manipulator. Of course, the manipulator's epistemic authority need not remain unchallenged and should not be considered unchallengeable.

as to believe something the manipulator wants him to believe or to act in a particular mode the manipulator wants him to act; b) the rational capacities of the persons involved and, in particular, the relating capacities of the manipulee; c) the overall harm done by means of manipulative actions; d) the impairment of the manipulee's autonomy. Contrary to standard accounts of manipulation, the CODI theory offers conceptual tools and empirical investigations to understand why manipulations operate efficiently, and particularly so in certain types of cases. Among these are those that have been associated with the so-called "battered women syndrome" which refers to a range of symptoms manifesting an acquired helplessness and co-dependence, low self-esteem or even self-denigration, or proneness to strong feelings of guilt and shame. As argued above, there are cases in which manipulative interventions lead to dynamic social interactions in terms of an unintended and unacknowledged tacit complicity on part of the manipulee. This very peculiar response helps to reduce the manipulee's cognitive dissonance which, in the first place, occurred owing to the manipulator's encroachment. Undoubtedly, interaction of this kind impairs the manipulee's autonomy and leads her to go far beyond a usual understanding of rational belief formation and decision-making. It nonetheless is rational at least in terms of self-preservation.[19] On these conditions, it is not surprising that in such cases the manipulee tacitly supports the manipulator's policy of concealment or opaque communication which is part of the definition of manipulation (see part I). If this is a correct description of the relating cases, we may assume that BOLI cases of manipulation will even be more "efficient" or successful than can a broad range of coercive actions that lack this additional aspect of complicity though they do not operate on physical threats either (and are akin to manipulative actions in this latter respect). To be sure, from a practical and a political point of view, acknowledging complicity as part of BOLI cases of manipulation is a sensitive and highly problematic issue. If it is true that manipulees, at some deep (unconscious) and hidden level of interaction, can have a paradoxical and ambivalent stake in cooperating with their manipulators, it surely becomes more difficult to publicly take responsibility for manipulative actions, criticize and restrain them or get rid of them. As the above discussion of the idea of an expertise in manipulating ("expert manipulator") has shown, it is worth pondering whether the topic of manipulation calls for moral considerations in terms of the difference between strangers and non-strangers. In any case, analyzing BOLI cases of manipulation requires taking note of individual circumstances and social relations holding between those who manipulate and those who are subject to manipulative actions. At this juncture, it is helpful to take up those conceptual distinctions I introduced in the first part, namely the distinction between objective (epistemic or moral) authority, on the one hand, and social authority, on the other. We may do so by responding to an objection Moti Gorin advanced against standard accounts of manipulation recently. This objection is directed towards the so-called *Bypass or Subvert View of manipulation* which is introduced as "the dominant view of interpersonal manipulation" (Gorin 2014, p. 59). It holds that manipulation necessarily involves the bypassing or subversion of the manipulated agent's rational capacities (Gorin 2014, p. 51). In challenging this approach the author construes several exemplary countercases that are meant to show that it is not true that manipulation necessarily undermines the manipulee's rational capacities (Gorin

---

19 What appears "irrational" from the point of view of the immanent standards of an idealized notion of human reason need not be irrational in terms of psychological self-preservation. In this vein, one may argue that on certain conditions it can appear rational not to insist on rational standards. There are different levels of self-preservation which comprise (for only mentioning the bottom-up and top-down operating extreme poles), deep-level instinctive behavior and highest-level self-determination with a view to certain ideals of acting reasonably and morally responsible.

2014, pp. 55-57). There are situations in which the former even enhances the latter as is evident, for instance, from certain paternalistic decisions physicians may feel inclined to (Gorin 2014, pp. 55f). In concluding, the author states that the provision of good reasons and sound arguments can be used manipulatively (Gorin 2014, p. 59). Depending on the given situations and the overall ends an agent may try to realize this seems to be a plain truth. Take, for instance, the following situation: I utter a statement (upon a certain matter of fact) which I believe to be true in the presence of another person who, as I know for sure owing to the deliberate indiscretion of a common friend, does not trust me, rather expects me to deceive him whenever possible. In this situation I have used good reasons and sound arguments manipulatively. Gorin comments on this possible track as follows: "One implication of this result [that it is possible to make use of good reasons and sound arguments in a manipulative manner, SR] is that insofar as manipulation is thought to be morally problematic, providing others with good reasons and sound arguments can sometimes be morally problematic." (Groin 2014, p. 59) I consider this statement to be misleading if not straightforwardly false. The trouble arises owing to the reduced form of its utterance. Providing others with good reasons and sound arguments can, of course, be morally problematic if they are utilized in such a way that one defies rational demands and, instead, plays on contingent or merely subjective aspects that one considers suitable to accomplish one's goal by non-rational means. What is morally problematic therefore is not the plain usage of good reasons and sound arguments but the fact that they are advanced on certain conditions *that are independent of their content.* These are: i) the reasons and arguments function merely as a means for realizing the manipulator's end which goes beyond the (supposed) truth-conduciveness of the reasons and arguments in question; ii) the manipulator takes care to conceal i) from the other person he is communicating with. On these conditions, it is a merely arbitrary fact whether the reasons are good or bad and whether the arguments are sound or faulty. The difference does not matter *in the situation at issue* in which the manipulator, ultimately, utilizes the other person's rational capacity by subordinating it to an arbitrary role of social authority that he ascribes to himself. Without mentioning the opposing notions of epistemic authority and social authority Claudia Mills appropriately summarizes the crucial difference: "A manipulator judges reasons and arguments not by their quality but by their efficacy. A manipulator is interested in reasons not as logical justifiers but as causal levers. For the manipulator, reasons are tools, and a bad reason can work as well as, or better than, a good one" (Mills 1995, p. 100 f).

If interpreted in a properly cautious way, we may nonetheless consent to Gorin's statement that it is possible to manipulate others *without* bypassing their rational capacities, actively interfering with these capacities, exploiting an inherent flaw in them or otherwise hindering them from functioning in a proper way (Gorin 2014, p. 58). There is, of course, a difference between abstaining from addressing someone's rational capacity and actively undermining it. However, I do not agree with the author's view that, the above (bypassing-, interfering- and exploiting-) options excluded, we should describe the situation by holding that "while I manipulate you, I fully engage your rational capacities" (Gorin 2014, p. 58). Though this is how things may appear from the point of view of the manipulee thanks to (or at least supported by) her successful reduction of cognitive dissonance,[20] it does not accurately describe the

---

20 This explanation is by no means merely psychological as one may object. It touches upon the intentional structure of the relevant actions, the mutual perceptions of the persons involved, mutually displayed emotions and mutually accessible judgements; and it also includes the socially acquired self-understanding of the agents which clearly is at stake in manipulative actions.

situation from the manipulator's point of view. The manipulator is in a privileged position. He has additional information at his disposal which concerns the overall (deceiving!) nature of the interpersonal relation at hand. To be sure, it could turn out that what I want you to believe or what I want you to do is a true belief and is (morally) right to do. Even in this case it would, however, be inadequate to talk as if you acquired a justified true belief (or knowledge) or as if you effectuated a morally justified mode of acting. To be sure, this is what could be expected if I actually fully engaged your rational capacities. Yet I did not do so. I merely relied upon your (habitual practice of your) rational capacities as a proper means to attain my arbitrary (non-rational) end. From the point of view of rational and moral agency, this is a fundamental difference – whether or not it remains unnoticed in the practical context at issue. Proceeding like this, I do not acknowledge you as a reasonable and morally responsible co-agent. When treating others in this way, I, however, risk to lose my own claim to reason; I think and act as if it were insignificant whether others acknowledge me as a reasonable and morally responsible agent. If my previous considerations are right this is a totally flawed idea. Reason works in a strictly non-exclusionary manner, both in theoretical and practical contexts. It is, in principle, sharable though not always actually shared.

REFERENCES

Baron, M. (2014), "The *Mens Rea* and moral status of manipulation", in Ch. Coons & M. Weber (eds.), *Manipulation. Theory and Practice*, Oxford University Press, Oxford, pp. 98-120;

Barnhill, A. (2014), "What is manipulation?", Ch. Coons & M. Weber (eds.), *Manipulation. Theory and Practice*, Oxford University Press, Oxford, pp. 51-72;

Coons, Ch. & Weber, M. (2014), *Manipulation: Theory and Practice*, Oxford University Press, Oxford;

DeGeorge, R.T. (1985), *The Nature and Limits of Authority*, University Press of Kansas, Lawrence, Kansas;

Festinger, L. (1957), *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford, Calif.;

Gorin, M. (2014a), "Towards a theory of interpersonal manipulation", in Ch. Coons & M. Weber (eds.), *Manipulation. Theory and Practice*, Oxford University Press, Oxford, pp. 73-97;

----. (2014b), "Do manipulators always threaten rationality?", *American Philosophical Quarterly*, 51, no. 1 (January), pp. 51-61;

Hampton, J. (1998), *The Authority of Reason,* R. Healey (ed.), Cambridge University Press, Cambridge;

Mills, C. (1995), "Politics and manipulation", *Social Theory and Practice*, Vol. 21, No. 1 (Spring), pp. 97-112;

O'Neill, O. (2000), *Bounds of Justice*, Cambridge University Press, Cambridge;

Rinofner-Kreidl, S. (2010), "Selbsttäuschungen. Zur Sozialität pseudo-reflexiver Handlungen", in Ch. Brünner et al. (eds.), *Mensch – Gruppe – Gesellschaft*, Neuer Wissenschaftlicher Verlag, Wien, pp. 281-298;

----. (2012a), "Self-deception: Theoretical puzzles and moral implications", in D. Lohmar and J. Brudzinska (eds.), *Founding Psychoanalysis Phenomenologically. Phenomenological Theories of Subjectivity and the Psychoanalytical Experience*, Springer, Dordrecht et al., pp. 213-233;

----. (2012b), "Selbsttäuschung und Flucht in paternalistische *Lebensformen*. Über ambivalente Formen des Umgangs mit Angst", in D. Goltschnigg (ed.), *Angst. Lähmender Stillstand und Motor des Fortschritts.* Grazer Humboldt-Kolleg 6.-9. Juni 2011, Stauffenburg Verlag, Tübingen, pp. 97-105;

Rosenbaum, A.S. (1986), *Coercion and Autonomy. Philosophical Foundations, Issues, and Practices*, Greenwood Press, New York, Westport, Connecticut and London.

SESSION 2

SESSION 2

GROUPS' INTENTIONALITY

JACOB HEIM
University of California, Irvine
jheim@uci.edu

COMMITMENTS IN GROUPS AND COMMITMENTS OF GROUPS

abstract

I argue that a group can have normative commitments, and that the commitment of a group is not merely a sum or aggregate of the commitments of individual group members. I begin with a set of simple cases which illustrate two structurally different ways that group commitments can go wrong. These two kinds of potential failure correspond to two different levels of commitment: one at the individual level, owed to the other group members, and one at the group level, which the group as a single body owes either to itself or to some third party. I distinguish the content of a commitment (what must be done for the commitment to be fulfilled) from the holder of that commitment: the party to whom the content is owed. I then discuss examples which support the two-level view of group commitment and show that, even when individual-level and group-level commitments have the same content, they are understood to have different holders. Finally I return to my original cases and argue that a two-level structure of group commitment allows us to make sense of the problems that occur in them.

When we talk about the commitments that exist in a group setting, it seems natural to ask either one or both of the following questions: (1) What is the group committed to; and (2) What are the group's members committed to? I will be arguing that a complete answer to (1) will involve more than a mere list or aggregation of the answers to (2). To put it another way, the commitments *of* a group cannot be fully reduced to the commitments that exist *within* that group. When we closely examine the ways that group membership constrains our activities, we find that these constraints entail the existence of group-level commitments that can only be honored by keeping them distinct from our commitments as individuals.

While I will not be assuming some particular theory of group commitments from the outset, the considerations that I raise will pose more of a challenge for some views than for others. As should be clear by the end of the paper, accounting for two distinct levels of joint commitment will be much easier within a framework like Margaret Gilbert's "plural subject" theory than it would be for a more reductive, individualistic view.[1] If we want to take seriously the standards to which we hold one another in practice, we are going to need to make room in our social ontology for the existence of group-level commitments that cannot be reduced to anything on the individual level.

1.
Group and personal failures

Though we talk easily enough about what this or that group is committed to, we might think that such talk is merely a convenient shorthand for what the *members* of the group are committed to.[2] After all, groups are made up of individuals, and what a group does is in some sense done by those individuals: when someone says that the team wears a yellow uniform, we might interpret her as saying that the *members* of the team wear yellow uniforms. If properties of groups can be fully analysed into properties of individuals, then we should expect that any case in which a group fails to fulfil its commitment can be fully described in terms of individual commitments and individual failures to fulfil them. We will see whether this turns out to be the case.

Let us begin by looking at a case of individual commitment. To avoid unnecessary complications, we will consider the simplest kind of situation in which a commitment can arise: one in which there is only a single agent, making a commitment to himself.

---

1 For the initial statement of her view, see Gilbert (1989). As far as I can work out, everything I say here is consistent with Gilbert's overall theory, but does not presuppose it.
2 Bratman (1999), Brewer (2003), and Wilkins (2002) all seem to be thinking along these lines.

Case 1: I am sitting on my couch watching TV. It occurs to me that I need to go to the store before it closes, and I decide to go as soon as this show is over. When the show ends, I get up and go to the store.

One might wonder whether a case like this counts as a commitment at all. I would answer that there are (at least) two senses of the term *commitment*, which refer to related but distinct concepts. One, the stronger sense, is what we mean when we say, for example, that Martin Luther King, Jr. was *committed* to the cause of racial equality.[3] The other, weaker sense is the one that I am using here. To be committed in this sense does not require any connection to one's deep-seated beliefs or goals; if I tell you that I will meet you at 2:00 then I have a commitment to do so, regardless of how much or little I care about it. This is the sense of commitment that we generally use when we talk about what so-and-so is committed to *doing.* Importantly, one can have a commitment in this sense without having made that commitment *to* anyone else: if I decide to do something then I am committed to do it.[4] My deciding to do *x* is in this way normatively equivalent to telling you that I will do *x*, except that in the former case I owe the performance of *x* to myself rather than to you. In the terminology I will be introducing later, I am the *holder* of the former commitment, and you are the holder of the latter.

Many of the commitments discussed throughout this paper could also be properly described as obligations. I will stick to the term *commitment*, however, as it seems to be the more general concept, and it is at that level of generality that my arguments (and conclusions) are intended to operate. While I see no reason not to assume that all obligations are commitments of some kind,[5] there do seem to be commitments (of a kind relevant to the topic of this paper) that are not, at least obviously, also obligations. In particular, it would be slightly odd to say that in Case 1 I have an *obligation* to myself.

Case 2: I am sitting on my couch watching TV, and I decide to go to the store after this show. But when the show ends, I do not move. I continue to watch TV until long after the store has closed.

Here I have failed in my commitment to myself. I would be entitled to rebuke myself in much the same way that you could rebuke me if I had told you I would go to the store for some purpose of yours.

Now let us move from a simple individual case to a simple group case.

Case 3: You, me, Amy, and Chad are sitting on the couch watching TV. It occurs to one of us that we need to go to the store before it closes, and after a brief discussion we agree to go together as soon as this show ends. But after the show, instead of moving, we continue to sit there.

Something has clearly gone wrong here. Perhaps each of us is waiting for the other to make a move, or perhaps we have all gotten distracted. But for whatever reason, we have made an agreement – the interpersonal analogue of a personal decision – and then failed to comply with it. The failure here is comparable to the one in Case 2. But not all group failures align so neatly with the individual case. Consider the following:

---

3 This example is borrowed from Talbot Brewer (Brewer 2003, p. 562).
4 See Gilbert 2006 (pp. 127-134) for her argument that a decision is a form of commitment to one's self.
5 Since nothing in my argument hangs on this one way or the other, I will not bother arguing for it.

Case 4: You, me, Amy, and Chad are watching TV. We agree to go to the store together as soon as this show is over. When the show ends, you, Amy, and Chad get up and move towards the door while I continue to sit there. You remind me of our agreement, and I vaguely reassure you that I intend to get up any minute now. After a while the three of you grow tired of waiting, and go to the store without me.

Once again, there has obviously been a misfire of some kind. But this is the kind of misfire that can only happen in a group setting; a single person cannot simultaneously perform and fail to perform the same action.

Pre-theoretically, it seems that the most natural way to describe what has gone wrong differs significantly from Case 3 to Case 4. In Case 3, it may be appropriate in some sense to say that each of us failed the others: after all, none of us did what we told the others we would do. But that does not seem to get at the heart of the problem. No one of us was under any particular obligation to get up first, so none of us is in a position to blame our own inaction on the inaction of another. Each of us has a decent excuse, even if it is not quite a good one, for not getting up first – after all, why should it have to be *me?* What complicates matters is that the rest of us have an identical excuse. One could imagine the four of us discussing this very issue, after the fact, and finally coming to the conclusion that the most appropriate thing we could say about it is simply that *we* failed. There need not be anything more to say as far as who failed whom on an individual level. Indeed, we may come to realise that to say anything at all, on that level, would be unfair.

Case 4 is quite different. Strictly speaking, it would still be correct to say that we did not do what we said we would; the plan was for the four of us to go to the store together, and that did not happen. The responsibility for that, which could not be assigned in Case 3, now falls squarely on me. But not only is it my fault that *we* fell short of *our* plan. Additionally, and more importantly, *I* failed the three of you by not doing what *I* agreed to.

What we have, it seems, are two different and apparently incommensurable ways in which group commitments can fail. In one, a group fails as a group, whether or not responsibility for that failure can be traced back to particular members. In the other, an individual group member fails as a group member, whether or not the group itself ends up failing as a result. Both kinds of failure occur regularly in our social lives. Of course, they can and do coincide often enough. But there are many cases in which they do not, and in practice we readily distinguish between them in the ways that we respond. Thus a satisfying account of group commitments should be able to explain failures of both kinds in a way that does justice to our ordinary way of thinking about them.

2.
A distinction about commitment

Given that our task is to explain two distinct kinds of commitment failure, a natural first step would be to try to identify two distinct ways of being committed, to which the failures might correspond. One of the failures, Case 4, is fairly straightforward. It involves a commitment that arises in the context of a group, but as we have seen, the nature of the failure suggests that the relevant commitment belongs to a particular individual only. If failures like Case 4 were all we needed to account for, we might be able to get by with personal commitments alone.

The other failure calls for a more nuanced approach, however, as Case 3 certainly *appears* to be a failure by the group as such, which in turn suggests a commitment by the group as such. This appearance must either be explained, or explained away. In the absence of compelling reasons to view group commitments as personal commitments in disguise, we ought to at least take seriously the possibility that a group can operate as a unified normative entity, capable of having its own rights and obligations.

Once we have admitted the possibility that our everyday talk of group commitments might

refer to something real and irreducible, we can begin to ask meaningful questions about the ways in which a particular group member can relate to a commitment of the group. In order to pursue these questions, I will set aside the question of how joint commitments are formed and will focus instead on some of the ways that such commitments work themselves out in situations where they are understood to exist.

When a group of people undertake a joint commitment, the group acquires a commitment which is different in kind from the aggregated commitments of the individual group members. This difference in *kind* of commitment is one that holds even when the *content* of the group-level commitment does not contain anything over and above the sum of personal commitments within the group. By "content" I mean the actions to be performed, or the conditions that must obtain, in order for the commitment to be considered fully met. Distinguish this from what I will call, for lack of a better term, the *holder* of the commitment: the party to whom the content of the commitment is owed, and (in general) whose right it is to declare the commitment satisfied or not.

Commitments can be differentiated with respect to either one of these features, even if they are identical with respect to the other. If I promise you that I will wash the car and walk the dog, I have two distinct commitments, even though they have the same holder, because they have different content. If I tell you that I will make us something good to eat, and I tell myself that I will make us something good to eat, there are two commitments as well, because they have different holders despite identical content. Admittedly, this is less obvious than in the case of differing content, but consider: if I make food that you greatly enjoy but that I find revolting, you will be satisfied that I have done what I said I would do, while I would not think the same. In this case I have met my commitment to you, but not to myself.[6]

It may be worth mentioning at this point that there is a common type of occurrence in which a group can have a commitment that differs from the commitments of its members, not just in terms of holder, but in terms of content as well. This is the case whenever a group has taken on a commitment but has not yet determined which group members will be responsible for which components of the larger commitment. Opponents of plural subject theory might want to explain this phenomenon away by arguing that each group member does have a personal commitment to do her part in the larger project, even before it has been determined what that part is, and that necessarily the sum of the as-yet-undefined personal commitments will add up to the group's commitment. I do not find this explanation satisfying, partly because, as I understand it, this is just what it means to be party to a commitment larger than one's own. Still, our imagined opponent is onto something, in that whatever actions the group is committed to do must ultimately be made up of the actions of its members. An orchestra plays a symphony, but each note of that symphony is played by a particular member of the orchestra. But even when the content of the group's commitment is nothing more than the aggregated content of the individual commitments – indeed, even if we were to grant that such is always the case – a group-level commitment nevertheless differs from the related individual commitments in that it has a different holder. An example will help to illustrate.

3. Two levels of commitment in practice

If a group of friends go out to a pizzeria and order a large pizza together, the cost of the pizza creates parallel commitments on two levels. As a body, the group sitting at that table has a commitment to pay the full cost of the pizza, while each person at the table has a commitment to pay a fair portion of that bill. However the individual shares get divided, they will add

---

6 In Roth's terminology, I would have fulfilled my *contralateral* commitment but not my *ipsilateral* commitment (Roth 2004).

up to the cost of the pizza which the group is committed to pay for. So the content of the commitments at the two levels are identical.

They do not have the same holder, though. The individual friends each owe it *to each other* to pay a fair share, and it is only the group as a whole that owes it to the restaurant to pay for the whole pizza. If one person is bent on paying less than her share, it is the other friends at the table who have standing to rebuke her; the waiter does not. As long as the full amount is paid, the waiter has no standing to complain about who pays what. And if the stingy friend wants to defend her position – perhaps she ate less pizza than the others and so feels entitled to pay less – she would naturally address herself to the other members of her party rather than to the waiter.

Suppose that she does otherwise: she ignores her friends and explains herself to the waiter directly, insisting that she has contributed a reasonable share and that someone else is going to have to make up the difference. This would not be merely odd, it would be out of line. It would be natural, in such an instance, for one of her friends to protest, "Don't bring *him* into this!". Even the waiter himself could say as much, politely insisting that they settle it among themselves. We would normally regard this as the proper way for him to respond: it is not his business whether a particular group member pays a fair portion of a shared bill. That is a commitment internal to the group, not a commitment between individual group members and the restaurant. She cannot individually make good her commitment with the restaurant, because she has no such commitment. Among the set of her commitments, there is not one whose content is that she pays for her share of the pizza and whose holder is the restaurant. In the extreme case, if she were to simply hand five dollars to the waiter and firmly insist that she will not pay another penny, this would rightly be regarded as a serious offense – not against the restaurant, but against the other members of her party and against the group that they constitute. By trying to isolate a single share of what is properly a group obligation, and presuming to settle it individually with the holder of that group obligation, she has directly violated one of the most basic norms of group membership.

Similar cases are easy to come by. You hire a construction crew to build a shed, and the finished product falls over the first time it rains. Naturally, you demand a refund. It would be outrageous for one of the workers to come to you and say that *his* work on the shed was not the cause of the collapse, and so you are not entitled to a refund on *his* share of the money. He cannot claim to have fulfilled his individual commitment to you, because he has no such commitment. As far as you are concerned, there was only ever a *group* commitment to build a complete shed, and in that commitment the entire group failed as one body. Whose handiwork is to blame, and who ought to be penalized and how, is a conversation for *them* to have.

What these examples show is that we do in fact distinguish, in ordinary practice, between individual-level and group-level commitments. There are real social consequences for failing to respect these distinctions. Commitments on the individual level are held by the group, and are owed to one's fellow group members. Commitments at the group level are held by external third parties, or in the case of a group decision that does not involve outward commitments, the commitment at that level is held by the group as a single entity. Social groups, both formal and informal, tend to be sensitive to these matters and act so as to discourage members from acting as if the group were not there as an intermediary. Outsiders are not to meddle with internal affairs, even if they have a stake in their resolution, except by addressing themselves to the group as a single entity – from the outside, as it were. Likewise, group members are not to act as their own agents where group-level commitments are involved. If a group fails to keep its commitment, but a single member did his part successfully, his defense is to be made in the context of the group and not directly to the third party. Only his group has the standing to accept his excuse as valid.

One kind of case stands out as a possible exception, though I think that in fact it can be most naturally explained on the two-level model I am arguing for. This is the case of a "whistleblower": a member of a corrupt organization (often an employee of a corporation or government office) who passes information about illegal or unethical activities to legal authorities, journalists, or other outsiders. The proper way to think about the whistleblower, I think, is that she is not held to share responsibility for the group's actions precisely because her own actions set her outside the group. She is, in effect, acting for the outsiders as an agent who has infiltrated the group structure. While her whistleblowing activities are generally admirable, it does not make sense to think of her as a member of the group in good standing. Recent history confirms this. American whistleblower Edward Snowden, for example, has been called "a traitor" by government officials of both parties,[7] while Snowden himself claims he was acting on behalf of "the world I love" (Greenwall, MacAskill, & Poitras 2013): just the kind of shift in identification that the two-level model would lead us to expect.

It is generally understood that the whistleblower, in virtue of her subversive activities, is no longer a party to the group-level commitments, or at least not all of them. She may in fact remain devoted to the group's ends, and continue to identify with the group itself, but strongly oppose the means which the group has adopted in pursuit of those ends. If she affirms this after the fact, or if her behavior otherwise suggests that she wishes to retain her group membership, she risks an especially sharp rebuke from within the group. The rebuke is justified, to an extent, because her actions can be seen as involving an implicit claim that she can remain a member of the group in good standing while abiding by some but not all of the group-level commitments.[8] They will say that if she disagreed about means but not about ends, she should have raised those issues internally rather than getting outsiders involved.[9] Such a rebuke would effectively be appealing to a two-level structure of commitments such as I have been describing, albeit not under that name. U.S. Secretary of State John Kerry, for instance, had the following to say about Snowden after Snowden expressed a desire to return to the United States: "A patriot would not run away and look for refuge in Russia or Cuba or some other country. A patriot would stand up in the United States and make his case to the American people" (Serrano 2014).

The fact that her actions open the whistleblower to such a rebuke does not preclude the possibility that she may still be doing the right thing, all things considered. But it is one of those unfortunate circumstances in which doing the right thing comes at a normative cost: she may have to violate the norms of group membership in order to conform with the norms of honesty, civic responsibility, or morality more generally.

I now want to return to my original cases of you, me, Amy, and Chad sitting on the couch. I believe that the two-level view of group commitment that I have been presenting can offer an illuminating explanation for the two ways in which the commitments in that group fail to be met. Recall that in Case 3, we all agree to go to the store together, and then no one moves when the time comes. We are now in a position to recognise this as a failure regarding a group-level commitment: a commitment the group has to itself as a single body. This diagnosis also gives content to the analogy, mentioned earlier, between Case 3 and Case 2, where I am

**4.
The
"whistleblower"
case**

**5.
Group and
personal failures
revisited**

---

7 For a small sample of reactions along these lines, see Weisman (2013) and Serrano (2014).

8 More likely, a rebuke of this kind will fit a pattern identified by Gilbert: someone has the standing to issue a rebuke but, all things considered, it is not morally appropriate to do so. See for example Gilbert (1993, p. 702) and Gilbert (2008, p. 7).

9 Of course, she may have already tried to do this, or it may for other reasons have been unfeasible. There are a number of delicate issues here that invite further exploration.

alone on the couch and fail to go at the time when I had previously decided to go. In both these cases the entity which formed the commitment (and later failed to act on it) was also the holder of it.

The disanalogy between Cases 2 and 3 is that in Case 3 there are also individual-level commitments which each of us owe to the others, to each do our part in pursuing the joint goal of going to the store together. On the one hand, since our individual roles were never clearly determined, none of us is in a position to rebuke anyone else in particular, since there is no determinate fact about who individually failed how or to what extent. On the other hand, each of us as a group member necessarily has at least the minimal commitment to act *in some way* so as to help bring it about that our group-level commitment is fulfilled. Given that none of us did *anything*, it is clear that each of us failed more or less equally in that individual commitment. So while it may not be appropriate to say "I failed you all" or "You all failed me", it would be appropriate to say either "We failed each other" (individual-level) or "We failed ourselves" (group-level). But the most salient of these is the group level: "We failed ourselves", or simply, "We failed".

Case 4 is the reverse of Case 3 in that the most salient failure is on the individual level. As far as the group-level commitment, it is unclear to what extent we ought to say that *we* failed. As mentioned earlier, the original goal of all four of us going was not achieved. Still, three of the four did go to the store, and they did go together. Perhaps the three of you took my inaction as a signal that I was no longer invested in the group commitment, and by leaving without me you signaled in response – as you all had a right to – that you no longer regarded me as part of the group that was committed to going to the store together.[10] In that case, any one of the three of you could say "We went to the store", but I could not: I am no longer a part of that particular "we". This in itself is not problematic, since groups regularly have members come and go without altering the identity of the group. On the present account this general phenomenon would be explained, at least in part, in terms of group-level commitments remaining unchanged while individual-level commitments are more responsive to the shifting boundaries of membership.[11]

In any case, the important thing is that I have failed the three of you by not going. The two-level view confirms and explains this: however it works out at the group level, I have broken my individual-level commitments. I broke my commitment to each of you individually by not acting as a committed group member should, and I broke my commitment to the group by not doing my part in pursuit of the collective goal. Remember that while these commitments have overlapping or identical content, they have different holders: one is owed to the group members, viewed as individuals, and one is owed to the group viewed as a single unit. You could therefore rebuke me either as yourself, or on behalf of the group; it would be equally appropriate for you to say "He failed each of us" (individual-level), and "He failed all of us" or simply "He failed us" (group-level). What these rebukes have in common is that, whatever their source, they are directed to me in my capacity as an individual group member and do not implicate the group itself in what is ultimately *my* failure.

Conclusion   Social life involves commitments, and some of these commitments can only be met when we acknowledge that they are not *ours* in an individual sense. Instead they belong to a group, and

---

10 There are interesting questions here about the ways in which a group can adjust its goals and its membership. Regrettably, I will have to leave them for another time.

11 This, too, seems to me an area worth examining in greater detail; however, to do so here would distract from the main argument of the present paper. A closer examination of the interplay between group- and individual-level commitments in a group setting only makes sense once we have reason to believe that there *are* such levels of commitment, which is what I take myself to be establishing here.

it is only to that group and its members that we can be committed on a personal level. I have argued that everyday social practices reflect a tacit but real recognition of this distinction. It is true that a two-level structure of commitments can sometimes limit our actions in uncomfortable ways. On the other hand, it is just this structure of commitments that allows one, in a not purely metaphorical sense, to be a part of something larger than oneself.

REFERENCES

Bratman, M. (1999), "Shared intention and mutual obligation", in M. Bratman, *Faces of Intention*, Cambridge University Press, Cambridge, pp. 130-141;

Brewer, T. (2003), "Two kinds of commitments (and two kinds of social groups)", *Philosophy and Phenomenological Research,* 66(3), pp. 554-583;

Gilbert, M. (1989), *On Social Facts*, Routledge, London;

----. (1993), "Agreements, coercion, and obligation", *Ethics*, 103(4), pp. 679-706;

----. (1997), "Group wrongs and guilt feelings", *The Journal of Ethics*, 1(1), pp. 65-84;

----. (1999), "Obligation and joint commitment", *Utilitas*, 11(2), pp.143-163;

----. (2006), *A Theory of Political Obligation*, Oxford University Press, Oxford;

----. (2008), "Social convention revisited", *Topoi*, 27(1), pp. 5-16;

Greenwald, G., MacAskill, E. & Poitras, L. (2013), "Edward Snowden: the whistleblower behind the NSA Surveillance Revelations", *The Guardian*, June 11, 2013;

Roth, A. (2004), "Shared agency and contralateral commitments", *The Philosophical Review,* 113(3), pp. 359-410;

Serrano, R. (2014), "Former NSA contractor Edward Snowden says he was a trained spy", *Los Angeles Times*, May 28, 2014;

Weisman, J. (2013), "Boehner calls Snowden a traitor", *The New York Times*, June 11, 2013;

Wilkins, B. (2002), "Joint commitments", *The Journal of Ethics*, 6(2), pp. 145-155.

FRANCESCA DE VECCHI
Università Vita-Salute san Raffaele
francesca.devecchi@unisr.it

THE PLURAL SUBJECT APPROACH TO SOCIAL ONTOLOGY AND THE SHARING VALUES ISSUE

abstract

I address the issue of the unity of social ontology despite the extreme variety of social entities: groups, money, promises, governments, laws, etc. Then, I focus on Gilbert's account of social ontology as an ontology specifically dealing with groups as plural subjects, and face with Gilbert's concept of shared values as values of plural subjects created by joint commitment. I argue that Gilbert's account of shared values is a cognitivist and extrinsic one: it neglects the specific role of values for the constitution of plural subjects and considers values neither as a necessary nor as a sufficient condition for social unity. I suggest that, unlike Gilbert and the main trend in the contemporary social ontological debate, phenomenology provides an axiology that can allow to account adequately for values and to understand values' crucial role for social unity. I discuss Scheler's dividing vs. sharing values thesis and mention Schapp's collective values thesis. Finally I address the question of the collective feeling value.

1.
Variety of social entities and unity of social ontology

Social entities, which pervade more and more our everyday life, are extremely various. Examples of social entities are: driver licences, taxes, euro notes, credit cards, bus tickets, judicial systems, health insurances, penal codes, theatres, governments, symphonic orchestras, walks together, football teams, friends groups, philosophical societies, marital couples, promises, marriage proposals, concert performances, elections, governing body sessions etc.

1.1
The extreme variety of social entities

All of these entities are very different from one another. Yet they represent the object of one discipline: social ontology. Social ontology's origins go back to the beginning of the last century – the expression "social ontology" occurs for the first time in Edmund Husserl's 1910 manuscript entitled "Soziale Ontologie und Descriptive Soziologie" [Social Ontology and Descriptive Sociology].[1] However, social ontology's flowering and development started just in the last thirty years in the domain of philosophical and social science research.

Today social ontology's reception in academia is quite accomplished: social ontology is a discipline which is taught in universities; there are social ontology societies, journals of social ontology have been founded, and the like.[2]

1.2
Two essential features of social entities

Despite the extreme heterogeneity of social entities we deal with in our everyday life, I claim that there are at least two features which essentially individuate social entities and distinguish them from other classical ontological types, such as natural entities and ideal entities. These two features are:
(i) Ontological dependence on intentionality;
(ii) Normativity.[3]

1.2.1
Ontological dependence on intentionality

Social entities existentially depend on individuals' intentionality (at least two individuals are required) both for their creation and for their maintenance and cessation. Without beliefs, perceptions, desires, intentions, memories, feelings and actions relating to social entities, social entities would not exist. Let us consider some examples.

---

1 Husserl 1910, pp. 98-104.

2 See the *International Social Ontology Society* (ISOS), the *Journal of Social Ontology*, the *European Network on Social Ontology* (ENSO), the *Collective Intentionality Conference.*

3 I delt with these two essential features of social entities (ontological dependence on intentionality and normativity) in De Vecchi 2012. On this topic see Searle 2010 and Thomasson 2009.

The promise I did, and its corresponding claim and obligation, are created by my social act of promising; and they will cease to exist only when I satisfy my promise by the corresponding-realizing action. The underground ticket I bought and validated to go to the university would not exist without our collective beliefs according to which that piece of paper is an underground ticket; if we would collectively stop to believe that that piece of paper is an underground ticket, then the underground ticket would cease to exist. Parliaments would not exist without parliamentarians, who let the parliaments live trough their actions and acts, and without citizens who elect other citizens as parliamentarians.

Thus, social entities ontologically depend on individuals' intentionality, just like psychic natural entities (e.g. feelings) and non-natural physical entities (e.g. artifacts), and unlike physical natural entities (e.g. trees) and ideal entities (e.g. numbers).

Psychic entities like the pain I feel in my leg for my skiing fall, or the joy I feel getting a good piece of news about a dear friend of mine, or *social entities* like the request I made to my friend or the taxes I paid, or *physical non-natural entities* like the seat of the underground on which I sit or the building where I give my lectures, all these are entities which depend on individuals and on their capacity to be bearers of intentional states and agents of intentional acts and actions.

Physical natural entities, like trees and mountains, exist independently of individuals and their intentionality about them: this Californian oak exists independently of the fact that I contemplate its beauty; Mont Blanc exists independently of the fact that I ski on Mont Blanc.

Social entities are essentially normative entities. Normativity essentially individuates social entities and distinguishes them from other entities, like artifacts and works of art which, like social entities, depend on individuals' intentionality.

1.2.2 Normativity

Now, the issues that need to be addressed are: what does "normativity" mean when referred to social entities? Why are social entities normative entities?

"Normativity" has many meanings and it may be predicable of other entities which are not specifically social entities, for instance and above all, moral entities like moral laws, moral judgments, moral actions. So, we have to inquiry into the specificity of social normativity: What do I mean by "normative" social entities? What kind of normativity distinctively characterizes social entities?

Social entities imply obligations, rights, duties, claims, permissions, authorizations, licenses, awards, commitments, requirements etc. Social entities are bearers of "deontic powers" – I will adopt here Searle's very convincing concept (and neologism).[4]

Let us consider some examples.

Promising implies a promisor's *obligation* and promisee's *claim*. A university professor has the *obligation* to give her lectures, according to her work contract with her university, and she has the *right* to examine her students and evaluate them. Your invitation to your party *authorizes* me to come to your party. Two friends who decide to go the Picasso Museum together are *committed* to go together to the Picasso Museum. Train tickets *entitle* us to travel with a certain train, directed to a certain destination, on a specific day; if we travel on this train without the corresponding ticket, we would be *liable to penalty*.

All of these kinds of obligations, authorizations, commitments, rights, etc. are not moral, as such. For instance, two criminals, who have committed themselves to bring about a certain criminal action, are consequently the bearers of such commitment, independently of the fact that the content of such commitment is immoral.[5]

---

4 See Searle 1995 and 2010.
5 On the issue of the relation between moral normativity and social normativity, see Gilbert 2013, pp. 5-9, and 2015 (in

1.3
Three kinds of
social entities

I suggest that it is possible and also needed to put the extreme variety of social entities into order. I try to do it, and identify three kinds of social entities:

(i) *Social objects:* they are specifically objects such as driver licences, train tickets, euro notes, credit cards, car insurances, laws, etc.
(ii) *Social subjects*: they are specifically collective entities such as symphonic orchestras, governments, societies, football teams, marital couples, groups of friends, etc.
(iii) *Social acts, events and processes*: social acts such as requesting, promising, promulgating legal provisions; social actions and events such as cooperation activities (e.g. concert performances); social processes such as election or the procedure of the constitution of a state.

These different kinds of social entities are essentially connected to one another. For instance, social objects such as laws would not exist without social acts such as law-making acts performed by the parliaments as collective subjects. Social subjects such as symphonic orchestras would not exist without social objects such as job contracts, which define the position of each single musician in the orchestra, without concert performances and without the cooperative activity of playing together.

Social ontological theories at our disposal on the market provide accounts of social reality which tend to deal specifically with just one of these kinds of social entities. Just to mention some few examples: Searle and his account of *social objects* such as money (Searle 1995, 2010); Bratman and his account of *social actions* such as shared intentions and cooperative activities (Bratman 2014); List-Pettit and their account of *social subjects* in terms of group agency (List & Pettit 2011); Gilbert and her account of *social subjects* in terms of plural subjects (Gilbert 1989, 2013).

2.
Gilbert's social
ontology of plural
subjects

I suggest that Margaret Gilbert' social ontology is therefore an ontology of social subjects, and, more precisely, of "plural subjects" – in her terminology.

2.1
An existential
approach to social
ontology

I argue that Gilbert's social ontology of plural subjects is impressive, and it is fundamentally and appropriately characterized by an original existential approach. Gilbert speaks of her social ontology as a "general project of understanding the terms in which human lives are lived" (Gilbert 2013, p. 2). "What puts you in a position to use the collective 'we'? One cannot hope fully to understand the human condition without an answer to this question" (Gilbert 2013, pp. 5-6). Such existential connotation, which is specific to Gilbert's social ontology, strikes positively because of its rarity in the contemporary social ontological debate as well as because of its appropriateness to "the thing itself" (to the proper matter of social ontology). In the end, social ontology is an ontology of the existence, i.e. of the human condition, since the experiences of the collective we (that is, what we live together with other human beings) are central for the existence of every human being.

2.2
The neglect of the
specific role of
values for social
unity and the
need of a different
account of values

Nevertheless, Gilbert's ontology of plural subjects does not adequately account for values. I argue that, astonishingly enough, Gilbert neglects the specific role of values for the quality of "the human condition" and for the quality of the existence of human beings in the social world. Values constitute neither a necessary nor a sufficient condition for the creation of plural subjects; only as (one of the possible) objects of a *joint commitment*, values become "shared collective values" or "collective values" which unify and bind people together constituting plural subjects (Gilbert 2005).

---

the present issue), Miller 2014 and 2015 (in the present issue), Searle 2010, chapter V, Reinach 1913, § 2.

In any case, the removal of values is, after all, a constant feature of the contemporary social ontological debate.

In the following part of my talk, I outline Gilbert's plural subjects account of shared values and then sketch some features of a different account of values, that of phenomenology; according to the phenomenological account of values, I try to show that values play a crucial role both for the creation and wellbeing of plural subjects.

I suggest that, unlike Gilbert and the main trend in the contemporary social ontological debate, social ontology should provide an account of values which focuses

(i) on the *specific nature of values* and

(ii) on *essentially collective values* such as solidarity or justice.

I also suggest that phenomenology could provide some important and fruitful insights for such account.

I outline Gilbert's account of *shared values.* It represents a perspective on values which is quite typical in the contemporary social ontological debate. Two features of Gilbert's account of shared values are typical of the way in which values are considered by social ontology today: a cognitivist account of values and an extrinsic account of values.

2.3 Gilbert's plural subjects account of shared values

Gilbert proposes a cognitivist account of values: values are the objects of beliefs.

2.3.1 A cognitivist account of values

> In order that one have values, then, one must have beliefs or opinions. More specifically, one must have beliefs or opinions to the effect that some item or items have a certain value (Gilbert 2005 (2013), pp. 183-184).

Consequently,

> The question about the nature of shared values will then be understood as a question about *shared beliefs or opinions to the effect that some item or items have a certain value* (Gilbert 2005 (2013), p. 184).

According to the cognitivist account of values, the specific nature or matter of values is not taken into consideration. Values are the object of beliefs just as anything else could be the object of beliefs. There is neither an inquiry into the specific being of values nor into a specific kind of intentionality that could grasp values directly (vs. the phenomenological intrinsic account of values: values are qualities which are directly grasped by "feeling value" [*wertfühlen*], that is, values are felt, see *infra* § 3).

2.3.2 An extrinsic account of values

Because of these two features, values do not play a significant role in the formation of social groups (and, consequently, in Gilbert's social ontology in general).

In her account Gilbert addresses the issue of whether shared values can be:

- either a necessary or sufficient condition,
- both conditions
- neither of the conditions

for the existence of social groups.

Gilbert's answer is that shared values are not a necessary condition for social groups, and that they can be a sufficient condition only "if sharing values [is] construed according to the plural subject account" (Gilbert 2005 (2013), p. 200), that is according to the joint commitment *modus.*

2.3.3 Values and social unity

2.3.4
Joint commitment:
on the sufficient
condition

All joint commitments are of the same form. X and Y (and Z and so on) are jointly committed *to do something as a body.* Here "doing something" is construed broadly enough so as to include believing that such-and-such and feeling thus-and-so. What is it to do something "as a body"?

The joint commitment in the case we are considering is a commitment to believe as a body that some item I has a certain value v (Gilbert 2005 (2013), p. 193)

Gilbert argues that only if individuals commit themselves to believe jointly that a particular item I has a certain value v, such value v is a sufficient condition for the creation of social unity and of a "plural subject".

In the case in which values are shared in the joint commitment *modus*, Gilbert speaks properly of "collectively shared values" or, more briefly of "collective values" (Gilbert 2005 (2013), p. 200). In other words values are here (cognitively) shared in a genuine collective way, that is in a we-mode, by a "plural subject"– and not in a summative or singularist way.

2.3.5
Joint commitment:
on the necessary
condition

According to Gilbert, people can be jointly committed in a variety of ways. Whatever this way is, the result of the joint commitment is always the creation of a "plural subject".

[People] can be jointly committed not only to believing as a body that such-and-such, but also to accepting or pursuing a goal as a body, to accepting as a body that A is to be done in circumstances C, and more. In my technical terminology those who are jointly committed to X-ing as a body constitute the *plural subject* of X-ing (Gilbert 2005 (2013), p. 197).

Therefore, concerning the necessary condition,

It would seem that such [collective] values are not necessary, however, for social unity. A plural subject, founded on a joint commitment, need not be the subject of values as opposed to other beliefs, rules of the fiat form, and so on" (Gilbert 2005 (2013), p. 204).

In other words, collective values can be a condition for social unity and for the creation of a plural subject, just as collective beliefs or collective rules can be a condition for it. According to Gilbert, the only fundamental condition for social unity and plural subject is that values, beliefs, rules, emotions etc. be shared in virtue of a joint commitment (see Gilbert 2014, on collective emotions). If values, beliefs, rules, emotions etc. are the object of a joint commitment, then, in virtue of such joint commitment such values, beliefs, rules, emotions, etc., are collective values, beliefs, rules, emotions of a plural subject, and they bring forth social unity.

2.3.6
The core
argument:
justifying
interventions
for unifying
individuals and
binding them
together

According to Gilbert, sharing values, beliefs, emotions etc. in the joint commitment *modus*" [...] *unifies* people, it *binds* them *together*, and it provides them with the standing to *intervene* in one another's lives"(Gilbert 2005 (2013), p. 206; italic mine).

Therefore, collective values, beliefs, emotions, etc. (i.e. values, beliefs, emotions etc. which are shared in the joint commitment *modus*) are a sufficient condition for the creation of social unity and plural subjects because they

- unify and
- bind the participants together

by giving them the standing to *intervene* in one another's lives, more precisely to make demands on one another and to rebuke one another in order that each participant acts conforming to the joint commitment in question.

Gilbert's account of shared values does not value values with regard to their specific nature. Values are considered as any other possible object of beliefs and as any other possible object of joint commitments. As objects of a joint commitment, values become shared collective values (or collective values) and constitute a sufficient condition for the creation of social unity and plural subjects.

This is just an extrinsic account of values and of shared values. Gilbert does not inquire into the specific nature of values and their role in the constitution of social unity and plural subjects. For instance, she does not address the question if there are *sharable values and not-sharable values per se.* Moreover, she does not take into consideration the role of *essentially collective values* such as solidarity for the wellbeing of a collective of persons, and consequently, for its stability and persistence. All of the "social unity game" played by Gilbert takes place on the field of what can entitle people to intervene in each other life in order to *unify* them and *bind* them *together.* This issue of intervention is not a specific issue of collective values, but rather a general issue, common to collective values, beliefs, emotions etc. In conclusion, the specific contribution of values in the creation, maintenance in existence (stability and persistence) and quality of the existence of plural subjects is not taken into account.

I argue that phenomenology provides an axiology (i.e. a theory of values which deals with their specific ontological nature) that can allow us to account adequately for values and to understand their crucial place in the social world.

I make two points:

(i) Scheler's axiology (Scheler 1913/1923, 1913/1927), his inquiry into the specific nature of values and *the dividing vs. sharing values thesis*: the more divisible values are, the less sharable they are; and the converse: the less divisible values are, the more sharable they are.

(ii) The *collective values thesis* (Schapp 1930): there are values which are essentially collective (vs. individual values).

These points should suggest a track for an alternative answer to *the issue of unifying and binding people together* which is, as we have just seen, Gilbert's main issue about the possibility conditions for social unity.

Here I limit myself to sketching some essential features of values which, according to Scheler's axiology, ground the thesis of dividing values vs. sharing values which I present.

(i)   **Feeling value:** Values are the specific and direct object of a particular kind of affective intentionality: *feeling value* [*wertfühlen*]. *Feeling value* is to be distinguished both from perceiving [*wahrnehmen*], which is a cognitive act, and from emotions or sensations, which the act of feeling values can give rise to.

(ii) **Realism about values**: Values are material qualities which exist independently of individuals' intentionality, and, therefore, are not reducible to mental phenomena like sensations, emotions or beliefs (as, instead, happens in Gilbert's account).

(iii) **Values and goods**: Values are not reducible to goods, the things which are bearers of values (valuable things). The distinction between values and goods is very important for preserving a transcendence of the type of the value with respect to its instantiations in goods: without such distinction, a value, for instance friendship, would be identified with a

certain instantiation of friendship in certain bearers.[6]

(iv) **Hierarchy of values**: There are "superior" and "inferior" values: values can belong to different spheres, which from the lower to the higher constitute a hierarchy of values:
– values of the "sensibly agreeable" [*Werte des "sinnlich Angenehmen"]*
– vital values [*Lebenswerte*]
– person values [*Personwerte*]
– values of the holy [*Werte des Heiligen*].[7]

Among others, one criterion for values hierarchy is the divisibility of values: the less values are divisible, i.e. the less values must be divided in participation by several, in order to be grasped by several participants, the higher values are; and the converse: the more values are divisible, i.e. the more values must be divided in participation by several, in order to be grasped by several participants, the lower values are.

There is no question about the fact that values are "higher" *the less they are divisible* [*teilbar*], that is, the *less* they must be *divided* [*geteilt*] *in participation by several* (Scheler 1913/1927, p. 110; En. Tr., p. 93).

3.2
Scheler: the dividing values vs. sharing values thesis

Now, I argue for the following thesis: the *more divisible* values are, the *less sharable* they are. And the converse: the *less divisible* values are, the *more sharable* they are.

3.2.1
Inferior values: the case of the values of the "sensibly agreeable"

The fact that the participation [*Teilnahme*] of several in "material" goods is possible only by dividing [*Teilung*] these goods (e.g., a piece of cloth, a loaf of bread) has this final *phenomenological* basis: the *values* of the *sensibly agreeable* [*Werte des sinnlich Angenehmen*] are clearly *extensive in their essence*, and their felt experiences occur as localized and as extensive in the body. For example, the agreeableness of sweet, etc., is spread over sugar, and the corresponding sensible feeling-state over the "tongue".
[...] It is therefore essentially impossible [*wesensgesetzlich ausgeschlossen*] for one and the same value of the values-series of the "sensibly agreeable" to be enjoyed by several beings *without* the division of its bearer and of the value itself. For this reason there are also, in the *essence* of this values-modality, "conflicts of interest" relative to the striving for a realization of these values, and relative to their enjoyment [...]. This also implies that it belongs to the *essence* of these values to *divide*, not to unite [*trennen, und nicht vereinen*], the individuals who feel them (Scheler 1913/1927, p. 110; En. Tr., p. 93).

The lowest sphere of values is that of the values of the sensibly agreeable [*Werte des "sinnlich Angenehmen"*], for example the "agreeableness of sweet". Such kind of values can be grasped by more individuals together only if their bearers (the material goods, in which they are embodied) are divided among the individuals because the values of the sensibly agreeable are essentially characterized by extension [*Ausdehnung*] and divisibility [*Teilbarkeit*]. Therefore, values of the lowest sphere (such as the sweet of a food or also the warmth of a blanket) essentially *divide and not unify the individuals*, because, in order to be enjoyed by a plurality of

---

6 See Scheler 1913/1927, First part, First Section «Materiale Wertethik und Guter-Respektive Zweckethik», §1 «Guter und Werte».

7 See Scheler 1913/1927, § 3, pp. 110-111. About these features of Scheler's account of values, see Mulligan 2009 and about the ontological status of values see De Monticelli 2015.

individuals, they have to be divided among the individuals and individuals have to compete with one another to grasp the values. So, the *more divisible* values are, the *less sharable* they are.

> In strict contrast to this [values of the sensibly agreeable] there stands a "work of art", for example, which is "indivisible" [*unteilbar*] and of which there is no "piece". [...] The most extreme opposite of these values [the values of the sensibly agreeable], the values of the *"holy"* [*das Heilige*], of *"cognition"* [*Erkenntis*], and of the *"beautiful"* [*das Schöne*], etc., as well as their corresponding spiritual feeling-states [*geistige Gefühlen*], have a totally different character. There is no participation in extension and divisibility with these values; nor is there any need to divide their bearers if they are to be felt and experienced by any *number* of beings. A work of spiritual culture can be simultaneously apprehended [*gleichzeitig erfasst*] by any number of beings and can be felt and enjoyed in its value, for it lies in the essence of values of this kind to be *sharable* [*mitteilbar*] *without limit* and without any division and diminution [...]. Nothing *unites* [*vereint*] beings more immediately and intimately than the common worship and adoration of the *holy*, which by its nature excludes a "material" bearer, though not a symbolic one. [...] It lies in the *essence* of the *intention toward the holy to unite* [*einen*] *and bind together* [*verbinden*] (Scheler 1913/1927, p. 111; En. Tr. pp. 93-94, slightly modified).

On the contrary, superior values (for instance "person values" [*Personwerte*] such as the beauty of a work of art, the knowledge, or, above all, values of the holy) can be grasped by a plurality of individuals without being participated in extension and divisibility and without their bearers being divided. The beauty of a work of art is not divisible and there is no piece of a work of art. A work of art can be simultaneously grasped by a plurality of individuals and its value can be felt and enjoyed by such plurality, since it belongs to the essence of superior values to be "sharable without limit and without any division and diminution". Moreover it lies in the essence of some superior values (especially the values of the holy) to *unify* and *bind together* the people who feel and enjoy them.

Therefore, the less divisible values are, i.e. the less values must be divided *in participation by several*, in order to be grasped and felt by several participants, the more they can be felt and enjoyed by several individuals together, that is the *more sharable* they are.

The main idea I argue for is that there are values which are sharable and values which are not. Sharable values are fundamental for the quality of the existence of human beings and for the creation of social unity (groups, societies, communities, etc.): sharing values, which are crucial for the development and flourishing of human beings such as the beauty of a work of art, the knowledge and the truth (think of a group of scientists), the justice of a law, *unifies* people and even *binds* them *together*.

3.3
**Sharing values
and the issue
of unifying and
binding people
together**

Therefore, this phenomenological insight into values provides a track for an account of the problem of unifying and binding people together (in order to bring about plural subjects: groups, collective of persons *etc. vs.* mere aggregates), which is different from Gilbert's account focusing just on the title people have to intervene in each other lives.

The phenomenological account implies that certain specific values–and not others–are sharable and that, once shared, they *unify* the people who share them: for instance sharing the justness of a law, brings about unity among people. Thus, people are not unified only by and through the standing to intervene in each other lives (Gilbert's account); rather, people are also unified by and through the experience of sharing some values which essentially and positively unify–and not divide–people.

3.4
The collective
values thesis

Moreover, the phenomenological account of values not only focuses on the distinction between values which are sharable and values which are not. It even suggests that, among sharable values, there are values which are essentially collective: as shown by Wilhelm Schapp (1930), we have to distinguish between values which are individual and values which are collective, that is between values which can be enjoyed by only one individual, on the one hand, and *values which are essentially collective*, on the other hand, for they need to be shared and enjoyed by individuals together in order to exist and to be realized. Think, for instance, to the solidarity of behaviour with respect to others, the justness of a law.[8]

The collective values thesis provides a promising path for answering the question of *binding* people *together*, the question of the social bondage, since it seems to me that collective values necessarily imply a bond among the people who share them: without such a bond, collective values can be neither enjoyed nor realized.

3.5
Ways of sharing
values

Now, the last issue I address concerns the very concept of sharing values: what does sharing values mean? Which phenomena do exemplify the sharing of values?

According to Gilbert's joint commitment account, sharing values implies that values are shared in the plural subject mode created by joint commitment. This is a genuine collective we-mode, i.e. not a singularist or a summative way of sharing.

In the phenomenological scenario I have just mentioned, in which sense of sharing sharable values (values which can be communicated and shared without limit and division of the goods in which they are embedded) can be shared? According to Scheler people can share values *lato sensu* in different ways corresponding to different kinds of co-feeling. However only one way of sharing values represents a genuine collective way of sharing values.[9]

(i) Collective feeling [*Miteinanderfühlen*] value: we-mode

The only one genuine way of feeling value together is collective feeling. It is, say, the case of two artists who feel together the beauty of the work of art they are creating, or of two parents who feel together the love for their child. In these cases, individuals share values in a genuine collective we-mode: it is not reducible to the I-form plus mutual or common knowledge (I feel the (sharable) value x and I know that you feel the same value x, etc.), that is, it is not reducible to merely individual experiences plus reciprocal knowledge. Why? Because collective feeling value is constituted by the deep interdependence and co-regulation of individuals' experiences and by their reciprocal relation to each other.[10] Therefore, collective feeling value is a very important moment of groups of individuals such as couples, families and communities which are deeply unified and bond together.[11]

(ii) Co-feeling [*Mitfühlen*]: I-mode

Simple co-feeling is the case in which, say, artist A feels the disvalue "grotesque" of the work of art she is creating together with the artist B; the artist B does not feels herself such disvalue, but she may co-feel the feeling the grotesque of the artist A: B sympathizes the feeling the grotesque of A.

---

8 On Schapp's account of values, see De Vecchi 2016.

9 Scheler presents his taxonomy of co-feeling in Scheler 1913/1923: Part I *Fellow Feeling*, Section II *Classification of the phenomena of fellow feeling*.

10 See Zahavi (2015: 116-117, 245) who discusses the specificity of "emotional sharing" in Scheler's account of co-feeling.

11 A further and deeper analysis of Scheler's collective feeling account would be needed here, but for reasons of lenght I have to limit myself to the few features I sketched. About this topic, see Schmid 2009, Zahavi 2015, Szanto 2015, Salice 2016, De Vecchi 2011.

In this case, values are shared not in a genuine collective way, not in a we-mode, but just in a singularist way. B co-feels the value felt by A, and so B shares the same value of A. But A may even not be aware of such co-feeling and sharing.

Co-feeling is grounded in the refeeling (*Nachfühlen*).

(iii) Refeeling [*Nachfühlen*] value: grounding co-feeling value

> It often happens that a value, e.g., the value of a work of art, comes to us in a more adequate form of giveness only trough a refeeling of the feeling [*Nachfühlen des fühlens*] of the value concerned; and we often realize through this that our ability to feel [*Fühlfähigkeit*] the type of value in question is very narrow indeed (Scheler 1913/1927, p. 249; En. Tr., p. 243)

I may grasp the value of the beauty of a work of art through my feeling of the feeling [*Nachfüehlen*] of the beauty value felt by another person: this is the case of empathy. In other words, in the case of superior values as the beauty (and not in the case of inferior values of the sensibly agreeable) I may increase my feeling of values through the capacity other people have to grasp certain values which I was not able to grasp.

In such case, I grasp the feeling value (e.g. the beauty) felt by the other. This does not imply that I endorse such feeling value. Refeeling value is just an act by which I learn the feeling value of others. Then, I may feel such value personally and share it. But I may also do not feel such value personally.

Therefore, in the case of refeeling value, values may be either co-felt or not-co-felt. If I personally endorse and feel the value felt by another person, I co-feel such value: i.e. I sympathize such value with the other person. Here values which are co-felt are shared in I-mode and not in we-mode. If I just limit myself to grasp the feeling value of the other person unless feeling such value personally, then no value is shared.

(iv) Contagion [*Ansteckung*]: pseudo-we-mode

Another possible case of co-feeling is the case of contagion: this is a pseudo collective sharing values. Contagion is a form of co-feeling which plays a very crucial role in the creation of social groups such as masses in which individuals identify themselves with impersonal collective forms of sociality. For instance: if I belong to a certain tradition, culture, etc., then I may be swayed by that tradition or culture, but I may be not aware of this contagion. In this case, I feel certain values characterizing that tradition or culture and I think that such values are personally mines and that they are grounded in me, because I identify myself with that tradition (but I am not aware of it). I may feel hate for a certain population because I belong to a population that traditionally hates that population; but I believe that such hate is grounded in me, and I am not aware that it is derived by the population I belong to. In other words, there is the creation of a pseudo-we, in which individuals are merged one in the other and the individuals' identities are not preserved, as instead it is the case in the genuine collective feeling.[12]

REFERENCES

Bratman, M. (2014), *Shared Agency. A Planning Theory of Acting Together*, Oxford, Oxford University Press;

---

12 See Scheler 1913/1923: Part I *Fellow-Feeling*, Section III *Genetic Theories of Fellow Feeling*.

De Monticelli, R. (2015), "Experience and value. A phenomenological approach to the dilemma of normative supervenience", *Metodo. International Studies in Phenomenology and Philosophy*, Vol. 3, n. 2 (2015), pp. 195-214;

De Vecchi, F. (2011), "Collective intentionality vs. inter-subjective and social intentionality. An account of collective intentionality as shared intentionality", *Phenomenology and Mind*, vol. 1, 2011, pp. 95-118.

----. (2012), "Ontologia sociale e intenzionalità: quattro tesi", *Rivista di Estetica*, n. 49 (1/2012), LII, pp. 183-201;

----. (2014), "Three types of heterotropic intentionality. A taxonomy in social ontology", in A. Konzelman Ziv & H.B. Schmid (eds.), *Institutions, Emotions and Group Agents. Contribution to Social Ontology*, Heidelberg, Dordrecht, London, Springer Verlag, Studies in the Philosophy of Sociality 2, pp. 117-137;

----.(2016), "A priori of the law and values in the social ontology of Wilhelm Schnapps and Adolf Reinach", in A. Salice & H.B. Schmid (eds.), *The Phenomenological Approach to Social Reality*, Heidelberg, Dordrecht, London, Springer Verlag, Studies in the Philosophy of Sociality 6, pp. 279-316;

Gilbert, M. (1989), *On Social Facts*, London, Routledge and Kegan Paul;

----. (2005), "Shared values, social unity and liberty", *Public Affairs Quarterly*, 19, 2005, pp. 25-49 (then in Gilbert 2013);

----. (2013), *Joint Commitment: How We Make the Social World*, Oxford, Oxford University Press;

----. (2014), "How we feel: Understanding everyday collective emotion ascriptions", in *Collective Emotions*, M. Salmela, Ch. Van Scheve (eds.), Oxford, Oxford University Press;

----. (2015), "Joint commitment: What it is and why it matters", *Phenomenology and Mind*, Vol. 9, 2015, pp. 18-27;

Husserl, E.(1910), "Soziale ontologie und deskriptive soziologie", in *Husserliana XIII, Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlaß. Erster Teil: 1905-1920*, ed. by I. Kern, Martinus Nijhoff, Den Haag, 1973, pp. 98-104;

List, Ch. & Pettit, Ph. (2011), *Group Agency*, Oxford, Oxford University Press;

Mulligan, K. (2009), "On being struck by value", in B. Merker (ed.), *Lebenmit Gefuhlen. Emotionen, Werte und ihre Kritik*, Paderborn, Mentis, pp. 141-161;

Miller, S. (2014), Review of M. Gilbert, J*oint Commitment: How We Make the Social World*, *Notre Dame Philosophical Reviews*, https://ndpr.nd.edu;

----. (2015),"Joint political rights and obligations", in *Phenomenology and Mind*, Vol. 9, 2015, pp. 138-146;

Reinach, A. (1913), *Die apriorischen Grundlagen des burgerlichen Rechts*, in A. Reinach, *Samtliche Werke*, ed. K. Schuhmann & B. Smith, pp. 141-278, München, Philosophia Verlag, 1989. English translation, "The Apriori foundations of the civil law" (trans: Crosby, J.), *Aletheia*, III, pp. 1-142, 1983;

Salice, A.(2016), "Shared emotions – A schelerian approach", *Thaumazein* (forthcoming);

Schapp, W. (1930), *Die neue Wissenschaf tvom Recht. Eine phänomenologische Untersuchung.* Berlin: Rothschild;

Scheler, M. (1913/1923), *Wesen und Formen der Sympatie*, in M. Scheler, *Gesammelte Werke*, vol. VII, ed. by M. Frings, Bouvier Verlag, Bonn 2005. English Translation: *The Nature of Sympathy*, London, Transaction, 2008;

----. (1913/1927), *Der Formalismus in der Ethik und die material Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus*, in M. Scheler, *Gesammelte Werke*, vol. II, ed. by M. Frings, Bouvier Verlag, Bonn 2009. English Translation: *Formalism in Ethics and Non-Formal Ethics of* Values, Northwestern University Press, Evanston, 1973;

Searle, J.R. (1995), *The Construction of Social Reality*. New York, Routledge;

----. (2010), *Making the Social World. The Structure of Human Civilization.* Oxford, Oxford University Press;

Schmid, H.B. (2009), *Plural Action: Essays in Philosophy and Social Science*, Dordrecht, Springer;

Szanto, Th. (2015), "Collectivizing persons and personifying collectives: Reassessing Scheler on group personhood.", in Th. Szanto, D. Moran (eds.), *The Phenomenology of Sociality: Discovering the 'We',* London/New York, Routledge, 2015;

Thomasson, A.L. (2009), "Social Entities", in R. Le Poidevin (ed.), *The Routledge Companion to Metaphysics*, Routledge, 2009, pp. 545-554;

Zahavi, D. (2015), *Self & Other. Exploring Subjectivity, Empathy, and Shame*, Oxford, Oxford University Press.

GLENDA SATNE
Alberto Hurtado University –
University of Wollongong
gsatne@uahurtado.cl
gsatne@uow.edu.au

ALESSANDRO SALICE
Center for Subjectivity Research
University of Copenhagen
alessandro.salice@hum.ku.dk

HELPING BEHAVIOR AND JOINT ACTION IN YOUNG CHILDREN

abstract

An important idea due to Tomasello and others is that the human capacity as the human capacity for social cooperation is at the heart of the species' capacity to understand others' mental states and behavior. Furthermore, they argue that this idea allows for an explanation of how humans came to share thoughts and language. While this is a promising idea, the special attempt to pursue this hypothesis in developmental studies and evolutionary theory developed by Tomasello and his research group faces several problems. This is especially apparent in their attempts to explain helping behavior and joint action in young children. In this paper, we argue that many of these problems result from assuming that the right explanation of joint action and simple forms of shared intentionality is given by Bratman's theory of shared intentions.

<div style="margin-left:2em;">

1. Introduction

One important application of theories of collective intentionality is in explaining the evolution of social understanding – and even human thinking (Tomasello 2014). The promising idea behind this enterprise is that insofar as the human capacity for social cooperation is at the heart of the species' capacity to understand others' mental states and behavior, it leads to an explanation of how humans came to share thoughts and language. While this is an important idea, the special attempt to pursue this hypothesis in developmental evolutionary studies developed by a dominant trend in developmental studies due to Tomasello and others (Tomasello 2014, Warneken & Tomasello 2006, 2007) faces several problems, or so we argue in this paper.

Such problems are especially apparent in Warneken, Tomasello et alia's attempts to explain helping behavior in young children. Their explanation can be characterized as a twofold explanation as it involves two explanatory factors: the first is children's *altruistic* tendency to cooperate, and the second is the ability to act jointly and share intentions, which is independent of the former factor and in turn explained in terms of theory of mind skills. In this paper, we identify three main problems with said explanation and argue that they result from assuming that the right account of joint action and simple forms of shared intentionality is given by Bratman's theory of shared intentions.

2. Bratman on joint action

According to Bratman's account, agents that engage in joint action have a number of individual and interlocked intentions that steer, guide and monitor their joint activity. In this section, we proceed to present this model of joint activity, analyzing some of its key features and especially the notion of helping behavior that flows from it. In the subsequent section, we show that assuming a model like Bratman's leads Warneken, Tomasello et alia to overlook important features of children's helping behavior and to wrongly asses its relation to their evolving interest and engagement in joint activities with adults.

Bratman identifies a set of jointly sufficient conditions that qualify an action as a case of shared agency. Given Jones and Jane and an action J, say, building a house, the conditions are:

</div>

(1) the agents pursue *interdependent* goals[1,2], that is, Jones intends that Jane and Jones J together, and Jane intends that Jones and Jane J together[3].

(2) the agents pursue the goal in an *intentional* and non-coercive way, that is, Jones intends that Jane and Jones J together in accordance with and partly because of Jane's intention in (1),and Jane intends that Jones and Jane J together in accordance with and partly because of Jones's intention in (1).

(3) the agents intend that their strategies towards the goal mesh: Jones intends that Jane and Jones J together in accordance with and partly because of the meshing sub-plans of (1), and Jane intends that Jones and Jane J together in accordance with and partly because of the meshing sub-plans of (1).

(4) all the conditions are out in the open, i.e., are common knowledge for Jones and Jane.

(5) among the agents, there is public mutual responsiveness that tracks the goal.

(6) there is a commitment to mutual help and support among the agents.

The interdependence between Jane's and Jones's goals, it can be argued, has to be characterized as "weak". It is weak because the intentions in (1),which identify such interdependent goals (Jane's and Jones's intentions *that* they J),*derive from or are sustained by* individual intentions, which are independent. The idea of "weak" interdependence should be understood in contrast to the "robust" form of interdependence that authors like Tuomela and others want to secure for common or collective goals, i.e., for goals intended by a group where individual intentions are not based upon individual intentions but are themselves dependent on collective intentions (cf. Tuomela 2005).
Bratmanian shared intentions can derive from independent and individual intentions in the sense that the formation of intentions in condition (1) temporally follows the formation of intentions of the form

(0) Jane intends to J, and Jones intends to J.

But, even if shared intentions are not preceded by intentions like (0), they are in any case *sustained by* intentions of such form in the following twofold sense. *First*, it is not possible for an agent to intend that she J with someone else without intending to J: if the agent's intention to J lapses, then the intention that the agent Js with someone else has to lapse, too. However, *second*, it can be that the intention that the agent Js with someone else expires and, yet, the agent still intends to J. That is, whenever agents have intentions of the form (1), they also have intentions of the form (0) – the latter have to accompany the former and, in certain cases, they temporally precede them.

---

1 No commitments need to be involved in this scenario, except for those private (non interpersonal) commitments towards one's goal that are generated by the agents' individual intentions.
2 By "goal", we mean a state of affairs that the bearer of the intention is committed to bringing about (in this case: J).
3 To avoid circularity, Bratman denies a robust reading of "together". Instead, "together" should be read as meaning that "each of us intends that we J by way of the intentions of each that we J" (Bratman 2015, p. 3). This view denies the need to postulate a plural subject, neither is a plural subject *part of the content* of the intentions nor are such intentions those *of* a plural subject.

We have said that the agents' goals in (1) are weakly interdependent: Jane intends to J with Jones only to the extent that Jones intends to J with Jane. But how can one characterize the goal in (0)? This can be said to be shared *distributively*, i.e., an outcome O is a distributive goal of two or more agents' actions if, *first*, O is one to which each agent's actions are individually directed and, *second*, each agent's actions are related to the outcome in a way that it is possible for all of them together to succeed in bringing about O (Butterfill 2012, p. 849).

The form of help in condition (6) can be said to directly depend on the notion of an interdependent goal as presented in condition (1) and on the notion of a distributive goal that defines the nature of the overarching goals in (0). If the agents pursue a distributive goal, then the help they are willing to provide is not on behalf of the other agent(s). Consider the following example. Imagine that at a given point Jane notices that Jones's behavior may put the possibility of building the house at risk, e.g., because he is not using the right amount of sand to prepare the concrete. She decides to do something to fix this problem. In this case, help is triggered by the fact that one agent, say, Jane, realizes that one of the *means* that she chooses to reach her goal (namely Jones's expected contribution, cf. Bratman 2014, p. 100f) does not work out appropriately.[4] Since the expected contribution is missing, the achievement of her goal is endangered. *Therefore*, Jane intervenes to fix the problem, which she conceives of as *her* problem. She fixes it in such a way as to supply the missing contribution. This, i.e., fixing her problem, can be contended to be what Jane *directly* intends.[5] Or, in other words, this is what is steered and guided by her intention. Now, according to the ordinary concept of help (Schmid 2010), helping seems to involve directly intending the other's goal, i.e., the other's goal is the *end* of the action. By contrast, in a scenario like the one framed by Bratmanian joint intentions, Jane does not have an intention of this sort towards Jones' goal, as Jones' goal is only a *means* towards her goal, the actual end of her action. To illustrate this contrast, we can introduce the notion of *indirect or oblique help* where by an event is intentionally brought about as a consequence of another action, but is not intended as such.[6] Helping would be instrumental in this case insofar as the other's goal is taken as a means for pursuing one's own goal and not as an end itself.

Thus, Bratman's model seems to provide an adequate account of the kind of joint action that involves instrumental helping among interactants motivated by individual reasons. While this model may be suited to account for many cases of cooperation, we argue that his analysis is not successful in accounting for children's cooperative behavior, as Tomasello et al. meant it to be. In the following section, we review Warneken et al.'s studies on the early development of joint intentionality. We present three major challenges that this account faces and show how these shortcomings spring from the authors' use of Bratman's model of joint intentions. The solution to these problems, we conjecture, has to pass through an

---

4 The other's contribution figures as a *means* to one's own goal insofar as it is "by way" of the other's intentions that the goal is pursued. This does not mean that the intention of the other is treated on equal footing with other means to the same end, means that are not intentional contributions to the same end. The conditions (1) – (6) spelled out above identify the other important features that the contribution of the other needs to meet in order to qualify as an intentional contribution.

5 This distinction between direct and indirect help is developed at length in Salice & Satne (*under review*).

6 In previous work on the philosophy of action, Bratman (cf. 1984) convincingly relaxes the condition that, for an action to be intentional, it has to be intended. He argues for the idea that, if an action's outcome A is salient and expected by the agent, then A qualifies as intentional. To illustrate with our example, Jane intends to fix the situation, and she knows that fixing the situation (given that Jones's goal overlaps with hers) will result in helping Jones. She then intentionally helps Jones. But she did not intend this in the sense of having an intention that refers to Jones's goal. She did not *directly* intend to help Jones; she obliquely or indirectly helps Jones. For a full-length explanation, see Salice & Satne (*under review*).

altogether different conception of joint intentionality, one that seriously considers the idea of genuinely collective intentions. Arguing in favor of this idea would nevertheless exceed the purposes of this paper.

Warneken, Tomasello et alia have studied children's engagement in joint action and helping behavior at a young age. Such studies analyze children's early understanding of shared intentions by focusing on two different kinds of behavior prominent in children between 14 and 24 months: 1. children's tendency to help others achieve their goals, which is manifested in their engagement in further pursuing the incomplete actions of adults; 2. children's ability to engage in shared cooperative activities in Bratman's sense, where the roles of each agent's actions are complementary and where all the interactants (i.e., children and adults) are together pursuing the goal (Warneken & Tomasello 2006, 2007). In this view, only the latter is an example of shared intentionality.

3. Children's engagement in joint action: Warneken & Tomasello's view

According to these studies, the first kind of behavior is to be found in children at around 14 months of age (Warneken & Tomasello 2006, 2007). The experiment that Warneken and Tomasello developed to test this kind of helping behavior ran as follows: the experimenters performed several tasks in front of children where the experimenter encountered some kind of hindrance – his marker fell while drawing, a clothespin was dropped while hanging towels, a spoon accidentally fell into a box on top of which the experimenter's teacup was placed. In each case, the experimenter expressed annoyance when this happened ("Oops! I dropped my marker", etc.). This was contrasted to control conditions where the experimenter did not seem to care about what happened, or where he intentionally caused the hindrance (e.g., intentionally throwing the spoon into the box after stirring the tea). While in the former cases the children spontaneously engaged in the activity by helping the experimenter (picking up the clothespin or marker, etc.), in the latter they did not. According to the authors, the outcomes of the experiment showed the presence of an altruistic tendency in young children, a tendency to act upon others' incomplete or impeded goals, and a correlated capacity to understand someone else's intentional actions. Importantly, in Tomasello et alia's view, this is not yet a case of shared intentionality because – even if children are acting upon another person's goal – they are not coordinating their activities with the adult.

The second kind of behavior – the one that can properly be called "shared intentional activity" according to these authors – is argued to satisfy Bratman's conditions spelled out above. It was shown that children as young as 18 months can successfully deal with joint problem-solving tasks in which two agents (the child and the experimenter) must perform complementary roles in order to achieve a joint goal, e.g., pulling opposite ends of a tube to retrieve stickers that are hidden inside, or placing a ball into a tube and catching it from the other side. Warneken et alia (2006) analyzed the behavior of children between 18 and 24 months in similar experiments. When comparing both age groups, it was observed, they surmise, that this sort of activity is consistently and spontaneously carried out by children aged 18 to 24 months, and becomes more skillfully and expertly performed over time. Moreover, in carrying out these experiments, the authors found that children over 18 months of age consistently protest (verbally or non-verbally) if the adult disengages before the shared activity is completed and, significantly, such protesting still arises even if the activity is successfully completed and the goal attained. In addition, they also observed a tendency in the children to repeat the action even when it was completed successfully. In their view, this showed that, in so doing, the children were not pursuing an individual goal (e.g., get the ball) but that, rather, they were interested in the shared activity as such.

Warmeken, Tomasello et al. understand shared intentional activity in terms of Bratman's analysis of shared intentional activity and explicitly so (see Warneken et al., 2007, p. 291;

see also Tomasello 2014). Accordingly, they characterize spontaneous helping behavior as an example of individual intentionality: contributing to another person's goal manifests an understanding of the other person's goal-directed behavior, but it is not yet a case of sharing a goal with another, as the conditions spelled out by Bratman are not fulfilled. The authors conclude that helping behavior in young children is a precursor of the kind of shared activity that is exhibited when they later engage with adults in coordinating activities, such as in the experiments above. Nevertheless, as we argue below, the specific features of the analysis they provide of shared intentional activity and helping behavior cast doubt on whether their conclusion really follows from the studies. Several features of children's behavior are not really explained by the Bratmanian model they use. Moreover, using such a model to explain children's engagement in shared activity leads to what could be taken as important shortcomings in their analysis.

4.
Some problems with Warneken and Tomasello's account

In this section, we identify three problems that arise for this theory, and we argue that their origin lies in the authors' attempt to employ Bratman's theory of joint action to explain helping behavior in infancy. If our arguments are sound, they show the need for an alternative account of joint intentionality that might provide us with a better explanation of how shared intentionality and social understanding develop in infancy.

The *first* problem concerns the inability of the theories under scrutiny to account for the continuity and overlap in the development of two sorts of behavior: helping others complete their tasks (over 14 months of age) and cooperating with others to perform joint tasks (over 18 months of age). While children seem to move naturally from one to the other at 18 months, the model that these authors use gives us no motivation to see these two kinds of behavior as interconnected in any essential way.

Let us assume, as the authors do, that the earliest form of shared intentional activity that children display can basically be captured by Bratman's analysis. If so, then children's tendency to help could not be brought to bear in their evolving interest and engagement in shared intentionality. Remember that, in terms of paving the way for obtaining my own individual goal, Bratman's individualistic account of helping behavior falls under what we have called oblique help. This shows that behavior that fits the ordinary notion of help, according to which one seeks the other's goal for its own sake, is at best understood in a Bratmanian model as an extra motivating factor to engage in joint action, but not as part and parcel of the very same joint engagement. For example, it is not the case that a child conceives of his helping an adult as pursuing a joint activity with the adult, since the other conditions for Bratmanian joint action are not met. Conversely, when in the context of engaging in joint action, the help the children might provide is motivated by their individual goals, rather than by the adult's goal (unless such help is based upon an extra motivating factor). Accordingly, helping behavior of the kind studied in the experiments is to be understood as being independent of joint intentions, although it could still constitute evidence that the child understands the adult's behavior as goal-oriented. This being so, that kind of behavior *would not contribute to explaining* children's early engagement in joint activities with others and their inclination to want to participate in and repeat these joint activities. Moreover, the tendency of children around 14 months old to help adults complete their actions is, according to this analysis, totally independent of any individual motivation to pursue the specific actions the adult is engaging in. This is indeed a feature of the situations explicitly controlled by the experimenters: in control situations identical to the experimental ones in which adults do not pursue the goals, children do not pursue the goals in question either. The experiments are designed precisely to avoid such interference. But, if this is so, then the experimental conditions exclude cases in which the children have the same goal as the adults do. For this

reason, the helping behavior in question does not seem to be related to the kind of helping behavior that is displayed when coordinating activities, as this would be a different kind of helping behavior, oblique or indirect, that presupposes independent individual goals, as explained above. This is why helping behavior in children requires a completely different explanation in this model. The authors' preferred strategy in explaining its origins (Tomasello 2014, Warnecken & Tomasello 2006) is to attribute it to an innate altruistic tendency. But, even if this is so, it would still be the case that instances of a spontaneous tendency to help would not be best understood as early precursors of distributive intentionality in Bratman's sense, since they do not involve the child's having the goal independently of the adult's having it. These would be concomitant types of behavior and not internally articulated developments.[7] *Second*, the authors also take as important evidence of the child's ability to share goals in a Bratmanian sense that they will protest both if the adult disengages before the action is completed and if he does so after it has been completed. This is important, indeed crucial, because it suggests that children are not acting individually, but that they understand the key contribution of the partner to what they are pursuing. Moreover, the fact that they want to repeat the sequence even when it has been completed successfully can also be considered as evidence of the special engagement and interest children have in these activities. Now, this evidence is not very congenial to Bratman's conception of joint action. In his theory, there would be no motivation to continue trying to reengage the adult in the action once it has successfully been completed. The key driving force of each individual's behavior is the goal he is pursuing: he acts to the effect that X (say to retrieve the ball from the tube) is reached by means of the other's intention to do the same. But, when the ball is retrieved, the goal is successfully attained and, arguably, there is no reason to continue engaging in the same activity. This seems to be even more cogent once one considers that, according to Bratman, no commitments are required for shared intentions. According to a Bratmanian framework, children would protest only until their individual goal was achieved. Once the goal is attained, no additional motivation to protest or reengage is expected to emerge.

The *third* worry concerns the appeal to a theory of mind model that underlies Tomasello et alia's account of joint action. Many authors (Tollefsen 2005, Brownell et al. 2006, Pacherie 2013, Michael et al. 2014, Zahavi & Satne 2015) have argued that we need a simplified account of joint action if we want to make sense of children's ability to engage in shared activities with others. In accepting Bratman's account of shared intentionality, Tomasello et al.'s analysis assumes that, in order to engage in joint intentional action, children have a minimal theory of the mind: at least one that involves a concept of another person as an intentional agent (with perceptions, intentions, emotions, etc.), if not the concept of a mental agent (that involves a belief-desire psychology, see Tomasello & Rakoczy 2003). If we follow Bratman's conception of shared intentional action, children who are capable of participating in joint action must be said to at the very least have both an understanding of the concept of intention as applied to themselves and others and common knowledge of their own intentions as well as those of others. This would of course involve higher order propositional attitudes; second and third order attitudes would at least be involved. It seems sensible to think that this is too much

---

7 That this is so is recognized by Warneken & Tomasello 2007: "[i]ntraindividual comparisons of infants' performance on helping and cooperation tasks [Bratman's shared intentional activity] revealed no straightforward associations between the two suggesting that these activities differ in important ways" (p. 291, see also pp. 291-292). What we suggest is precisely that, while this should be expected when comparing helping behavior with Bratmanian coordinated activities, the results might be different if the demands for engaging in joint action were provided by a different theoretical model. In an alternative setting, the temporal and conceptual links between the two might become apparent.

to ask of children in order for them to be able to engage in joint intentional action, not only because they engage in this sort of activity quite early, but also because engaging in some shared activities and joint actions may be essential for acquiring such sophisticated knowledge of other persons as agents – instead of presupposing it.

**5.
Conclusion**

If the considerations developed so far are correct, then they suggest that this two-factor explanation of the development of social understanding is not completely plausible vis-a-vis the behavior young children display when helping others and cooperating with them in different coordinated activities. We have argued that this is so because Bratman's theory of shared intentions, which is used to model children's understanding of others and their helping behavior, does not seem to square with the experimental observations conducted by Warneken, Tomasello et alia.

The suggestion we would like to put forward as a conclusion to this paper is that such difficulties might be overcome if one uses an altogether different notion of shared intention, one in which collaboration flows from the very structure of the intentions shared by the agents without the need to postulate any further factor like an innate pro-social attitude. But which notion of shared intention could be of use here? It seems promising to look into those authors (e.g. Gilbert 2014 and Toumela 2005) that describe collective intentionality not as merely distributive, but as genuinely collective, as the intentionality of a group or of a *we*. A strategy like this might help address the three difficulties presented above, i.e., (i) showing that there is an internal articulation between helping behavior and shared intentions, (ii) explaining children protesting and reengaging behavior in joint activities, and (iii) providing a less cognitively demanding explanation of children's ability to cooperate with others than Bratman's model offers. Arguing in favor of these ideas, however, clearly exceeds the purpose of this paper.

REFERENCES

Bratman, M. (1984), "Two faces of intention", *The Philosophical Review*, 3, pp. 375-405;
----. (2014), *Shared Agency. A Planning Theory of Acting Together*, Oxford University Press, Oxford;
----. (2015), "Précis of shared agency: A planning theory of acting together", *Journal of Social Ontology*, 1(1), pp. 1-5;
Butterfill, S. (2012), "Interacting mindreaders", *Philosophical Studies*, 165(3), pp. 841-863;
Brownell, C., Ramani, G. & Zerwas, S. (2006),"Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds", *Child Development*, 77(4), pp. 803-821;
Gilbert, M. (2014), *Joint Commitment. How we Make the Social World*, Oxford University Press, Oxford.
Michael, J., Sebanz, N., & Knoblich, G. (2015). "The sense of commitment: A minimal approach", *Frontiers in Psychology*, 6, 1968. doi: 10.3389/fpsyg.2015.01968
Pacherie, E. (2013),"Intentional joint agency: Shared intention lite", *Synthese*, 190(10), pp. 1817-1839, (DOI) 10.1007/s11229-013-0263-7;
Salice, A. & Satne, G. (*under review*),"Two ways of sharing goals";
Schmid, H.B. (2010),"Philosophical egoism: its nature and limitations", *Economics and Philosophy*, 26, pp. 217-240;
Tollefsen, D. (2005),"Let's pretend! Children and joint action", *Philosophy of the Social Sciences*, 35(1), pp. 75-97;
Tomasello, M. (2014), *A Natural History of Human Thinking*, Harvard University Press, Cambridge;
Tomasello, M. & Rakoczy, H. (2003), "What makes human cognition unique? From individual to shared to collective intentionality", *Mind & Language*, 18(2), pp. 121-147;

Tuomela, R. (2005), "We-intentions revisited", *Philosophical Studies*, 125, pp. 327-369;

Warneken, F. & Tomasello, M. (2006), "Altruistic helping in human infants and young chimpanzees", *Science*, 311, pp. 1301-1303;

----. (2007), "Helping and cooperation at 14 months of age", *Infancy*, 11(3), pp. 271-294;

Zahavi, D. & Satne, G. (2015), "Varieties of shared intentionality: Tomasello and classical phenomenology", in J. Bell, A. Cutrofello & P. Livingston (eds.), *Beyond the Analitic-Continental Divide: Pluralist Philosophy in the Twenty-First Century*, Routledge, London, pp. 305-325.

GIAN PAOLO TERRAVECCHIA
University of Padua
gp.terravecchia@libero.it

A PHENOMENOLOGY OF SOCIAL STANCES*

abstract

The paper develops a phenomenology of social stances, trying to show that Margaret Gilbert's work on joint commitments can be understood as a special case of what here presented. The offered conceptualization shows that "to accept" is an important moment of social reality (as in Gilbert's work), but also that there are many more stances to be discussed which are important. These are "to refuse" (or "to rebel against"), "to suffer", "to assent" and "to make something one's own". The last part of the paper tests the explanatory power of the sketched theory, trying to show that it provides valuable elements for an account of convention.

**1.
From joint
commitment to
social stances**

The present paper will try to show that, through a phenomenology of social stances, it is possible to give a better account of some of the philosophical problems Margaret Gilbert tries to solve in her social philosophy and that in this way it is possible to better explain the phenomena of joint commitment and acting together.

First of all, let us consider Gilbert's core notion of joint commitment, in order to prepare the background for the following parts. Gilbert's social theory is built on the notion of joint commitment. According to Gilbert, a personal decision is a case of commitment. This means that the commitment as such does not need to have a social dimension; indeed "a personal commitment is brought into existence by one person alone" (Gilbert 2014, p. 31). Consequently, a personal commitment is rescinded when the person changes her mind. A joint commitment, instead, comes into existence when two or more people commit themselves to do something as a body, or as a unit, or as one. Gilbert expresses her idea with the formula: "to endorse a certain goal as a body" (Gilbert 2008, p. 33). The commitment brings out obligations, so that after the joint commitment is in place the parties are reciprocally obliged to act according to such joint commitment. Each of the parties cannot rescind the commitment unilaterally, under normal circumstances. On this ground, Gilbert defines a collective action as follows:

> Persons X, Y, and so on, collectively perform action A if and only if X, Y, et. al. are jointly committed to intend as a body to perform A and, in light of this joint commitment, relevant persons from among X, Y, et al. act accordingly (Gilbert 2014, p. 70).

So, for instance, Jane and Hilda go hiking to the top of the mountain; for this reason they jointly commit themselves to intend as a body to hike to the top of the mountain (Gilbert 2014, pp. 70-71). Each of the two acts doing such things as not leaving the companion behind, encouraging her to move, or rebuking the other if she does not comply with the commitment. The notion of collective action is meant to work both for simple cases and for complex ones. As one can see the joint commitment creates obligations and collective actions: Gilbert introduces the "we", the collective action, as something new in social philosophy, not reducible to the acts of the individuals.

In what follows I will develop a phenomenology of social stances (§ 2.) trying to show that Gilbert's work on joint commitments can be understood as a special case of what here presented. The offered conceptualization will show that "to accept" is an important moment

of social reality, but also that there are many more stances to be discussed which are important, such as "to refuse"(or "to rebel against"), "to suffer", "to assent" and "to make something one's own". Finally, (§ 3.) I will test the explanatory power of the sketched theory, trying to show that it provides valuable elements for an account of convention.

I will now present a phenomenology of social stances made up of five irreducible elements: to refuse (or rebel against), to suffer (or to be subject to), to accept, to assent and to make something one's own.[1] I will try to show that each of them has its own essence so that in order to give an adequate account of social reality it is not possible to reduce their number. On the other hand, it would be possible to make further distinctions: for example, the case of refusal could be distinguished from that of rebelling (as I will point out).

For a better understanding of the importance of such distinctions it can be useful to consider the problems that their absence causes to other accounts of acting together. For this reason I develop my ideas keeping Gilbert's work on the background. In so doing, I will try to show that Gilbert's theory tends to overestimate the case of acceptance and does not see the importance of other relevant stances such as suffering, nor it is able to show the role played by refusal in social life.

**2.
A discussion of the social stances**

Refusal is a special case in the scale of social stances. Of course, refusal is not exactly rebellion. However, they have something important in common: the interior attitude and the manifest attitude of resistance and/or of contrast to an X they are against. For this reason and for the sake of brevity I will discuss them here together. If not specified, what will be said about one can be considered valid also for the other.

Rebellion is the zero level of the social relationship, being effective even when not actualized, because it is in principle always possible (apart from Orwell's dystopia described in *1984*, and even in that case as a tragic result of a struggle of the system against individual free will). Every relationship is already conditioned by the mere possibility of a refusal or, worse, of a rebellion. It is possible to fully appreciate acceptance and all the further stances (and eventually to feel gratitude) only on the background of the possibility of a refusal.

At first sight, it might seem strange to consider the case of rebellion as one step, even if quite low, on the ladder of social stances. Indeed, rebellion seems to be the negation of such a ladder. On second thought, however, we should consider the enormous difference between simply ignoring (I add "simply", because ignoring could hide a true and subtle form of rebellion), on the one hand, and rebelling, on the other. The former is a case of an unconditioned action, as it is not a reaction to something else: the action has its own reasons. Rebellion, on the contrary, entails the intentional action which has in its agenda the opposition to a social reality, such as a social status, a norm, an institution, which has a claim (or a set of claims). Rebellion fights the reasons of what is its target (X), or the very existence of X. It is indeed a social stance, even though extreme: its goal is to transform the social situation radically. The rebel, in her fight, acts "together" with her opponent: the fight is the resultant of the collective action.

There is a second way social agents can act together in rebellion. People can act together, having a common goal: they can go storming the Bastille, for example. Each rebel, in such a case, does not have any commitment to the others. If one simply leaves the ground, no one

**2.1
To refuse (to rebel against)[2]**

1 The expression "stance" means the same as the german "*Stellungnahme*". It refers to the intentional act of taking a position about something.

2 I do not use the notion of opposition ("to oppose") here: refusal or rebellion are discussed as social stances, while "to oppose" has more to do with an activity and therefore with a possible outcome of the social stance.

has the right to address her any complaint. At most, one can ask her: "Where are you going?", "Why are you leaving?", "Don't leave" (which in this case cannot be a command), but this is far from a rebuke. The single rebel, for psychological reasons is expected to keep the pressure against the fortress, but he or she does not have any joint commitment. In the same way, mobs can perform collective actions, such as destroying a fortress, or killing the tyrant, without any joint commitment between the social agents performing the joint action. This gives some troubles to judge flash mobs performing violent acts. Accepting this leads to see that some collective actions, though being joint actions, do not have any collective "we" responsible for what happened; and yet each single agent is responsible for his or her own taking part in the collective action that emerged.

Students of social sciences, like Gilbert, tend to restrict the acting together to the acting in accordance with someone. As we have just seen, this is a limit that must be corrected. Here is Gilbert's somehow romantic account of acting together:[3]

> Two or more people are acting together (doing something together) if and only if: (1) they are jointly committed to espousing as a body the appropriate goal; (2) they are fulfilling the behavioural conditions associated with the achievement of the goal; (3) their satisfaction of these conditions is motivated in each case by the existence of the joint commitment (Gilbert 2008, p. 146).

In some cases of rebellion, however, people act together with others, as we have seen, but there is no joint commitment of any kind, nor any common goal (at least among opponents). At this level we have collective actions of opposing and allied agents, but we do not have any social obligation.

2.2
To suffer (to be subject to)

While rebellion, acceptance, and the other social stances have been widely studied before (see e.g. Camus 1951; Gilbert 2014, pp. 131-162), the second step has been unfortunately neglected. To suffer or to be subject to the initiative of someone else does not entail being able to refuse it, or to rebel against it. Such stance is a midpoint between rebellion and acceptance. In this case, the social agent is passive, intimately not accepting what she considers to be wrong, or she dislikes, or she simply does not have reasons to accept. Yet she is not ready or able, or convinced, to contrast it. The strange thing is that apparently the social agent could be quite active in complying with what it is asked. This can be understood only by distinguishing social stances from social activities: one can be passive on the level of the interior personal stance and quite active on the level of the activity performed.

Let us take an example from Gilbert, whose discussions of going for a walk are possibly her major contribution to social philosophy. According to one of her examples, Paula and James ended up walking together, so performing a joint action. Let us see how she created and described the situation:

> Perhaps Paula said, "Shall we go for a walk?" thus proposing a walk to James, and he accepted her proposal by saying "Yes, let's". In short, they agreed to go for a walk together. Having thus agreed, they set off their walk. Such an agreement is, clearly, a standard way of initiating joint action (Gilbert 2008, p. 116).

---

3 Gilbert's defence is that her stipulative constructed theory explains the situations it explains. Her constructivist approach and her rigorous work give consistency to the theory. Here I try to show that the phenomena she discusses deserve a wider approach: consistency is important, but explanatory power is important too.

Here it seems that acceptance is essential: James accepted the proposal and after such acceptance the joint action began. Let us now consider some new background details and, after that, let us reconsider what is happening. We now know that Paula is a terrorist, she wants James far away from the other prisoners, for some reason. James does not want to look weak and, above all, he knows that it is useless to resist. Here James is not accepting any commitment, he is just suffering an imposition, though performed without brutality. He complies with the request and even declares that he will comply, but he does not have any commitment. It would be odd to say that if he tries to escape during the walk, Paula has the right to rebuke him, because of his saying "Yes". That agreement was just a sign of non rebellion. Again, also in this case, we have a collective action, but there is no joint commitment. We have a *prima facie* "we" acting coordinately, jointly, but nothing more, so that the agent suffering the action is active in performing the action itself. He is however not properly responsible for the doing, unless rebellion were a true option.

To suffer the initiative of others without rebellion, nor acceptance, is more common that one could imagine. I will try to show it later, when discussing Gilbert's account of conventions (see § 3.). I just want to anticipate here that to suffer the social reality does not necessarily entail feeling pain, being afraid or feeling sorrow; it is not necessarily a matter of negative feelings of any kind. It is rather a stance of the social agent not rebelling, nor accepting.

2.3 To accept

To accept the other's initiative is a further step: in this case the agent does not just suffer the activity. Here we are two degrees far from rebellion (and refusal): with acceptance the social agent is not passive and intimately against what is done. More generally, acceptance is a stance in which the social agent is outwardly and inwardly open to what he has to deal with (whether an act, a status, an institution etc.). The point here is that the agent does not comply unwillingly, he is merely concessive. He has a neutral stance which is open to start a course of action: to be opened to something is quite far from giving one's own assent to it. The initiative is of an X (someone or something) outside the social agent, so that the accepting social agent performs the minimal effort possible. Acceptance is the mildest stance.

To go back to the discussion of the walk of Paula with James presented above in its Gilbertian form, to accept the proposal, in the technical meaning here discussed, James will answer something like: "Why not?": he does not have reasons to say "no", and accepts the initiative of the other. Accordingly, one can see that Gilbert speaks of "acceptance" about James' answer, "Yes, let's", because of the lack of a phenomenology of social stances in her thought. In a "Yes, let's", given under normal circumstances, we can find much more than a simple acceptance, as this stance is here explained.[4]

It is important here to notice that from this stance onwards, social obligations arise. Now James cannot leave Paula alone, taking a new direction, without any explanation or excuse. He has a social obligation to walk with her, and even a moral duty not to harm her sensibility leaving her alone or, worse, leaving her in a difficult situation such as walking alone in a dangerous neighborhood. Acceptance, however mild it may be, is the starting point for a collaborating activity, so that it is possible to say that it is the initial moment of an aware buildup of the social reality. Nevertheless, with sole acceptance, no social reality would be initiated, since acceptance is a passive stance. To find something more proactive one must go to the next social stances.

2.4 To assent

Assent can be treated in many ways: from a cognitive perspective, from a psychological point of view, with the approach of the philosophy of language. From the point of view of social

---

4 For more on this, see the next paragraph.

philosophy, assent has to do with the stance of social agent *a* in respect to an X (an act, a proposal, an institution, an emerging social reality, etc.) the agent *a* is dealing with. The assent is a social stance which consists in an interior "yes" given without reserve and manifested. The "Yes, let's" of James, at length already discussed above, seems to be an instance of assent, rather than of acceptance. This is true, if the act is what it seems to be. In fact, we have already seen that a "Yes, let's" could actually be a case of a suffered initiative.

The discussion so far can be summarized with the following table, built on the intuition of Adolf Reinach according to which social acts have two irreducible moments, an interior one and an exterior one (Reinach 1913):

	To refuse	To suffer	To accept	To assent
Interior attitude	–	–	0	+
Exterior manifestation	–	+	+	+

There are three values about the stance of the agent: negative (–), neutral (0) and positive (+). The table helps to understand why it may be easy to misinterpret, from an external perspective, a case of suffering situation X, with a case of giving assent to X.

The table also helps to grasp the reasons why assenting implies commitment: the intimate "yes" and its manifestation create social obligations. The social agent will be under a more severe judgement if she leaves the joint action assented, than she would by leaving a joint action she merely accepted. There are several levels of commitment. Social agents normally understand the social stance of the others and they judge and act accordingly. This fixes also the levels of trust from a degree that goes from assent to incorporation, the stance discussed in the next paragraph.

Indeed a good part of Gilbert's analysis, rather than with acceptance, deals with assent, so that her theory of joint commitment can be considered a special case of a wider theory sketched through the analysis offered in the present paper.

2.5 To make something one's own

It may seem that a social stance more positive than assent could not exist; and indeed the fifth stance I am about to discuss does not differ in degree from the previous one. In order to make something one's own the agent must have a positive stance and also manifest it. The difference has rather to do with the distance of the X accepted from the social agent *a* which takes her personal position. All previous cases entail a certain distance of agent *a* from the X judged and refused, accepted and so on. In the present case, on the contrary, the distance is zero: the agent makes of X her own mind. She decides to identify herself with it, somehow.

Let us try to understand how it works, using a comparison with a joint action derived from the assent stance. James and Paula are going for a walk, as we already know. We can imagine the two being happy about it and manifesting their feeling. In this case they will behave as a body, walking together for some time. We have here a body given by two distinct members, each doing their part. They behave as "a we", but I would not say "as one". Each is doing his or her part and is responsible for that alone.

Let us now imagine that the two are colleagues, cops for instance. They are not just giving their assent, they are now part of a patrol. While having the walk, they see a well known thief and they decide to capture her. Paula blocks the way of escape and James goes straight to the thief. They capture her, performing a joint action in which they acted as "a we", and indeed "as one". They are responsible for the action as a whole and probably they will be praised. They both identify themselves as cops, they act as such and in this joint action the social

bond between them made them be one thing: the police making an arrest. To be incorporated into something (e.g. family, group of friends, a social institution) gives reasons to act as a we, performing collective actions "as one".

Discussing David Lewis' famous book on conventions (Lewis 1969), Margaret Gilbert presents her own explanation. It seems to be plausible to read her text as I will do here, but I leave it opened if this is a correct way. For the sake of my argument, it is enough to show that in comparison with the perspective presented, my account of social stances gives a contribution. Now, Gilbert defines conventions as jointly accepted *fiat*. Here is her definition of a joint behavior according to a given convention: "Members of a population, P, jointly accept a given *fiat* if and only if (by definition) they are jointly committed to accept as a body that *fiat*" (Gilbert 2014, p. 218).

**3.
About conventions**

I want to stress three points about this approach to conventions. First, it is not true that any possible convention is the result of a *fiat*. Under the Nazis, during the Holocaust, Jewish people were supposed to wear the star of David. This convention was the result of an odious *fiat*. However, there are and there have always been many other conventions – such as "wearing a top hat", "using in English the word 'convention' to mean convention", "nodding to indicate acceptance" –which have not been instituted by any *fiat*. The *fiat* is an intentional act of institution, while the mentioned conventions simply emerged. There is a second and perhaps more dramatic trouble within such understanding of conventions: it seems to require acceptance. To keep with the example of the star of David, even if a few Jewish, at the beginning, accepted proudly to wear it, in the end they all realized that it was an instrument of discrimination and hate. We see here a convention which cannot and should not be accepted and indeed was suffered by the majority of the Jewish people and at end by all of them. Finally, when one follows a convention, she does not bother to do it as a "we", if not in special cases. Again, as I said, some Jewish people were initially proud to wear the star of David, because it gave them the feeling of being a "we". But, to make a different example, when we use the word "convention" we are not thinking of our being a "we". Which kind of "we" would be at stake? The degrees of stances I have discussed above are useful to understand the mechanism of following a convention. People can just suffer a convention or they can accept it. These two cases allow us to understand why it is quite easy to change conventions: there is a significant number of people that do not feel any particular commitment to follow a given convention. This fact is not explicable within Gilbert's account. Some people may give their assent to a convention, for various reasons (they like it, they have a special role etc.). These are the supporters of the convention and they usually tend or are in duty to give social sanctions to those who do not follow it, if this is expected. A teacher, for instance, will underline with a red pen her student's text if it reads "convintion" instead of "convention". Finally, there are those who think that a certain convention expresses the "we" of their group. In this case, they will be ready to apply some kind a social rebuke, from a nasty look to more severe sanctions such as ostracism. Indeed, to enter the tatami in jeans and sweater is not advised, not even to a beginner.

REFERENCES

Camus A. (1951), *L'homme révolté*, Gallimard, Paris; English translation: *The Rebel: An Essay on Man in Revolt*, Knopf, New York, 1956;
Gilbert M. (2008), *A Theory of Political Obligation. Membership, Commitment, and the Bonds of Society*, Oxford University Press, Oxford;
----. (2014), *Joint Commitment. How We Make the Social World*, Oxford University Press, Oxford;
Lewis D. (1969), *Convention: A Philosophical Study*, Harvard University Press, Cambridge MA;

Reinach, A. (1913), "Die apriorischen Grundlagen des bürgerlichen Rechts", in Reinach, *Sämtliche Werke. Texkritischeausgabe in 2 Bänden*, hrsg. von k. Schuhmann - B. Smith, Philosophia Verlag, München-Hamden-Wien, 1989, pp. 141-278; English translation: "The apriori foundations of the civil law", in J. Crosby (ed.) (1983), *Aletheia*, pp. 1-142.

SESSION 3

SESSION 3

SHARED NORMS

FRANCESCO GUALA
Università degli Studi di Milano
francesco.guala@unimi.it

THE NORMATIVITY OF INSTITUTIONS

abstract

Some philosophers have claimed that normativity is an essential feature of social institutions, and have identified the source of normativity in collective intentionality. In this paper I argue that (1) normativity is not essential for social institutions; (2) normativity has many sources and it is unlikely that a single theory is able to account for them all; (3) a powerful conception of institutions – the "rules in equilibrium" account – provides an adequate framework to represent the effects of normativity and to explain its social function; (4) it is a mistake to ask a theory of institutions to do more than that. While normativity is an important element of institutions, we should not build our social ontology on a single, specific conception of normativity.

1. Introduction

The topic of my talk is broader and narrower than the topic of the Spring School.[1] It is broader because I will focus on normativity, which is a more general phenomenon than commitment. But in another respect the topic is narrower because I will focus specifically on the normativity of institutions. This choice is determined mainly by my research interests, but what I will say has implications for social ontology in general – or at least I hope so.

By way of an introduction, it is useful to place the problem in a wider framework. Normativity has become an increasingly important topic in contemporary philosophy. According to some scholars it is *the* most important question of philosophy. Partly this is because the domain of philosophy has been shrinking under the pressure of science, and as a consequence philosophers have retrenched in those domains where science does not seem to be able to make much progress. But one reason why science has made little progress is that normativity is a complex, even puzzling phenomenon. So philosophers have been lured toward a topic where progress may be difficult to obtain.

I should like to clarify at the outset that I do not mean to criticise those philosophers who work on normativity, and especially the normativity of collective or joint intentions. Collective intentionality is a fascinating topic in itself, and may turn out to be of utmost importance in ethics and the philosophy of mind. My view is simply that its role in social ontology has been misunderstood, and that it is important that we look at it from a different perspective. This position is not particularly original – I share it with those naturalists who believe that current debates in social ontology have become too detached from contemporary social science.[2] I fear that the divide between naturalists and anti-naturalists is growing deep (again) in social ontology, to the detriment of fruitful conversation, and this talk is a modest attempt to bridge the gap.[3]

Some of the thoughts that follow are admittedly only half-baked. I am addressing a huge topic,

---

1 This paper is a revised transcript of the talk given at the Spring School on Joint Commitment, held at San Raffaele University in June 2015. Some of the material upon which the talk was based, and which is included in this written version, comes from my forthcoming monograph *Understanding Institutions*, as well as from papers written jointly with Frank Hindriks (see the reference section). Frank does not necessarily agree and is certainly not responsible for the views expressed in this paper, however. I am grateful to the members of the audience for their feedback, and especially to Margaret Gilbert and Christian List for subsequent conversations on the topic of this talk.

2 See e.g. Kincaid (2012), Turner (2010).

3 I have written about this in Guala (2015).

and I am concerned that social ontologists do not spend enough time thinking about it. If the talk will be at least a little provocative, I will have succeeded.

Social ontologists often make use of a classic philosophical argument that ought to be handled with care. The argument begins by claiming that current social science cannot explain X, where X is some interesting entity or phenomenon the existence of which we do not normally question. It then proceeds by saying that social science cannot explain it satisfactorily because Y is necessary to explain X, and our best scientific theories have no room for Y. Therefore, current theories must be revised or replaced by better theories that are able to account for Y. Raimo Tuomela, for example, has claimed that "central social notions as cooperation, social institutions, and the evolution of institutions, as well as collective and group responsibility, do require the full we-perspective as an underlying notion" (Tuomela 2002, p. viii), a perspective that standard theories in the social sciences lack. Therefore, his own research in social ontology "is meant in part to critically analyze the presuppositions of current scientific research and [...] to provide a new conceptual system for theory-building" (2002, p. 7). The structure of Tuomela's argument is the following: (1) Y (collective intentionality) is necessary to explain X (cooperation, conventions, institutions, responsibility). (2) Current social science has no room for Y. Therefore, (3) current social science cannot explain X, and ought to be reformed accordingly.

2. Collective intentions and normativity

What has collective intentionality to do with normativity? Seminal theories of collective intentions typically begin with a critique of standard accounts of coordination and cooperation such as those that are commonly used in rational choice theory. The idea is that institutional behaviour cannot be explained as the aggregate outcome of a multitude of actions guided by individual intentions ("I want X, I believe Y", "you want X, you believe Y", etc.), plus some reflexive or recursive notion such as common knowledge. An adequate account must be built on the observation that the intentional states of human beings engaged in institutional activities are usually formulated in a collective mode ("we want X, we believe Y"), and such intentions have an intrinsic normative component.

I am simplifying a lot here, and I do not mean to say that anyone subscribes exactly to this simple version, but it gives the gist of the idea. Here are some examples to it back up. The first two are from Margaret Gilbert's seminal book *On Social Facts*:

> Our conception of a social convention is the conception of a quasi-agreement which is generally regarded as grounding an ought judgment, where it is common knowledge in the population that this is so (Gilbert 1989, p. 369).
> The link is a conceptual one: normative attitudes are to social conventions as femaleness is to sisterhood (Gilbert 1989, p. 350).

John Searle similarly identifies "deontic powers" as fundamental building blocks of social ontology:

> An institution is any collectively accepted system of rules (procedures, practices) that enable us to create institutional facts. [...] The creation of an institutional fact is, thus, the collective assignment of a status function. The typical point of the creation of institutional facts by assigning status functions is to create deontic powers. So typically when we assign a status function Y to some object or person X we have created a situation in which we accept that a person S who stands in the appropriate relation to X is such that (S has power (S does A)). The whole analysis then gives us a systematic set of relationships between collective intentionality, the assignment of function,

the assignment of status functions, constitutive rules, institutional facts, and deontic powers. (Searle 2005, pp. 21-22)

Finally, according to Tuomela,

if there was an agreement to perform the joint action, there is also a non instrumental, intrinsic "ought", due to the very notion of agreement, to paint the house together and to perform one's part (Tuomela 2002, p. 142).
Collective acceptance in the strict we-mode sense entails that the participants have as their joint intention to do X in C and are collectively committed to this (Tuomela 2007, p. 207).

3.
Is normativity necessary for institutions?

The claim that the normativity of collective intentions is a necessary constituent of human sociality is not very plausible. The main reason is that the distinction between collective and individual mental states does not mirror the distinction between institutional and non-institutional, or even social and non-social phenomena. In other words, a taxonomy based on collective intentions does not cut the social world where it should.

To see why, it is sufficient to realise that many social institutions do not rely on normative commitments engendered by a joint intention. Consider racist institutions such as those that used to be common in many countries until not so long ago. Such institutions were constituted by many rules, some of which were encoded formally in the legal system and some of which were enforced by other means. In South Africa for instance, during the apartheid era, black people were expected to use different public transportation, to live in segregated townships, and white people could decide how to allocate resources (such as land) among the various "races" or ethnic groups. Informal conventions were also an important part of the system. In his autobiography, *Long Walk to Freedom*, Nelson Mandela for example recalls that at tea-time the black lawyers of the firm where he worked were expected not to use the same cups used by the white employees.

Now, to say that these rules were based on a collective agreement seems very bizarre. The rules certainly carried deontic force: the black person who refused to conform was breaching a norm. Those who conformed in turn did it because they were expected to, and did not dare contradicting these expectations – because they feared the punishment inflicted on deviants. But the norm did not rest on a collective agreement or a group intention: the whole point of apartheid in fact was to undermine the notion that black and white people belonged to the same community.[4]

The underlying problem is that normativity is an ambiguous notion. On one reading, a normative behaviour is just a behaviour that is expected and sanctioned in the community. This is the standard social science interpretation and does not require any notion of collective intentionality. On another interpretation, a normative behaviour is mandatory in virtue of a stronger binding relation – like a pact or a promise – that has been forged in the community. The normative force, in this case, is not based merely on expectations but on some underlying contract that provides legitimacy to the sanctions.

Returning to our example, it seems clear that the racist institutions of South Africa were endowed with normative power in the former, but not in the latter sense. And yet surely they were genuine institutions. But then we are forced to admit that the normativity of joint

4 White people may have collectively agreed to enforce apartheid, of course. But the black did not, and nevertheless conformed most of the time. So clearly conformity with an institution does not require collective agreement.

commitment is not a necessary feature of institutions. Whether collective intentions are involved or not in social action seems to be a contingent rather than a conceptual issue, and we should better not build social ontology on this premise.

Some collective intentionality theorists have recognised the problem, and over the years have modified their position accordingly. For example, Searle now claims that institutions only require collective "recognition", instead of collective acceptance. This means that some institutions, like apartheid, exist even if some people are not bound by a joint commitment and do not take the rules to be legitimate: "one can recognize and act within institutions even in cases where one thinks that the institution is a bad thing" (Searle 2010, p. 57). But with commitment, collective intentionality also has to go. "This is an important point, because it shows that there are some forms of collective intentionality which are reducible to I-intentionality plus mutual belief" (Searle 2010, p. 58).

This line of argument, if correct, indicates that collective intentionality is not necessary for the existence of institutions. It does not demonstrate that collective intentionality does not exist or that it is useless, of course. We can happily concede that joint intentions and commitments are probably important for the functioning of *some* institutions.[5] But we should not make them the starting point of a controversial transcendental argument. As Stephen Turner has pointed out,

**4.
An alternative
theory of
institutions**

> One oddity of transcendental arguments is this: they work only when there is no under determination, which is to say where the logical conditions for some possibility are univocal. But in the kinds of cases discussed by normativists, there are ordinarily a variety of theories. (Turner 2010, p. 22)

In this section I will argue that there is a viable alternative theory of institutions, and that normativity plays a different role in this theory than the one it plays in the collective intentionality programme. In various papers and in a forthcoming book I have argued that a satisfactory account of institutions can be found at the core of contemporary social science.[6] According to this account institutions are sets of *rules in equilibrium.* The main function of the rules is to indicate actions that promote coordination and cooperation. And, crucially, each individual has an incentive to choose those actions, if the others do the same.[7]

For example, here is a sample of rules belonging to some familiar institutions:

- if the light is red you must stop, if it's green you can go (traffic);
- if the land has been registered in your name, you can use it (property);
- if the bill has been printed by the Central European Bank, you can use it as a medium of exchange (money).

Each rule codifies a behaviour – a set of actions – that solves a problem of coordination. A coordination problem is a strategic game with multiple equilibria. An equilibrium in game theory is a profile of strategies (or actions), one for each player participating in a strategic interaction. Each action may be described by a simple sentence of the form "choose X" or "do Y". The defining characteristic of an equilibrium – what distinguishes it from other

---

5 In fact there are good reasons to believe that reasoning in collective mode is useful for coordination, and hence for the existence and resilience of institutions, even though it is not necessary (see Guala 2016, Chs. 8-9).
6 Cf. Guala and Hindriks (2015), Hindriks and Guala (2015), Guala (2016).
7 For similar views, see Aoki (2011), Binmore (2010), Greif and Kingston (2011), Smit et al (2014).

profiles – is that each strategy must be a best response to the actions of the other players or, in other words, that no player has an incentive to change her strategy unilaterally. If the others do their part in the equilibrium, no player can do better by deviating. A classic example is the "driving game": drivers do not particularly care about keeping right or left, provided everybody does the same (Figure 1). So there are two possible equilibria in pure strategies, LL and RR. The theory, however, can be easily generalised to other cases, where the payoffs are asymmetric and the players have different preferences about the outcomes.
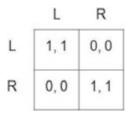
	L	R
L	1, 1	0, 0
R	0, 0	1, 1

Figure 1. The driving game

It may be pointed out that the rules that govern our social interactions, from traffic to property, do not simply *indicate* an action ("do X"): usually, they have a normative element built into them ("you *must* do X"). The rules of traffic for example say that you *ought* to stop if the light is red, or that you *ought* to drive on the right-hand side of the road when you are in Italy. The rules of private property assign a right ("if it's registered in your name, you can use it"), indicating something that you have the possibility to do, and possibility is a deontic notion. Your rights, moreover, are typically my duties: to say that you can use a piece of land is tantamount to say that I must not use it without your permission. So deontic powers must be accounted for by the theory of institutions as rules in equilibrium.[8]

The observation is correct, and it is important to realise what it implies and what it does not. It would be silly for a theory of institutions to deny that norms play an important role in the regulation of human behaviour. But one thing is to say that normativity must be allowed some room in the theory, quite another is to build the theory on a specific notion of normativity. The sensible strategy is to make room for different kinds of normativity in the theory, without making a strong commitment to a single account.

How can normativity be represented in the framework of the rules-in-equilibrium theory? A convenient way of modelling the constraints introduced by norms is in terms of costs. As long as the latter are understood in a sufficiently broad way, we can say that compliance with a norm carries costs while conforming to a non-normative rule or convention (a rule of thumb for example) does not. For example, a littering norm imposes the cost of looking for garbage bins. A norm against cheating imposes the opportunity cost of not having extra-marital affairs, and so forth. Such costs call for compensation (counter-reasons) in order to be overridden. And counter-reasons may be represented as negative incentives – other costs, effectively – that deter norm violations.

The representation in terms of costs is compatible with various enforcement mechanisms. Social scientists distinguish between internal and external costs, corresponding roughly to a distinction between "internalised norms" and "externally sanctioned norms". The first type of norm has been central in the sociology tradition, dating back to the work of Talcott Parsons: social norms according to this interpretation are internalised prescriptions that people follow

---

8   See Searle (2015) for a critique of this kind.

partly out of habit, and partly because they do not want to feel bad about it. The costs may be seen as a form of self-inflicted punishment, in the form of guilt feelings. Another equally influential tradition sees social norms as supported by external punishment instead. In this case, the costs are inflicted by other members of society, either professional enforcers (judges, policemen) or fellow citizens who just happen to witness the transgression.

Elizabeth Crawford and Elinor Ostrom (1995) introduce a delta parameter to capture the costs involved in violating a norm. In the case of externally sanctioned norms, the cost of transgression is generally known to the members of the community, and can often be quantified objectively. A century ago, for example, the cost of refusing to marry a Sicilian woman after a one-night stand was equal to the expected value of death. The cost of littering the floor in Singapore is three hundred dollars weighed by the probability of being fined. When the costs are internalised, deltas are more difficult to measure of course. But measurement is not impossible, if we have enough information about the desires that the norm is supposed to trump. The key is to know what people would have preferred to do if the norm had not been in place. The abiding citizen who regularly clears the pavement in front of her house from the snow invests precious time that she might have liked to spend differently (by watching TV in her warm living room, for example) had she not internalised the norm. Representing normative power by means of delta parameters (costs) facilitates the extension of the rules-in-equilibrium theory beyond the realm of coordination games. Many social theorists have pointed out that institutions improve the performance of players in games where there is an individual incentive to deviate from the socially optimal rule. A classic case is the prisoner's dilemma game (Figure 2). Instead of multiple equilibria – as in the coordination games examined so far – there is only one equilibrium in the one-shot prisoner's dilemma (DD). Augmenting the game by means of conditional strategies does not help, so there is no way to solve a prisoner's dilemma using an external correlation device. Defecting strictly dominates rule-following. By means of a suitably large delta, however, one can transform the dilemma into a game with two equilibria, CC and DD. Normative rules thus can turn a cooperation problem into a coordination problem: norms change the games that people play. And, of course, they also work as coordination devices in the new games that have just come about.

	C	D
C	2, 2	0, 3
D	3, 0	1, 1

Figure 2. The prisoner's dilemma game

	C	D
C	2, 2	0, 3-δ
D	3-δ, 0	1, 1

Figure 3. The prisoner's dilemma with delta parameters

Suppose there is a rule in the population that says "if the other player cooperates, then you ought to cooperate, otherwise defect", and the rule has normative force. This means that an extra cost (delta) must be subtracted from the payoffs, as in Figure 3. The payoffs in these two games may be interpreted in various ways: one possibility is to take the numbers in the original game (Figure 2) as representations of pre-normative desires. In the new game (Figure 3), then, the numbers may represent individual goals after the internalised desire to comply with the norm has been taken into account. Alternatively, the modified payoffs of Figure 3 may incorporate new information that has become available to the players, for example when they discover that a system of punishments has been set up to deal with those who transgress

the rule. In any case the deltas represent the force of normative rules, taking a given (pre-normative) game as a benchmark.

Depending on the force of the rule, the second game may turn out to be quite different from the first one. If delta is at least as large as one unit of payoff, then the prisoner's dilemma is transformed into another simple game, where DD and CC are both equilibria. If delta is equal to three, for example, we obtain the hi-lo game of Figure 4.
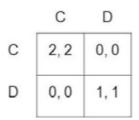
	C	D
C	2, 2	0, 0
D	0, 0	1, 1

Figure 4. From a prisoner's dilemma to a hi-lo game

Notice that norms then "solve" dilemmas of cooperation only in a peculiar sense. Games such as the one-shot prisoner's dilemma cannot be properly solved, because there is no way to escape the disturbing conclusion that players ought to defect, without changing the rules of the game. The only "solution" is to change the game itself, and this is precisely what institutions endowed with normative power can do. They create new equilibria introducing costs that make defection unattractive, at least within a certain range of payoffs.

But this is not the only function of norms. Deontic powers are useful in games of coordination as well. This may sound strange, since the players do not have any reason to deviate from a coordination equilibrium. Adding deontic power seems gratuitous, if there is no selfish desire to counter-act. But deontic power is unnecessary only if the players have full information and never make mistakes. In real life people have the unfortunate propensity to make mistakes, to misinterpret the rules, and sometimes they are also uncertain about the payoffs of the game. In such cases it is useful to have some extra mechanism that helps enforce conformity with a rule. People pay more attention to the signals they receive and to the structure of the game, if they know that they will pay an extra cost in case of non-compliance. So it is not surprising that many conventions have a tendency to turn into norms.[9]

Norms thus help fulfilling two key functions of institutions: they stabilise behaviour and make it more predictable in situations of uncertainty; but they also create behaviours that did not exist before, by changing the payoffs of a game. Thus norms facilitate not only the persistence of institutions, the fact that they are followed in the face of incentives to deviate; sometimes they also facilitate the emergence of new equilibria, when they are introduced ex novo. A central authority – like a government or a recognised leader – can reshape a game with "bad" equilibria, such as a prisoner's dilemma, and turn it into a better game, such as hi-lo. New rules may be introduced by decree: if the government announces that all cars will drive on the other side of the road starting from tomorrow (as it happened in Sweden in 1967); and if the announcement is supported by credible formal and informal sanctions, then the players will recognise that the game has changed and that a new equilibrium has been created. The punishment mechanisms will make the new equilibrium more salient than it would have been in virtue of the announcement alone. So the function of normativity can be accounted for within the framework of the rules-in-equilibrium theory.

---

9 See Guala and Mittone (2010) and Guala (2013) for empirical evidence and philosophical discussion.

Notice that the rules-in-equilibrium theory only provides a formal apparatus to represent normativity, but it remains neutral about the nature of normativity or where it comes from. And I think that this is just as it should be. Normativity is one of the thorniest issues in contemporary philosophy, and it would be foolish to make a theory of institutions depend on a specific account. Part of the problem is that normativity may not be a single thing. When we say that "you ought to do X" we usually mean that we expect you to do X, even if you have reasons (especially selfish reasons) to do otherwise. But the "ought" may have several different sources. For example, when I say that "you ought to slow down" I can mean any (perhaps more than one) of the following:

1. if you don't, you will be fined;
2. if you don't, I will be angry at you;
3. If you don't, your reputation will be ruined;
4. If you don't, you will not belong to this community anymore;
5. If you don't, you will contradict yourself and behave irrationally;
6. if you don't, and you cause an accident, you will become a bad person;
7. if you don't, and you cause an accident, you will feel bad about it;
8. if you don't, and you cause an accident, God will punish you.
(And so forth – the list is certainly not exhaustive.)

Each of these sentences refers to some mechanism that is a plausible candidate for behavioural regulation. Scholars disagree on which one (or ones) is more important. Some philosophers and social scientists believe that normativity can be explained in terms of mutual beliefs and the feeling of resentment that we experience when our expectations are frustrated (Lewis 1969; Sugden 1998); others believe that normativity requires a stronger notion of collective or joint intention (Gilbert 1989); some philosophers and social scientists argue that normativity depends on emotions (Frank 1987; Gibbard 1990; Nichols 2004); and still others believe that normativity has to do with the possibility of justifying our actions by means of rational arguments (Broome 2013; Raz 1999; Skorupsky 2010). Whether any of these accounts is able to capture "true" normativity – whatever it may be – is an open issue that we do not need to settle. In fact choosing among them may not be very sensible: if normativity is important for institutions, then it is likely to take several different forms. As an analogy, consider the multiple ways in which an organism tries to accomplish a goal that is important for its survival. If perceiving the existence of a prey is important, it is likely that a predator has more than one way to accomplish that task (by vision, hearing, and smell for example). Nature likes redundancy. Similarly, normativity probably has different sources and many facets, which means that more than one account is likely to be right. So instead of asking what normativity is, it is more sensible to ask what it does, or what its function is. One drawback of this approach is that it does not allow one to make any substantial normative assessment of institutions. It does not allow to distinguish good from bad institutions – dictatorship from democracy, capitalism from socialism, or poligamy from monogamy, for example. My own view is that judgments of this sort belong to the realm of ethics, rather than social ontology, and that it is best to keep the two projects separate. Other philosophers disagree and have tried to construct more robust theories of institutions, but I will let the readers make up their own minds on this point (see e.g. Miller 2010).

5. Conclusion

By way of a conclusion, let me summarise the overall argument of this talk. The twin notions of collective intention and joint commitment lie at the core of a hugely successful research programme in social ontology. In fact the project has grown so large that it constitutes the bulk of the research that is currently done on in this subfield of philosophy.

There is nothing wrong with this situation, in principle, because collective intentions and commitments are interesting phenomena that deserve philosophers' attention. But ever since its inception the programme had higher ambitions: the goal was not simply to better understand a peculiar kind of mental state, but to build social ontology (a whole theory of institutional reality) on collective intentions and their deontic force.

This programme, I have argued, is bound to fail because institutional reality does not require collective intentions, joint commitments, and the like. It is easy to find counterexamples (institutions that do not depend on collective intentions) as we have seen in section three. I have been careful to say that this does not mean that commitments, normativity and related notions are unimportant for the existence and persistence of institutions. On the contrary, they are among the mechanisms that facilitate coordination and cooperation in situations of strategic interaction. But these mechanisms differ from one another, and normativity is unlikely to have a single source. For this reason, a theory that aims at providing a general account of institutions should be flexible enough to make room for several types of normativity without committing a priori to one of them.

In section four I have argued that the account of institutions that is implicit in contemporary social science – the theory of institutions as rules in equilibrium – meets this criterion. Philosophers may dislike it because it does not shed light on the "true" nature of normativity, but if I am right this is a virtue rather than a defect. An entirely general theory of institutions can only highlight the *function* of norms and commitments.[10] The way in which the function is fulfilled is probably context-dependent and is a matter for psychologists and sociologists to discover by means of empirical research, rather than by armchair theorizing that is prevalent in social metaphysics.

REFERENCES

Aoki, M. (2011), "Institutions as cognitive media between strategic interactions and individual beliefs", *Journal of Economic Behavior and Organization*, 79, pp. 20-34;

Binmore, K. (2010), "Game theory and institutions", *Journal of Comparative Economics*, 38, pp. 245-252;

Broome, J. (2013), *Rationality Through Reasoning*, Blackwell, Oxford;

Crawford, S. E. S. and Ostrom, E. (1995), "A grammar of institutions", *American Political Science Review*, 89, pp. 582-600;

Frank, R. (1987), *Passions Within Reason*, Norton, New York;

Gibbard, A. (1990), *Wise Choices, Apt Feelings*, Harvard University Press, Cambridge, Mass.;

Gilbert, M. (1989), *On Social Facts*, Princeton University Press, Princeton;

Greif, A. and Kingston, C. (2011), "Institutions: Rules or equilibria?", in N. Schofield and G. Caballero(eds.), *Political Economy of Institutions, Democracy and Voting*, Springer, Berlin, pp. 13-43;

Guala, F. (2013), "The normativity of Lewis conventions", *Synthese*, 190, pp. 3107-3122;

----. (2015), "Philosophy of the social sciences: Naturalism and anti-naturalism", in P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*, Oxford University Press, Oxford;

----. (2016), *Understanding Institutions*, Princeton University Press, Princeton;

Guala, F. and Hindriks, F. (2015), "A unified social ontology", *Philosophical Quarterly*, 65, pp. 177-201;

Guala, F. and Mittone, L. (2010), "How history and conventions create norms: An experimental study", *Journal of Economic Psychology*, 31, pp. 749-756;

Hindriks, F. and Guala, F. (2015), "Institutions, rules, and equilibria: A unified theory", *Journal of Institutional Economics*, 11, pp. 459-480;

10   For an instructive example, see Michael and Pacherie (2015).

Kincaid, H. (2012), "Introduction: Doing philosophy of the social sciences", in H. Kincaid (ed.), *The Oxford Handbook of Philosophy of Social Science*, Oxford University Press, Oxford, pp. 3-20;

Lewis, D. K. (1969) *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Mass.;

Michael, J. and Pacherie, E. (2015), "On commitments and other uncertainty reduction tools in joint action", *Journal of Social Ontology*, 1, pp. 89-120;

Miller, S. (2010), *The Moral Foundations of Social Institutions*, Cambridge University Press, Cambridge;

Nichols, S. (2004), *Norms with Feelings*, Oxford University Press, Oxford;

Raz, J. (1999), "Explaining normativity: On rationality and the justification of reason", *Ratio*, 12, pp. 354-379;

Searle, J. R. (2005), "What is an institution?", *Journal of Institutional Economics*, 1, pp. 1-22;

----. (2010), *Making the Social World*, Oxford University Press, Oxford;

----. (2015) "On status functions and institutional facts: Reply to Hindriks and Guala", *Journal of Institutional Economics*, 11, pp. 507-514;

Skorupski, J. (2010), *The Domain of Reason*, Oxford University Press, Oxford;

Smit, J. P., Buekens, F. and du Plessis, S. (2014), "Developing the incentivized action view of institutional reality", *Synthese*, 191, pp. 1813-1830;

Sugden, R. (1998), "Normative expectations: The simultaneous evolution of institutions and norms", in A. Ben-Ner and L. Putterman (eds.), *Economics, Values, and Organization*, Cambridge University Press, Cambridge, pp. 73-100;

Tuomela, R. (2002), *The Philosophy of Social Practices*, Cambridge University Press, Cambridge;

----. (2007), *The Philosophy of Sociality*, Oxford University Press, Oxford;

Turner, S. (2010), *Explaining the Normative*, Polity Press, Cambridge.

JOSHUA KETON
The Graduate Center, CUNY
jketon@gradcenter.cuny.edu

THE SOCIAL IMPACT THEORY OF LAW*

abstract

Margaret Gilbert's work on sociality covers a wide range of topics, and as she puts it "addresses matters of great significance to several philosophical specialties – including ethics, epistemology, political philosophy, philosophy of science, and philosophy of law – and outside philosophy as well" (Gilbert 2013, p. 1). Herein I argue that Mark Greenberg's recent call to eliminate the problem of legal normativity is well motivated. Further, I argue that Gilbert's work on joint commitment, and more specifically obligations of joint commitment, allows us to move beyond the problem of legal normativity while cashing out H.L.A. Hart's thesis that moral and legal obligations are distinct.

keywords

legal normativity, Hart-Dworkin debate, legal positivism, interpretivism, joint commitment

Intro   Margaret Gilbert's work on sociality covers a wide range of topics, and as she puts it "addresses matters of great significance to several philosophical specialties – including ethics, epistemology, political philosophy, philosophy of science, and philosophy of law – and outside philosophy as well" (Gilbert 2013, p.1). I want to explore the implications of Gilbert's work for an ongoing debate in philosophy of law, and more specifically a recent call to move beyond the debate by eliminating the problem at its heart.

Herein I argue that Mark Greenberg's recent call to eliminate the problem of legal normativity is well motivated. However, I argue that Greenberg's own solution, the Moral Impact Theory of Law (MIT), is only a likely candidate if one falsely relies on the assumption that only moral facts can provide the type of normativity needed to fix legal content. Instead, I argue that Gilbert's work on joint commitment, and more specifically obligations of joint commitment, allows us to move beyond the problem of legal normativity while cashing out H.L.A. Hart's thesis that moral and legal obligations are distinct.

To be clear, it is not within the bounds of a paper of this scope to offer a robust defense of a Gilbertian-inspired theory of law based on obligations of joint commitment. Nevertheless I want to lay the groundwork for how Gilbert's work on the normativity of joint commitment can help resolve an intractable debate in philosophy of law. My hope is that this will both re-enforce the importance of Gilbert's work and also point to a fruitful new avenue of research in philosophy of law that holds the promise of advancing a new solution to the intractable problem of legal normativity.

I.
The two systems theory: the problem of legal normativity

In this section I give a brief overview of the problem of legal normativity as it arises in the Hart-Dworkin debate. As such, I shall give rough and programmatic accounts of the two positions that form the main participants in the debate: positivists and interpretivists. My goal is to simply motivate the thought that the problem of legal normativity is rather intractable and to provide a general background for understanding the motivation for leaving the debate behind rather than finding a solution.

I.A
Positivism

For centuries the established view of law was natural law theory. Natural law theorists held, in varying degrees of sophistication, that law was derived from morality. On such a theory, the law of a given community is the application of moral principles to particular material and historical circumstances. One can see that the problem of legal normativity need not arise if a debate is occupied entirely by natural lawyers because legal obligations simply are moral

obligations according to natural law theories.

By contrast, positivists argue that law is a special class of norms that is determined entirely by social facts – namely facts about what officials in the legal system have done, thought, said, etc. For example, Hart argued that these social facts could be viewed from two points of view: the internal and the external (Hart 1997, pp. 89-91). From the external point of view, at its extreme, one could accurately describe the law in terms of the behaviors those engaged in the legal system. For instance, the law's claim that subjects ought to phi could be reduced to the claim that officials in the legal system declare guilty and punish those who fail to phi. Thus, the external point of view remains thoroughly descriptive.

Of course, the external point of view fails to capture a basic fact about legal systems. Practitioners view the content of the legal system as giving themselves and others *reasons* for acting in certain ways. However, this positing of a normative, internal point of view raises the question of what the relation is between these reasons and one's moral reasons. Hart famously contended that there is no connection between the two on the grounds of a need for critical distance between the law as it is and as it ought to be (Hart 1958). Natural law, Hart thought, ran the serious danger of conflating the two. Hart, however, never seems to provide a *reason* for officials to adopt the internal point of view. For Hart it is a fact about legal systems that they obtain when a group does take such a stance toward a rule of recognition (1997, esp. pp. 115-117). Thus, the problem of legal normativity is the problem of explaining what is distinctive about the type of normativity that is endemic to legal systems.

Interpretivism as elaborated by Dworkin rests on the idea that law is characterized by fundamental disagreement, and yet is simultaneously used to justify the use of state power. Dworkin's main issue with positivism is his contention that moral facts are necessary in order to determine the content of the law. Thus, he takes a strong stance against the positivist position that social facts alone can determine legal content.

I.B Interpretivism

Dworkin criticized Hart's thesis that the law is totally dependent on social practices of rule following by pointing out that theoretical disagreements in the law abound. Legal practice is rife with examples of what appear to be disagreements over how the law is constituted. For instance, judges disagree about whether or not they are to interpret law based on the plain meaning of the texts, the intentions of the legislators, or the purpose of the law (Dworkin 1986, pp. 31-35). How can it be argued that officials agree on a rule of recognition when they so patently seem to disagree?

Dworkin argues that laws should be interpreted such that they best justify the social practices that surround them – especially the use of coercive force. In order to decide a case, Dworkin argues, judges must find the legal decision that best "fits" and "justifies" the legal practice at issue (1986, p. 52). An interpretation fits when it corresponds to past legal practice. Justification consists in assigning the value that makes the practice the most worth pursuing. Thus, interpretation requires that judges assign each legal practice a purpose, then argue that that purpose both best fits past legal practice and also presents the law in the best possible light. According to Dworkin, law generally has the purpose of justifying the use of state power-coercion. Thus, judges must evaluate past legal practices and justify them through the use of principles of political morality. What makes a practice distinctively legal is its particular purpose and the way it is tied to past practice. While interpretivism argues that we must present law in its *best* light, the possibility of immoral or substandard practices which the law must fit means that it remains possible that the best interpretation of a law is not necessarily the interpretation one would reach when arguing purely through principles of political morality, much less morality proper – thus maintaining a distinction between law and morality, but showing that legal obligations are a species of moral obligation.

I.C
Intractability

Perhaps the most striking feature of the forgoing debate is that neither set of reasons seems to directly confront the other. The positivist claims that the purpose of law is to guide action – and thus that the purpose of law is to authoritatively settle disputes over what subjects are obligated to do or forbear doing. Given widespread moral disagreement the positivist will claim, if law's normativity rests on moral principles it will be unable to serve this function. By contrast, interpretivists' claim that widespread disagreement over the facts that could determine legal content – e.g., the proper interpretive methodology – prevents the law from guiding action in this way. Instead the interpretivist claim is that law is used to justify coercion rather than to guide action – and this requires the invocation of moral principles. The crucial thing to note is that to a degree neither directly confronts the main claims of the other – in the end the question is more a matter of style or intuition. If one sees disagreement as the primary phenomenon of law then one is likely to agree with interpretivists. However, if one sees the law's distinctive feature as its functioning authoritatively then one is likely to agree with the positivists. Nothing prevents one from acknowledging both that there is deep controversy in the law, and that it manages somehow to guide conduct by functioning authoritatively. Consequently, neither side can claim to have solved the problem of legal normativity.

II.
Eliminating the problem

In a recent turn in the literature there is an attempt to move beyond the Hart-Dworkin debate by eliminating, rather than solving, the problem of legal normativity. One can do so, these authors suggest, by rejecting the thought that law creates its own distinctive kind of normativity in addition to moral or prudential concerns. The question of legal normativity is eliminated if the obligations created by legal practices are not of a distinctive kind. In this section I canvass two types of arguments for eliminating the problem of legal normativity – one from positivists and the other from interpretivists.

II.A
Oughts from the legal point of view

Hart has come under serious criticism for his theory of internal and external reasoning. While reference to the internal point of view can explain how it is that an official who internalizes the system's secondary rules takes herself to be obligated by them, it cannot explain how it is that she can view *others* as having those obligations. That is, one can sensibly speak of being obligated to act a certain way based on one's own self-interest or desires. By contrast, it does not seem that one could take one's own self-interest, inclinations, or desires as giving others reasons for acting in the same way. It is plausible that my acceptance of the authority of the legal system as legitimate gives *me* a reason to do as commanded, but I cannot claim that *my acceptance* likewise gives *you* a reason – and not because that would not be fair, but because that is not how reasons *work*. But that is just what one who adopts the internal point of view must do when directing legal judgments toward those who do not adopt the internal point of view. According to Shapiro, moral norms are the only types of norms that one can take to give others reasons (Shapiro 2011, pp. 186-188). Thus, Shapiro argues that those who approach legal rules from the internal point of view must either see legal norms as moral norms, or at least insincerely presume them to be moral norms.

Thus, a certain stripe of positivist claims, a legal norm just is a moral norm from the legal point of view (2011, pp. 184-185). On this picture legal obligations are simply the moral obligations that the law takes those subject to have. Since the nature of law is such that the law takes itself to be authoritative, it is the case that the law takes others to be under obligations to obey it. Whether or not the obligations the law takes one to have are the obligations one actually has is beside the point. How then do we account for all the folk talk of *legal* obligations?

Shapiro argues that when we distinguish *legal* obligations from *moral* obligations we are using the term 'legal' "perspectivally" rather than "adjectivally" (2011, pp. 184-188). Adjectival use,

according to Shapiro, attributes a property to a term, whereas perspectival use attributes the claim that the truth of the statement is relative to a certain perspective or point of view. Thus perspectival claims about what we are *legally* obligated to do need be no more mysterious than claims about what one ought to do from the perspective of Christian morality, utilitarianism, Marxism, etc. We do not assume that each of these perspectives generate new *types* of obligations, only (sometimes conflicting) sets of the same types of obligations – namely, moral obligations.

The point is that in order to be authoritative something must *give others* reasons. Since the law claims authority, one who accepts an authority must thereby take it to be giving others reasons, which one cannot account for solely by reference to the internal point of view. Thus the law comprises, for Shapiro at least, a set of putative moral norms that have been singled out by past social practices as legal norms – i.e., the moral obligations, rights, duties, and powers that the law attributes to its subjects.

Over the course of the last decade or so Mark Greenberg has developed an intriguing and novel account of the law (Greenberg 2014). Greenberg argues that for any given set of legal practices, there will be more than one set of rights, duties, powers, etc. that it can be said to generate. This is partly because, Greenberg argues, social facts cannot determine their own rules of application. For instance, given some legislative practice why do some actions contribute to the law and not others? Presumably committee notes could either be taken or not taken to contribute to the law. Purely social facts about what people have said, thought, and done will not be enough to determine which facts count as contributions to the law's content and which do not because they cannot determine their own significance. To determine which facts contribute to the content of the law and which do not contribute, we need normative facts (Greenberg 2004). For instance facts about democratic legitimacy that rule out the private communications of legislators because of the value of transparency and predictability in democratic lawmaking.

II.B
The Moral Impact Theory of law

If this is so, then positivist accounts lack the resources to choose among the different sets of moral obligations that are compatible with the social facts to settle on one set of obligations that comprises the legal point of view a given legal system. Greenberg however is able to meet this challenge in a straightforward and intuitive way by arguing that it is the legal point of view of any legal system is comprised by the *actual* moral obligations that are imposed by the legal practices of that system. Thus, for Greenberg the content of the law just is the difference (or as he puts it, the *impact*) legal practices make to our moral obligations.

Greenberg call his theory the Moral Impact Theory (MIT). MIT holds that the content of the law is the change in the total profile of moral rights and duties that a legal system makes through the practice of law (2014, p. 1323). MIT is ecumenical with respect to how laws impose these obligations. In the case of simple coordination problems like traffic regulation the law may impose this obligation not through any special moral credentials but rather simply by making a particular solution salient – at least, assuming people generally obey the law and that it is common knowledge among subjects of the legal system that this is so. The moral value of just or democratically legitimate legislation will also serve to contribute to a change in our moral obligations on most understandings and so can explain the importance of legal solutions to problems that do not fit standard coordination dilemmas.

The advantage here is that one is no longer presented with the problem of legal normativity, as traditionally understood. Instead of having to spend energy making sure that we are able to carefully derive legal obligations from social facts plus moral obligations we are able to engage in straightforward moral reasoning from start to finish. The moral upshot of our legal practices will also be the legal content of those practices.

Positivists will argue that the point of law is to guide action and that a legal system according to MIT cannot do that, or at least not well, because of moral disagreement. Greenberg could respond that moral disagreement notwithstanding the law is what it is, and sometimes we will simply disagree about that because some people draw the wrong moral conclusions for whatever reasons they do. Still, there is something in this argument that if laws are meant to guide action, then forcing people to engage in straightforward moral reasoning is likely to lead to a considerable amount of disagreement, and so considerable difficulty. We have then, a reason to look beyond MIT for a different theory.

III.
The Social Impact
Theory of law

What would an alternative to MIT look like? It is worth considering that both eliminativist positions gone over in the last section make essentially the same assumption. Shapiro assumes that there is no alternative to moralized concepts of obligation that will overcome the fact that legal officials have to see the law as giving reasons not just to themselves, but also to those subject to the law. Greenberg likewise argues that normative facts must play a role in fixing legal content since descriptive facts alone cannot do so. He seems convinced that moral facts are the only normative facts available, or at least the only ones fit to do the job of yielding determinate legal content. However, there is an alternative: Margaret Gilbert's account of the obligation of joint commitment, which I propose to call *social* obligations, as opposed to legal or moral obligations.

III.A
Obligation and
joint commitment

Margaret Gilbert has argued for introducing a type of obligation that has many of the features of moral obligations. This account relies on the notion of joint commitment. Gilbert applies this analysis to the notion of political obligation, but I think it usefully characterizes legal obligations (Gilbert 2006, 2013, ch. 17). I give a brief overview of her account before turning to its implications for the law.

A decision gives one a sufficient reason for acting as one decides. Gilbert notes that in the case of personal commitment the reason provided by a decision is amenable to acts of one's own will. That is, in the case of personal commitment it is only an act of one's own will that is needed to do away with the obligation – one need do no more than "change one's mind" in order to divest oneself of the reason for acting that followed from the decision.

This however, does not make sense in the case of a joint commitment. It must not be up to oneself to change the commitment, but rather up to the group – which Gilbert calls a plural subject – to change the commitment, including releasing individuals from the commitment. A particularly salient difference between the decision of one person and the decision of a group seems to be the fact that the decision of a group is not amenable to the will of its individual members – that is, the reason for acting in a certain way provided by a group decision is recalcitrant to removal by individual acts of will. Certainly individuals can decide to act contrary to their obligations in light of the group decision, but it is not within the power of individuals to "change the group's mind" on some issue.

Gilbert suggests that we cash out this notion of the recalcitrance of group decisions to individual acts of will in terms of parties to a joint commitment *owing* one another the object of that commitment (2013, pp. 392-394).This sense of owing captures the idea that one is not at liberty to simply change the joint commitment in the same way that one is with a personal commitment. Thus, according to Gilbert, a joint commitment gives rise to reasons for acting both for oneself and the other parties to the joint commitment. These reasons have the form of directed obligations insofar as one seems to own the object of the commitment to the others and vice-versa. That is each seems committed to all the others to the object of the joint commitment.

It is a short step to the idea that a particular political community might be the object of a joint commitment. Further, it should be noted that the obligations thus generated are not moral

obligations, and so they do not seem to suffer from objections about voluntarism that abound in the literature on political obligations. Instead, the obligations are simply generated by the relation between members toward a particular community or way of life. They exert a rational pressure in the same way that decisions exert rational, rather than moral, pressure.

Good or just decisions might gain some normative force from the moral consequences or justifications for those decisions, but morally neutral or even bad decisions seem quite capable of generating obligations even if they fail to provide all-things-considered reasons because of their moral deficiencies. Similarly, while some joint commitments to just or worthy causes or groups might generate moral obligations, they need not do so. The obligations generated by joint commitments are a distinct type of normativity. They form, as it were, a part of the social-ontological landscape in which parties to joint commitments find themselves, and these circumstances and commitments are open to moral appraisal. We may find, upon reflection, that many of our joint commitments fail to generate all-things-considered reasons for acting in certain ways, but we should not blind ourselves to the social reality of these obligations lest we lose the capability to critically appraise these commitments, which might be the root cause of much that is morally undesirable.

What then are the prospects for the project of eliminating the problem of legal normativity given the above argument for the addition of a class of obligations that are neither moral nor legal, but rather social? On this view then, which I call the Social Impact Theory (SIT), joint commitments to settle disputes focus on the authorizations of institutions committed to the articulation, explication, adjudication and enforcement of our social obligations. The rules that such officials recognize, promulgate, enforce, and practice give rise to legal obligations, which are just various descriptions and entailments of our pre-existing social commitments – legal normativity is just an expression of rational commitment to group decision-making. Thus SIT, like MIT, has no need to posit a unique type of normativity.

With respect to solving Shapiro's problem, the social obligations SIT posits can play the role of giving others reasons, since they constitute directed obligations. Thus, if such social obligations of joint commitment were the obligations that we see as legal obligations, then they could serve to explain the normativity of legal obligations that officials address to subjects of the legal system. While a given subject's all-things-considered reasons might conflict with her social obligations as specified by the legal system, judges nevertheless are authorized to enforce only the social obligations, and thus can avoid charges of irrationality or insincerity of taking other to be bound by reasons they do not recognize. More often perhaps the conflict is not one of the subject morally obliged to act contrary to law, but the subject who chooses to ignore the law out of self-interest or disinterest. Here too SIT fairs better than MIT insofar as SIT builds in the notion that personal commitments (or preferences, desires, decisions, etc.) and joint commitments might contain different, even conflicting, content while remaining valid for the same individual.

What of Greenberg's contention that social facts alone cannot fix the content of the law? It seems to be a brute capacity of humans that we are able to form commitments, and also joint commitments (Gilbert 1992, pp. 392-407). However, these are not moral commitments, they are social commitments. Is it possible that any number of obligations might attend these joint commitments, such that only moral facts have the ability to determine the content of our social obligations?

A full exploration of this issue is beyond the bounds of this paper, but the following thoughts indicate a negative answer. To return to an earlier point, it seems possible that one can form morally repugnant joint commitments. For example, one can form a joint commitment to execute a murder, theft, etc. While one may be morally obligated *not* to commit these acts it

nevertheless seems strange to argue that fellow thieves owe nothing to one another At the least such view is contrary to common sense judgments against turning on compatriots even when one compact involves morally dubious or repugnant goals. We might agree that all-things-considered reasons entail that compatriots to a morally repugnant group should turn on one another, but we are likely to retain certain judgments against the betrayal despite thinking it justified on the balance of reasons.

Second, the fact that there is this disagreement does not seem to hold the same force that it does against positivism. Disagreement might make it difficult to ascertain the content of the obligations but on SIT it is not conforming behaviors that gives the legal content their normative force, but the obligations of joint commitment. Just because one can reasonably posit a number of different obligations for some given joint commitment is no greater an obstacle than that one can posit a different number of moral obligations as arising from some legal practice. To the contrary, MIT requires convergence in moral reasoning to overcome disagreement, whereas SIT only requires ascertaining which social facts obtain among the possibilities. If I were betting, I would bet on the sociological task being completed before the moral one.

REFERENCES

Dworkin, R. (1986), *Law's Empire*, Harvard University Press, Cambridge MA;

Gilbert, M. (1992), *On Social Facts*, Princeton University Press, Princeton NJ;

----. (2006), *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*, Oxford University Press, Oxford;

----. (2013), *Joint Commitment: How We Make the Social World*, Oxford University Press, Oxford;

Greenberg, M. (2004), "How facts make law", *Legal Theory*, 10(03), pp. 157-198;

----. (2014), "The moral impact theory of law", *Yale Law Journal*, 123, pp. 1288-1607;

Hart, H.L.A. (1958), "Positivism and the separation of law and morals", *Harvard Law Review* 71(4), pp. 593-692;

Hart, H.L.A. (1997), *The Concept of Law* (2nd ed.), Oxford University Press, New York;

Shapiro, S. (2011), *Legality*, Harvard University Press, Cambridge MA.

SEUMAS MILLER

Centre for Applied Philosophy and Public Ethics at Charles Sturt University (Canberra) and 3TU Centre for Ethics and Technology at Delft University of Technology (The Hague)
semiller@csu.edu.au

JOINT POLITICAL RIGHTS AND OBLIGATIONS

abstract

In this paper it is argued that: (1) political rights and obligations are a species of institutional (moral) right and obligation (respectively) and are not, therefore, natural rights and obligation; (2) political rights and obligations in a given polity are not simply aggregates of individual rights and obligations rather they are joint political rights and obligations; (3) the exercise of these joint rights, and the concomitant discharging of these joint obligations, is (i) a collective good in itself; (ii) productive of the collective good of legitimate government, and (iii) productive of the collective good of the coordination and regulation of other social institutions (government is a meta-institution), and (4) the procedure of voting in a democratic polity is a joint institutional mechanism – understood as a particular construction out of the notion of a joint action – and a specific expression of the joint right and obligation to engage in political participation.

Political rights, such as the right to vote in elections and the right to stand for political office, are also moral rights. Other more general moral rights, such as the right to freedom of speech and to freedom of association, are also critical in political settings. While all of these rights, and many others, tend to be lumped together under the heading of human rights, there are good reasons to insist on some distinctions and, in particular, the distinction between natural rights and institutional rights.[1] Political rights are, I suggest, a species of institutional right and are not, therefore, natural rights (in my favored sense of this latter term – see section 1). Likewise political obligations are institutional rather than natural obligations.[2] Moreover, as I argue in section 3, political rights and obligations in a given polity are not simply aggregates of individual rights and obligations. Rather they are best understood as *joint* political rights and obligations. As elaborated in section 2, a joint right is an individual right which each possessor has, jointly – and, therefore, interdependently – with others. As such, it is not merely an aggregate of individual rights. Likewise a joint obligation is not simply an aggregate of individual obligations. Moreover, the exercise of these joint rights, and the concomitant discharging of these joint obligations, is a collective good in itself. It also produces the collective good of *legitimate* government and the collective good of the coordination and regulation of other social institutions. In the fourth and final section I provide an analysis of voting in a democratic polity. I do so in terms of my notion of a joint institutional mechanism. I argue that the joint institutional mechanism of voting is a specific expression of the joint right and obligation to engage in political participation.

**1.
Institutional
rights and duties**

Some moral rights, such as the right to life, the right not to be tortured, the right to (clean) air and water and the right to freedom of bodily movement are logically prior to social institutions.[3] These *natural* rights are ones we possess simply by virtue of properties we have qua human beings, such as the capacity to suffer physical pain and to enjoy conscious life. Moreover, some natural rights are needs-based rights. For example, the right to clean air and water is based on biological need. However, many moral right, duties, values, principles and

---

1 There is a vast philosophical literature on human rights, moral rights, and so on. For a useful general discussion see Nickel (2007).
2 There is a vast philosophical literature on political rights, principles and obligations. For one influential discussion see Simmons (1979).
3 See, for example, Griffin (2008). Griffin's distinction between rights per se and practicalities is salient here.

so on are *not* logically prior to social institutions. Consider in this connection the moral right to vote, the moral right to a fair trial, the right to buy and sell land, and the moral right to a paid job. The first of these rights presupposes institutions of government of a certain kind (democratic government), the second, criminal justice institutions of a certain kind (e.g. courts of law that adjudicate alleged crimes), and the third and fourth, economic institutions of a certain kind. Let us refer to such institution-dependent moral rights as "institutional moral rights" (as opposed to natural moral rights) (Miller 2010, ch. 2). Since political rights and, for that matter, political obligations presuppose political institutions, they are a species of institutional right and of institutional obligation, respectively.

Evidently, institutional moral rights depend in part on rights-generating properties possessed by human beings *qua* human beings, but also in part on membership of a community or of a morally legitimate institution, or occupancy of a morally legitimate institutional role.

Such institutional moral rights and duties include ones that are: (a) derived at least in part from collective goods and (b) constitutive of specific institutional roles, e.g. the rights and duties of a fire officer. They also include moral rights and duties that attach to all members of a community because they are dependent on institutions in which all members of the community participate, e.g. the duty to obey the law of the land, the duty to contribute to one's country's national defense in time of war, the right to vote, the right of access to paid employment in a particular economy, the right to own land in a particular territory, the right to freely buy and sell goods in a particular economy.

By the lights of my favored teleological account of social institutions (Miller 2010, ch. 2), these institutional moral rights and duties are institutionally relative in the following sense. Even if they are in part based on an institutionally prior human right (e.g. a basic human need, the natural right to freedom) their precise content, strength, context of application (e.g. jurisdiction) and so on can only be determined by reference to the institutional arrangements in which they exist and, specifically, in the light of their contribution to the collective good(s) provided by those institutional arrangements. So, for example, a property regime, if it is to be morally acceptable, must not only reward the producers of goods (e.g. by protecting the ownership rights of the producers of goods to the goods that they produce) it must also ensure that consumers are benefited and not harmed (e.g. producers are required to meet health and safety standards). More particularly, a property regime, if it is to be morally acceptable, must satisfy the requirements of institutionally prior human rights; specifically, it must ensure that the needs-based rights of consumers are fulfilled (e.g. producers are required to compete under conditions of fair competition, or are otherwise constrained, to ensure that their products are available at prices the needy can afford).

In respect of institutional moral rights and duties we need to distinguish between *general* institutional moral rights and duties, on the one hand, and *special* institutional moral rights and duties on the other (Miller 2010, ch. 2). General institutional moral rights/duties depend in part on properties we possess as human beings and in part on membership of a community with social institutions of the relevant kind. General institutional moral rights/duties attach to all members of a given nation-state because these moral rights/duties are dependent on, and in part constitutive of, institutions in which all members participate. Examples of general institutional moral rights are the right to vote in a given polity and the obligation to obey the law in a given jurisdiction. Some general institutional moral rights transcend the nation-state by virtue of being in part constitutive of trans-national social institutions. Examples of such rights/duties are rights to buy/sell in international financial and other markets.

Special institutional moral rights/duties are in part constitutive of particular institutional roles, e.g. a police officer's right to arrest. As such, they derive in part from the collective good which is the raison d'être for the institution in question. Thus a police officer's right

to arrest derives in part from the institutional purpose served by police organizations of crime reduction; police have a right to arrest offenders as a precursor to their prosecution, punishment and/or deterrence. Moreover, these special moral rights and duties do not necessarily attach to ordinary persons. For example, the moral right to perform open heart surgery attaches to surgeons but not to ordinary citizens or other non-surgeons.

Finally, we need to make a distinction between: (a) institutional moral rights; and (b) institutional rights that are not moral rights. The right to vote and the right to stand for office embody the human right to autonomy in the institutional setting of the state; hence to make a law to exclude certain people from having a vote or standing for office, as happened in apartheid South Africa, is to violate a moral right. But the right to make the next move in a game of chess, or to move a pawn one space forward, but not (say) three spaces sideways, is entirely dependent on the rules of chess; if the rules were different (e.g., each player must make two consecutive moves, pawns can move sideways) then the rights that players have would be entirely different. In other words, these rights that chess players have are *mere* institutional rights; they depend entirely on the rules of the 'institution' of the game of chess. Likewise, (legally enshrined) parking rights, such as reserved spaces and one hour parking spaces in universities are *mere* institutional rights, as opposed to institutional *moral* rights.

2. Joint moral rights and obligations

Thus far we have distinguished between natural and institutional moral rights, and between *general* and *special* institutional moral rights. Political rights and obligations of the relevant kind are *moral* rights and obligations (respectively). Moreover, they are also institutional (moral) rights and obligations since they are logically dependent on the social institution of government. Further, some of these political rights and obligations are general institutional moral rights and some are special institutional rights and obligations. For example, the right to vote is a general institutional moral right. However, the rights and duties that attach to political office holders (e.g. prime ministers) are special institutional moral rights. So far so good. However, there is an important conceptual category of moral rights and obligations that now needs to be introduced, namely, joint moral rights and obligations. It will turn out that political rights and obligations are in large part *joint* (moral and institutional) rights and obligations (Miller 2010, ch. 12).

Joint rights (and obligations) stand in contrast with individual rights (and obligations).[4] An individual right is one possessed wholly in virtue of properties one has as an individual. For instance, the right to life is an individual right since life is something one possesses independently of its possession by others. By contrast, a joint right is a right which each individual has, jointly – and, therefore, interdependently – with its possession by others. For example, if A, B and C author a book together then each has a right to be named as an author jointly with the others (and no-one else has this right).

Consider a couple of types of example from the political sphere: rights of exclusion and rights of secession. Most nation-states hold that exclusion is a legal, indeed, a moral right. Accordingly, exclusion from Australian territory is a joint right of all Australian citizens (but not of non-Australians). Exclusion from Italian territory is a joint right of all Italians (but not of Australians and other non-Italians). Secession, supposing it to be a right of some social groups, is a joint right. Many hold that the Kurds have a right to secede from Turkey. If so, this right is a joint right: a joint right of all Kurds in Turkey (but not of non-Kurds).

Joint rights (and, likewise, joint obligations) need to be distinguished from both *merely* individual rights that attach to individual human persons, as well as from (alleged) rights that

---

4 Miller (2001, ch. 7). For an earlier discussion of joint rights see Miller (1999).

attach to social groups and other collective entities *per se*. For, on the one hand, joint rights are *relational* individual rights as opposed to non-relational individual rights. As we saw above, the right to life is and individual but not a joint right. On the other hand, joint rights are not rights that attach to collective entities such as do the legal rights of a corporation.

Moreover, joint rights are not merely conditional individual rights. A conditional right (in this sense) is one that a person has conditionally on the actions of someone else. For example, A might promise B that A will pay B $X on condition that B wins the race that B is about to compete in. Here B's right to $X is a conditional right: a right to $X conditional on B winning the race. However, it is not a joint right.

Further, in accordance with the above-mentioned teleological account of social institutions, joint rights in my target sense are rights to collective goods (or parts or aspects thereof) (Miller 2010, chs. 2 and 10). For example, each citizen in some well-ordered jurisdiction has a right to security jointly with his or her fellow citizens. Here security is a collective good in the economists' sense; the good is non-rival and non-excludable.[5] Again, members of a community might each have a right, jointly with the others, to graze their cows on a particular parcel of pasture land. Here, the quantum of grass consumed by the cows of each individual member of the community is a part or share of the collective good. However, as tragedy of the commons scenarios illustrate, this good is a rival good.

Let us now consider in more detail this notion of a collective *good*. Here the related notion of a collective *end* is crucial. A collective end is an end that is achieved by two or more agents acting together, i.e. performing a *joint* action. Each of the two or more individual participating agents has the collective end in question, and each intentionally performs a singular action in order to contribute to the realization of this collective end. So a collective end is an individual end that each agent is aiming at. I note that ends are not necessarily intentions, and collective ends, in particular, are not, in general, intended. Importantly, some collective ends are also collective *goods*. Thus, the workers in a farm jointly act to produce and harvest a much-need food crop such as, say, wheat. The wheat crop is a collective good in my sense.

Some joint actions which produce collective goods generate joint rights to the good jointly produced. Here John Locke's idea that mixing one's labor with pre-existing material can generate moral rights is salient. Just as an individual may well have a moral right of possession of a good which that individual produces by his or her own labor, so two or more individuals may well have a joint right of possession to a good which they jointly produce by means of their labor. The individuals in question may have a joint right to use or consume that good or, indeed, to transfer it to a third party in return for a benefit. Consider a team of artisans who make furniture, for example, or a team of scientists who invent a strain of wheat that is resistant to a prevalent disease.

Some joint rights are not, or not entirely, based on joint production of a collective good. Joint moral rights can be based on promises or agreements. For example, the joint ownership rights to a house might be conferred on a newly married couple as a wedding gift by one of their wealthy parents.

Some joint rights are based at least in part on membership of a social group. Consider, for example, the joint right of the citizenry of some polity to establish (e.g. by legislation) a particular system of land rights (e.g. a freehold system or a system of common ownership) within part of the territory of the social group. I refer to such joint rights – joint rights based at least

---

5 The definitions of these terms are problematic. However, roughly speaking, a non-rival good is one such that its enjoyment by one does not reduce the amount available for enjoyment by others, and a non-excludable good is one such if it is available to some then others cannot be effectively excluded from enjoying it.

in part on membership of social group or institution – as *collective rights.* The above-mentioned secession rights of members of the Kurdish social group and exclusion rights of Australian citizens, supposing they exist, are collective rights in this sense. That is, they are a particular species of joint right; they are joint rights based in part on membership of a social group.

I note that some such collective rights are what is often referred to as group rights or minority rights.[6] For example, the fishing, hunting and minority voting rights enjoyed by indigenous people in countries such as Australia and Canada are minority rights. That is, they are joint rights held by members of the indigenous community in question, but not by members of the larger community; moreover, they are joint rights possessed in part by membership of the minority group in question.

Having provided ourselves with serviceable accounts of the notions of institutional moral rights and joint moral rights, let us now turn to a consideration of joint political rights and obligations, bearing in mind my claim that such rights are simultaneously moral, institutional and joint rights.

3.
Joint political rights and obligations

The rights of political participation include the rights to vote, to stand for political office, to engage in public political discussion, to organize a political party, and of political assembly. Concomitant with political rights there are political obligations. These include the obligation to accept the results of legitimate political processes, to obey the lawful directives of duly elected leaders, and to comply with the laws enacted by the duly elected legislature.

My claim, then, is that these political rights and obligations are not simply institutional moral rights and obligations, but also *joint* moral rights and obligations. So in general terms the moral rights and obligations of individual citizens to participate in the political decision-making process in the polity of which they are members by deliberating, voting, standing for office and so on, and by accepting the results of the legitimate political processes in that polity, (e.g. by complying with laws enacted by the duly elected legislature), are joint moral rights and obligations.

If this is correct then the moral legitimacy of the institution of government is not based on the exercise of a set of mere *aggregates* of individual human rights and the concomitant discharging of a set of mere *aggregates* of individual obligations, but rather on the exercise of a set of *joint* moral rights and the concomitant discharging of a set of *joint* moral obligations. Let me explain.

Representative government in liberal democracies is a social institution. By the lights of the teleological theory of social institutions, it therefore has a constitutive collective end or, at least, collective ends, and these collective ends are to secure a number of collective goods. Perhaps the first and most obvious of these is the provision of legitimate leadership, i.e. of an institutionally and morally legitimate government. Naturally, it is a constitutive functional feature of any social institution of government that it provides political leadership, i.e. government. However, a distinctive feature of the institution of representative government, in particular, is that this leadership is morally legitimate.[7]

A further (related) collective good secured (especially) by contemporary institutions of representative government is the coordination and regulation of other social institutions,

---

6 See, for example, Kymlicka (1995).

7 There is a vast philosophical literature on normative theories of political legitimacy. See, for example, Simmons (1979). For a collectivist account that contrasts with Simmons' individualist account and my own relational individualist account see Gilbert (2014, ch. 17). Also, and by contrast with Simmons and myself, Gilbert posits a sui generis notion of commitment which is neither moral nor institutional arising from joint action.

such as the education system, the public health system and the criminal justice system, in a manner that ensures that they realize their (respective) collective ends and, thereby, provide the collective goods that justify their existence. In this respect representative government is a *meta-institution*: an institution concerned with other institutions.[8]

A third collective good provided by representative government is political participation; political participation in the form of the exercise of joint moral rights and the concomitant discharging of joint moral obligations. In this form political participation confers legitimacy on governments. The exercise of the joint right to political participation by the members of the citizenry, (e.g. by voting), is itself a collective end of this institution. As such, it is an end that is realized by the actions of many; not by one person acting alone. Moreover, this collective end is a collective good. For it is a collective end that *consists in* the fulfillment of moral rights – joint moral rights. Accordingly, the exercise of the joint right of political participation is an end in itself; and not merely a means to some further end. That said, it is also the case that the exercise of the joint right of political participation is also a means to other ends, such as the above-mentioned coordination and regulation of other institutions.

This joint moral right to political participation is in turn *in part* based on the prior *natural* individual right of autonomy. So my account of political participation stands within the dominant Enlightenment tradition of John Locke, Immanuel Kant and so on.[9] It is also based *in part* on membership of the relevant political community. So my account – cast, as it is, in terms of *joint* rights – eschews atomistic accounts of political rights and obligations.[10] In short, the joint moral right to political participation is a *collective right* in the sense adumbrated in section 3 above. Finally, as already noted, the joint right to political participation goes hand in glove with the joint obligation to participate, including by accepting the results of legitimate political processes.

Let me now turn to a specific institutional expression of the joint moral right and obligation to engage in political participation, namely, voting in a democracy. Here the notion of a joint institutional mechanism is crucial.

4. Voting and joint institutional mechanisms

Elsewhere I have introduced and analyzed the notion of a joint mechanism.[11] An example of the use of a joint mechanism is two friends tossing a coin to resolve a dispute as a one-off action. Some such mechanisms are institutionalized, e.g. the practice of tossing a coin to decide who is to bat first in an international game of cricket between England and Australia. Let us refer to these as joint *institutional* mechanisms. I now offer an analysis of joint institutional mechanisms of which the institutional practice of voting in a democracy is a very important example.

Joint institutional mechanisms consist of: (a) a complex of differentiated, but interlocking, intentional actions (the input to the mechanism); (b) the result of the performance of those actions (the output of the mechanism), and (c) the mechanism itself. Here the mechanism itself is to be understood as consisting of an operation on the inputs that yields an output; so strictly speaking inputs and outputs are constituents of the mechanism only in the sense of being placeholders or variables. Thus, the notion of a joint mechanism does not collapse into the notion of a joint action.[12] For in the case of a joint action the individual actions are not

---

8 For the argument for this see Miller (2010, ch. 12).
9 See Griffin (2008) for an attempt to reconstruct an account of human rights in the Enlightenment tradition.
10 See Charles Taylor's essay on atomism (essay 7) in his (1985) *Philosophy of the Human Sciences: Philosophical Papers (Volume 2)*, Cambridge University Press, 1985.
11 *Ibid.* Chapter 1. See also Miller (2001, ch. 2).
12 See, for example, Bratman (2014) and Miller (1992).

inputs upon which an operation is performed; rather they are simply actions directed at a collective end and, as such, are constitutive of the joint action. Accordingly, while joint actions are individuated by their constitutive individual actions, joint mechanisms are individuated by their various operations.

Consider the joint institutional mechanism of voting. Citizen A votes in an election and A does so only if others, B, C, D etc. also vote, and only if there is someone to vote for. So in addition to the actions of voting there are the actions of the candidates, X, Y, Z etc., in standing for political office. That they stand as candidates is (in part) constitutive of the input to the voting mechanism; after all, voters vote *for candidates*. So there are interlocking and differentiated actions (the inputs). Further there is some result of the operation of the mechanism: some candidate, say, Smith is voted in by virtue of having secured the most votes (the output). What of the mechanism itself? A key constitutive feature of this voting mechanism is as follows: to receive the most number of votes *is* to win the election.[13] Importantly, that Smith, in particular, is voted in is not something aimed at by all of the participants; specifically, those who voted for Jones were (obviously) not aiming at getting Smith elected!

How does joint action figure in this, given that voters who voted for Jones were not participants in the joint action to vote in Smith? Each voter, of course, performs an intentional individual action of voting and believes others are doing likewise. However, being a species of joint action there must be, on my analysis, a collective end which *all* the voters have. Here we need to be careful. Naturally, it is not an end of *all* the voters (and, therefore, not a collective end of all the voters) that Smith is voted in; for a number of voters voted for other candidates, such as Jones. Rather it is only a collective end of those who vote for Smith that he be voted in; each member of this sub-group of voters votes for Smith in the belief (or, at least, hope) that others will also vote for Smith. Since we are assuming Smith did in fact receive the most votes it follows that those who voted for him have realized the collective end of their joint action. Likewise it is a collective end of those who voted for Jones that she be voted in. However, since Jones did not receive sufficient votes to win the election theirs is an unsuccessful joint action. So at the level of sub-groups of voters there may be multiple joint actions, only one of which is successful.

Importantly, there is also a collective end of *all* the voters and *all* the candidates (or at least all those voting and standing for election in good faith). This is the collective end that the one who gets the most votes – whoever that happens to be – is the winner. This is a collective end of all *bona fide* participants in the joint institutional mechanism and reflects the commitment of the participants to the above-mentioned key constitutive feature of the mechanism, i.e. that the candidate with the most votes wins the election. Accordingly, participants in this joint institutional mechanism perform the individual actions of casting a vote and/or standing as a candidate having as a collective end that the one who gets the most votes – whoever that is – wins the election. So voting is a species of joint action and, more specifically, a joint institutional mechanism.

There are, however, two important salient normative features of the joint institutional mechanism of voting in democracies that need to be rendered explicit at this point. These features go hand in glove. Firstly, each has a *moral right* to vote and, in the light of our analysis of the voting mechanism as a *joint* mechanism, this is a *joint institutional and moral right*. As such, it is a right possessed by each interdependently with the others in the service of a collective end which is also a collective good or, at least, a telescoped set of collective goods, namely,

---

13 There are, of course, any number of alternative voting systems in democracies. However, this does not materially affect the analysis on offer here.

political participation (first collective good) in the provision of a legitimate government (second collective good). Secondly, the one who gets the most votes is not simply the winner of the election. For that person is also, by virtue of winning the election, the morally and institutionally legitimate occupant of the political office in question. Indeed, there is a *joint institutional and moral obligation* on the part of all voters – irrespective of which candidate each might have voted for – to accept the election result and, thereby, confirm this legitimacy.[14] So the obligation to accept the result is a moral obligation possessed by each interdependently with the others in the service of a collective end which is also a collective good, or, at least, telescoped set of collective goods, namely (and as is the case with the closely related joint rights), political participation in the provision of a legitimate government.

5. Conclusion

In this paper I have argued that: (1) political rights and obligations are a species of *institutional* (moral) right and obligation (respectively) and are not, therefore, natural rights and obligation; (2) political rights and obligations in a given polity are not simply aggregates of individual rights and obligations rather they are *joint* political rights and obligations; (3) the exercise of these joint rights, and the concomitant discharging of these joint obligations, is (i) a collective good in itself; (ii) productive of the collective good of *legitimate* government, and (iii) productive of the collective good of the coordination and regulation of other social institutions (government is a *meta-institution*), and (4) the procedure of voting in a democratic polity is a *joint institutional mechanism* and a specific expression of the joint right and obligation to engage in political participation.

REFERENCES

Bratman, M. (2014), *Shared Agency: A Planning Theory of Acting Together*, Oxford University Press, New York;

Gilbert, M. (2014), *Joint Commitment: How We Make the Social World*, Oxford University Press, New York;

Griffin, J. (2008), *On Human Rights*, Oxford University Press, New York;

Kymlicka, W. (1995), *Multicultural Citizenship: A Liberal Theory of Minority Rights*, Oxford University Press, New York;

Miller, S. (1992), "Joint action", *Philosophical Papers*, 21(3), pp. 275-299;

----. (1999), "Collective rights", *Public Affairs Quarterly*, 1(4), pp. 331-346;

----. (2001), *Social Action: A Teleological Account*, Cambridge University Press, Cambridge;

----. (2010), *The Moral Foundations of Social Institutions: A Philosophical Study*, Cambridge University Press, Cambridge;

Nickel, J. (2007), *Making Sense of Human Rights* (2nd ed.), Blackwell, Hoboken;

Simmons, J. (1979), *Moral Principles and Political Obligations*, Princeton University Press, Princeton;

Taylor, Ch. (1985), *Philosophy of the Human Sciences: Philosophical Papers* (Vol. 2), Cambridge University Press, Cambridge.

---

14 This is, of course, subject to various other conditions being met, e.g. that the person is eligible to hold the office in question, the electoral and subsequent processes were not flawed in their operation, and so on.

LORENZO PASSERINI GLAZEL
Università di Milano - Bicocca
lorenzo.passerini@unimib.it

SHARED NORMS AND NOMOTROPHIC BEHAVIOUR

abstract

The paper proposes to construe the "interventions in the lives of others when one thinks that the others' behaviour is wrong" (Devlin and Gilbert), and the "reactions to disappointment of normative expectations" (Niklas Luhmann) under the notion of "nomotrophic behaviour", i.e. a behaviour that aims at the maintenance of a norm in the event of its infringement. Subsequently, some of the implications of nomotrophic behaviour at an epistemological level (as a clue for inferring norms from action), and at an ontological level (with respect to the existence of norms) are investigated. The paper then examines a possible norm-generating (nomogenic) effect of "nomotrophic behaviour am Phantasma", and eventually inscribes the different forms of nomotrophic behaviour in the superordinate category of "meta-normative behaviour".

0.
Introductory remarks

In *Shared Values, Social Unity, and Liberty* (2005, ²2014),[1] Margaret Gilbert deals with the question of social unity, that is to say with the question of when a plurality of persons constitutes a social group.

0.1.

Gilbert moves from Patrick Devlin's thesis that "society is not something that is kept together physically; it is held by the invisible bonds of common thought. If the bonds were too far relaxed the members would drift apart. *A common morality is part of the bondage*" (Devlin 1959, ²1965, p. 10; italics mine).[2]

Gilbert then reformulates the question of social unity in the following terms: "Is there a way of sharing values such that such sharing is sufficient for social unity?" (Gilbert 2014, p. 182).

Following Devlin, Gilbert focuses on a peculiar phenomenon which she considers revealing for the question of social unity, a phenomenon connected with her salient notion of "joint commitment". This phenomenon is the "intervention" in the lives of others when one thinks that the others' behaviour is wrong with respect to a presupposed value or commitment (or, I would say, with respect to a presupposed norm).

In the present paper I will focus on this phenomenon.

0.2.

I suggest to call "nomotrophic behaviour" the intervention in the lives of others when one thinks that the others' behaviour is wrong with respect to a presupposed value, commitment, or norm.

By "nomotrophic behaviour" I mean a behaviour which aims at the maintenance of a norm in the event of its infringement: it typically consists in a reaction to the (actual or possible) infringement of that norm.[3]

The idea underlying the concept of nomotrophic behaviour is that a social norm that is repeatedly infringed with no reaction may slowly "atrophy" and vanish (by "desuetude"); and that its atrophy may be countered through different forms of *nomotrophic behaviour*.

---

1 The chapter "Shared values, social unity, and liberty" in the book *Joint Commitment* (2014) by Margaret Gilbert was first published in *Public Affairs Quarterly*, 19 (2005), pp. 25-49.

2 Quoted by Gilbert (2014, p. 205).

3 See Passerini Glazel (2013a, 2013b and 2015). The adjective '*nomotrophic*' (with 'ph') is derived from Greek 'νόμος' '*nómos*' ("norm") and 'τρέφω' '*tréphō*' ("to nourish").

In principle, any norm can be the object of a nomotrophic behaviour, be it a norm deriving from a shared value or joint commitment, a moral norm, a legal norm, a conventional norm, a rule of a game, etc.

Nomotrophic behaviour (nomotrophic with 'ph') is a peculiar form of what Amedeo Giovanni Conte proposed to call "*nomotropic* behaviour" (nomotropic with 'p'): *nomotrophic* behaviour is indeed a *species* of the *genus* of *nomotropic* behaviour.

By "nomotropic behaviour" Conte means "acting with-reference-to a norm", which does not imply complying with that norm.[4] In other words, nomotropic behaviour is a behaviour that is oriented to a norm, without necessarily being in conformity with the norm itself.[5]

Here are three examples of nomotropic behaviour.

0.3.

(i)   The stealthily behaviour of a thief. As Max Weber (1922) remarked, a thief, in concealing his action, acts with-reference-to the norms of the criminal code that punish theft (even though he does not comply with those norms).

(ii) The behaviour of a cheater illegally extracting an ace from his sleeve in the card game of poker. The cheater infringes the rules of poker, but he still acts with-reference-to those rules, given that it is only in virtue of those rules that an ace is an ace (and has its specific ludic value) in poker.

(iii) Tax avoidance. Tax avoidance is a behaviour oriented to tax legislation, even though it is intended to avoid its application.

In the present paper I suggest to construe Devlin's and Gilbert's "interventions in the lives of others", as well as Niklas Luhmann's "reactions to disappointment of normative expectations", as forms of (that peculiar kind of nomotropic behaviour that is) *nomotrophic behaviour* (§ 1.). Subsequently, I will show some of the implications of nomotrophic behaviour at an *epistemological level*, as a clue for inferring norms from action (§2.), and at an *ontological level*, with respect to the existence of norms (§ 3.). I will then examine a possible norm-generating (nomogenic) effect of what I suggest to call "nomotrophic behaviour *am Phantasma*" (§ 4.), and finally I will propose to inscribe these different forms of nomotrophic behaviour in the superordinate category of "meta-normative behaviour" (§ 5.).

0.4.

Devlin's and Gilbert's main concern in their analysis of the "intervention" in the lives of others (when one thinks that the others' behaviour is wrong with respect to a presupposed value, commitment, or norm) is the question of the "standing to intervene", that is the question of the *legitimacy* of such an intervention.

On my part, I will leave aside the question of the *standing* to intervene, as well as the conditions of the *legitimacy* of nomotrophic behaviour, and I will focus on the phenomenon itself, and on its relation to norms.[6]

**1.
Intervention in the lives of others as a form of "nomotrophic behaviour"**

---

4 Complying with a norm, and acting in conformity with a norm (as well as nomotrophic behaviour), are only peculiar cases of nomotropic behaviour. See Conte (2011).

5 The adjective '*nomotropic*' (with 'p') is derived (on the model of 'heliotropic') from Greek 'νόμος' '*nómos*' ("norm") and 'τρέπω' '*trépō*' ("to turn to"). On the concept of "nomotropic behaviour", see Conte (1975, 2000a, 2002, 2003, 2004, 2011, 2012a). Conte's investigations on nomotropism are closely intertwined with Di Lucia's investigations on "acting with-reference-to a norm" and on "effectiveness without fulfilment": see Di Lucia (1996, 2002, 2003 and 2007). See also Passerini Glazel (2012b). Two very fruitful implementations of the concept of "nomotropism" can be found in Chiodelli & Moroni (2014) and in Lauer (2015).

In another context, Eric L. Santner employed the term 'nomotropism' in psychoanalysis: by nomotropism he means "the obsessive-compulsive preoccupation with *nomos*, with matters of law, justice, and ethics" (which for Freud "also comprised the compulsive dimension of the search for scientific truth, the *Zwang* internal to *Wissenschaft*"): see Santner (1999, 22000, p. 68).

The term 'nomo-tropism' is employed in yet another context by the cultural anthropologist Allen Feldman in correlation with the term 'trauma-tropism': see Feldman (2002).

6 I equally suspend any *value judgment* on nomotrophic behaviour itself.

What do Devlin and Gilbert mean by "intervening in the lives of others" though?
According to Gilbert, "Devlin would surely include as interventions *rebuking* others for certain actions and, in advance, *demanding* that they act in a particular way" (Gilbert 2014, p. 185). Gilbert considers rebukes and demands as paradigmatic interventions.

I propose to construe *rebukes* as a form of backward-looking reaction to an actual (or supposed) infringement of a norm, and *demands* (in Gilbert's sense) as a form of forward-looking reaction to a possible infringement of a norm.[7]

Both rebukes and demands, in Gilbert's sense, are forms of what I propose to call "nomotrophic behaviour".

By "nomotrophic behaviour" I mean a behaviour aimed at the maintenance of a norm in the event of its infringement, that is to say a behaviour by which one reacts to the (actual or possible) infringement of a norm in order to prevent the relevant norm from being neglected, forgotten, abandoned, and eventually atrophying and vanishing.

Besides rebukes and demands, there are many other possible forms of nomotrophic behaviour.[8]

The imposition of a sanction is, of course, one of the possible forms of nomotrophic behaviour, and a prominent one. However, we are not always in the position of imposing a sanction to someone (in Devlin and Gilbert terms, we do not always have the "standing" to impose a sanction to someone);or we may consider that a sanction would not be appropriate to the situation.

This point has been emphasized by the German sociologist Niklas Luhmann in *A Sociological Theory of Law* (1972). Luhmann criticizes those who define the concept of "norm" uniquely through the inclination to impose sanctions "in the event of disappointment". According to Luhmann, "the repertory of possibilities is thus too strictly limited and often it is misunderstood that the retention of expectation is more important than being able to impose it" (Luhmann 1972, trans. 1985, p. 47).[9]

Besides the imposition of sanctions, there are many other forms of reaction to what Luhmann calls the "disappointment of normative expectations".[10]

With regards to the question of the *legitimacy* of nomotrophic behaviour it may be noted that, in some cases, nomotrophic behaviour may bring about an *a posteriori* auto-legitimation (this can be the case, for instance, in what in § 3. I propose to call "nomotrophic behaviour *am Phantasma*").

With regards to the possible *value judgments* on nomotrophic behaviour, I assume a scientific and neutral (anaxiological) point of view. It is worth noting, by the way, that any value judgment on a particular nomotrophic behaviour (which appears mainly as a contingent axiological question) seems at least in part related to a preliminary value judgment on the norm it is intended to maintain, or on its practical consequences. I am grateful to Helen Lauer for having drawn my attention on this point.

7 On the one side, a demand can be construed as a before-the-fact reaction to a possible infringement of a norm; on the other side, as Gilbert suggests, a rebuke can be construed "as an after-the-fact demand" (Gilbert 2014, p. 397) (or at least, I would add, as a demand not to reiterate the infringement of a particular norm in the future).

An icastic example of nomotrophic demand is the famous passage from Saint Catherine of Siena's Letter 8 to Pope Gregory XI: "*Voi dovete venire: venite dunque* [You ought to come: come, then!]". I am grateful to Amedeo Giovanni Conte for this example.

I assume that both rebukes and demands presuppose a norm, with reference to which the others' behaviour is qualified as wrong.

8 Gilbert remarks this point, and mentions as an example commending for conformity to a commitment in difficult circumstances: see Gilbert (2014, p. 199). Gilbert responds here to a comment by Jennifer Nadelsky.

9 See also Luhmann (1969).

10 For his sociological determination of the concept of "norm" Luhmann refers to Johan Galtung's paradigm: "*cognitive* expectations vs. *normative* expectations" (Galtung 1959). A *cognitive* expectation is an expectation one is disposed to change or redefine in case of dissonance with respect to reality (following G.E.M. Anscombe (1957) and John R. Searle (1975, 2010) I would speak of an expectation-to-world direction of fit); on the contrary, a *normative*

The variety of forms of reaction that are alternative to sanctions is illustrated by Luhmann through an example:

> If I arrange to meet a friend in a café and do not meet him there, I do not only feel hurt in my cognitive, but also in my normative, expectations. He should be there! Some kind of 'treatment' of disappointment and expectation is now required, but there are various possibilities at my disposal which do not all have the character of a sanction. For example, I can ask the waiter about the friend and express my norm of expectation by undertones of disappointment, annoyance and worry [...]. However, I can also turn to him personally by telephoning him or reproaching him during a later meeting. As a consequence [an] apology may be forthcoming: I can accept an apology from my friend without imposing any type of sanction, which presumes that my expectation was justified in principle. [...]
> A different type of strategy operates with the non-verbal characteristics of the given situation. I may leave the café immediately and expose the late-comer to his own injury. [...] On the other hand, I can remain sitting in the café to prove the meaning of the norm by the extent of my sacrifice. I can let it turn into scandal in order to enjoy to the full the social resonances of the scandal, if not the norm.
> Techniques of making known and spreading about the case of disappointment, the escalation into scandal and the enjoyment of one's own set-backs, the techniques that enjoin fulfilment of norms, of hurt or the tactful acceptance of excuses, techniques of self-effacement and enduring pain or techniques of innocent enlargement of injury and justifiable pleasure at the other's injury [*Schadenfreude*] – there is a series of possibilities to give the old norm the expression which is adapted to a new situation (Luhmann 1972, trans. 1985, pp. 46-47).[11]

All the techniques mentioned by Luhmann "give the old norm the expression which is adapted to a new situation", "so that even the less robust natures are capable of carrying on life with their norms, even if they are not capable of imposing sanctions" (Luhmann 1972, trans. 1985, p. 47).
How do these techniques give "expression" to the infringed norm? It is important to remark that they do not do it necessarily in an *explicit*, or in a *linguistic* form.[12] My interpretation is that these techniques "give expression" to the infringed norm *modo obliquo*, in virtue of the fact that the reaction *presupposes* the existence of the infringed norm. The infringed norm is implied in the pragmatical presuppositions of the reaction to the infringement of a norm.
A brief analysis of the respective presuppositions of *conviction* (verdict of guilty) and *forgiveness* may contribute to clarify this point.[13]
Despite their opposite effects, conviction and forgiveness share three identical presuppositions:[14]

---

expectation is an expectation one is *not* disposed to change or redefine, so that, in case of dissonance, it is reality that has to be changed or redefined in order to match the expectation (world-to-expectation direction of fit).
11 See also Luhmann (1969,p. 39).
12 Gilbert, too, speaks of a "proposal and acceptance of a joint commitment", which "need not be verbal" (see Gilbert 2014, p. 196).
13 My analysis follows Maria-Elisabeth Conte's analysis of the presuppositions of forgiveness. See Maria-Elisabeth Conte (1992). See also Conte (1977), Fillmore (1971), Silvi (2004) and Passerini Glazel (2015).
14 "Great and honest kind of revenge is forgiveness". Fra' Bartolomeo da San Concordio, in *AmmaestramentidegliAntichi*, a work written presumably at the beginning of XIV century, attributes this maxim to Seneca, and quotes Seneca's treaty *De quatuor virtutibus cardinalibus*; this treaty (probably modelled on a lost work

(i)   *factive* presupposition: the existence of the fact being sanctioned or forgiven respectively;

(ii)   *axiological* presupposition: the fact has a negative value with reference to a norm, whose existence is presupposed;

(iii) presupposition of *responsibility*: the person respectively convicted or forgiven is responsible for the fact.[15]

The *axiological* presupposition is the essential presupposition for nomotrophic behaviour: any behaviour with this kind of presupposition implies a reference to a norm (or a value), and so does any nomotrophic behaviour.[16]

It is in virtue of this axiological presupposition that nomotrophic behaviour "gives expression" to the (actually or possibly) infringed norm.

2.
Nomotrophic behaviour as a clue for inferring norms

The fact that nomotrophic behaviour (explicitly or implicitly) gives expression to the infringed norm has a particular relevance for the *epistemological* question concerning the conditions of possibility of the inference of norms from action.[17]

The inference of norms from action is made clearly easier when the relevant norms are *explicitly* and *linguistically* formulated. However, this is frequently not the case, not even within the law: the norms that operate within a given social group are not always *verbal* norms, nor are they always *verbalized* norms.[18]

What means can be used to infer *non-verbal* and *non-verbalized* norms from the behaviour of people belonging to a social group?

This question acquires even greater importance if we take into account the two following remarks.

First remark: An empirically observed *regularity* of behaviour (a regular pattern of behaviour) is not necessarily a *normative* regularity: it is not necessarily a regularity determined by a norm (or by a rule).[19]

The fact, for instance, that a family regularly eat fish on Fridays may well depend on a norm of the Catholic religion, but it may also be a mere (non-normative) habit (possibly derived from that particular religious norm being followed by past generations, although it was subsequently abandoned as a norm).

Second remark: Even when a behaviour is determined by a norm, it may be "semiotically mute" about the norm: it does not necessarily tell anything about the norm (or norms) which determined that particular behaviour: acting in compliance with a norm does not necessarily imply the (explicit or implicit) expression of that norm, nor does it imply the ability to express that norm in linguistic form.

---

by Seneca) is now attributed to Saint Martin of Braga (c. 520-580 A.D.), under the title *Formula vitae honestae, or De differentiis quatuor virtutum vitae honestae.*

15 It may be helpful to recall that in some cultures, and in some legal systems, one can be responsible for an action performed by another subject.

16 As I remarked in note 4, any norm, in principle, can be the object of a nomotrophic behaviour, be it a norm deriving from a shared value or joint commitment, a moral norm, a legal norm, a conventional norm, a rule of a game, etc. The axiological aspect of this presupposition lies in the fact that it refers to an *infringement* of a norm.

17 This question is strictly related to Ludwig Wittgenstein's considerations on rule-following. Inspired by Wittgenstein, but partly in contrast with Wittgenstein, Amedeo Giovanni Conte investigated the presuppositions of the inference of a norm from action in many of his works: see, for instance, Conte (1990), Conte (2000b) and Conte (2002); see also Carcaterra (2002).

18 On the distinction between *non-verbal* norms and *non-verbalized* norms, see Sacco (2015), Caterina (2009), Passerini Glazel (2009).

19 The German sociologist Theodor Geiger calls *Regelhaftigkeit* a *non-normative* regularity, and *Regelmäßigkeit* a *normative* regularity (see Geiger 1947, trans. 1969, p. 44). See also Conte (1990), Conte (2000b) and Conte (2002).

Eating fish regularly on Fridays, for instance, is in itself semiotically mute about its possible determination by a norm.[20]

On the contrary, a person who reacts to the infringement of a norm uses his nomotrophic behaviour as a manifestation of his thought.

On the one side, indeed, nomotrophic behaviour explicitly or implicitly qualifies the broken regularity as a *normative* regularity: it is a *reaction* to the *breaking* of that regularity.

On the other side, nomotrophic behaviour is semiotically pregnant, since (as I have shown in § 1.) it gives (explicit or implicit) expression to the norm implied in its presuppositions.

Thus, if a person in a family, for instance, rebukes another member of the family who has prepared meat for lunch on Friday, this is quite a relevant clue of the existence, in that family, of a norm on eating fish on Friday (and consequently of the *normative* character of the relative regularity).

Nomotrophic behaviour, in comparison to behaviour in conformity with a norm, is thus a more salient clue to infer the existence (at least from the point of view of the person who acts nomotrophically) of a norm, especially in informal normative systems, where norms are generally not explicitly formulated.[21]

Nomotrophic behaviour, thus, has a particular relevance at an *epistemological* level; but it also has implications at an *ontological* level, in particular with reference to the issue of the *existence of norms.*

I said that "nomotrophic behaviour" is a behaviour by which one reacts to the (actual or possible) infringement of a norm in order to prevent that norm from being neglected, forgotten, or abandoned, and eventually atrophying or vanishing: it aims at the *maintenance* of a norm (or a value, or a normative expectation), by contrasting its possible atrophy.

Let me try to clarify this point.

According to Luhmann, a normative expectation "that is continuously disappointed and is without expression fades away. It is inadvertently forgotten, and it is not believed any more" (Luhmann 1972, trans. 1985, p. 46). The risk, thus, is that a continuously infringed norm loses its vitality, becomes inoperant, atrophies, and fades away by desuetude.

The aim of nomotrophic behaviour is to counter the possible atrophy and desuetude of the norm.[22]

The phenomenon of nomotrophic behaviour emphasizes then, *ex negativo*, the correlative phenomenon of atrophy and desuetude of norms.

Desuetude (*desuetudo*)[23] is a particular way (alternative to explicit derogation) in which norms

**3.
The incidence of nomotrophic behaviour at the level of the existence of shared norms**

---

20 This remark has been made, for instance, by the Italian legal anthropologist Rodolfo Sacco with reference to customary norms: "The quality of the psychological act preceding or concurring with the applying of a customary norm is the same in modern man (who often cannot verbalize the norm), in man who had no articulated language, as well as in animals with a developed social organization [...]. A man who is not able to verbalize his own norm, our ancestor lacking speech, an animal studied by an ethologist, all of them do not use their action as a manifestation of their own thought" (Sacco 2000, p. 121).

In contrast to Sacco's claim is the claim by Emilio Betti according to which "any form of practical activity implies an implicit representative value" (Betti 1971, p. 7).

21 Gilbert seems to be aware of this epistemological salience when she focuses on the "standing to intervene" in her account of shared values based on the notion of joint commitment.

22 In many legal systems there is a phenomenon that closely resembles nomotrophic behaviour. If a right of a certain kind (for example: a right to the compensation of damage deriving by a tort, in Italian law) is not exercised or claimed in a given (generally statutorily determined) lapse of time, that right may decay, may prescribe, and vanish. The holder of the right has to exercise or claim that right before the given period of time is elapsed to keep that right alive and to prevent that right from prescribing and coming to non-existence.

23 On desuetude, and on the possible relations between *desuetudo* and *consuetudo*, I am working in a book to come (*Il farsi e il disfarsi delle norme*).

pass from *existence* to *non-existence*. Just like not every norm comes to *existence* in virtue of an explicit speech (verbal) act of enactment, not every norm comes to *non-existence* in virtue of an explicit speech (verbal) act of derogation.

In philosophy of law the existence of a norm is sometimes construed as its *validity* (by most legal positivists and normativists, for instance), sometimes as its *effectiveness* (by most legal realists, for instance); but, in principle, *validity* and *effectiveness* are two distinct phenomena.[24] Desuetude is a tricky phenomenon for this distinction: the *validity* of a norm ceases in virtue of the lack of *effectiveness* of that norm.[25]

An interesting account of desuetude is suggested by Hans Kelsen. In the second edition of the *Pure Theory of Law* (1960) Kelsen writes:

> A legal norm may lose its validity by never being applied or obeyed – by so-called desuetude (Kelsen, 1960, trans. 1967, p. 213).

Recalling his theory of the "basic norm" as the origin of the validity of every norm within a legal system, Kelsen writes:

> In the basic norm the fact of creation and the effectiveness are made the conditions of the validity – "effectiveness" in the sense that it has to be added to the fact of creation, so that neither the legal order as a whole nor the individual legal norm shall lose their validity (Kelsen, 1960, trans. 1967, p. 212).

Kelsen clarifies that "effectiveness is a *condition* for the validity – but it's not validity":

> A condition cannot be identical with that which it conditions. Thus, a man, in order to live, must have been born; but in order to remain alive other conditions must be fulfilled, for example, he must receive nutrition. If this condition is not fulfilled, he will lose his life. But life is neither identical to birth, nor with being nourished (Kelsen, 1960, trans. 1967, p. 212).

Also the italian legal philosopher Norberto Bobbio makes use of the metaphor of "nourishment" with reference to norms. In *Consuetudine e fatto normativo* (1994), Bobbio writes:

> In the long run, a normative system can survive only if the majority of its norms gives rise to corresponding customs. Custom is not only *optima legum interpres*, but also the nourisher and feeder of the statute law: it maintains statute law alive (Bobbio 1994, p. 45).

My notion of nomotrophic behaviour is precisely the notion of a behaviour which aims at nourishing norms against their possible atrophy; nomotrophic behaviour seems thus to have an incidence at the ontological level of the *existence* of norms, at least in so far as it may prevent a norm from coming to *non-existence*: it may play an important role in maintaining norms and normative systems (especially informal ones) alive.[26]

---

24 See, for instance, Bobbio (1993).
25 Desuetude is a puzzling phenomenon for the philosophy of law, which has not been thoroughly studied yet. Incidentally, many legal systems tend to exclude the possibility of desuetude, at least with reference to statute laws.
26 A specific phenomenon that could be fruitfully investigated under the concept of nomotrophic behaviour is, in anthropology of law, the phenomenon of self-protection (e.g. vendetta) in primitive law.

To norms may fit the following words from Johann Wolfgang Goethe's play *Torquato Tasso*: "*Was gelten soll, muß wirken*".[27]

However, nomotrophic behaviour may not only contribute to the maintenance of the existing norms of a normative system: it may also give rise to new norms, it may have a norm-generating (nomogenic) effect.
This may be the case with what I propose to call "nomotrophism *am Phantasma*".[28]
Nomotrophism *am Phantasma* occurs when one acts nomotrophically with reference to a non-existing norm, i.e. with reference to a mere representation of a norm, to a deontic noema,[29] which is not (yet) a shared or valid norm.
In some cases, one may act nomotrophically with reference to a non-existing norm because he erroneously thinks that that norm exists.
In certain situations, though, one may intentionally act nomotrophically with reference to a non-existing norm because he thinks that that norm would be appropriate for the situation. Human rights movements may be an example: by reacting to the violation of what they believe to be human rights (in accordance with their normative representations, with their deontic noemata), human rights defenders act nomotrophically with reference to norms that are not (yet) valid in some legal systems, but which they promote, and resolutely propose to be recognized, shared and enacted within all legal systems.
This last case of nomotrophic behaviour *am Phantasma* implies a peculiar "normative will", such as an *animus inducendi consuetudinem* (a will to arouse a customary norm).
However, this normative will is quite different from the normative will implied in the formal enactment of a norm: it is not an immediately effective thetic (performative) will, which immediately creates a new norm. The normative will implied in this kind of nomotrophism *am Phantasma* is an indirect will: it consists in a *proposal* of a norm, appealing for a shared *a posteriori* recognition and validation of that norm.[30]

**4.
The norm-generating effect of nomotrophism *am Phantasma***

In § 4. I stated that nomotrophic behaviour *am Phantasma* implies a peculiar indirect normative will: the will that a non-existing norm be recognized and shared, or enacted.
Even ordinary nomotrophic behaviour implies a kind of indirect normative will, though: the "normative will" that an existing norm remain operant and valid.
Thus, following a suggestion by Margaret Gilbert and Seamus Miller, nomotrophic behaviour, as well as nomotrophic behaviour *am Phantasma*, may be called "*meta-normative* behaviours".[31]

**5.
Meta-normative behaviour**

---

27 Goethe (1790, act I, scene IV). I suggest two possible translations: "What ought to be valid, must be effective", and "What is to be valid, has to be operant". I interpret Goethe's passage taken out of context and of co-text.
28 My idea of "nomotrophism *am Phantasma*" is inspired by the concept of "deixis *am Phantasma*", introduced in linguistics by Karl Bühler (1933). An example of deixis *am Phantasma* is when a speaker, while uttering a sentence, points at something that is not there, behaving as if it was there, by hands and gestures, for instance. See also the concept of "*praxis am Phantasma*" in Conte (2003).
29 "Deontic noema" is a term proposed by Amedeo Giovanni Conte for a mental normative representation (in contrast with an actual normative state-of-affairs, a "deontic status"). See Conte (2012b).
30 With reference to John Langshaw Austin's (1962) triadic paradigm "locutionary vs. illocutionary vs. perlocutionary", the immediate "nomothetic" effect of an act of enactment is evidently an *illocutionary* effect; the mediated nomogenic effect of nomotrophism *am Phantasma* may be considered a *perlocutionary* effect.
The normative will implied by nomotrophic behaviour *am Phantasma* may be compared to the will implied in Kant's categorical imperative: "Act only according to that maxim by which you can at the same time will that it should come a universal law" (Kant 1785, tr.1993, p. 30).
31 I am grateful to Margaret Gilbert and to Seamus Miller for this suggestion. Another possibility is to speak of "para-normative behaviour".
A different kind of meta-normative (or para-normative) behaviour is the phenomenon known in some places as

REFERENCES

Anscombe, G.E.M. (1957), *Intention*, Basil Blackwell, Oxford;

Austin, J.L. (1962), *How To Do Things with Words*, Clarendon Press, Oxford;

Bartolomeo da San Concordio (1963), *Ammaestramenti degli antichi*, P.G. Colombi (ed.), Cantagalli, Siena;

Betti, E. (1971), *Interpretazione della legge e degli atti giuridici*, second revised edition, G. Crifò (ed.), Giuffrè, Milano;

Bobbio, N. (1993), *Teoria generale del diritto*, Giappichelli, Torino;

----. (1994), *Consuetudine e fatto normativo*, in N. Bobbio, *Contributi ad un dizionario giuridico*, Giappichelli, Torino, 1994;

Bühler, K. (1933), *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Fischer, Jena;

Carcaterra, G. (2002), "Indizi di norme", *Sociologia del diritto*, 29, 3, pp. 123-139;

Caterina, R.(ed.) (2009), *La dimensione tacita del diritto*, Edizioni Scientifiche Italiane, Napoli;

Chiodelli, F. & Moroni, S. (2014), "The complex nexus between informality and the law: Reconsidering unauthorised settlements in light of the concept of nomotropism", *Geoforum*, 51, pp. 161-168;

Conte, A.G. (1975), *Validità*, in *Novissimo digesto italiano*, UTET, Torino, vol. XX, 1975, pp. 418-425;

----. (1977), *Aspekte der Semantik der deontischen Sprache*, in A.G. Conte, R. Hilpinen & G.H. von Wright (eds.), *Deontische Logik und Semantik*, Athenaion, Wiesbaden, 1977, pp. 59-73;

----. (1990), *Validità athetica*, in S. Castignone (ed.), *Studi in memoria di Giovanni Tarello*, Giuffrè, Milano, vol. II, 1990, pp. 163-176; seconda edizione in A.G. Conte, *Filosofia del linguaggio normativo. II. Studi 1982-1994*, Giappichelli, Torino, 1995, pp. 409-424;

----. (2000a), "Nomotropismo: agire in-funzione-di regole", in *Sociologia del diritto*, 27, 1, pp. 1-27;

----. (2000b), *Regole di linguaggio e linguaggio delle regole*, in M. Carapezza & F. Lo Píparo (eds.), *La regola linguistica. Atti del VI congresso della Società di filosofia del linguaggio*, Novecento, Palermo, 2000, pp. 41-63; second edition in A.G. Conte, *Filosofia del linguaggio normativo. III. Studi 1995-2001*, Giappichelli, Torino, 2001, pp. 921-945;

----. (2002),"Unomia. Il luogo delle regole in un mondo di fatti", in *Sociologia del diritto*, 29, 3, pp. 9-44;

----. (2003), "Filosofia del baro", in *Rivista internazionale di Filosofia del diritto*, 4, 4, pp. 679-746;

----. (2004), "*Kanon*. Filosofia della regolarità", in *Sociologia del diritto*, 31, 3, pp. 5-22;

----. (2011), *Sociologia filosofica del diritto*, Giappichelli, Torino;

----. (2012a), "Nomotropismo", in G. Lorini & L. Passerini Glazel (eds.), *Filosofie della norma*, Giappichelli, Torino, 2012, pp. 307-316;

----. (2012b), "Norma: cinque referenti", in G. Lorini & L. Passerini Glazel (eds.), *Filosofie della norma*, Giappichelli, Torino, 2012, pp. 57-65;

Conte, M.-E. (1992), "Semantica del predicato 'perdonare'", in G. Galli (ed.), *Interpretazione e perdono*, Marietti, Genova, 1992, pp. 13-21;

Devlin, P. (1959, ²1965), "The enforcement of morals", *Proceedings of the British Academy*, 45; second edition: *Morals and the Criminal Law*, in P. Devlin, *The Enforcement of Morals*, Oxford University Press, Oxford, 1965, pp. 1-25;

"Italian strike", or "*sciopero bianco*" (white strike), in English work-to-rule. Work-to-rule may be construed as an "hypertrophic nomotrophic behaviour", a behaviour which (on the contrary of ordinary nomotrophic behaviour) aims, by contrast, at a nomo-annihilating effect, that is at the derogation of a norm ('nomo-annihilating' is a term proposed by Edoardo Fittipaldi 2012).

Di Lucia, P. (1996), "Agire secondo una norma, agire per una norma, agire in funzione di una norma", in P. Comanducci & R. Guastini (eds.), *Struttura e dinamica dei sistemi giuridici*, Giappichelli, Torino, 1996, pp. 37-45;

----. (2002), "Efficacia senza adempimento", in *Sociologia del diritto*, 29, 3, pp. 73-104;

----. (2003), "Norma *in actu*", in P. Di Lucia, *Normatività*, Giappichelli, Torino, 2003, pp. 183-215;

----. (2007), "Agire *in-funzione-di* norme", in A.G. Conte, P. Di Lucia, A. Incampo, G. Lorini & W. Żełaniec, *Ricerche di Filosofia del diritto*, L. Passerini Glazel (ed.), Giappichelli, Torino, 2007, pp. 164-178;

Feldman, A. (2002), "Strange fruit. The South African Truth Commission and the demonic economies of violence", in B. Kapferer (ed.), *Beyond Rationalism. Rethinking Magic, Witchcraft, and Sorcery*, Berghan Books, Oxford, 2002, pp. 234-265;

Fillmore, Ch.J. (1971), "Verbs of judging", in Ch.J. Fillmore & D.T. Langendoen (eds.), *Studies in Linguistic Semantics*, Holt, Rinehart & Winston, New York, 1971, pp. 272-289;

Fittipaldi, E. (2012), *Everyday Legal Ontology. A Psychological and Linguistic Investigation within the Framework on Leon Petrazycki's Theory of Law*, LED, Milano;

Galtung, J. (1959), "Expectations and interaction processes", in *Inquiry*, 2, pp. 213-234;

Geiger, Th. (1947),*Vorstudien zu einer Soziologie des Rechtes*, Munksgaard/Universitetsforlaget, Aarhus/København, [Acta Jutlandica, 19, 2 (1947)]. Partial English translation by E.R. Peck, in Th. Geiger (ed.) (1969), pp. 39-122;

----. (1969), *On Social Order and Mass Society. Selected Papers*, R. Mayntz (ed.), Chicago University Press, Chicago/London;

Gilbert, M. (2005), "Shared values, social unity, and liberty", in *Public Affairs Quarterly*, 19, pp. 25-49; second edition: "Shared values, social unity, and liberty", in M. Gilbert, *Joint Commitment. How We Make the Social World*, Oxford University Press, Oxford, 2014, pp. 181-206;

----. (2014), *Joint Commitment. How We Make the Social World*, Oxford University Press, Oxford;

Goethe, J.W. von (1790), *Torquato Tasso. Ein Schauspiel*, Georg Joachim Göschen, Leipzig;

Kant, I. (1785), *Grundlegung zur Metaphysik der Sitten*, Harknoch, Riga; J.W. Ellington (trans.), *Grounding for the Metaphysics of Morals*, third edition, Hackett, 1993;

Kelsen, H. (1960), *Reine Rechtslehre*, Deuticke, Wien; Max Knight (trans.), *Pure Theory of Law*, University of California Press, Berkeley, 1967;

Lauer, H. (2015), "The anatomy of collusion", *Phenomenology and Mind*, 9, pp. 182-191;

Luhmann, N. (1969), "Normen in soziologischer Perspektive", *Soziale Welt*, 20, pp. 28-48;

----. (1972), *Rechtssoziologie*, Rowohlt Taschenbuch, Reinbekbei Hamburg; E. King & M. Albrow (trans.), *A Sociological Theory of Law*, M. Albrow (ed.), Routledge & Kegan Paul, London, 1985;

Passerini Glazel, L. (2009), "Decriptare criptòtipi", in R. Caterina, *La dimensione tacita del diritto*, Edizioni Scientifiche Italiane, Napoli, 2009, pp. 139-145;

----. (2012b), "Operanza di norme", in G. Lorini & L. Passerini Glazel (eds.), *Filosofie della norma*, Giappichelli, Torino, 2012, pp. 237-244;

----. (2013a), "Disused norms: norm atrophy and nomotrophic behaviour", in C. Stancati, A. Givigliano, E. Fadda & G. Cosenza (eds.), *The Nature of Social Reality*, Cambridge Scholars, Newcastle upon Tyne, 2013, pp. 55-64;

----. (2013b), "Norm atrophy and nomotrophic behaviour. On a mode of inexistence of norms", in S. Colloca (ed.), *The Value of Truth/The Truth of Value*. *Proceedings of the International Seminar*, Nomologics 1. Pavia, Residenza Golgi, 14th-16th July 2011, LED, Milano, 2013, pp. 159-168;

----. (2015), "Il perdono come atto nomotrofico", in S. Biancu & A. Bondolfi (eds.), *Perdono: negazione o compimento della giustizia?*, FBK Press, Trento, 2015, pp. 49-67;

Sacco, R. (2000), "Lingua e diritto", *Ars interpretandi*, 5, pp. 117-134;

----. (2015), *Il diritto muto. Neuroscienze, conoscenza tacita, valori condivisi*, Il Mulino, Bologna;

Santner, E.L. (1999, ²2000), "Freud's 'Moses' and the ethics of nomotropic desire", in *October*, 88,

pp. 3-41; second edition in R. Salecl (ed.), *Sexuation*, Duke University Press, Durham, 2000, pp. 57-105;

Searle, J.R. (1975), "A taxonomy of illocutionary acts", in K. Gunderson (ed.), *Language, Mind, and Knowledge*, University of Minnesota Press, Minneapolis, 1975, pp. 344-369;

----. (2010), *Making the Social World. The Structure of Human Civilization*, Oxford University Press, Oxford;

Silvi, M.Q. (2004), *Struttura giuridica del perdono*, Franco Angeli, Milano;

Weber, M. (1922),*Wirtschaft und Gesellschaft*, Mohr, Tübingen;

Wittgenstein, L. (1953), *Philosophical Investigations* (*Philosophische Untersuchungen*), G.E.M. Anscombe (trans.), Basil Blackwell, Oxford.

SESSION

4

SESSION 4

JUST JOINT COMMITMENTS

JOHN HORTON
Keele University, UK
j.horton@keele.ac.uk

RYAN WINDEKNECHT
University of Tennessee, USA
rwindkn@utk.edu

ASSOCIATIVE POLITICAL OBLIGATIONS AND THE DISTRIBUTIVE OBJECTION

abstract

In this article we assess the distributive objection to the claim that we have associative political obligations. We deploy three kinds of strategy in our response: avoidance, mitigation and confrontation. Avoidance and mitigation are accommodating responses, in which we argue that associative political obligations often need not fall foul of the distributive objection in any very damaging way. But they may not always be enough, and thus confrontation may sometimes be necessary. This response involves rejecting the more extreme claims of proponents of the distributive objection, while accepting that occasionally serious tensions between the demands of our associative political obligations and those of global redistribution may remain, both of which have moral standing.

In this paper, we consider an important criticism of the idea of associative political obligations. This criticism, following Samuel Scheffler, has become known as "the distributive objection" (Scheffler 2001), and its principal proponents tend to be theorists of distributive justice who are firmly committed to extensive global redistribution (e.g. Caney 2005, 2008; Tan 2004). The structure of our argument is fairly straightforward. In the first section (**I**), we begin by saying a little about associative obligations in general and associative political obligations in particular. The purpose of this discussion is to fix the *kind* of view that we defend. In the second section (**II**), we set out the distributive objection. Again, we concern ourselves with the generic features of this objection, although we distinguish between stronger and weaker versions of it. In the third, fourth and fifth sections, we present our response to the distributive objection. We deploy three kinds of strategy in this response: avoidance (**III**), mitigation (**IV**) and confrontation (**V**). Avoidance and mitigation are accommodating responses. They involve arguing that associative political obligations often need not fall foul of the distributive objection, because either there is no real conflict between them and global redistribution, or, where there is conflict, it can be rendered significantly less troublesome. These responses do a good deal to undermine supposed tensions between associative political obligations and global redistribution. But they may not always be enough. Confrontation, therefore, may sometimes be necessary, which involves rejecting the more extreme claims of proponents of the distributive objection.

To anticipate the direction of the argument, the view that we defend does not hold that the global poor have *no* legitimate significant moral claims against the rich, or that such claims *never* take priority over our associative political obligations. However, the view that we reject is the obverse of this. That is, we want to insist on the validity and robustness of our associative political obligations, and to deny that they are spurious or necessarily subordinate to universal principles of global redistribution. These are two, independent sources of moral claims. Both associative obligations in general and associative political obligations in particular, on the one hand, and global poverty, on the other hand, matter. Neither can be silenced or consigned to permanent inferiority by the other. Moreover, much of the time, they can comfortably coexist, as long as extravagant and unjustifiable claims are not made on behalf of either.

I. The idea of associative obligations is derived from a commonplace feature of ordinary moral opinion. Namely, that there is a dimension of our moral life that is based on neither universal

principles, such as equality or utility, nor voluntary transactions or commitments, such as promises or individual choices. Instead, it depends upon and grows out of patterns of relationships within particular associations. What is distinctive about associative obligations is that they derive in large part from being socially embedded within specific relationships. These relationships are presented as either independent of any choices, such as relations with parents, or inexplicable in terms of any choices that may have been involved in the process of acquiring them, such as relations between friends. Similarly, they cannot be adequately explained in terms of universal moral principles, such as maximising utility or respecting human rights. These relationships are often of great importance to us, sometimes being among the most valuable features of our lives, and they make serious moral demands on us. Moreover, we commonly think that it is enough to cite these relationships as justifications of our actions. They play a central role in justifying partiality in our behaviour and explaining why we have good reason to favour those who stand in a special relationship to us. It is frequently argued, however, that the idea of associative *political* obligations is less firmly grounded in ordinary moral opinion than, for example, the belief that we owe associative obligations to our family. Admittedly, our relationships with our compatriots is different in significant respects from our relationships with our family, most obviously in the absence of the kind of intimacy that is characteristic of the latter. Nonetheless, the belief that we owe associative political obligations to our compatriots that we do not owe to others is widespread. Our communal political life matters to us in a variety of ways and for a variety of reasons.[1]

What, then, is the distributive objection and how is it relevant to associative political obligations? Samuel Scheffler explains very well how this objection bears on associativism:

II.

> The distributive objection challenges the idea that members of affluent societies have special responsibilities to their associates that they do not have to other people. The objection need not deny that there are important differences of character and motivation between those who take such responsibilities seriously and those who act out of crudely self-interested motives. Nevertheless, it insists that special responsibilities serve to validate a natural tendency to partiality or favouritism within groups, and the effect of this form of validation is to confer unfair advantages on the members of wealthy groups while placing other people at an unfair disadvantage (Scheffler 2001, p. 85).

The distributive objection holds that associative obligations are a mask for unfairness and favouritism, whether the belief in them is held in good faith or as an excuse for selfishness. Associative obligations are always exclusionary. In addition to creating the benefits that are part of the relationship, from which outsiders are excluded, they also provide reasons further to benefit those who are parties to the relationship, rather than those who are excluded. The criticism of associative political obligations, in particular, is that such obligations, when given independent weight, detract from our universal duties to bring about global redistribution. If we have a countervailing duty, say, to address relative, but lesser, poverty in our own country, then there will be less resources to redistribute to those in countries where people are considerably worse off. Such a conception of associative political obligations, it

---

1 We have developed our account of associative political obligations more fully in Horton 2006, 2007a, 2007b, 2010; Horton & Windeknecht, forthcoming; Windeknecht 2010, 2012.

is argued, favours the citizens of prosperous countries. Political membership is irrelevant to moral consideration, and individuals do not have a legitimate claim to more resources merely because they happen to have the good fortune to live in a wealthier country.

Theories of global redistribution come in many different shapes and sizes, and it is impossible to explain let alone to respond to the range of diversity here. The theories in which we are primarily interested are those that are seriously sceptical of political membership having any moral relevance and therefore are strongly antagonistic towards associative political obligations. As we are only defending associative political obligations from the distributive objection, insofar as a theory does not embrace the distributive objection we have no argument with it.

III. We now move on to consider whether there is a necessary antagonism between associative political obligations and global redistribution. In arguing that there often need not be, we adopt the first of our conciliatory strategies: avoidance. This strategy involves showing that associative political obligations and global redistribution can coexist without there being any necessary conflict between them.

The first point is that distributive concerns do not necessarily bear on the *content* of associative political obligations. An obligation, say, to observe laws regulating education, traffic, local commerce, public health and many other areas; to give due cognisance to the political affairs of their own country; or to support its political institutions does not necessarily conflict with global redistribution. The point is that there is no *need* for theorists of global redistribution to deny that there are *any* associative political obligations, as many such obligations will be irrelevant to the distributive objection.

A further way in which in the content of associative political obligations does not necessarily conflict with global redistribution is that the former may sometimes be supportive of the latter. Admittedly, this is a contingent matter, but it is at least plausible that global redistribution may be advanced best through associative political obligations. As Richard Vernon has argued (Vernon 2010), given the world as it is and is likely to be for the foreseeable future, the only political agencies that can effectively bring about global redistribution are states or state-backed multi-national institutions. Assuming this is the case, it is vital for citizens to recognise associative political obligations to their own states, and to encourage and support their own state's role in actively promoting global redistribution.

A second point concerns the *demandingness* and the *indeterminateness* of associative political obligations. Often these obligations are demanding and determinate: there are specific actions that we are morally required to perform. But sometimes they are more like permissions, although this does not mean that they should be thought of as supererogatory. For instance, in many circumstances, they give us good reasons for favouring our compatriots, but do not require us to do so. And, even where they do, they can be quite open-ended in what they require. Thus, to focus on a non-political obligation, the defender of associative obligations does not need to believe that it is wrong for a mother to be concerned more about alleviating global poverty than about providing material goods to her children, as long as such concern does not involve denying her children what properly belongs to them. While theorists of global redistribution and defenders of associative obligations may disagree about the point up to which favouring one's own children is permissible, they need not, and even when they do, neither is thereby committed to wholly rejecting the legitimacy of the other's concerns.

It may be conceded, however, while this is true of parental obligations, it is more problematic in relation to *political* obligations, which are more demanding and determinate. Two points, though, can be made. First, political obligations are not always so demanding or determinate that they need conflict with global redistribution. Once laws are obeyed, including taxes paid,

people have considerable scope to decide how to dispose of their remaining resources. They can still, without impinging on their political obligations, devote much of such resources to alleviating global poverty. Secondly, an obligation to play some role in the political life of one's own community or to support its political institutions, for example, is consistent with interpreting such an obligation in a way that involves encouraging and cajoling both one's government and one's fellow citizens to do much more to promote global redistribution. So, again, with respect to their demandingness and their determinateness, conflict between associative political obligations and global redistribution can in some circumstances, although not always, be avoided.

A final point concerns the *limits* of associative political obligations. Even the most determined advocate of associative political obligations must acknowledge that they have their limits, although there may be no agreement about exactly where these limits lie. Associative political obligations do not imply that individuals are permitted, let alone required, to do absolutely anything to promote the interests of their compatriots,and certainly not to do whatever their state may demand. Associative political obligations are an independent source of moral claims, but they do not exist apart from the rest of morality.

IV. Having explored the first, we now move to the second of our conciliatory strategies: mitigation. There is, perhaps, no sharp line between the strategies of avoidance and of mitigation, and there is, actually, a tendency for them to shade into one another. The distinguishing feature of mitigation, however, is that it comes into play when conflict between associative political obligations and global redistribution cannot be avoided. But this strategy contends that such conflicts need not necessarily be a serious concern. For example, the strategy of mitigation often works especially well where there is a putative conflict between associative political obligations and global redistribution, but where it is obvious that one has priority over the other. Thus, when the cost to political community is small, but the contribution to alleviating global poverty is great, the claims of the latter will be weightier than the former. But the reverse may be true as well.

There are also other cases where mitigation can play a role in reducing the tensions between associative political obligations and global redistribution. This is where there is no clear priority of one over the other, but an imperfect compromise or partial accommodation is possible. In these cases, we cannot fully satisfy both sets of demands, and one does not have priority over the other, but we can partially fulfil both sets of demands through a compromise or mutual adjustment between the two. Again, this will not always be possible, but sometimes by accepting that we cannot do all that we should to promote global redistribution, as well as all that we should do to fulfil our associative political obligations, we can nonetheless do something to meet both sets of demands. And, as far as possible, such a strategy might represent the best opportunity to give each their due, while acknowledging that neither is fully satisfied.

Taken together, the strategies of avoidance and of mitigation support the claim that it is possible to combine robust views about associative obligations with concerns for global redistribution. In short, it is possible to be committed to both without rejecting either. This is not to claim that all tensions between them can be eliminated – which is where we disagree with theorists, like Kok-Chor Tan (Tan 2004), who allow associative political obligations, yet subordinate them to universal principles of global redistribution – but such tensions are part and parcel of difficult moral choices with which we are all familiar.

V. So far, our responses to the distributive objection have been conciliatory, avoiding or mitigating conflicts between global redistribution and associative political obligations. In

doing so, we have shown that there is sufficient moral space for both kinds of considerations to co-exist, without them coming into serious conflict. However, some proponents of the distributive objection, especially those committed to a strong form of global egalitarianism, have combined it with a thoroughgoing critique of the whole idea that associative political obligations make serious moral claims on us.

The argument that we are most concerned to address focuses on the contention that any distributive distinctions which are based on 'mere' membership of a polity must be morally arbitrary. Simon Caney provides a particularly trenchant statement of this view:

> [T]he thesis that the borders of some, or all, principles of distributive justice are defined by the borders of the state needs to show why state membership is morally relevant … For it is hard to see how state membership could have the type of normative significance ascribed to it … Which state someone belongs to is, in very many cases, a matter of luck. It is a matter of fortune whether one is born into Berkshire or Bihar and it seems highly perverse to argue that such facts should affect what people are entitled to. Why, one might ask, should being born into one state have such a tremendous impact on people's prospects in life? (Caney 2008, p. 505).

This is the argument that we are most concerned to address. In particular, we are most concerned to show that 'such facts', properly understood, have the moral relevance that Caney denies them.

First, it is worth noting that, in the sense deployed by Caney, virtually everything about us involves some measure of fortune. When and where we were born; whether we were born healthy, intelligent, charming or none of these; that we were born at all; and so on. All are arbitrary in the sense that they are contingent facts that could have been otherwise. They are also arbitrary in the sense that we cannot be said to deserve any measure of fortune which we may have had because of these facts. This is an important point in resisting the temptation to argue that we do somehow deserve our good or bad fortune. However, even if we resist this temptation, and even if we are dealing with sheer 'brute luck', it is important to note that nothing necessarily follows about how we should regard these facts. To draw any moral inferences from these facts requires the introduction of highly controversial, moral premises. Notwithstanding Caney's apparent claims to the contrary, his own preferred explanations for moral relevance are also open to charges of arbitrariness. So, his first explanation, "that principles of distributive justice apply to all those who belong to a common system of interaction and interdependence" (Caney 2008, p. 493), leaves open the question of how individuals come to belong to such systems in a morally relevant way. Surely, from Caney's viewpoint, the fact that we are born into particular systems of economic interaction is just as arbitrary as the fact that we are born into particular systems of political relations. Both involve patterns of social and economic interaction, but the chains of economic interdependence are no more obvious or easier to identify than those between members of a particular political community. In fact, if we are concerned with interaction and interdependence, there is a more straightforward tie between the well-off and less well-off in Berkshire, to take up Caney's example, than between the well-off in Berkshire and the less well-off in Bihar. This is not to deny that there is a connection between the global rich and poor, but it is to ask why more local and circumscribed patterns of interaction between compatriots should be dismissed as irrelevant or trivial by comparison.

Furthermore, his second explanation, "that persons should be included in the scope of distributive justice in virtue of their humanity" (Caney 2008, p. 491), similarly leaves open the question of the moral relevance of belonging. Beyond our trivial inability to choose the

species into which we are born, it is far from obvious what moral relevance the *bare fact* of belonging to the same species has for distributive justice. What, it may be asked, about our common humanity makes us special? Why, for example, is reason or language morally relevant? Our point here is that on his own account, Caney's attempt to draw the boundaries of distributive justice along the lines of our common humanity, ignoring the broader category of sentient beings and the narrower category of social groups, is problematic and stands in need of justification. To avoid any misunderstanding, we are not denying that our common humanity can serve as a basis for moral treatment, but are making an essentially *ad hominem* point, drawing attention to the implications of the justificatory standards he uses to attack associativism.

In short, global egalitarianism does not just fall out from acknowledging the arbitrariness that affects our lives. Nor does it simply fall out from some principle of moral equality, which is what is supposed to fill the gap between the arbitrary facts and the moral claims that Caney (and others) want to make. Of course, additional arguments to support such claims have been proffered, but the least that can be said is that they are all controversial. However, we cannot here show that such principles cannot be justified, as it is beyond our remit to assess the merits of the various arguments for global egalitarianism. We must rest content with observing that global egalitarianism needs support from further argument, which is a more difficult task than setting out some good reasons for a duty of general concern for the suffering of the global poor. Moreover, such principles, if they are to have purchase on the real world, must pay attention not only to the realms of ideal theory, but also to the consequences of trying to implement such principles in situations where the vast majority of people do not subscribe to such principles or anything like them.

In fairness, Caney is not unaware of this last point, and he does give some attention to practical political agency. Following an argument noted earlier, he suggests that the state "may have normative significance as an instrument of cosmopolitan justice, as a source of duties and as an object of loyalty, pride or shame" (Caney 2008, p. 515). Specifically, he claims that "an egalitarian cosmopolitan can, and should, recognise the instrumental importance of political institutions, including the state ... [It] can pursue cosmopolitan policies ... [It] can construct cosmopolitan institutions and reform existing international institutions in a more cosmopolitan direction" (Caney 2008, pp. 510-511). Regarding political obligation, he allows that "one might adhere to cosmopolitan principles of distributive justice ... but also hold that as a member of the state one is under a special duty to uphold the cosmopolitan entitlements of one's fellow citizens", as well as "duties to members of other countries in cases where the latter are disadvantaged ... because of the unjust foreign policy of one's state" (Caney 2008, pp. 511, 514). And, regarding political identity, he writes that in "very many cases people's membership of the state affects their identity", and accepts that "an egalitarian cosmopolitan approach can recognise ... pride and patriotism are sound and justified when they are grounded in their state's successful pursuit of cosmopolitan ideals" (Caney 2008, p. 512). However, Caney's concession that the state may play a role in bringing about global justice is entirely instrumental, and it is not easily reconciled with his critique of what he considers to be an arbitrary mapping of the borders of justice onto the borders of the state, as well as the intrinsically unjust "system of state borders" or "global statist system" (Caney 2008, pp. 500, 508).

Against this purely instrumental account, we insist that states are, or at least can be, genuine political communities, which have a meaning for their members that is far from exhausted by the promotion of global justice. A polity is not *just* an arbitrary arrangement of borders, although borders are in some respects arbitrary. There is a common life – in the sense that citizens share a set of institutions and laws, as well as a diverse set of cultural and historical

values, all of which have meaning for them and play an important role in structuring their world. The idea of being a citizen of a particular state is a morally significant idea. The polity, at least when well-functioning, is a strong source of stability and security, and provides a rich source of identity. Moreover, to be a member of a polity is to share not only the benefits of membership, but also the corresponding burdens. Pretending that a polity is no more than an arbitrary arrangement of borders or lines on a map is, therefore, a false and misleading form of reductivism. Membership matters, to different degrees and in different ways, for those who are members; they are neither rogues nor fools for thinking that it does and for attaching moral significance to their special relationship with other members.

As we have seen, Caney does at times acknowledge some features of the picture that we have just presented, but his doing so raises more questions than it answers. With respect to political obligation, Caney's attempt to ground it in global egalitarian duties runs afoul of the "particularity requirement", the condition that political obligation binds individuals to *their* polity and not to any other. As A. John Simmons, who, for different reasons, is sympathetic to the challenge that Caney poses to "statism", writes:

> Suppose we accepted ... that we have an obligation or duty to support just governments, and that this is what our political obligation consists in. And suppose that I am a citizen living under a just government. While it follows that I have an obligation to support my government, it does not follow that there is anything special about this obligation. I am equally constrained by the same moral bond to support every other just government (Simmons 1979, p. 31).

Similarly, if we suppose, with Caney, that one has a "duty to uphold the cosmopolitan entitlements of everyone", it would follow that one has a "duty to uphold the cosmopolitan entitlements of one's fellow citizens", but it would not follow that there is anything 'special' about this duty, contrary to Caney's claims (Caney 2008, p. 511). The latter adds nothing to the former. Caney contends that the content of the special duty may affect the content of the general duty, but it is difficult to see how (Caney 2008, p. 511). Likewise, it is difficult to see how the content of his other special duties, specifically duties of compensation, which are owed to those wronged by actions performed by our state, affect the content of the general duty in any significant way (Caney 2008, p. 514). For instance, we might suppose that special duties would magnify general duties, that if we owed a fellow citizen a duty to uphold her cosmopolitan entitlements, or if we owed a foreign national a duty of compensation for past wrongdoings committed by our state, then we would have a stronger reason for discharging the general duties that we owed to such individuals. However, Caney rejects such a view (Caney 2008, p. 514). Special duties do not add any weight to general duties, but may affect those to whom we are bound. According to the general duty, though, and herein lies the point, we are required to uphold the universal egalitarian entitlements of everyone, both those with whom we share citizenship and those we do not, and those with whom we share compensatory relationships and those we do not. On Caney's view, *we are always already bound* by the "duty to uphold the cosmopolitan entitlements of everyone".

There is also a more fundamental tension in Caney's attempts to ground political obligation in global egalitarian duties. For the question is: why is political membership morally relevant? If it is morally arbitrary to map the borders of justice onto the borders of the state regarding entitlements, then surely it is morally arbitrary to map the borders of justice onto the borders of the state regarding requirements. Similar issues arise in relation to Caney's attempts to account for the significance of political identity, issues which are further compounded by a misunderstanding about the phenomenology of belonging. First, consider identity. What is

puzzling is how our *identification* with our political community can shift from being morally arbitrary to being morally relevant, simply because that with which we identify becomes instrumentally useful; how our *identification* with our political community can shift from being "as arbitrary as [our] class origin or social status or ethnic identity" (Caney 2008, p. 505) to being "sound and justified", simply because our polity pursues a set of "cosmopolitan ideals" (Caney 2008, p. 512). Either we belong to our polity in a morally relevant way or we do not. And, *a fortiori*, either it makes sense for us to identify with our polity in a morally relevant way or it does not.

Now, consider belonging. On the one hand, Caney wants to say that the state, membership of it and identification with it, is merely contingent and morally arbitrary. On the other hand, he wants to acknowledge that:

> In very many cases people's membership of the state affects their identity. They think of themselves as members of that state and it is an integral part of their self-understanding. In virtue of this, persons often feel sentiments like shame for the injustices committed by their state ... Citizens may feel ashamed of their state's colonial record or their state's waging of an unjust war. They may take pride in the fact that it has stood up against an evil oppressor ... And if we accept the cosmopolitan conception of political morality this means that a state acts 'in our name' when it discharges its (cosmopolitan) duty well and is a fit object for our loyalty. When it does not, and it jeopardises the rights of others, then its members are entitled to regard it with shame or guilt and to deny that it acts in 'their name' (Caney 2008, p. 512).

There is much in this description of how people experience political belonging that we would readily endorse. However, it is difficult to see how Caney can square this characterisation with the charge of arbitrariness. If our polity is an integral part of our self-understanding, then it is at best misleading to present political belonging as completely contingent and arbitrary, and without meaning and significance. Furthermore, if our political belonging were completely contingent and arbitrary, it would make little sense for us to feel sentiments such as pride and shame with regard to the conduct of our polity.

One possible response would be simply to bite the bullet, and to allow that the state, membership of it, and identification with it, carry an independent and non-reductive moral relevance. As we have argued, acknowledging our associative political obligations and promoting a substantial measure of global redistribution need not be mutually exclusive. We can, for example, recognise that, as citizens, we have responsibilities towards our fellow citizens, to ensure levels of social welfare, public education and the like, and yet also recognise that we also have responsibilities towards our fellow human beings more generally, to help alleviate world hunger, provide sufficient drinkable water, minimum levels of education and the like. It is certainly possible to link these global responsibilities to justifiable feelings of national pride or shame, depending upon how effectively our polity contributes to global redistribution. Such feelings, though, would be parasitic upon having the relevant sense of political identity that made thinking in terms of *our* polity meaningful in the first place. Political identity necessarily precedes such feelings. Because of this, accounts such as Caney's are inherently confused, in their attitude toward the state, and in their understanding of what it means to be a citizen of a state.

VI. We have sought to assess, in general terms, the merits of the distributive objection to associative political obligations. Part of our assessment has taken the form of showing how it is possible to avoid or mitigate the seriousness of the claim that being genuinely concerned

with global poverty is incompatible with a commitment to a robust conception of associative political obligations. We have argued that the conflicts between the two sets of claims canbe easily overstated. Arguing that associative political obligations make substantial moral claims on us is consistent with accepting that there are global duties, such as to alleviate poverty and suffering, even if the more extensive claims of some global egalitarians are more challenging in this regard. It is clear that associative political obligations do not justify the pathetic levels of resources that affluent countries currently devote to alleviating global poverty. But doing more need not cast a shadow over associative political obligations. In fact, associative political obligations may shed some light on why and how we could and should do more.

The other part of our assessment has been markedly less conciliatory. In addressing theorists of global redistribution who have tried to deny the moral significance of associative political obligations, we have sought to vindicate their meaning and value. We have argued that presenting them as morally arbitrary fundamentally misconstrues their nature. We acknowledge that this means that there can be circumstances in which tensions between associative political obligations and global redistribution can arise. But these tensions must be lived and negotiated as best they can, in a world which does not always offer easy, or sometimes any, ways to resolve them. We do, nonetheless, insist that associative obligations in general and associative political obligations in particular are meaningful, and therefore occupy an important place in the moral geography of this world.[2]

REFERENCES

Caney, S. (2005), *Justice Beyond Boundaries: A Global Political Theory*, Oxford University Press, Oxford;

----. (2008), "Global distributive justice and the State", *Political Studies*, 56, pp. 487-518;

Horton, J. (2006),"In defence of associative political obligations: Part one", *Political Studies*, 54, pp. 427-443;

----. (2007a), "In defence of associative political obligations: Part two", *Political Studies*, 55, pp. 1-19;

----. (2007b), "Defending associative obligations: A response to Richard Vernon", *Political Studies*, 55, pp. 880-884;

----. (2010), *Political Obligation*, 2nd revised ed., Macmillan, London;

Horton, J. and Windeknecht, R. G. (2015), "Is there a distinctively associative account of political obligation?", *Political Studies*, 63 pp.903-908;

Scheffler, S. (2001), *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought*, Oxford University Press, Oxford;

Tan, K.C. (2004), *Justice Without Borders: Cosmopolitanism, Nationalism and Patriotism*, Cambridge University Press, Cambridge;

Vernon, R. (2010), *Cosmopolitan Regard: Political Membership and Global Justice*, Cambridge University Press, Cambridge;

Windeknecht, R.G. (2010), *Rethinking Political Obligation: An Associative Response to Philosophical Anarchism*, Ph.D. thesis, Keele University;

Windeknecht, R. G. (2012), "Law without legitimacy or justification? The flawed foundations of philosophical anarchism", *Res Publica*, 18, pp. 173-188.

2 This is a shortened version of a much longer paper. We are grateful for discussion at the San Raffaele Spring Philosophy School; the Colloquium on Ethics, Politics, and Society at LUISS, Rome in December, 2014; and also for the comments of colleagues at the University of Tennessee. A distant forbear was presented as the keynote lecture to the 'Postgraduate Conference on Law and Philosophy' at the University of Stirling in 2008.

ROBERTA SALA
Università Vita-Salute san Raffaele
sala.roberta@unisr.it

MAY JOINT COMMITMENT STABILIZE MODUS VIVENDI?

abstract

This contribution intends to extend my previous attempt to defend modus vivendi as an alternative way to include people who do not share the essentials of a liberal society. The idea was to respond to a claim to realism: besides loyal citizens, whose doctrines overlap on the basics of a fair society, there are people whose loyalty towards institutions is not be wholehearted – since they do not concur to their public justification – but who may endorse them in a stable way. I consider now a further way to deal with inclusion, compliance and stability: besides division on fundamental commitments and disagreement about values, peaceful coexistence may find strength in an alternative way to conceive the attitude of cooperation as rooted in a joint commitment. My argument will be presented as follows: a) I recall my conclusions about my idea of stable modus vivendi; b) I try and improve the wished outcome of stability in spite of partial political loyalty by reinforcing this with the argument of joint commitment; c) I draw some interlocutory conclusions.

The title of my contribution requires to be explained. I previously dealt with modus vivendi in the light of the Rawlsian political liberalism. My aim was to rehabilitate modus vivendi as an alternative way to include the so-called "unreasonable". The idea was to respond to a claim to realism: besides loyal citizens, whose doctrines overlap or will overlap on the basics of a fair society, there are quasi-loyal citizens, those people whose loyalty towards institutions is not and will not be wholehearted but who may endorse them in a stable way. Their inclusion is grounded in a *stable* modus vivendi. The core question was to indicate reasons and motives backing their compliance and assuring stable cooperation. I consider now a further way to deal with inclusion: besides division on fundamental commitments and disagreement about values, peaceful coexistence may find strength in an alternative way to conceive the attitude of cooperation as rooted in a joint commitment. The bet is to defend a joint commitment as a sort of allegiance to one's political community. My argument will be presented as follows: a) I recall my idea of stable modus vivendi; b) I try to improve the wished outcome of stability in spite of partial political loyalty by reinforcing this with the argument of joint commitment; c) I draw some interlocutory conclusions.

I. Modus vivendi has generally been dealt with in regard to the idea of toleration in a plural society. In this perspective modus vivendi has seen as the dark side of toleration as a moral notion roughly grounded in the respect for the others' freedom and self-determination. When there is no room for toleration as a positive attitude of acceptance of persons who disagree with us against their ideas and beliefs, modus vivendi does its work to forbear those persons despite their unbearable beliefs with the only prudential aim to reach a balance or equilibrium among respective powers as long as possible. In this negative light modus vivendi has been rejected by John Rawls. According to Rawls, modus vivendi is a sort of political final disposition based on an unstable balance of political forces, rather than being grounded in a set of moral principles. For that, modus vivendi shows to be unfair: as Rawls puts it, modus vivendi is "political in the wrong way" (Rawls 2005, pp. 39-40). Thus, unfairness and instability are strictly connected: any modus vivendi is unstable since it is based on unfair reasons, i.e. mere convenience and fear. The only reason people agree to any modus vivendi is that it seems the best choice for them at an acceptable cost. Rawls contrasts this prudentially-motivated modus vivendi with the idea of an overlapping consensus, that is, an agreement based on a sort of moral (political) values. Only when this consensus obtains, stability may be assured for the "right" reasons.

Such an overlapping (moral) consensus among comprehensive doctrines may be achieved only among that class of comprehensive doctrines that are reasonable, meaning that reasonableness qualifies the group of comprehensive doctrines that can find common ground within a liberal-democratic regime (Sala 2013, p. 255). By contrast, those people who do not share this common terrain of values cannot reach any moral consensus. Their possible adherence to institutions may be attained only on the ground of a modus vivendi: i.e., it cannot be warranted at all. In face of them, Rawls trusts the benefits of living under liberal institutions will enable those people to come to support those same institutions, at first by a modus vivendi, then transforming it into an overlapping consensus (Rawls 2005, pp. 158-168). The Rawlsian thought about consensus is arguable, first of all because of the unrealistic dichotomy between moral consensus and modus vivendi. Rawls maintains that consensus is moral or it is not. I believe that, on the contrary, it is possible to imagine a consensus established on prudential reasons: the idea is that one may consent to a settlement for reasons placed in non-moral or prudential reasoning. Speaking so is coherent with the statement of value pluralism: moral reasons may appear non-moral when they are seen from another point of view. The idea is that, crudely put, we may admit further models of coexistence, although not all of them are "technically" forms of consensus. They may be accounted as modus vivendi in a specific sense (Sala 2013): it is not a mere modus vivendi, as Rawls depicts it. It is a special modus vivendi in so far as it is more stable than the ordinary one, as it is reached from different points of view or world visions that are not necessarily connected to mere balances of forces at risk of overturning.

The inclusion of people *via* this special modus vivendi is to be understood as based on a "partial" loyalty to society, as it is not referred to a full endorsement of its fundamentals. Nonetheless, they show to be ready to support liberal institutions although not for sharing the liberal values that are supposed to ground a fair society (equal respect for others, freedom of conscience, toleration, justice and so on). They may have other reasons relying on a view of the world in which people are not free and equal, but in which persons are viewed – for instance – as divided into saved or sinners, elect or damn. The way in which they see the others as fellows to be saved – although it does not appeal to an ideal of equal respect for persons – does not imply any direct infringement of the rights of others. They live peacefully with others in a collaborative way that takes the form of a special modus vivendi. It is special because it is not doomed to be precarious, or to be more precarious than possible morally grounded institutions. To conclude: we should admit that not all "regular" citizens abide by the "right terms" of cooperation, nor comply with liberal-democratic institutions by agreeing *with* them: sometimes their compliance corresponds to a mere agreement *to* them on the basis of their reasons, be they moral or prudential, or on the basis of mere motives, habits, traditions or mere (non reflected) adoption of a shared practice of cohabitation (Scheffler 1994).

The above reflection stems from a realist claim: besides loyal citizens, whose moral positions overlap and,in doing so, enable them to sustain a fair society, there *actually* are "partially-loyal" citizens, people whose loyalty towards institutions is not wholehearted – since they do not endorse them on the basis of liberal moral values – but who may adhere to them in a stable way. *Pace* Rawls, nothing wrong happens if liberal institutions find stable compliance of people who are divided at the level of their fundamental commitments.

My current aim is to explore further how fair society may be supported by people, whose adherence to liberal institutions contribute to their legitimation, hence the social stability, despite their disagreement on basic values of social coexistence (here legitimation is about what reasons there might be to justify the exercise of political power). Taking realism seriously, my intent is to probe possible legitimation through joint commitment, in the

II.

circumstances in which liberal institutions are not and will not be endorsed. Possibly institutional arrangements *prove* to be legitimate since they obtain compliance from citizens, independently of her specific reasons or motivations. Modus vivendi as I depicted it before, as a stable modus vivendi among people differently motivated to sustain liberal institutions, may be further defended as an alternative way to coexistence, according to a theory of joint commitment.

My attempt to go deeper into my idea of a stable modus vivendi consists of experimenting two ways: (1) first, approaching an alternative account of loyalty inspired by Margaret Gilbert's theory of joint commitment; (2) second, reaffirming modus vivendi as a reasonably stable pattern of peaceful coexistence among liberal citizens and not-liberal-nor-democratic people. (3) In conclusion, my will is to renovate my idea of modus vivendi as a moral engagement of liberal institutions towards those people, a way for their inclusion into citizenry in spite of their being unable to share liberal values.

Before entering the debate on joint commitment, I would like to stress the following point. I understand this "moral engagement" on the side of liberal institutions to include not-liberal-nor-democratic people as their overarching aim. The underlying belief is that as liberals we should care of the ones who disagree with us, whose ideals do no overlap with ours. I understand liberalism as committed to ensuring that the political order is to be justifiable or, at least, to appear legitimate in face of all people, be they liberal or not. To appear legitimate means to be accepted for whatever reason or motive or, even, for no reason at all and for motives only, but in a stable way (Rawls 2001, p. 33). A minimum acceptance of institutions may be attained even unwittingly, for example, as involuntary result of a joint commitment. This justificatory project of modus vivendi is morally based upon the reference to the moral status of all individuals as free and equal: the overall idea is that all individuals are entitled to the right not to be coerced unless they overtly menace liberal order and infringe others' rights. In maintaining this, I part company from the defenders of the internal conception of liberalism (Quong 2011). Their objection is a well-known one: the justificatory work of liberal theory – it is often emphasized – is not addressed to those people who do not take basic liberal values for granted. In the light of this objection the goal of political justification is not to solve the crucial question "why be liberal?", but more modestly, to understand what kinds of arguments citizens already committed to certain liberal values can legitimately offer one another. Those who do not share basic liberal values, and especially the moral claim that persons are free and equal, are not entitled to any justification. I reply to this objection by arguing that a moral consensus – specifically, an overlapping consensus on liberal moral values – is too abstract an ideal to be reached by all citizens in a plural society, inhabited not only by reasonable, but also by non-liberal-nor-democratic people. In the *real* circumstances in which a moral consensus is unattainable there is room for a modus vivendi to build on the two *facts* already mentioned above: a) that there are people within liberal-democratic societies whose lives are not inspired by liberal values; b) that these people may nonetheless sustain liberal-democratic institutions. That there are not-liberal-nor-democratic people within liberal society is a *fact* and it is probably a relevant one. Honestly I see that the matter is about the grounds of their inclusion: how may those people come not coercively to adhere to liberal institutions? What reasons may push them to support them, even in absence of a calculus of opportunities? My tentative answer takes the cue from two preliminary points: first, legitimation may occur in the absence of moral reasons; not moral causes or simple motives may act as factors of legitimation (Horton 2012). Second, legitimation may spring from relevant facts, that is, facts that turn out to be relevant to build up a theory of political legitimation. So conceived, legitimation is not a mere description of a matter of fact, something like a contingent acquirement as obtained by a specific group of individuals as

empirically identified. It finds a further ground on selected facts, the normativity of which is offered by joint commitment. I will hold that joint commitment is not to be claimed as a substitute for moral arguments or whatever further motives and motivations to defend social order as a legitimate and stable one. It is just an alternative argument to sustain the notion of legitimate social stability without appealing to external (mostly moral) sources of the mere circumstances of cohabitation. In this sense I will defend my idea of modus vivendi as corresponding to an enlarged *practice* of public exchange that is hospitable to various kinds of reasons and motives, which do not necessarily refer to liberalism as morals.

Before coming to discuss the theory of joint commitment, let me emphasize again the relevance of facts for political theory and legitimation (Horton 2010a, p. 435; Rossi 2010). To admit that a political theory should meet the criterion of the descriptive adequacy means to maintain that political theory should engage with the *phenomena* of politics as they are, as they happen, without indulging in idealizations (Scheffler 1994). There is no such thing as a clear-cut normative-descriptive distinction: desirable normative political theory has to be in dialogue with a phenomenological grounded understanding of a society's forms of legitimation (Rossi 2013). The underlying idea is that – besides theorizing – we have to acknowledge the world. We cannot but also accept that the world is only to a limited extent directly improvable in accordance with our ideals.

III.

Having said that, the question now is to detect such *facts*: when may we say to be in front of people actually sustaining institutions? Where to turn the eyes to see people adhering to institutions? How could the mere fact of compliance play a normative role in defending the social order? The challenge here consists in interpreting joint commitment as a *fact* about people when they live within a community, willingly or not. Joint commitment seems to provide a case of a realistic approach of legitimation.

In the remainder of this part I will introduce the idea of joint commitment with regard to the kind of bonds it implies among people jointly committed. I will consider the theory of joint commitment political obligations defended by Margaret Gilbert and the alternative way to deal with it by John Horton. This comparison should lead to focus on joint commitment as a relevant fact for politics: this *fact* has to be paid a special scrutiny as it is supposed to act as a factual basis of social cooperation. The challenge consists in interpreting joint commitment as a sort of allegiance to the community: it is not to be conceived as a mere subjective feeling nor as voluntary or contractual relationship, but as an expression of membership. The task is to understand the debate on joint commitment as a contribution to draw a realistic approach of political theory.

I intend now to address the above challenge as follows: a) I will sketch out briefly Margaret Gilbert's account of joint commitment as referred to political obligation; then, I will depict the objections risen by John Horton, with specific regard to the normativity of joint commitment. b) I will do some tentative reflections about the relationship among stability, legitimation, and political compliance. In the end, modus vivendi will be back again as a name for a way to be involved (and not coerced) in a social enterprise, more or less willingly, generally aiming at a peaceful coexistence.

III.A Joint commitment and political obligation

Let me start by recalling the general meaning of political obligation. To have a political obligation is to have among others a moral duty to obey the laws of one's country or state (Dagger & Lefkowitz 2014). Some questions arise immediately: why should one obey laws? Why should one be loyal to institutions? What about those people who live in a state, even born within its territory, but who live like aliens as they belong to, say, a cultural minority? How do they acquire – if they can – an obligation to comply with political institutions? A possible

answer may descend from Gilbert's theory of joint commitment political obligation. Gilbert focuses on how the sentence "I am a member of a political society" (that is a natural fact) can logically imply "I have obligation", that is a non-natural fact (Gilbert 2006, p. 9; Gilbert 2014). She refers to that sort of duties or grounds for obedience created by membership in a political society. The idea is that people in certain contexts have sufficient reasons to support and comply with certain political institutions. Political obligation is not to be referred to a voluntary or a contractual relationship. Instead, it springs from a joint commitment emerging out solely from an act or state of "wills". This concept of joint commitment stands outside a "singularist" conceptual scheme. To exemplify: walking together is a joint action that does not need any agreement. It is wrong to say that two friends walking together have agreed to walk together: to walk is what they are doing. The mutual expression of readiness to engage in a joint activity is to be understood as a *common knowledge* between the parties. Common knowledge is a *fact* out in the open between them. At the same time, what occurs at the collective level is enough to motivate the individuals who make up the collective. That is, our goal is sufficient to motivate each one of us. Further, the obligations of joint commitment are *not moral requirements.* They do not count as moral obligations. A joint commitment obliges by virtue of its structure, that is, by virtue of "jointness". The *fact* that one owes another person an action is not the same to say that she is morally obliged to perform this action.

To discuss Gilbert' position I avail myself of Horton's critique to it, as unlike Gilbert he investigates the moral aspects of political obligation (Horton 2006, 2007). According to Horton a commitment to the institutions is not the same than a commitment to walk together. In Horton's associative account of political obligation the relevant idea is that there is some sort of moral relationship that holds by virtue of membership to a polity, the members of which mostly do not voluntarily choose to join. Even if we may agree with the idea that if one stops walking she is under the obligation to explain why she did so – Horton says – this obligation is very weak, and may rise a very weak right on the side of the partner. If political obligations are deprived of any moral dimension, the force of what makes them compelling remains obscure. Horton's account of political obligation, though sharing some aspects of Gilbert's one, like the emphasis on membership and association, tries to capture the normativity implied by being members of the same community. The focus is both on the idea that membership of association gives rise to obligations and on which kind of obligation one is dealing with. According to Horton, membership comes up as an important part of a *sort of phenomenology of our moral-political experience*: in some circumstances – especially those in which we feel shame or dishonour – we acknowledge our being a member of a polity and that being so has a moral meaning for us. The idea that we acquire memberships that we have not chosen is simply *the way things are.* Horton comments that this is a *fact* sufficient to support the claim that we can understand ourselves to be ethically bound to polity as non-voluntary group. He finds reasonable to think that we could have obligations to the polity because its distinctive element is the need for an effective coercive authority to provide order, security and some measure of social stability. A polity as a form of association holds as a generic value the good of order and security. This conclusion permits to avoid a reductionist account of political obligation as that depicted by Gilbert (political obligations as stemming from a not moral joint commitment) and to identify an independent source of normativity, since associative relationships intrinsically compel individuals to sustain (overtly or tacitly, wittingly or unconsciously) their own community. It is remarkable here to stress how that kind of associative obligation does exist independently of people's endorsement and how people are however requested to acknowledge it. Horton concludes by saying that political obligations are a concomitance of membership to a particular polity, a polity being a form of association that has as its generic value the good of order and security (Horton 2010b).

The exchange on political obligation between Gilbert and Horton as sketched above makes some suggestions about how to revise modus vivendi in order to strengthen its stability. My idea now is to develop a case for further stabilizing modus vivendi. This is only a tentative enterprise, and it is to be understood as more interpretative than explanatory. Consistently with this realist outlook an interpretative understanding of the *practice* of politics is favoured against any moral explication (Rossi & Sleat 2014; Newey 2010). It goes without saying that the descriptive element is not intended to substitute the normative one: notoriously, to be obliged by a commitment does not mean to be obligated to act consequently. But it also goes without saying that to be obligated by a norm does not imply to feel obliged to follow it as well. What I mean is that if the descriptive level and the normative level part company, we should admit that the normative level and the practicability of the norms do the same. Having said that, an interpretative understanding as the one being dealt with here is concerned with what sense can be made of the idea that people join a community, or they have some sort of relationship with it. The crucial point is about what kind of obligation the not-liberal-nor-democratic people may have in front of a fair society, in spite of their – so to speak – inability to share its constitutional essentials. In fact, if an associative account of political obligation may work to scrutinize how people feel their engagement possibly with their "native" community, the question is whether and how this account may possibly work.

With this question in mind, I find promising to my purpose of the stabilization of modus vivendi three elements of the above discussion between Gilbert and Horton: (a) the idea of a *common knowledge*, as claimed by Gilbert; (b) the idea of a *sense of identity*, as claimed by Horton; (c) the idea of a *shared practice*, as essentially claimed by both. Though their accounts of political obligation are distant, I see some common features that I would adopt and even adapt into my proposal. To sketch it out: the shared idea of non-voluntary or at least not-necessarily-voluntary commitment is especially attractive as it helps to figure the relationships among liberal institutions and not-liberal-nor-democratic people. After the reasons or motives they may have to comply with laws, the *practice* of cohabitation works undoubtedly to cement a peaceful community of citizens, be they committed for the "right reasons" or not. Non-liberal-nor-democratic citizens show themselves to be able to comply with liberal-democratic institutions not reluctantly, but as a kind of commitment to behaving politically together with their fellow citizens.

The point then is how to interpret the ideas of a common knowledge, of a sense of identity and of shared practices as further stabilizing modus vivendi. My perspective – but I may just announce it as the next step of this work – is that peace, the pursuit of peace and living in peace represent respectively the common knowledge, the sense of identity and the shared practice we are dealing with. Let us remind why people have always sought a modus vivendi: at least – this is the simplest answer – because they have always aimed at peace and, through peace, they have always aimed at safety. The special place of the goods of peace and safety does not mean that they are for everyone the supreme goods. That is, it is not necessary to conceive peace and safety as our final ends – or ideals – to feel committed to seeking them. People can, and sometimes will, have goals that they set above such goods. They will certainly have other aims as well as securing a stable and peaceful settlement. To speak so means to admit realistically that reasons for peace may include moral principles and general prudential considerations as well (Sleat 2011). It means also that though people (non-liberal-nor-democratic people specifically) join society to attain their purposes for their explicit or even implicit reasons, the outcome is still peace: and that is all we can be assured of in the real world (Horton & Windeknecht 2014). There may be a plurality of reasons why people value peaceful coexistence; some of these will be specific moral reasons internal to specific ways of life, others might be more instrumental. The idea is that

modus vivendi is openly universalistic in accepting any reason to be sustained. As Sleat emphasizes,

> if we were to pursue modus vivendi politics as a way of developing liberal realism, we would do well to disconnect it from any controversial justification that would only serve to replicate the disagreements about the normative foundations of politics that the realist challenge highlights. [...] What is potentially attractive about the politics of modus vivendi is that it allows us to overcome the realist challenge by finding common ground on minimal normative commitments, in particular peaceful coexistence (Sleat 2011, pp. 488-489).

Be peace and safety final ends or not, such goods have at least instrumental value to almost everyone, as they are an essential precondition for the achievement of almost any other good. If peace may be conceived as a realistic goal, we need now to understand in which sense it should be conceptualized as being possible without consensus on morality (Wendt 2013). This is not an easy task: the notion of peace as non-violent coexistence based on modus vivendi has been accused to be too weak to be an attractive ideal (Rawls 1993). In response to this objection, pursuing a demanding but realist conception of peace, we need a sort of consensus, but not a substantial moral consensus. A non-moral consensus may be conceived as a compromise that is an agreement reached by motivations, rather than reasons, understandable as a genuine willingness to abide. We can keep silent on the kind of reasons for accepting a compromise: I dare say that sometimes there are no reasons, that is, explicit or conscious reasons, at stake. We have no reasons to exclude other non-moral commitments, perhaps shaped by jointness, by the mere fact to join a society. What compromise entails is a long debate, but it exceeds the length of this brief contribution.

REFERENCES

Dagger, R. & Lefkowitz, D. (2014), "Political obligation", *The Stanford Encyclopaedia of Philosophy* (Fall 2014 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/fall2014/entries/political-obligation/;

Gilbert, M. (2006), *A Theory of Political Obligation*, Oxford, Oxford University Press;

----. (2013), *Joint Commitment*, Oxford, Oxford University Press;

Horton, J. (2006), "In defence of associative political obligations. Part one", *Political Studies*, 54, pp. 427-443;

----. (2007), "In defence of associative political obligations. Part two", *Political Studies*, 55, pp. 1-19;

----. (2010), "Realism, liberal moralism and a political theory of modus vivendi", *European Journal of Political Theory*, 9, pp. 431-448;

----. (2010b), *Political Obligation*, New York, Palgrave MacMillan;

----. (2012), "Political legitimacy, justice and consent", *Critical Review of International Social and Political Philosophy*, 15, pp. 129-148;

Horton, J. & Windeknecht, R. G. (2014), "Is there a distinctively associative account of political obligation?", *Political Studies*, doi: 10.1111/1467-9248.12129;

Newey, G. (2010), "Two dogmas of liberalism", *European Journal of Political Theory*, 9(4), pp. 449-465;

Quong, J. (2011), *Liberalism Without Perfection*, Oxford, Oxford University Press;

Rawls, J. (2001), *Justice as Fairness: a Restatement*, Cambridge (MA), Harvard University Press;

Rawls, J. (2005), *Political Liberalism*, New York, Columbia University Press;

Rossi, E. (2010), "Modus vivendi, consensus and (realist) liberal legitimacy", *Public Reason*, 2(2), pp. 21-39;

----. (2013), "Consensus, compromise, justice and legitimacy", *Critical Review of International Social and Political Philosophy*, 16(4), pp. 557-572;

Rossi, E. &Sleat, M. (2014),"Realism in normative political theory", *Philosophy Compass*, 9/10, pp. 689-701;

Sala, R. (2013), "The place of unreasonable people beyond Rawls", *European Journal of Political Theory*, 12(3), pp. 253-270;

Scheffler, S. (1994), "The appeal of political liberalism", *Ethics*, 105(1), pp. 4-22;

Sleat, M. (2011), "Liberal realism. A liberal response to the realist critique", *Review of Politics*, 73(3), pp. 469-496;

Wendt, F. (2013), "Peace beyond compromise", *Critical Review of International Social and Political Philosophy*, 16(4), pp. 573-593.

HELEN LAUER
University of Ghana, Legon, Ghana
helenlauer@yahoo.com

THE ANATOMY OF COLLUSION*

abstract

Received models of cooperative action overlook the way that the social environment itself features in collective intentions; recognising the divergent ways that norms function can help explain how people collaborating covertly are not considered irrational even though the things they are doing are totally contrary to the principles and convictions they sustain as individual agents.

Introduction   "Collusion" is often used to express opprobrium when describing couples or groups cooperating covertly or secretly to achieve aims that those agents themselves may recognise as morally dubious. Collusion of the sort analysed here is inherently social: the individuals involved would not venture to pursue such ends entirely on their own, because the particular strategies employed could not be executed rationally except as joint efforts or through surreptitious teamwork, nor would these endeavours be reasonable to attempt without the expectation of cooperation or tacit consent of others. Collusion occurs within and across hierarchical divisions; for example suppose the candidacy of a worthy applicant with short stature is undermined by a secretary's intentionally misplacing the applicant's dossier and then disregarding its absence, because the secretary is colluding passively with her boss, whose indefensible aversion to short employees is signalled surreptitiously through pointedly timed grimacing or sighing. The secretary's vertical collusion becomes assertive when she initiates, of her own accord, preparation of the interviewing schedule in a way that delays attention to an excellent short candidate until after the deadline, and prioritises the application of the boss's tall nephew whose skill set is grossly inadequate. The search committee chair, aware of this manoeuvre, colludes vertically and passively by doing nothing to avert the sabotage. Then he colludes horizontally and assertively with other committee members by orchestrating their conduct during interviews in a way that undermines the performance of the short applicant with a barrage of intimidating questions and humiliating banter. The committee's scrutineer, witnessing the discriminatory treatment, colludes passively by not reporting it. In consequence of the collusion overall, the short applicant's prospects are defeated.

My aim here is to illuminate and account for the apparent divergence of collusion from other sorts of rational collaboration, but not so as to cast it in a unique light or as an isolated category of collective endeavour. On the contrary, I hope to show how the social environment itself may contribute directly to rationally cooperative intentional behaviour[1] of many sorts, and how the environment in turn gets shaped by accretive repeated episodes of dogmatic

---

1 Throughout this discussion I use the terms "intention" and "intentional" always and only to connote episodes of action that are motivated or attempted in ways that would be wrong to characterise as unwitting, reflexive, automatic, absent-minded. Of course it does not follow that in this sense of acting intentionally, an agent can intend or know everything or even anything that actually comes about as the causal effect of what he or she thinks or intends to be doing.

norm-following (both covert and overt). Understanding the intentionality that underlies collusion may help to illuminate socially acceptable hypocrisy, and to explain why sometimes it makes so little difference to the recalcitrance of a social injustice whether the individuals who perpetuate it approve explicitly of doing so or not. The considerations assembled here may also help to account for the typically sluggish pace of a social reform despite its enjoying ostensibly wide commitment and support.

The account proposed follows from suggestive leads of researchers in cognition, anthropology (Burge 2014; Cohen 1981, 1992; Cancian 1975), stigmergic epistemology, swarm theory (Cazangi et al. 2006; Marsh and Onof 2007; Valckenaers et al. 2006) and Amedeo Conte's notion of nomotropism as it has been applied to urban policy analysis (Chiodelli and Moroni 2014).

Models of collective action received over the last three decades have characterised individual agents participating either in (a) large scale, ongoing, institution-dependent coordination – the stockmarket or football tournaments, or in (b) small scale, episodic collaborations – taking a walk, painting a room.

The content of beliefs constitutive of collusion

For instance "structured social groups" (Bratman 1993, p. 98) feature participants' knowledge of institutionalised norms, customs, traditions, and bureaucratic procedures as essential to the kind of habituated and ritual activity that "contribute to the maintenance and renewal of the institutions involved" (Tuomela 2003, p. 159). Examples include "teleconferences, the stock market" (Kutz 2000), academic departments (Bratman 1993), team sport tournaments (Turner 2003), groups that annually go berry-picking (Tuomela 2003). Some theorists argue that these activities require the occurrence of specific kinds of "group commitment" (Tollefsen 2002), or propositional "we-attitudes", "we-beliefs", or "we-modes" of intention (Tuomela 2003, pp.153, 162). Others posit that these large group processes entail the existence of special "plural subjects" (Gilbert 1990, p. 9, 2014).[2]

In stark contrast, theorists have analysed smaller scale projects which "are not embedded in institutional structures and authority relations" (Bratman 1993, p. 98). Examples of such coordination include a couple's taking a walk (Gilbert 1990; Velleman 1997) or painting a house (Bratman 1993), or investigating a murder (Gilbert 1990). Only some of these accounts require a lattice of agents' immediate "interactive knowledge" that directly acquaints and associates all the participants involved with each other's intentions and expectations (Bratman 1993, p. 109, 1999; Chant and Ernst 2008; Gilbert 1990 and 2014; Kutz 2000, p. 6; Searle 2008; Tuomela 2003). On other accounts, participants are attributed with knowledge about the occurrent events that they actually achieve through their coordinated "sub-plans" (Bratman 1993, p. 105) or through their "participatory intentions" (Kutz 2000, p. 20). Principal agents' knowledge is deemed essential according to these models in order to capture just those cases where an outcome is achieved through genuine collaboration, rather than coercive subjugation (Bratman 1993, pp. 104-106), or as the mere coincidental effect of a confluence of simultaneously occurrent but "unshared" intentions (Kutz 2000, p. 20, n. 41).

As different as their views are from one another, the sorts of clandestine collaboration featured in this essay are not captured straightaway by these theorists. Contrary to the models depicting cooperation on a massive scale – like football tournaments and the European Union – the number of passive or active agents participating in a covert operation is usually small. Nonetheless collusion definitely belongs among the institutionally-dependent collaborations that require of participating agents an intimate, working knowledge of the bureaucratic

---

2 Margaret Gilbert (1990) regards plural subjects as presupposed by any joint activity, not just those whose participants identify themselves as belonging to a structured group.

arrangements and structures which they exploit. Yet unlike the essential allegiance to norm-following attributed to collaborators in models of the large scale (like teleconferencing, football tournaments and the European Union), the mark of colluders is to presuppose a prevailing habit of fidelity to protocol which they themselves flout. To be feasible, collusion relies upon the resiliency of the very institutional structures which participants in the collusion betray (actively or passively). For example, it is only because colluding committee members know that job interviews are generally regarded as reliable and fair threshold indicators of professional demeanour that they could rationally expect to destroy an applicant's chances by surreptitiously sabotaging his interview performance. Similarly, it is only because of the secretary's astute know-how in the conventions of office procedure, her confidence in co-workers' adherence to office protocol, and her belief that she enjoys tacit respect in the office environment, that she can rationally assume the power to disqualify or dissolve an unsuspecting applicant's competitive edge by tampering with his interview date. On the other hand, in the models of group projects on a smaller scale (like painting a room or taking a walk together) some theorists assume that collaborators must have each other's intentions and expectations in view, in order to capture their activities as joint efforts (Bratman 1993; Kutz 2000). But with respect to passive colluders it would be wrong to suppose that knowing what to do must involve caring or having beliefs about the specific expectations and sub-plans of other individuals involved in a collective covert operation at a particular time, or indeed to hold impersonal generalisations about others involved in a collusion. Consider the overseeing scrutineer who played the most crucial part in the sabotage of the short interviewee. He knew that looking the other way was just what he was expected to do, but not because he held true propositional beliefs describing any specific expectations or intentions of those individual committee members who orchestrated the irregularities which occurred at that particular interview, nor because he held more general beliefs about his co-conspirators as a group. This point applies just as well to those who pro-actively initiate a collusion. The secretary who undermined one short interviewee's chances by perverting his appointment date, and who destroyed the application process of another by suppressing his file, executed her sub-plans with the sole intention of winning a promotion for herself and a permanent place in her boss's good books. Even if her co-workers failed to sideline these short applicants, or were never able to get the tall nephew hired, the outcome of these specific collusions that she herself intentionally helped to instigate need not have mattered to her; for in the event of their failure, she believed she would try something else to win the boss's favour. Indeed, she might rationally sustain a calculated disinterest in those outcomes, to absolve herself of blame for any untoward consequences of the office staff's covert mischief. Toward this end, it would be rational for her to remain wholly indifferent to the interests of other colluders – either individually or impersonally as a group – upon whose cooperation the success of her part in the subterfuge was fully dependent (contra Bratman 1993 and Gilbert 2014, p. 66). Nonetheless commitment to collaboration is the essence of successful collusion: given the antecedent conditions in this case, a highly qualified short applicant getting rejected and the inept nephew's getting hired would not be interpreted correctly as the convergent effect of merely coincidental events resulting from independent parties simultaneously pursuing their isolated self-interests.

Yet that is exactly how a third analytic approach would treat these disappointments and rejections: not as the orchestrated handiwork of collective behaviour of the large scale type (football matches), or as the one-off small scale type (solving a crossword puzzle). Instead, some nominalistic theorists depict the unfortunate happenstances of agents routinely disadvantaged in a biased meritocracy as being mislabelled when they are cited as the outcome of specific cases of group discrimination. These nominalists object that

the complaint is based upon misleadingly aggregating stochastic evidence of coincidental occurrences. Michael Levin (1981) attempts to demonstrate that losing the prize, no matter how often, is the probabilistic outcome of discrete incidents where fair competition occurs between self-interested individuals operating in isolation, legitimately relying upon their own atomised potentials gained through rightfully earned or inherited personal virtues and capacities. Levin's account of manifest inequities in a biased meritocracy obfuscates the causal link between prevailing effects of apparent patterns of inequity and instances of successful collusion which do involve individuals' "intentional participation". It is not readily obvious that the intended goals involved must constitute a specifiable intention of a "plural subject" (Gilbert 1990, 2014). Nor is it obvious that participating intentionally in collusion requires individuals' personal commitments to a shared "group-intention" (Kutz 2000, pp. 7, 22, 24). People in collusion may be committed to an outcome that none of them would pursue without the others' commitment or endorsement, yet not because they share thoughts with any specifiable content which constitutes the object of shared commitment or tacit consent. Nor would it be rational for colluding agents to be prepared in every case to mutually *express* or otherwise to concede what they are up to (contra Gilbert 2014, p. 119). To show this, first we need to consider in more detail how social facts may function directly in causal sequences constituting rational collective agency. In particular, we need to appreciate the variety of ways that a collective intentional action is gauged or calibrated as rational in light of many different kinds of norms. These considerations independently support Chiodelli and Moroni (2014) in replacing the familiar concept of "rule-following" with the more inclusive trait "nomotropic" (so dubbed by A. Conte 2012) to characterise an episode of behaviour as rational because the intention to act has taken into account existing rules, norms, formal statutes, conventions and protocols – and here quoting Chiodelli and Moroni: "without necessarily[...] following or acting in conformity with them" (2014, p. 162).

Momentarily I will spell out the variant ways that beliefs may contribute to the reasoned intentionality of collaborating agents, for which I introduce a "collective utility" function that warrants a belief based on its popularity rather than its probability. The label "collective utility" is meant to dub just one of a variety of norms of reason that work in tandem. Recognising that a variety of norms is effective in causing collaborative intentions allows us to explain how rational agents can embrace dogmatic precepts which are patently false. And it allows us to explain how fully rational, enkratic people can do things collectively that blatantly confound the values they profess sincerely as individuals.

To account for this apparent paradox consider that an outcome, which in fact issues from acting collectively in light of prevailing norms, is not always a matter which can be read off from any description of the norm itself. Nor can the effect of a norm follow by inferring from the propositional content of any agent's thoughts about what that norm dictates. The causal effects of a norm are indeterminate from the point of view of fallible agents. One has to consider the empirical circumstances particular to each situation at hand independently of what agents know, to learn retrospectively the outcome of any instance of norm following.[3] We know this from the moral of Chisholm's story about deviant causal chains. But norms can be causally affective in ways not foreseeable, because they affect behaviour perversely. For instance in the story just told, the colluding individuals anxious to protect their job security

---

3 As Francesco Chiodelli and Stefano Moroni say, "case by case" (2014, p. 163). With reference to Di Lucia (2002) they observe (2014, p. 162) "that there are in fact diverse ways in which a rule may causally influence an action". For a trivial example, consider that young people in modernity habitually defy conservative dress codes. So nowadays in allegiance to the dictates of covert prestige, young men walk around with their trousers falling down – a practice which counts as rational only insofar as it flouts the convention of securing trousers snugly around one's waist.

diligently took account of the requisite protocols constituting their company's recruitment procedure. But they did *not* do so in compliance with the formally, ostensibly intended purposes of the company policy. They did so in order to mimic compliance, hoping for a quite different outcome than the one prescribed by statutory protocol. As for the secretary, only by assiduously contouring her handling of applicants' portfolios in light of statutory rules was she effective in meeting her own ends, commensurate with the covert in-house legacy of excluding short people – the mischief was made possible because of, yet in defiance of, the very antidiscrimination measures that her company sustains to ensure its hiring record remains beyond rebuke. This example demonstrates colluding agents within a complex social environment – characterised by Chiodelli and Moroni following McFarlane (2012) as a "meshwork" or "entanglement" (2014, p. 162) of covert and overt, formal and informal, norms, rules, and conventions.

L. Jonathan Cohen (1981, p. 321; 1992) has argued for a comprehensive treatment of cognitive attitudes that distinguishes between our *believing* an expressible proposition, and our *accepting* something as a policy in the course of practical reasoning. To this we can add Chiodelli and Moroni's (2014) depiction of the Contean nomotropic nature of rationality. Collusion seems to require agents collectively *countenancing* a policy without abiding by or conforming to that policy, but rather by acting in defiance or in resistance to it, or in divergence from it. But the effect would be impossible to achieve, indeed it would be impossible to conceive, were it not for the causally affective role of the policy.

Treating covert norms as features of the social environment

We have just observed that the rational coherence of professing a principle T while colluding with others in behaviour that defeats T, rests on the fact that an agent may make choices that are causally affected by prevailing norms without necessarily formulating propositional beliefs about doing so. Cohen (1981) has shown how testable theories of cognitive competence presuppose that logical norms – not just intuitions or facsimiles of them in the language of thought – are attributable directly to the episodic reasoning of individual human subjects engaged in problem-solving. Following from empirical results such as these, consider that social norms themselves can enter into one's rational choices as well, bypassing truth-functionally contoured beliefs. A social norm, or convention, or perceived expectation of others, can indicate what to think and how to behave without one's formulating distinguishable propositional beliefs about the information conveyed.

Suppose that in conformity with the custom of keeping short people out of my neighbourhood and office, I intentionally refuse to share my work schedule or to rent my house to an applicant because she is short – not because I dislike short people or believe that short people are lazy or that they make bad neighbours. Instead my refusal expresses my concern to be regarded as a good neighbour or employee, together with my awareness that – where I live or work – being a good member of the neighbourhood or the office team entails deterring short people from acquiring residence or employment. This is an assessment I make about my neighbourhood or my workplace itself; it need not be a belief that refers to any specifiable neighbour or colleague, nor to any differentiable group existing as an entity distinct from the social environments that we happen to share. Sometimes, knowledge that contributes essentially to a successful collusion may be about facts constituting the social environment in which the collusion is executed – yet not about the agents engaged in the collusion. In a different part of the country, or at a different workplace or given a different global economic structure, the same neighbours or office staff might well behave very differently towards short people. This consideration suggests that the causal or other determinacy entailed by intending to collude is in some sense "context sensitive" (Gilbert 2014, p. 122) – yet without necessarily being sensitive to facts about the particular other agents whose cooperation is essential to one's plan.

In formulating and carrying out my intention to nonverbally convey that I am a good neighbour or co-worker, I need not believe an articulable proposition that shunning short people and facilitating nepotism have the desirable characteristics of increasing one's popularity and chances of promotion. If I use speech instead to convey the same message, I need not have any beliefs about the sounds or movements I am making in my larynx. I just go ahead and produce the phonemic sounds as I have learned to make them in the appropriate speech context. Similarly I may execute an unmistakable rebuff by turning my back abruptly upon receiving a direct query or request from a short person, without holding a specifiable propositional belief or value judgment about what I think I am doing. Thereby my gesture of disdain exhibits the image of myself that I intended, and in the process I am participating in the collusion required to keep short people at bay. I do a bit more than that; my exhibition reinforces the status quo of the neighbourhood and office environment that I came to meet, wherein it is acceptable – indeed, expected – to exclude short people. Together with my avowed convictions that our company is an equal opportunity employer, and that people of short stature are just the sort with whom I want to work and send my children to school, I commit no *logical* inconsistency through my intentional display of rudeness or my covert sabotage, because the component thoughts comprising my intentionally nasty behaviour need not carry properties that could enter into logical relations with each other or with the propositional content of my sincere commitment to principles of fairness, civil rights and equal opportunity.

Using L. Jonathan Cohen's (1992) distinction, these sketches suggest that I may accept *as a policy* – or as a covert norm guiding my behaviour – the correlation between ostracising short people and being a good neighbour or co-worker, without believing any proposition which expresses that correlation. In this respect, Hector-Neri Castañeda contrasts propositions with "practitions" (1975, pp. 131-141, 207-208, 240, 331-332). Endorsement of a behavioural pattern or norm as rational is always relative to distinct "contexts of legitimacy" that preclude the agent's individual encounter with what she is doing or with the norms she is following (or flouting) in their propositional formulation.[4] The force of legitimacy of a norm may vanish with a change in the social environment, not because anyone has changed their propositional beliefs about short people. For example the covert legitimacy and so too the incentive for shunning short people may disintegrate if there is a change in the economy and there emerge plenty of jobs to go around.

We noticed already that it can be the factual features of the social environment itself that a rational agent must negotiate and which impel conformity on rational grounds. And we saw that this need not involve the agent's own propositional thoughts and judgments *about* those facts. Typically when a person stops en route to where he is going in order to avoid passing in front of a moving vehicle, it is because he fears getting hit by a moving vehicle, not because he fears the experience of getting hit by a moving vehicle. Conversely, it is unlikely that by correcting a foolish belief about short stature I will be sufficiently equipped either to detonate or to bypass or to impede the effects of a dogma that renders it normal and convenient to comply with prevailing covert practices which ostracise short people.

Insofar as an agent's intention is to conform to the dictates of political fashion by doing "the done thing", be it covert or otherwise, cooperative behaviour counts as consummately sensible, demonstrating that social facts themselves can be dominant in the primary reason

---

4 Romane Clarke (1980) has neatly laid down the basis for showing that the thoughts comprising an agent's intentions which might participate in a causal sequence cannot all be distinguishable in propositional form without threat of paradox.

for acting intentionally, and covert norm-following is such a fact. So consider colluding as a rational response to prevailing covert norms. Then since a rational agent need not approve of the propositional formulation of a norm in order to follow it, colluders may be causally contributing to the perpetuation of a certain kind of social environment without committing themselves to any truth-functional thoughts about doing so. Covert norm-following may sometimes entail acceptance of a policy which provides the agent a rationally discernible advantage – not because the content of beliefs summoned as the overt basis for that policy are *true*, nor because such beliefs are derived from the agent's avowed principles for right living, but because those ratifying beliefs are *popular.*

Alasdair MacIntyre (1986, pp. 66-69, 79) details how normality of a social environment precludes reason-giving for the bulk of our intentional actions. I must be aware of a social norm, whether it be official or illicit, in order to follow it or to circumvent it or to defy it intentionally. But the content of my awareness of how to do this may be images of activity, or slogans, or emotional states of relative tension or relaxed ease learned through the body language and facial expressions of parents and playmates. People often know how to do what is expected without ever being told explicitly. This may help to explain why it appears to cause so little conventional stir when, as individuals, we avow our belief in principles and personal values on Sunday that diverge from what we practice throughout the week. Rarely is such inconsistency assessed as irrational. Perhaps in a better world it would be; perhaps in this world it should be; but usually it is not.

In and of itself, the social fact that a belief is widely held can be a reliable guide for individual agents choosing to adhere to it, subscribing to "the way things are". A belief whose cognitive value rests not on its truth functional content but on its stochastic frequency of occurrence throughout a population may be referred to as *dogmatic.* Perhaps analogously to pheromones (the olfactory deposits) chemically effecting a social insect's environment, complex phoneme strings may function as intangible fixtures in the social spaces in which we live and work, constituting a plethora of slogans and stereotypes, prejudices, myths, interwoven with (and sometimes masquerading as) epistemically justified beliefs.

Dogma variously provides protection and cohesion, political leverage, consumer incentive, competitive advantage, consolidating influence, solace and social recognition. If dogmatic beliefs form part of the linguistic architecture that shapes the locality in which agents make plans, set goals, and collaborate, then typically an individual's rational intention to conform is composed of both her personal preferences and also dogmatic posits knitted into the immediate social environment upon which her existence at least partly depends. This collective inter-knittedness can be depicted without attributing preferences or commitments to social group entities as specially designed ontological referents.

Explaining the recalcitrance of the social status quo

To spell out more formally the utility of accepting a dogmatic belief, consider a reactionary neighbourhood group G notorious for colluding against short people. Being a hard-nosed conservative as my membership in G attests, I accept the belief P (that short people are inherently inferior); but I do so as a die-hard bigot, i.e. my warrant for believing that P holds only to the extent that I thereby endorse the views of at least one other member of my group G. If no one else believed P, then I would have no rational incentive or warrant for subscribing to P myself. And suppose all members of G use this same standard for assigning subjective credence or collective utility values to their dogmatic beliefs. This would not be the case for other beliefs attributable to members of G, of course, only for those beliefs that might be called dogmatic. To call P a dogmatic belief implies that individual G-members uphold P provided they perceive it to be the consensus view of G, not because P is true. In appraising P for its utility as a "membership norm" (Cancian 1975, pp. 143-147, 152), its truth value need not enter

into the assessment. As a member of G, the only evidence I ever need to warrant my accepting P is the consensual status I have conferred upon it; in other words P is part of the dogma I embrace because doing so affirms my group identity and solidarity. Because I need not have any propositional beliefs about a policy in order to act in light of it, the same sound string or sequence of marks can function to signal a membership norm and also as a propositional belief content. Hence I can act in accord with the slogan that all short people are unscrupulous and undermine office productivity; while simultaneously embracing the belief that my maid, who is short in stature and looks after my children and my silver, is trustworthy and industrious. P bears a value I am labelling its "pure" collective utility if there is *no* incentive for individual members of G to accept P other than the fact that they believe other members do. In the absence of new evidence relevant to P pro or con, the pure collective utility of P accumulates merely with the passage of time, as members of group G who share an acceptance of slogan P also meet and interact with each other without cognitive dissonance.[5] If P serves in this way as bonding material to enhance and strengthen membership in group G, then the (objective) factual status of P need not contribute directly to calculating its collective utility. On the contrary, if there is any relation between P's Bayesian probability and its value in terms of collective utility, it might be inversely proportional: the *less* there is of neutral and impartial evidence available in P's favour, the more its collective utility increases. P's collective utility increases as the objective counterevidence and growing opposition mounts in its disfavour, because the more improbable P is at face value, the better it serves as a testament to the group-fidelity of a member who accepts it. And so, generally, the more improbable P appears, the better it serves to indicate the resilience of G's group solidarity demonstrated by G-members' adherence to P in the face of all opposition.

Thus, dogma is tenacious; but it is also fickle. Being a good neighbour or office mate – like being a good American – may entail despising short people one year and courting their favour the next. The membership norms of my group will change as people sense that others have a new image of what good G-neighbours and employees do, not because anyone has learned new facts about short people.

If the considerations assembled here are tenable, they should help us to analyse how social injustice gets tolerated and perpetuated through covert means by members of privileged groups who individually regard principles of fairness and equity as sacrosanct. This is not to say that membership norms are indelible, nor that collusion is rooted in genomic preference, nor that historically inherited social norms are evolutionarily optimal. On the contrary, this analysis suggests that correcting unjust and inequitable collaboration requires a change in the social architecture comprised of linguistic noise influencing "normal" conduct and partly shaping the rational choices that individuals find advantageous.

REFERENCES

Bratman, M. E. (1993), "Shared intention", *Ethics*, 104(1), pp. 97-113;

Burge, T. (2014), Abstract for "Perception and default: Warranted perceptual belief" Master Class, *Philosophy: Language, Mind and Practice,* Doctoral Programme, Institute of Philosophy, Zurich, Switzerland, October 24-25; http://www.philosophie.uzh.ch/doktorat/meisterkurse.html;

Cancian, F. (1975), *What are Norms? A Study of Beliefs and Action in a Maya Community*, Cambridge University Press, New York;

---

5 This analysis of dogmatic beliefs is consonant with Iñiguez, et al. (2014) who analyses in detail the social utility of different species of lying.

Casteñeda, H. N. (1975), *Thinking and Doing: The Philosophical Foundations of Institutions*, Reidel, Dordrecht;

Cazangi, R. R., et al. (2006), "Stigmergic autonomous navigation in collective robotics", *Studies in Computational Intelligence*, 31, pp. 25-63;

Chant, S. R. & Ernst, Z. (2008), "Epistemic conditions for collective action", *Mind*, 117(467), pp. 549-573;

Chiodelli, F. & Moroni, S. (2014), "The complex nexus between informality and the law: Reconsidering unauthorised settlements in light of the concept of nomotropism", *Geoforum*, 51, pp. 161-168;

Clarke, R. (1980), "Not every act of thought has a matching proposition", in P.A. French (ed.), *Midwest Studies in Philosophy* V: *Studies in Epistemology,* University of Minnesota, Minneapolis, pp. 509-524;

Cohen, L. J. (1981), "Can human irrationality be experimentally demonstrated?", *Behavioral and Brain Sciences*, 4, pp. 317-370;

----. (1992), *An Essay on Belief and Acceptance,* Clarendon, Oxford;

Conte, A. G. (2012 [2011]), "Nomotropismo" repr. in G. Lorini & L. Passerini Glazel (eds.) *Filosofie della norma*, G. Giappichelli Editore, Torino, pp. 307-316;

Di Lucia, P. (2002), "Efficacia senza adempimento", *Sociologia del diritto,* 29, pp. 73-102;

Gilbert, M. (1990), "Walking together: A paradigmatic social phenomenon", in P.A. French (ed.) *Midwest Studies in Philosophy* VI: *The Foundations of Analytic Philosophy*, pp. 1-14;

----. (2014), *Joint Commitment: How We Make the Social World,*Oxford University Press, Oxford;

Iñiguez, G. et al. (2014), "Effects of deception in social networks", Accessed July 25, 2014. <http://arxiv.org/ftp/arxiv/papers/1406/1406.0673.pdf>;

Kutz, Ch. (2000), "Acting together", *Philosophy and Phenomenological Research*, 61(1) July, pp. 1-31;

Levin, M. E. (1981), "Equality of opportunity", *Philosophical Quarterly*, 31(123) April, pp. 110-125;

MacIntyre, A. (1986), "Intelligibility of action", in J. Margolis, M. Kraus & M. Burian (eds.), *Rationality, Relativism and the Human Sciences*, Nijhoff, Dordrecht, pp. 66-79;

Marsh, L. & Onof Ch. (2007), "Stigmergic epistemology, stigmergic cognition", *Cognitive Systems Research*, doi: 10,016/j.cogsys.2007.06.009;

McFarlane, C. (2012), "Rethinking informality: politics, crisis, and city", *Planning, Theory and Practice*, 13(1), pp. 89-108;

Searle, J. R. (2008), "Language and social ontology", *Theory and Society*, 37(5), pp. 443-459;

Tollefsen, D. (2002), "Collective intentionality and the social sciences", *Philosophy of the Social Sciences*, 32(1), pp. 25-50;

Tuomela, R. (2003), "Collective acceptance, social institutions, and social reality", *American Journal of Economics and Sociology*, 62(1), pp. 123-165;

Turner, S. P. (2003), "What do we mean by 'we'?", *ProtoSociology*, 18-19, pp. 139-162;

Valckenaers, P. et al. (2006),"Emergent short-term forecasting through ant-colony engineering in coordination and control systems", *Advanced Engineering Informatics*, 20, pp. 261-278;

Velleman, J. D. (1997), "How to share an intention", *Philosophy and Phenomenological Research,* 57, pp. 29-50.

50,00 €