Savas L. Tsohatzidis
Editor

Theory and Decision Library A

Intentional Acts and Institutional Facts

Essays on John Searle's Social Ontology

Springer

INTENTIONAL ACTS AND INSTITUTIONAL FACTS

ESSAYS ON JOHN SEARLE'S SOCIAL ONTOLOGY

THEORY AND DECISION LIBRARY

General Editor: Julian Nida-Rümelin (*Munich*)

Series A: Philosophy and Methodology of the Social Sciences

Series B: Mathematical and Statistical Methods

Series C: Game Theory, Mathematical Programming and Operations Research

SERIES A: PHILOSOPHY AND METHODOLOGY
OF THE SOCIAL SCIENCES

VOLUME 41

Assistant Editor: Thomas Schmidt (*Göttingen*)

Editorial Board: Raymond Boudon (*Paris*), Mario Bunge (*Montréal*), Isaac Levi (*New York*), Richard V. Mattessich (*Vancouver*), Bertrand Munier (*Cachan*), Amartya K. Sen (*Cambridge*), Brian Skyrms (*Irvine*), Wolfgang Spohn (*Konstanz*)

Scope: This series deals with the foundations, the general methodology and the criteria, goals and purpose of the social sciences. The emphasis in the Series A will be on well-argued, thoroughly analytical rather than advanced mathematical treatments. In this context, particular attention will be paid to game and decision theory and general philosophical topics from mathematics, psychology and economics, such as game theory, voting and welfare theory, with applications to political science, sociology, law and ethics.

The titles published in this series are listed at the end of this volume.

INTENTIONAL ACTS AND INSTITUTIONAL FACTS

Essays on John Searle's Social Ontology

Edited by

SAVAS L. TSOHATZIDIS

Professor of General Linguistics and the Philosophy of Language
Aristotle University of Thessaloniki
Greece

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4020-6103-5 (HB)
ISBN 978-1-4020-6104-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved
© 2007 Springer
No part of this work may be reproduced, stored in a retrieval system,or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

CONTENTS

PART III: ASPECTS OF INSTITUTIONAL REALITY

CONTRIBUTORS

Professor ***Margaret Gilbert***, Department of Philosophy, University of California, Irvine, California, USA

Professor ***Kirk Ludwig***, Department of Philosophy, University of Florida, Gainesville, Florida, USA

Professor ***Steven Lukes***, Department of Sociology, New York University, New York, USA

Professor ***Anthonie Meijers***, Subdepartment of History, Philosophy and Technology Studies, Eindhoven University of Technology, Eindhoven, The Netherlands

Professor ***Seumas Miller***, Centre for Applied Philosophy and Public Ethics, Australian National University, Canberra, Australia

Dr ***Hannes Rakoczy***, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Professor ***Ignacio Sánchez-Cuenca***, Center for Advanced Study in the Social Sciences, Juan March Institute, Madrid, Spain

Professor ***John R. Searle***, Department of Philosophy, University of California, Berkeley, California, USA

Professor ***Barry Smith***, Department of Philosophy, State University of New York at Buffalo, Buffalo, New York, USA

Professor ***Michael Tomasello***, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Professor ***Savas L. Tsohatzidis***, Department of Linguistics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Professor ***Robert A. Wilson***, Department of Philosophy, University of Alberta, Edmonton, Alberta, Canada

Professor ***Leo Zaibert***, Department of Philosophy, University of Wisconsin, Parkside, Wisconsin, USA

INTRODUCTION

Savas L. Tsohatzidis

John Searle is famous for his contributions to two fields with long and distinguished traditions within analytic philosophy—the philosophy of language and the philosophy of mind—, but his interests and achievements extend beyond these fields. From the early 1990s he has added to his research agenda a theme that was not only largely new to his philosophical preoccupations, but also largely absent from the concerns of analytic philosophy as a whole: the systematic examination of the mode of being of a particular kind of facts, *institutional* facts, that appear to be no less *objectively* knowable than ordinary physical facts, yet seem to be essentially dependent for their existence on the *subjectivity* of human minds (to recall one of his favourite examples, one can know that something is a piece of paper as objectively as one can know that it is a twenty-dollar bill, but something's *being* a piece of paper does not depend on anyone's *taking it* to be a piece of paper, whereas its *being* a twenty-dollar bill crucially depends on a lot of people *taking it* to be a twenty-dollar bill). Searle's attempt to give a systematic account of the combination of epistemic objectivity and ontological subjectivity that, in his view, characterizes institutional facts has led to a full-blown theory that he presented in his 1995 book, *The Construction of Social Reality*, and further developed in his 2001 book, *Rationality in Action*. The present work, prefaced by a new contribution by Searle, is the first book of original essays specifically devoted to the critical examination of central aspects of Searle's theory of institutional facts, whose interdisciplinary relevance is evident from the extensive attention it has already received in specialist journals representing an unusually wide range of academic subjects.[1]

According to Searle, institutional facts are a subclass of social facts, and social facts are all and only those facts that are manifestations of *collective*, as distinct from individual, intentionality. A proper account of institutional facts, then, could not, in Searle's view, be developed unless a satisfactory account of the nature of collective intentionality could become available. That larger topic was for independent reasons coming to the centre of attention of some analytical

[1] Symposia on *The Construction of Social Reality* have been published in *Philosophy and Phenomenological Research*, 57(2): 1997; *History of the Human Sciences*, 10(4): 1997; *Philosophy of the Social Sciences*, 28(1): 1998; and *Journal of Economic Methodology*, 9(1): 2002. Special issues devoted to the book have been published in *The American Journal of Economics and Sociology*, 62(1): 2003 and in *Anthropological Theory*, 6(1): 2006.

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 1–10.
© 2007 Springer.

philosophers at about the time when Searle's interest in the nature of institutional reality emerged, and Searle has contributed to discussions of that larger topic by offering a distinctive and controversial view of collective intentionality that he later used as the foundation of his theory of institutional facts. The critical examination of that view of collective intentionality, and its relation to rival views, is the object of the first part of the volume, 'Aspects of Collective Intentionality'.

Although the capacity for collective intentionality, and so for the creation of *social* facts is, according to Searle, a capacity that many kinds of animals besides humans possess, the particular form of collective intentionality that is deployed when a specifically *institutional* fact is brought into existence is, in his view, a form of collective intentionality that only human animals are capable of deploying (since, among other things, it essentially involves symbolization abilities of the sort that only human language clearly exemplifies). In Searle's view, then, institutional facts, unlike other kinds of social facts, are in an important sense *uniquely human*. Consequently, if his account of them is right, the processes leading to their creation and maintenance should be expected to figure prominently in an adequate account of human evolution and development. The second part of the volume, 'From Intentions to Institutions: Development and Evolution', examines that aspect of Searle's theory in the naturalistic spirit that its evaluation demands, and that his own formulations clearly invite.

The distinctive feature of institutional facts is, according to Searle, that they only exist because they are collectively *recognized* as existing; the collective recognition in question takes the form of the collective acceptance of linguistically expressible constitutive rules through which entities are assigned functions of a special kind, called *status-functions*; and the characteristic effect of the assignment to entities of such status-functions is that the entities in question come to thereby posses certain *deontic powers* that their physical constitution would not, by itself, enable them to possess. (Thus, a piece of paper is a twenty-dollar bill because it is collectively recognized as a twenty-dollar bill, and its being so recognized endows it with certain deontic powers—e.g., the power of being legally exchangeable against valuable commodities—that, considered merely as piece of paper, it would not have.) The third part of the volume, 'Aspects of Institutionary Reality', examines questions raised by the wide-ranging theory of institutional facts that Searle has constructed on that simple basis, in particular the question whether it can give a fully satisfying account of the relevant features of the institutional facts that it recognizes, and whether there are kinds of such facts, or relevant features thereof, that it cannot recognize.

The volume's three parts are preceded by a new essay by Searle that offers a uniquely authoritative and characteristically lucid synopsis of his theory of institutional facts, revisits aspects of its original formulation with the purpose of clarifying their content and implications, and indicates directions that should be explored with a view to its further development. Since that essay is more than sufficient for introducing readers to the details of Searle's position, I shall devote the remainder of this introduction to some remarks on each of the ten

critical essays of the volume and their relations to Searle's work. The remarks follow the order of presentation of the essays in the volume's three parts.

* * *

Though the broad notion of intentionality used by Searle covers all sorts of cases in which an entity is 'about' entities other than itself, the particular variety of intentionality that is typically under discussion when he and others raise the question of the distinctive character of collective as opposed to individual intentionality is the intentionality of action: is there something distinctive that takes place when a group of agents are intentionally doing something *together*, which is more than the sum of what each individual member of the group does? Searle's answer is that there is indeed something distinctive that takes place, but that what is distinctive can be elucidated without admitting that there are any minds other than individual ones, and, furthermore, without supposing that anything external to an individual mind can be constitutive of its contents. All that is required, according to Searle, is to assume that each *individual* mind can contain two mutually irreducible *kinds* of intentions, *I*-intentions and *We*-intentions, which are such that realizing the latter can be an end for which the realization of the former can be a means; two or more individuals will then be doing something *together* when, and only when, they satisfy a type-identical *We*-intention that each of them individually has, by means of satisfying appropriate *I*-intentions, which might or might not be type-identical, that they also have.

The authors of the four essays of the first part of the volume agree with Searle that, when a group of agents are intentionally doing something *together* (when, in other words, there is *joint* or *collective* action at work), there is something distinctive that takes place that is more than the sum of what each individual agent does. They argue, however, that Searle's attempt to specify what is distinctive about joint action in terms of a special kind of individually held *We*-intentions is not successful (either because the *We*-intentions in questions are not necessary or because they are not sufficient for collective action), and develop alternative suggestions as to how a proper analysis should proceed.

Margaret Gilbert argues that there are situations where all the conditions that Searle would take to be definitional of a joint action are met, but where it is evident that no joint action exists, and that, therefore, whether or not Searle's conditions are necessary, they are certainly not sufficient in order for two or more agents to be doing something *together*. Gilbert also suggests that the crucial property of joint actions that is missing from Searle's account is that they require a kind of antecedent *agreement* between their participants, and that it is only if such an agreement is *already* in place that their participants could form (assuming that they do form at all) the sorts of *We*-intentions that Searle's account posits. Since, furthermore, an agreement *between* individuals could hardly be an episode internal to each one of a set of non-interacting minds, Gilbert's suggestion implies that the distinctiveness of joint actions is unlikely to be captured, as Searle and others have been assuming, in purely mental terms,

and that it can certainly not be captured in the explicitly *internalist* mental terms that Searle's account requires.

Kirk Ludwig accepts, with Searle, that it is the internal structure of its participants' intentions that determines the special character of a collective action (and that therefore accounts for the difference between someone's merely doing things that others do as well and someone's doing things *together* with others), but argues that, contrary to what Searle claims, that special character can be acknowledged without having recourse to any conceptual resources other than those required for understanding individual action—and, in particular, without supposing that individual minds contain a special mode of *We*-intending besides the ordinary mode of *I*-intending. Instead of assuming that individuals have these two distinct *modes* of intending, one should only posit, Ludwig argues, a single *I*-mode of intending, and, by exploiting Searle's distinction between the *mode* and the *content* of a mental state, one should capture the difference between situations where an individual merely does things that others do as well and situations where it does things *together* with others as a difference in the *contents* to which individual *I*-modes of intending are attached in the two cases. Assuming that, in both cases, the content to which an *I*-mode of intending is attached is expressible by a plural action sentence (in other words, that, in both cases, the individual intentions involved have the form "I intend that we VP"), the difference between the two cases corresponds, according to Ludwig, to the difference between the 'distributive' and 'collective' readings that plural action sentences can have, and so can be elucidated only if the logical difference between these two kinds of reading is properly analysed. In order to provide such an analysis, Ludwig adopts the influential Davidsonian account of singular action sentences as involving quantification over events, and sets out to investigate how that account should be extended from the singular to the plural case, in a way that would be sensitive to the distinction between distributive and collective readings of plural action sentences. Concerning the distributive reading of a plural action sentence, he proposes that it should be handled by treating the plural subject term as, in effect, a restricted quantifier that takes wide scope over the event quantifier introduced by the action verb. The collective reading of the same plural action sentence can then be generated, Ludwig argues, by reversing the order of the quantifiers, so that it is the event quantifier introduced by the action verb that now takes wide scope over the restricted quantifier represented by the plural subject term. Given this purely structural account of the difference between distributive and collective readings of plural actions sentences, and assuming that such sentences express the contents to which individual *I*-modes of intending are attached in the cases under consideration, Ludwig contends that the difference between an individual's merely doing things that others do as well and an individual's doing things *together* with others can be readily explained. In the former case, the individual realizes an *I*-intention whose singular mode attaches to a content that has the logical form of a plural action sentence *in its distributive reading*. In the latter case, the

individual realizes an *I*-intention whose singular mode attaches to a content that has the logical form of a plural action sentence *in its collective reading*. And that explanation, if correct, makes it superfluous to assume, as Searle does, that, when an individual participates in genuinely collective activity, that individual's mind contains a special mode of *We*-intending besides the ordinary mode of *I*-intending.

Seumas Miller also rejects Searle's fundamental assumption that individual minds can contain special *We*-modes of intending besides ordinary *I*-modes of intending, and specifically addresses the reason Searle has given for making that assumption—namely, that the special character of collective action appears incapable of being accounted for just by reference to sets of individual actions and to mutual beliefs of their agents to the effect that those actions occur. Miller argues that Searle's scepticism about the prospects of an individualistic analysis of collective action underestimates the power and flexibility that particular individualistic analyses can have, and outlines his own individualistic analysis, where collective actions are held to be fully accountable in terms of individual actions, relations of *dependence* among those individual actions, collective *ends* that individual actions can serve (where the notion of a collective end is construed individualistically as a single state of affairs that each of a group of agents has as its individual end to realize), and, finally, mutual beliefs among the participants about the occurrence of those individual actions, the relations of dependence holding between them, and the collective ends served by them. Miller then argues that that analysis, though strictly individualistic, is not open to counterexamples of the kind that Searle and others have proposed against individualistic analyses of collective action, and so undermines Searle's main reason for positing special *We*-modes of intending in the analysis of the latter. Miller finally suggests that the analysis can be naturally extended to provide illuminating accounts of a variety of complex social phenomena, and that it is unclear how, if at all, these phenomena could be systematically accounted for on anti-individualist premises.

While Miller claims, like Ludwig, that *We*-intentions, as analysed by Searle, are not necessary for collective action, Anthonie Meijers claims, like Gilbert, that, even if they were necessary, they would not be sufficient for collective action. Meijers focuses on the common but almost entirely unexplored phenomenon of collective speech acts—that is, speech acts whose agents or addresses, or both, are groups rather than individuals (e.g., acts of collective acceptance or collective rejection performable by utterances such as "Our team accepts the committee's decision" or "This committee rejects the union's allegations of unfairness")—and adduces various reasons why such collective speech acts cannot be held to be equivalent to conjunctions of corresponding individual speech acts of the relevant group members. The most important of these reasons, according to Meijers, is that collective speech acts generate special types of inter-individual entitlements and commitments that could not possibly be generated by multiple individual speech acts performed by each one of the relevant

group members, *even if* the members in question were individually holding the kinds of *We*-intentions posited by Searle: for, an individually held *We*-intention is, in Searle's view, an intention that is purely *internal* to the individual holding it (in the sense that it would exist even if no other individuals existed), and *We*-intentions that are purely internal to each individual in that sense cannot possibly ground, Meijers contends, *inter*-individual entitlements and commitments. Meijers further claims that collective speech acts not only show that Searle's account of collective action is unsatisfactory, but that they create an independent problem for Searle's influential theory of speech acts, which has been constructed by taking only individual speech acts into consideration. Meijers proposes that the latter problem should be addressed by recognizing that each one of the five major categories of speech acts that Searle posits (assertives, directives, commissives, expressives, and declarations) includes not only, as Searle was supposing, individual but also irreducibly collective members, and concludes by suggesting that the irreducibly collective members of one among those five categories (the category of declarations) is especially significant in articulating the crucial role that, in Searle's view, language plays in the creation of extra-linguistic institutional facts.

* * *

As already noted, Searle believes, on the one hand, that institutional facts, unlike other social facts, are uniquely human phenomena, and, on the other hand, that, just like all other social facts, they are products of collective intentionality, which is not a uniquely human phenomenon, but is, on the contrary, widespread in the animal kingdom. The question then arises as to why collective intentionality has led to the creation of institutional social facts in the human case, whereas it has only issued in social facts that are not institutional in the non-human case. Searle does not address that question in great detail, apart from suggesting that it is the existence of language that, in the human case, has allowed collective intentionality to move beyond the merely social and towards the institutional level. The two essays in the second part of the volume address from different (and in part opposite) perspectives certain wider developmental and evolutionary aspects of the question, each noting respects in which Searle's position is weak and each proposing ways in which its weaknesses could be remedied.

Hannes Rakoczy and Michael Tomasello begin by arguing (partly on the basis of experimental investigations of the behaviour of chimpanzees in the domains of cooperation, communication and social learning) that, contrary to what Searle assumes, genuinely collective intentionality *cannot* be attributed to non-human animals, since, even those among non-humans animals that can be supposed to be capable of attributing mental states to others, do no interact with those others in ways that satisfy certain minimal requirements (such as commitment to the attainment of a common goal or commitment to mutual support in attempted realizations of such a goal) that should be taken to be the

marks of distinctively collective intentionality, if the notion of distinctively collective intentionality was to be appropriately operationalized. Assuming, on the basis of that evidence, that it is not only institutional reality, but collective intentionality itself, that should be taken to be uniquely human, Rakoczy and Tomasello propose that the problem of the ontogenesis of institutional reality can be narrowed down to the problem of identifying the particular type of joint human intentional activity within which children can be supposed to have primarily acquired the aptitudes that later enable them to enter the complex world of institutional facts that characterizes much of adult human life. The most important such activity, Rakoczy and Tomasello then argue, is the activity of *joint pretence* that children engage in when playing games. For it is primarily in the game context that children immerse themselves in the process of collectively assigning to objects special functions that the physical constitution of those objects does not equip them to have (of collectively pretending, e.g., that a pen is a knife). And it is also primarily in the game context that they use proto-performative utterances ("This is now our knife", said of what is known to be a pen) in order to represent those special function assignments. But the functions thus assigned and the utterances through which they are assigned, Rakoczy and Tomasello observe, are of precisely the same kind as the *status-functions* and the *constitutive rules* that, according to Searle, are the building blocks of collectively created human institutional reality. Rakoczy and Tomasello therefore suggest that the activities of joint pretence in which children participate when playing games should be taken to be the most significant ontogenetic precursors of the collective activities through which adult humans create the complex institutional structures that Searle was primarily interested in investigating.

While Rakoczy and Tomasello regard Searle's overgenerous attribution of collective intentionality to non-human animals as unduly burdening his account of human institutional facts, Robert A. Wilson regards Searle's heavily intellectualized account of human institutional facts as resulting in an unduly impoverished conception of the social lives of non-human animals. Wilson first argues (relying in part on ethological evidence) that certain non-human animals are clearly involved in the assignment to environmental elements of certain functions that the physical properties of those elements do not equip them to have (and so, in the assignment of *status-functions* in Searle's sense), but that these assignments are evidently not manifestations of language-facilitated higher-order intentionality of the sort that Searle deems to be necessary for the emergence of status-functions, and therefore suggest that a proper account of status-function assignment may not require the sophisticated variety of intentionality that Searle postulates in the course of his analysis of human institutional reality. Wilson argues, furthermore, that there are many varieties of both human and non-human animal sociality that do not require *any* kind of intentionality (and so, cannot be supposed to be *constructed* in Searle's sense of that term), thus undermining Searle's identification of social facts with all and only those facts whose existence is a manifestation of collective intentionality. Wilson

then suggests that the underlying evolutionary reason for that state of affairs is that social aggregation, which is a direct outcome of the way most organisms reproduce, is a more pervasive feature of the living world than intentionality, and that Searle's primary focus on human institutional facts, valuable though it has been in highlighting the clearly intentional basis of that particular manifestation of sociality, could not be taken as supplying the prototype along which all, or even most, social phenomena should be conceptualized.

* * *

At the heart of Searle's account of institutional facts, and independently of developmental or evolutionary considerations that might be brought to bear on its evaluation, there are, as already noted, two basic theses: first, that such facts are due to the collective recognition of (linguistically expressible) constitutive rules through which status-functions are assigned; and second, that the assignment of such functions has distinctive normative consequences, in the form of deontic powers (in particular, rights and obligations) that it brings into existence. The four essays of the third part of the volume examine these core features of Searle's account of institutional facts, acknowledging the insights they offer, but expressing doubts regarding either their sufficiency or their necessity in dealing with aspects of institutional phenomena that a fully satisfactory account should cover.

Leo Zaibert and Barry Smith begin by arguing that the distinction between 'regulative' and 'constitutive' rules, which is fundamental to Searle's theory of institutions, closely parallels the distinction between 'primary' and 'secondary' rules in Hart's philosophy of law, and the distinction between 'summary' rules and 'practice' rules in Rawls's political philosophy. The common but unacknowledged effect of all three distinctions, Zaibert and Smith then argue, is to marginalize discussion of substantive moral issues in the philosophical analysis of core social phenomena, since substantive moral issues evidently belong to those that can only be settled by adopting 'regulative', 'primary' or 'summary' rules, and not by acknowledging the 'constitutive', 'secondary' or 'practice' rules that, according to Searle, Hart and Rawls, are the only rules on which the philosophical analysis of the relevant social phenomena should primarily concentrate. But the marginalization of substantive moral considerations thus operated, Zaibert and Smith contend, has the implication that the main kind of normativity relevant to the philosophical analysis of social, legal and political institutions is the normativity characteristic of games; and since that implication is, they argue, obviously unacceptable, the methodological significance of the distinctions from which it follows should be questioned. Zaibert and Smith then propose, focusing specifically on Searle's case, that his theory of institutional facts would gain in both depth and comprehensiveness by analysing not only the conventionally generated norms on which the functioning of human institutions depends, but also the non-conventionally generated (and, in particular, morally grounded) norms that the functioning of many of them requires. And they

conclude by noting several important differences between these two kinds of norms on which institutions often rely (in particular, the fact that norms of the latter kind are far less defeasible and far less subject to change than norms of the former kind), arguing that the source of many of these differences is the fact that the latter norms are reflections of the inherent normativity of certain mental states (in particular, of intentions) whereas the former tend to be no less arbitrary than the linguistic conventions thanks to which they can often be formulated.

The idea that not all institutional phenomena can be adequately analysed by exclusive reference to the categories Searle provides is also developed, from a different perspective, by Ignacio Sánchez-Cuenca. Sánchez-Cuenca argues, on the one hand, that, among the phenomena that Searle describes as institutional, some do but some others do not owe their existence to the recognition of constitutive rules (and so, are not really *constituted* by those rules, contrary to what Searle claims), and, on the other hand, that, among the deontic powers that institutional facts generate, some do but some others do not derive their normative force from a community's generalized acceptance of rules, but rather (and contrary to what Searle claims) from a community's submission to brute power (or simply from a community's generalized expectation that brute power would be applied if behaviour discordant with the rules were to be manifested). These shortcomings of Searle's account, Sánchez-Cuenca argues, derive from his exclusive concern with the analysis of institutional arrangements that function as coordinating devices in the sense that none of their participants would have any interest in deviating from them, and from his neglect of other kinds of institutional arrangements, whose participants are not indifferent to the possibility of deviating from them, and which are therefore potential sources of conflict. A proper theory of institutions, Sánchez-Cuenca therefore concludes, should be in a position to cover both of these kinds of institutional arrangement, and to elucidate the differences between them. And such a theory, he suggests, should take as its basic notion not the notion of *acceptance* of a rule system but rather the notion of *compliance* with a rule system, and, in the spirit of economic theory, should identify the existence of an institutional order with the existence of a compliance pattern (which might or might not be the manifestation of acceptance) that tends to maintain a rule system in equilibrium.

Steven Lukes's worries are not about the relevance of the notion of acceptance *per se* in understanding the creation of institutional facts, but rather about Searle's perceived failure to elucidate the complex processes through which such acceptance is actually secured in human communities, and about the implications of that perceived failure on Searle's thesis (whose examination is the primary topic of Lukes's discussion) that Durkheim (and classical sociology more generally) have not been able to provide a satisfactory account of institutional phenomena. Lukes grants that many points of Searle's critique of Durkheim are well taken, but argues that some others are not, and that those that are not reflect the need to avoid interpreting Durkheim's methodological writings in

abstraction from the detailed sociological analyses that they were intended to frame. When that connection is properly understood, Lukes claims, it becomes apparent that Durkheim's thesis, which Searle finds most objectionable, that social facts are *external* to the individuals they affect, was not meant to deny, as Searle is supposing, that they are not *also* internal to them, but was rather intended to emphasize the fact that they *become* internal to them through complex socialization processes over which individuals often do not have control (and which Durkheim and subsequent sociologists have tried to analyse). But it is precisely those socialization processes, Lukes argues, that must *already* be in place if the generalized *acceptance* that Searle takes to be essential to the creation of institutional facts is to be possible. And since the preconditions of such generalized acceptance are left mysterious in Searle's account, Lukes's conclusion is that Durkheim would not only have an effective response to some of Searle's charges, but that the kind of sociological analysis he had helped to initiate would provide an essential ingredient that is conspicuously missing from Searle's account of institutional facts.

The volume's third part concludes with my own contribution, where I focus on Searle's famous early derivation of 'ought' from 'is' (specifically, of statements about obligations from statements about promises), whose significance in the present context consists in the fact that, if correct, it would provide a particularly clear instance of Searle's general claim that extra-linguistic institutional facts, and the deontic powers they bring into existence, are *logical* consequences of the performance of linguistic acts according to established conventions. After arguing that, even if it were valid, Searle's derivation would not be sufficient, as he was supposing, to settle the issue of the 'naturalistic fallacy', I argue that the derivation is in any case not valid, since there is a class of cases where a statement ascribing a promise can be true even though a corresponding statement ascribing an obligation cannot be true. Since the cases in question are those in which the clause specifying the content of the ascribed promise and of the ascribed obligation contains vacuous names or vacuous descriptions, these counterexamples, I maintain, point to an important logical difference between promise-ascriptions and obligation-ascriptions, —namely, that the latter are *extensional* in a way in which the former are *non-extensional*. And it is that difference, I conclude, that accounts for the invalidity of Searle's proposed derivation of 'ought' from 'is', and that consequently undermines what he has presented as the clearest piece of supporting evidence for the claim that the existence of deontic powers is a *logical* consequence of the performance of linguistic acts.

* * *

I am grateful to the contributors, and especially to John Searle, for their cooperation in this project. I hope that the essays presented here will help readers acquire a deeper understanding of Searle's work on the philosophy of social phenomena, and will also suggest new directions of research in that area, whose current flourishing is in no small measure due to Searle's pioneering contributions.

SOCIAL ONTOLOGY: THE PROBLEM AND STEPS TOWARD A SOLUTION

John R. Searle

I have a quite specific conception of social ontology, both of the problem (or problems) and of a possible solution. This conception is under discussion in this volume. In this introductory essay I will try to state some general features of my approach.

1. THE PROBLEM OF SOCIAL ONTOLOGY

The problem of social ontology arises as part of a very general philosophical problem that is specific to, and I think the most important problem in, our present epoch. The problem arises out of advances in human knowledge. Beginning in the 17th century, we have developed an increasingly powerful conception of the basic structure of the universe. To put it very crudely, the universe consists of physical particles in fields of force; these are typically organized into systems. In at least one system, our solar system, certain carbon-based systems have evolved over a period of about 5 billion years, roughly speaking a third of the age of the universe, into our present plant and animal species. Let us call these two sets of facts, one from atomic physics and one from evolutionary biology, the basic facts. Now given that the structure of our universe is fixed by the basic facts, we are inevitably presented with a set of philosophical problems. The main problem is how to reconcile our self-conception with the basic facts. The problem has many different aspects. How can we reconcile consciousness—subjective, qualitative, first-person consciousness—with the basic facts? How can we reconcile rationality, language, free will, ethics and aesthetics with the basic facts? The specific problem of social ontology can be put in the form of a paradox. We are confronted with a social and institutional reality that is for us objective, yet exists only because people believe it exists. In a sense it exists, at least in part, in the minds of the individual participants in the society. Yet when we are actually dealing with social reality, it is as real to us as the brute physical facts. For the baseball player, the points scored are as real as the bat and the playing field; for the lawyer, victory and defeat in the court case are as real as the courthouse, and for the professor, the universities and seminars that he encounters are as real as the sky and the earth. How can there be an epistemically objective institutional and social reality when this reality appears to be ontologically subjective, in the sense that it is at least in part constituted by human attitudes?

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 11–28.
© 2007 Springer.

For the sake of this discussion, I am going to assume that the problems of consciousness and intentionality have been solved, that is that we have at least a broad outline of a conception of how it is possible for consciousness and intentionality to exist in physical systems like us. In broad outline the solution is to point out that all of these phenomena, from feelings of thirst to suffering the angst of postindustrial man under late capitalism, are caused by neurobiological processes and are realized in human and animal brains.[1] So let us assume that conscious intentionalistic beasts like ourselves are walking the earth, What more do they have to have to get social ontology?

2. COLLECTIVE INTENTIONALITY, STATUS FUNCTIONS, DEONTIC POWERS AND FREE WILL

The first step is to understand the human capacity for cooperation. Humans cooperate both in their activities and in sharing beliefs, desires, and attitudes generally. But if all intentionality has to exist in the brains of individuals, then how is it possible that there can be collective intentionality? Before we address this question head on, we need to set the boundaries that any answer to the question has to observe. The boundaries are that we have to show how any particular answer is constrained by the basic facts. In this case we have to show how *collective* intentionality can exist in the heads of *individual* human and animal agents. Standardly the answer to this question has been to suppose that somehow or other we have to reduce collective intentionality to individual intentionality, reduce "We" intentionality to "I" intentionality. The reason for this is that if all intentionality exists in brains, and if each brain is specific to the body in which it is housed, then it looks like there could not be any *irreducible* "We". Any "We" has to consist in "I" plus "I" plus "I", etc. I have objected to this analysis on the grounds that the way that the problem is posed reveals a fallacy. The fallacy is to suppose that, because all intentionality exists in the brains of human and animal individuals, that therefore it must always, in thoughts and utterances, take the grammatical form of the first person singular. There is no reason why you could not have intentionality in individual heads that took the form of the first person plural. I can have in my head and you can have in your head the answer to the question, What are we now doing? which is of the form, *W*e are doing such and such. Where "doing such and such" involves doing something that is *essentially cooperative*. We are playing Beethoven's Ninth Symphony. We are engaging in a public debate. We are campaigning for our candidate in an election. It is an important fact that a great deal of our lives is spent cooperating with other people. Every collective activity from fighting a war to engaging in a casual

[1] For details of this account see John R. Searle, *The Rediscovery of the Mind*, Cambridge, MA: MIT Press, 1992; and John R. Searle, *Mind: A Brief Introduction*, Oxford: Oxford University Press, 2004.

conversation requires collective intentionality. It is not enough that we share a common goal, we have to share common attitudes.

So let us suppose that we have established that collective intentionality is a fundamental and irreducible fact of human and some animal psychology. In order to fully understand how collective intentionality can change and even create reality, we have to understand a fundamental distinction between those features of the world that are observer-independent and those that are observer-relative. In our most metaphysically hard-nosed moods, we think that only the brute facts—facts about such things as tectonic plates, molecular structures, and galaxies—exist. But there are many things that are important in our lives that exist only relative to our attitudes: things such as money, government, property, marriage, universities, and social institutions generally. All of these are observer-relative, where "observer" includes participants and not just outside observers. "Observer-relative" means intentionality-relative. Now this distinction between the observer-independent and the observer-relative relies on a more fundamental concept, which is the notion of *mind dependence*. States of consciousness are mind dependent in the sense that they could not exist without the mind, but they are not observer-relative because whether or not I, for example, have a certain state of consciousness does not depend on what any observer thinks. So though all observer-relative phenomena rely on conscious observers for their existence, mind dependence is not the same as observer relativity. Observer relativity is a special case of mind dependence, but not all mind dependent phenomena are observer-relative. States of consciousness, for example, are mind dependent but not observer-relative.

For the present purposes we need to insist that all of institutional social reality, all of money, cocktail parties, governments, football games, and the stock market are observer-dependent or observer-relative.

Now given this distinction between the observer-dependent and the observer-independent we need to notice a remarkable capacity of humans and many other animals: That is the ability to impose *functions* on objects. So a human being, even in a preliterate state can treat a piece of stone as a tool or can treat a log as a bench on which to sit. Some animals also have the capacity to impose functions on objects. Think of bird's nests and beaver dams. This leads to an important point: all functions are observer-relative. If we say the function of the heart is to pump blood, we are saying something in addition to saying that the heart causes the circulation of the blood. We are also saying that the pumping of the blood relates to the overall economy of the animal body in such a way as to contribute to its health, survival, flourishing, and reproduction. To put the point very succinctly, though roughly, a function is a cause that serves a purpose. The crucial argument that all functions are observer-relative is that the notion of function introduces a notion of normativity. To say the heart *causes* the circulation of blood is quite different from saying that the *function* of the heart is to pump blood. Causation, by itself, sets no normative criteria, whereas functions do. Once we have said that the function of the heart is to pump blood we

can talk about malfunctioning hearts, about heart disease, in a way that we cannot talk, for example, about malfunctioning stones. We can only talk about a malfunctioning stone if we impose a function on the stone. If we want to use it, for example, as a paper weight or a projectile, then given that assignment we can talk about its correctly performing its function or its "malfunctioning".

It bothers a lot of people to think that all functions are observer-relative, because they think that somehow or other, observer relativity implies nonreality or epistemic subjectivity or some other dreadful low ontological and epistemic status. But this conclusion does not follow. The piece of paper in my hand is a twenty-dollar bill, and that is an epistemically objective fact, even though, because money is observer-relative, it has an element of ontological subjectivity. Observer relativity does not imply epistemic subjectivity. It does indeed imply ontological subjectivity, but ontological subjectivity does not necessarily carry with it epistemic subjectivity. It is for this reason that we can have an epistemically objective science that deals with money (economics) even though something is money only because of certain attitudes that people have toward it.

With the concepts of intentionality, collective intentionality, and the assignment of function in hand, we can go to the next step in the explanation of social ontology, the introduction of the crucial notion of *status functions*. Many functions of objects and people are performed solely in virtue of physical properties. Thus an object can perform the function of a hammer, a watch, a car, or a pen solely in virtue of its physical structure. There is, however, a fascinating class of functions where physical structure by itself is not enough, rather people have to assign a certain *status* to the object in question. And with that status goes a function that can only be performed in virtue of the collective recognition and acceptance of the object or person as having that status. Think of what it is to be a twenty-dollar bill or the president of the United States and you will see that the objects and people in question are indeed capable of performing certain functions, but their having the status of money, or president, is crucial to their being able to perform the functions that go with those statuses. For the sake of simplicity, I have in the past said that, typically, the imposition of these status functions takes the form of X counts as Y in C. Thus this piece of paper (X) counts as a twenty-dollar bill (Y) in the United States (C). This person, in virtue of having been elected and sworn in, counts as President of the United States in the United States, and so on through a really huge number of other cases.

The "counts as" locution has several virtues. One is that it calls attention to the fact that it is only in virtue of the fact that we count something as having a status that it has the status. Some commentators have objected that "counts as" implies "not really", but that is an implication that we should welcome because nobody is intrinsically the president of the United States and nothing is intrinsically a twenty-dollar bill. Something is president or money only in virtue of our willingness to count it as having a certain status.

The "counts as" structure as a formal structure, and the existence of status functions as a social phenomenon, also have important formal features that enable them to provide the glue that holds society together. The "counts as" structure iterates upward more or less indefinitely, and spreads laterally across many different kinds of institutions. For example, in virtue of my birth I count as a citizen of the United States, but as a citizen of the United States of a certain age, I am eligible to become a registered voter. And as a registered voter I can become a member of the Democratic Party. As a member of the Democratic Party I can attempt to become a political candidate. And if I succeed in becoming a political candidate and get elected, I then become the holder of a certain office. This is the upward iteration of the "counts as" structure. Furthermore, it spreads out laterally. So I do not just have money in my bank account, but it is put there by my employer and I use it to pay my taxes, my credit card debts, and my utility bills. All of these are institutional notions. And no single institutional fact can exist all by itself. It could not be the case, for example, that for one instant in the history of the universe somebody possessed twenty dollars, but that instant soon passed and no one had any money before or since. Institutional facts are typically parts of huge and complex holistic structures.

Now why is all of this so important in our lives? Because it creates vast systems of powers. Some writers have mistakenly seen social facts as essentially a source of coercion and constraint,[2] but what they fail to see is that the constraints exist within systems of enormously increased power. And what exactly is the form of that power? I have described it as "deontic" power because it takes the form of rights, duties, obligations, authorizations, permissions, certifications, eligibilities, qualifications, etc. Now again, why are these so important? Their basic importance lies in the way they lock into human rationality. They provide us with that essential characteristic of all human civilization: desire independent reasons for action. It is in this respect that human beings differ dramatically from most, and perhaps all, other animal species. Even if I do not much feel like working on this article today, I have a desire independent deontic reason for doing so, because I have made a promise to finish it within a certain time. We are blinded to the existence of desire independent reasons for action because so many of our motivations are desire dependent. So we want to get rich, become elected to public office, get married and buy a house. But notice that each and every one of these selfish desires presupposes an institutional structure that provides people with desire independent reasons for action. There is no point in getting a lot of money if people do not recognize the validity of the money, no point in obtaining political power if the political power is not accepted and recognized, and so on through the other cases. So the subject of

[2] For example, Emile Durkheim writes, "A social fact is to be recognized by the power of external coercion which it exercises or is capable of exercising over individuals, and the presence of this power may be recognized in its turn either by the existence of some specific sanction or by the resistance offered against every individual that tends to violate it." See "What Is a Social Fact?", Chap. 1 of his *The Rules of Sociological Method*, Chicago, IL: Chicago University Press, 1938, p. 10.

social ontology is not just a branch of the study of what exists, it is also a branch of the study of human rationality.

Because most of the social sciences—history, political science, and economics, for example—are about institutional facts, most of them are concerned with the freedom of the will. Indeed one could say, in general the subject matter of the social sciences is the freedom of the will. This will seem surprising to many social scientists, so let me explain why it has to be the case. Social sciences deal with human actions, but in general the human actions in question are performed on the basis of reasons. Those reasons are subject to constraints of rationality, and rationality presupposes freedom of the will. It presupposes that the actions in question were not determined by antecedent causally sufficient conditions, for if they had been so determined, rationality would have no purchase; rationality would have no room to operate, and its apparent operation would be entirely an illusion.

So there are really two topics that I have raised about the study of the foundations of society.[3] One is the intentionalistic character of social ontology together with the creation of deontic powers, and the second is the study of rationality and how it locks into social ontology. We will understand these points better if we see the contrast between humans and the sorts of animals that lack deontic powers. As far as I can tell, we are the only animals who have deontic powers, because such powers require language and we are the only ones known to me to have a language in the relevant sense.

It is important also to see that language is not just the means for describing and classifying institutional reality, language is essential to the creation and maintenance of institutional reality. Why? Deontic powers, as enshrined in institutional facts, exist only insofar as they are represented as such, but that representation requires a means of representation which, in a broad sense, has to be symbolic or linguistic. Furthermore, and this is a point I did not make in *The Construction of Social Reality,* once you have a language you inevitably have entire systems of deontology because language is itself deontological through and through. Commitment is not a feature confined to promises, contracts, and the like, but is a feature of every type of speech act, and the speech acts in language inevitably generate nonlinguistic institutional facts.[4]

3. COMMITMENT

The notion of commitment is so crucial for understanding language, and indeed social ontology in general, that I need to say a little more about it. There are two closely related aspects in the notion of commitment, or one might say there are

---

[3] These two topics are discussed in more detail in John R. Searle, *The Construction of Social Reality*, New York: The Free Press, 1995; and John R. Searle, *Rationality in Action*, Cambridge, MA: MIT Press, 2001.

[4] For an explanation of this point, see John R. Searle, "What is Language: Some Preliminary Remarks". Forthcoming in Savas L. Tsohatzidis (ed.), *John Searle's Philosophy of Language: Force, Meaning, and Mind*, Cambridge: Cambridge University Press.

two components to the meaning of "commit", or even there are two senses of the word.[5] In one sense to be committed to something is to have *undertaken* something in a way that makes it difficult or awkward to change course. For example, in this sense we speak of a general as having committed his troops on the left side of the front. There is a type of irreversibility to this sense of commitment. The other sense of "commitment" involves an obligation, or other deontic requirement. If I have made a promise to come and see you, then I have undertaken a commitment to come and see you. The way in which these two senses often coalesce is that sometimes it is a result of some action of an irreversible kind that I place myself under an obligation. And having placed myself under an obligation I have in a sense created an irreversible course. Promising, for example, contains both irreversibility and deonticity, and promising is a paradigmatic form of commitment.

But not all commitments are deontic. An animal attacking another animal can commit itself to a certain line of attack. Once the dog jumps for the other dog's throat there is no turning back. But there is no obligation in this case. There is, oddly, a minimal desire independent reason created by the irreversible choice, because having undertaken the line of attack, the animal has thereby created a reason for continuing. Struck by the connection between commitment and desire independent reasons for action, I have often talked as if all such reasons are deontic. But that is not really true. All commitments do indeed create desire independent reasons, but not all commitments are deontic.

4. MORE ABOUT COLLECTIVE INTENTIONALITY

There are some interesting further problems about collective intentionality once we have accepted the fact that "we intend", "we believe", and "we desire", cannot be reduced to "I intend", "I believe", "I desire", etc. The deep question is: How can it be the case that We-intentionality can move individual bodies if the content of the We is not the same as the content of the I, which constitutes doing one's part of the collective effort? If we are playing a piano-violin duet, I might be doing my part by playing the piano and you might be doing your part by playing the violin. How could *our* collective intentionality move *my* body? How do we get from "We are playing the duet" to "I am playing the piano part" said by me and "I am playing the violin part" said by you?

In order to get completely clear about the structure of collective intentionality and collective action we have to say some things about the intentionality of human action in general. I am going to summarize some of the central theses of the philosophy of action, and please beware that I am leaving out lots of qualifications and details. For the full theory see *Intentionality* especially Chapter 3,

[5] For a good discussion of the distinction, see Seumas Miller, "Joint Action: The Individual Strikes Back", this volume.

and *Rationality in Action*.[6] Intentions divide into two kinds: prior intentions, which are those had prior to the onset of an action, and intentions-in-action, which are the intentional components of actions. The proof that there must be a distinction is that they have different conditions of satisfaction. If I now intend to raise my arm in fifteen seconds, the conditions of satisfaction are that I perform the whole action of raising my arm, and that that action be caused by this prior intention. But when I am actually raising my arm, the conditions of satisfaction are that the intention-in-action itself should cause the raising of the arm; and the whole action consists of these two components, the intention-in-action and the bodily movement, where the bodily movement is caused by the intention-in-action. So both prior intentions and intentions-in-action are causally self-referential.

We can represent these facts in the following notation, where the part inside the parentheses represents the propositional content of the intention and the part before the parentheses represents the type of intentional state it is.

Prior intention (this prior intention causes that I perform the action of raising my arm)
Intention-in-action (this intention-in-action causes that my arm goes up).

Using obvious abbreviations, "p.i." for prior intention, "i.a." for intention-in-action "b.m." for bodily movement, and "a" for action, we can say that the general form of these relations is given by the following schemata.

p.i. (this p.i. causes a)
i.a. (this i.a. causes b.m.)
a = i.a + b.m., where i.a. causes b.m.

Human actions also have some very special features deriving from the fact that, typically, I do not just perform a simple action like raising my arm, but I do something *by way of* or *by means of* doing something else. For example, I fire the gun *by means of* pulling the trigger and I vote in the committee meeting *by way of* raising my arm. Pulling the trigger causes the gun to fire, but raising my arm does not cause the vote to take place, it just constitutes voting. I call these two types of inner structure of action *the casual by-means-of relation* and *the constitutive by-way-of relation*. We can represent these relations in the following notation:

i.a. B by means of A (this i.a. causes A, which causes it to be the case that B)
i.a. B by way of A (this i.a. causes A, which constitutes B)

Now the question for our present discussion is: How does all this apparatus carry over to collective intentions and actions? If we are cooperating in some group endeavor, where our individual contribution causes some further effect, then we have the causal by-means-of relation. And if we are cooperating in an

---

[6] John R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 1983; and John R. Searle, *Rationality in Action*, Cambridge, MA: MIT Press, 2001.

endeavor where our individual efforts constitute the desired effect then we have the constitutive by-way-of relation.[7]

We can work out these cases by considering examples of collective intentionality involving both the causal relation and the constitutive relation. Consider an example of the causal relation where you and I are both trying to start a car engine by means of me pushing and you sitting in the driver's seat and letting the clutch out after the car has attained a certain amount of speed. Causally speaking we are starting the engine by means of me pushing and you releasing the clutch while the ignition is on.

Compare that with the case where we are performing a duet where I play the piano part and you play the violin part. Here our playing does not cause the duet to be performed. My playing and your playing simply constitute the performance of the duet. So from my point of view, where I take it for granted that you are doing your part, I have a collective intention-in-action that *we* play the duet by way of *me* playing the piano. How can we represent all of that in our canonical notation for representing the structure of intentionality? Here is the causal case:

i.a. collective B by means of singular A (this i.a. causes: A car moves, causes: B engine starts).

In English, this is to be read as: I have a collective intention-in-action, B, which I do my part to carry out by performing my singular act A, and the content of the intention is that, in that context, this intention-in-action causes it to be the case, as A, that the car moves which, in that context, causes it to be the case, as B, that the engine starts.

Notice that the assumption behind my collective intentionality is that if I make my contribution on the assumption that you make your contribution, together we will cause the car to start. Notice that in the part outside the parentheses the occurrence of "B" and "A" are free variables. They are bound inside the parentheses by the verb phrases that follow the respective letters, "car moves" and "engine starts".

Now we turn to consider the case of the constitutive by-way-of relation where I am playing the piano part and you are playing the violin part. In this case we have:

i.a. collective B by way of singular A (this i.a. causes: A piano plays, constitutes: B duet is performed).

In ordinary English again, this can be read as: I have a collective intention-in-action to achieve B by way of contributing my part, the singular A. The content

---

[7] What follows is a continuation of the discussion in John R. Searle, "Collective Intentions and Actions" in P. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*, Cambridge, MA: MIT Press, 1990. Reprinted in John R. Searle, *Consciousness and Language*, Cambridge: Cambridge University Press, 2002.

of that intention is: this intention-in-action causes it to be the case, as A, that the piano plays, which, in that context constitutes its being the case, as B, that the duet is performed.

Notice also that the content of my singular intentionality in both cases does not make essential reference to the content of your singular intentionality. I simply take it for granted, in that context, that if I do my part we will succeed, because I am operating on the assumption that you will do your part, and you are operating on the assumption that I will do my part. There is an epistemic basis for this: often one does not know what the individual intentionality in the minds of the other members of the collective is. I might have collective intentionality to achieve a certain goal, and I have that on the assumption that you are working toward the same goal as I am. But it need not be the case that I actually know the content of your intentionality. The offensive lineman blocking on a pass play does not necessarily need to know what routes are being followed by the wide receivers or how many steps backward the quarterback is taking before throwing the pass. All he has to know is what part he is supposed to do ("his assignment" in the jargon of football coaches), and this notation is an attempt to capture the condition of satisfaction of his intention.

If I am successful in this analysis then I will have succeeded in showing how it can be the case that collective intentionality can exist in the minds of individuals and at the same time that collective intentionality, though irreducibly collective, can nonetheless cause the movements of individual bodies. The action does not get performed if my body does not move, and this has to be reflected in the collective intentional content, even though the collective intentional content exists entirely in my brain and your brain and the brains of the other members of the collective.

5. SOME IMPROVEMENTS TO THE THEORY

Essentially I have so far summarized the account of institutional reality and collective intentionality that I gave in *The Construction of Social Reality* and in "Collective Intentions and Actions" over 10 years ago. Since that time there have a number of interesting developments and criticisms, and I want to explain some of these. First of all, some authors (notably Barry Smith and Amie L. Thomasson[8]) have pointed out that in certain cases you can have what Smith calls, free-standing Y terms, where you have an institutional fact but the status function is not imposed on any particular person or object. For example, if you create a corporation in the State of California there will be no physical entity which is identical with the corporation. You have to have officers, a mailing

[8] Barry Smith, "John Searle: From Speech Acts to Social Reality" in Barry Smith (ed.), *John Searle*. Cambridge: Cambridge University Press, 2003, pp. 1–33; and Amie L. Thomasson, "Foundations for a Social Ontology," *Protosociology*, 18–19 (2003): 269–290.

address, a board of directors and stockholders but you do not have to have, for example, a building. There is no physical entity which constitutes the corporation. The president and board of directors are essential but they are not identical with the corporation. Similar things could be said about money. Currency is common to money but it does not seem to be logically necessary to the existence of money that it take the form of physical currency. A representation of the amount of money on a computer disk is adequate for commercial transactions to take place.

What should we say about such cases? Actually if we understand them adequately I think they strengthen the theory. In *The Construction of Social Reality*, I pointed out that the main purpose of creating institutional facts is to increase human power, and that the types of power in question are deontic powers and that these bottom out in individuals. So the logical form of the creation of institutional power is:

We collectively accept (S has power (S does A)).

Thus if you have money you have a positive power, the "right" to buy things with it. If on the other hand you receive a parking ticket, you have a negative power, an "obligation" to pay the ticket fine. Rights are positive powers, obligations are negative powers. The hypothesis I put forward, and I have not seen any reason to modify it, is that all of institutional power can be reduced to a single power acceptance operator, and that the other forms of power, such as obligations, rights, responsibilities, and duties, are the result of Boolean operations on this fundamental power creation operator. For example, when we create a corporation, so to speak out of nothing, what in fact happens is that we assign a set of powers to actual people. The genius behind the idea of the corporation is that the individuals will have the powers without the corresponding liabilities that go with the assumption of individual fiduciary responsibility. That is why corporations are called "limited liability" companies.

What then is the relationship between the "X counts as Y" formula and the "We accept (S has power (S does A))" formula? It is only because S stands in a certain relation R to the Y status function that S has the power in question. Thus the president of the United States is identical with the object on which the status function is imposed. Thus the structure of his power is given by combining the two formulae. We can see this by spelling it out:

Because X, George W. Bush, has the Y status function, President of the United States, we accept that (X has power (X does A)).

And in the case of money (currency)

Because I am in possession of this piece of paper, X, which counts as having the Y status function of a 20 dollar bill, we accept that (I have power (I buy up to the value of 20 dollars)).

And so on through the other cases. For example, the president of the corporation stands in the relation of being president of the institutional entity and he

is identical with the object on which the status function of being president is imposed. In general we can say that status functions bottom out in human beings, and the assigned status functions create power relations among human beings. In cases where it seems that they might bottom out in nothing, as the case where you can have money without any physical object being the money or you can have a corporation without any physical object being the corporation, the actual operative notion is that human individuals stand in the appropriate relation to the Y term. So it is the *possessor* of money or the *officer* of the corporation who actually has the deontic power that goes with the Y status function. Where power is concerned, and the whole point of institutional facts is the creation and distribution of power, it is not the Y term as such but the people who stand in the appropriate relations to the Y term who are the bearers and objects of the powers in question. Sometimes, but by no means always, that relation is one of identity. The operative institutional facts are not that such and such is money, or such and such is a corporation, but rather that someone has money or owes money, or earns money, etc., and that someone owns shares in the corporation, or is a member of the board, or the president of the corporation, etc. Free-standing Y terms *appear* to bottom out in nothing, but they do not. They bottom out in people and power relationships among people.

6. TOWARD A TAXONOMY OF INSTITUTIONAL FACTS

The next step in the research enterprise is to get an adequate taxonomy of institutional powers and therefore of institutional facts. The difficulty is that there are a number of different crisscrossing criteria for distinguishing one type of institutional fact from another. I believe the first step is to identify certain general areas of human practice where under that description, the description given in the identification, the practice necessarily involves deontic powers. Thus money, marriage, private property, and government necessarily involve deontic powers. You could not have a system of government, marriage, money, or private property without some rights, duties, obligations, etc. But lots of other areas of human endeavor for example, science and recreation, do not under those two descriptions necessarily identify systems of deontic powers. There are plenty of institutions within science and recreation, The National Science Foundation and The Squaw Valley Ski Team, for example, but science and recreation do not, as such, identify systems of deontic powers. This is why it is important to emphasize *under what description* one is identifying the practice.

Now within these general areas of human practice there are subareas that also necessarily involve deonticity. Thus under government we have legislative, police, military, judicial, and executive activities, and all of these essentially involve deontic powers. Again under private property we have exchange, possession, transfer, sale, purchase, tax obligations, and inheritance. All of these are general patterns of human activity involving deonticity. Marriage and family also involve deonticity because they are systems of rights and obligations.

Going now from the more general to the less general, we have specific institutional structures that are explicitly designed to serve certain functions and are linguistically constituted. The United States Government, the University of California and the National Football League are typical examples.

By now the reader will have noticed that I have been talking about institutional structures and not about institutional facts. The reason for this is that in general institutional facts only exist within institutional structures, and in order to get a conception of institutional facts we have to get a conception of the structures within which they are facts.

In order to see the motley of institutional facts I want to make a more or less random list of the kinds of fact which constitute institutional facts before we begin any effort to classify them. Consider the following: Sam invited me to dinner, Sally said that it is raining, Bush is president of the United States, my sister got married in Phoenix, Arizona, the Detroit Tigers won last year's World Series, I own property in the city of Berkeley, Bill is captain of the ninth grade softball team, I am a licensed driver in the state of California, the piece of paper in my hand is a twenty-dollar bill, Jones received a Ph.D. degree, national debt of the United States increased last year, Sally and Bill are very close friends.

Now let us contrast these with a set of facts which, though often social, are nonetheless not typically institutional facts: Sally and Bill are walking together down the street, the crowd rioted on the streets of San Francisco, Sally is pregnant, Bill has a case of the flu, short skirts have become fashionable again, I am wearing a red sweater, the small snowfall this year is bad for skiing, global warming is a threat to civilization.

What is the difference between these two sets of facts? One crucial test is that institutional facts, as opposed to social facts in general, involve some form of deontic power: some right, duty, authorization, obligation, etc. The only exceptions I can think of to this general principle are the honorific status functions, such as receiving an honorary degree or becoming Miss Alameda County, where one acquires an honor but no power that goes with the honor. Sometimes institutional facts may be completely uncodified. For example, one boy may simply emerge as the captain of the softball team without there being any official election or other recognized mode of selection. He may simply be recognized by his fellows as the captain of the team. Similarly, friendship is typically uncodified and would lose much of its flexibility if it were codified. But as I understand the term, there are special obligations that go with friendship. In this respect, being a friend differs from being a drunk, a nerd, or an intellectual. There are no special obligations that go with being a drunk, and no special rights and duties and no generally recognized deonticity that go with being a nerd or an intellectual. I might personally feel that as an intellectual I have certain obligations, but unless these are collectively recognized by my society, the status of an intellectual is not an institutional status and the fact that I am an intellectual is not an institutional fact.

If we are going to get an adequate taxonomy of this vast array of different types of institutional facts, we need to begin by stating some principles that distinguish one type of institutional fact from another. I will simply list four such principles and then explain them in more detail:

1. Purely linguistic institutional facts versus the not purely linguistic. Thus, the fact that Sally stated that it is raining is purely linguistic. The fact that I own property in the city of Berkeley is not purely linguistic.
2. 'Natural' versus 'artificial' institutional facts. Some institutions, like private property, are an extension of a natural impulse to possession. Others like limited liability corporations are highly artificial in that they are not simply extensions of biologically given relations.
3. The sacred versus the secular. Some institutional facts like being Pope require a belief in the supernatural. Others like having money do not.
4. Certifications versus enablements. Being a licensed driver permits me to exercise a preexisting physical ability. Being president of the United States, however, grants new powers.

Let us now consider these in order.

6.1. The Distinction Between the Purely Linguistic and the not Purely Linguistic

Language is constitutive of all institutional reality. You can have language without having money, marriage, or government, but you cannot have money, marriage, or government without language. *All institutional facts are linguistic facts*. All the same, we need to distinguish those which are purely matters of uses of language, such as making a statement, asking a question, giving an order, or making a promise and those that involve some extra linguistic powers, such as owning property, becoming elected president of the United States, or being convicted of a felony. So any taxonomy has to make a distinction between the most fundamental institutional facts, namely those that record speech acts and other linguistic phenomena, and those that go beyond the phenomenon of linguistic meaning and include other sorts of deontic powers. The apparently nonlinguistic institutional facts involving money, property, and government, for example, all have to contain a linguistic component, because they all are what they are only because they are represented as being what they are, and those representations must be, in the broadest sense, linguistic or symbolic. You cannot have money, for example, if you do not have some way to represent something as money, and that form of representation is necessarily linguistic.

However, I do not see how to use the linguistic component of the not purely linguistic institutional facts as a device for making taxonomical distinctions among them, because the functioning of language in nonlinguistic institutional reality is always the same. It is always a matter of creating a set of deontic powers by using language to represent them as existing. Let us then simply state that in any good taxonomy there has to be a basic taxonomic distinction between the

purely linguistic and the not purely linguistic, and the purely linguistic is primary in the sense that you can have language without money, property, and all the rest, but you cannot have money, property, etc. without language.

6.2. The Distinction Between the 'Natural' and the 'Artificial'[9]

One distinction is between those institutional structures which are natural extensions of pre-institutional biological tendencies and those which, so to speak, involve specifically human creations that go beyond the biologically given tendencies. For example, pre-linguistic animals already have something close to private property when they have territorial instincts. I am reluctant to call these territorial *rights*, because of course the animals do not have any system that would enable them to create rights, but they do fight for what it is not unreasonable to think they regard as something like their property. Again pair bonding is not the same as marriage but it is perfectly natural to see it as the pre-institutional form of behavior out of which marriage grows. And obviously humans and other primates are extremely status conscious, and within the group or tribe there will typically be an alpha male, alpha female, and lower elements of the system of power relations. (But it is important to see that the pre-linguistic alpha male, cannot think "I am the alpha male" and thus being an alpha male cannot be an institutional fact. The terminology is in the theory of the ethologist, not in the social reality.)

The most primitive forms of institutions will be cases where status functions are simply imposed on these pre-linguistic phenomena, where the person is able to lay claim to this as "my property", refer to this person as "my husband" or "my wife", to that person as "our leader".

To the extent that these are collectively recognized, they are well on the way to being institutional facts, and once the deonticity is introduced, I think inevitably in the form of language, we have full-blown institutional facts without yet the development of full-blown institutions. How can there be institutional facts without full-blown institutions? Well the formula "X counts as Y" and the "We accept (S has power (S does A))"formula, are intended as systematically ambiguous between the universally quantified forms and their particular applications. We can just count this person as our leader, I can count this person as my spouse, and I can count this hut as my property without yet having a universal form, a genuine constitutive rule that would enable me to carry out the "counts as" relation in a systematic way. So I can think "This person is my woman (Frau, femme, wife)" in a particular application of the X counts as Y form. Or a society can have the general form: For each man M and for each woman W, if M and W satisfy such and such conditions, they are husband and wife.

---

[9] I put 'natural' and 'artificial' in sneer quotes because all institutions are artificial and all institutions must be consistent with human natural biology or they could not exist.

The most primitive human institutions are natural extensions of biologically more basic forms of social relationship. But some of the greatest human inventions are creations of institutions that go beyond the biologically determined predispositions. Money goes beyond the institutional structure of private property and even exchange. It introduces a universal medium for assigning value, storing wealth, and exchanging goods and services. Corporations are a strikingly original contribution to human civilization, because they run dead counter to our normal instinct of assigning responsibility to actual human beings or groups of human beings. Universities are an interesting marginal case. Any literate society, and indeed any human society, has to have some mechanism for educating its young in ways that enable them to cope. Even in preliterate hunter-gatherer societies, the boys have to learn how to hunt and the girls have to learn how to gather. And if writing is available, some people will have to learn to read and write or there is no use for the notation. But the idea that once the basic skills and knowledge necessary for coping have been imparted during childhood and adolescence there is still room for an institutional structure, a university, dedicated to advanced education and research, most of it of no apparent practical value whatever, and open to relatively large numbers of people on a collective basis, rather than just as individual apprentices to established scholars, is a remarkable development in human culture. I think museums and public parks are also striking inventions, but I will leave it to the reader to figure out why.

6.3. The Distinction Between the Supernatural and the Secular

Many institutional facts depend on belief in the supernatural. Thus, for example, a paradigm institutional fact is that such and such a person is the Pope of the Roman Catholic Church. This is clearly an institutional fact because it involves the collective acceptance of a certain class of deontic powers. However, according to Catholic dogma, the rights of the Papacy do not derive from collective acceptance by the membership of the Church, but are directly given by God in the person of Jesus Christ. Christ appointed Simon, as he was then called, as the first Pope and subsequent Popes derive power from the original statement that Christ made to Simon, called the "Petrine Guarantee." Christ said to Simon, "And I say also unto thee, that thou art Peter, and upon this rock I will build my Church, and the gates of Hell shall not prevail against it" (Matthew 16:18). This is a spectacular example of how institutional facts can be created by supernatural beliefs. It is worth going over this utterance in some detail. The first clause, "And I say also unto thee, that thou art Peter" is clearly a declaration. Christ makes it the case that Simon is henceforth Peter by renaming him as "Peter". The second clause, "upon this rock I will build my Church," I think is best construed as a declaration as well. It is not a mere expression of an intention, or even a promise, on Christ's part to subsequently build a church here when He gets around to it, but rather it *now* makes it the case that in the *future* there will be built a church here. The third clause, "The gates of Hell shall

not prevail against it," I think is also a declaration. It is not a prediction based on empirical evidence of some sort, but a case of making it the case that the gates of Hell shall not prevail against it.

The phenomenon of institutional facts based on belief in divine powers is widespread among supernatural religions. Thus, for example, every priest is assumed to have power that derives from God, and in the Catholic Church, marriage does not consist in a contract between man and woman. It is a sacred creation by God.

Notice that there is no neutral way to describe the phenomenon of institutional facts based on a belief in the supernatural. Either one accepts the supernatural beliefs or one does not. I am here only interested in stating how the system works in fact. It works because there is collective acceptance of these institutional facts, and in that respect they are like all other institutional facts. But that is a tendentious claim on my part. For the faithful the collective acceptance is based on a miraculous truth. I am saying that the actual work is done by the collective acceptance, and that is why the system continues to function even if there is no underlying truth.

6.4. The Distinction Between Certifications and Enablements

We need to distinguish those institutional powers which simply authorize, permit, or certify individuals to do something that they are able to do anyhow and those that actually enable or empower people to do things that they would not otherwise be able to do. In the first class we authorize people to exercise an ability they already have. Thus think of licensed drivers, Certified Public Accountants, members of the Bar Association and elevators with a Certificate of Safety. In these cases, the people and objects are physically able to perform the functions without the certification, but we require that they are subjected to tests or otherwise prove that they meet certain criteria, without which we will not allow them to perform the functions. So there are lots of people in the state of California who are physically able to drive cars, but for one reason or another, do not have driver's licenses, and thus are not legally permitted to drive a car. However, if they are driving without a license, they are nonetheless driving. We need to distinguish these cases from what I will call general *enablements*. In the case of an enablement for example, Congress has the power to pass legislation or the President has the power to command the army. The entity in question does not get its deontic power by the certification of some preexisting ability, rather, the creation of the institutional fact is a creation of the ability. You and I may draft legislation and issue commands to the army, all to no avail, because only if Congress does it, is it legislation, and in the military only a command by a qualified officer or the president of the United States counts as a command. There is a crucial philosophical distinction between those cases where the action can be performed illegally, without benefit of the constitutive rules that provide permission, certification, and authorization, and those that can only be performed

within a system of constitutive rules, because the constitutive rules define the act in question. Physical acts such as driving cars or pulling teeth can be done illegally, without a license. But acts which require an institutional structure which is both the result of and the source of cooperation and recognition by others, such as passing legislation and commanding the army, cannot be done by anybody without the appropriate institutional empowerment.

It will be clear to the reader by now that we have such a welter of different ways of characterizing institutional structures and institutional facts that it is hard to get a single elegant principle of taxonomy. I had better luck with speech acts, where the existence of illocutionary points, which determine the expressed psychological state and the direction of the fit, provided me with a neat taxonomy. The purpose of the speech act is to represent reality, and there are exactly five ways in which you can do that, depending on the point of the speech act in question. We can tell people how things are (Assertives), we can try to get them to do things (Directives), we can commit ourselves to doing things (Commissives), we can express our feelings and emotions (Expressives) and we can bring about changes in reality by declaring it to be so changed (Declarations). Why can't we get a similar simple taxonomy for institutional facts? The purpose of the institutional fact is to create or rearrange institutional powers and there are only so many ways in which we can do that. What are they? I think there must be a simple answer to this question, but so far I do not see what it is.

Part I
Aspects of Collective Intentionality

SEARLE AND COLLECTIVE INTENTIONS

Margaret Gilbert

1. INTRODUCTION

John Searle's 1990 article "Collective intentions and actions" is now a classic text in the burgeoning field within analytic philosophy that would aptly be named the philosophy of social phenomena.[1] The article represents the core of Searle's account of social reality or, as he sometimes puts it, social facts, and the foundation of the account of institutional facts on which his famous 1995 book *The Construction of Social Reality* focuses.[2] The present discussion revisits this article and notes some questions it provokes.[3] It focuses on the first section of Searle's article, in which its central ideas are laid out.

2. THE NEED TO GO BEYOND "I-INTENTIONS"

A phrase Searle prefers to "collective action" is "collective intentional behavior". This is a reasonable preference since the former phrase is often used in the literature of philosophy, economics, and political theory, in a sense other than that he has in mind.[4]

---

[1] The focus has tended to be on social phenomena in the human domain and this will be my focus here. Searle's article is part of what one might call the "first wave" of publications in this field, beginning in the 1980s, from which a growing body of literature is currently spreading out. Included in this initial wave are Raimo Tuomela's book *A Theory of Social Action* (1984), and my book *On Social Facts* (1989). Tuomela draws on the important earlier work of Wilfred Sellars. See, e.g., Sellars (1963); also Rosenberg (1980) and (on the consonance of Sellars's and Searle's position) note 37 below. My own work in the area started with my Oxford doctoral thesis (*On Social Facts* (1978) Bodleian Library: Oxford). I know of no earlier discussion of comparable extent. My 1989 book of the same title offered a different approach to the same general problem: what is a social phenomenon? One stimulus for that was Charles Taylor's important critical discussion of "common knowledge". See, e.g., Taylor (1980) and (1985: Chap. 10, Sect. 3), and, for discussion, Gilbert (2007). Searle has sometimes expressed impatience with the expression "analytic philosophy". I mean only to gesture in the direction of a particular tradition in terms of philosophical topics and methods. The discussions of Alfred Shutz, for instance, though of interest to those writing within that tradition, might be held to stand outside it (and within the different tradition of phenomenology). On the whole the analytic tradition has been relatively uninterested in understanding the social domain.

[2] The brief section in Searle's book (1995: 23–26) entitled "Collective Intentionality" references the treatment in his 1990 article as a more complete one. In a more recent discussion Searle again refers to his 1990 article as the place to go for amplification of his position (2006: 16).

[3] I have previously offered comments on Searle's article in Gilbert (1998), Gilbert (2000b), and Gilbert (2001). The present discussion incorporates some of the material there.

[4] On that other sense, see, e.g., Gilbert (2006a).

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 31–48.
© 2007 Springer.

What is collective intentional behavior? Examples, Searle indicates, include a football team executing a pass play, an orchestra playing a musical work, a corps de ballet in action, and one person pushing a car up the road with another. In everyday thought and talk we are indeed ready to ascribe an action to a population as a whole in such cases. We are ready to ascribe a single action to *those people*, or, indeed, to *it* (as in the case of the football team, the orchestra, and the corps de ballet). The action is *their* or *its* action. These are also cases where we might talk of people (the footballers, musicians, and so on) as doing something together or as a group.

Not surprisingly, the nature of collective intentional behavior among human beings has become a central topic in the philosophy of social phenomena.[5] If one does not understand what it is for one person to do something with another, one cannot have much of a grasp of the social domain.

Traditional philosophical action theory concentrates on what it is for a particular human individual to do something. Perhaps the central concept invoked by this theory is that of an intention. In everyday life, indeed, we often ask of a human being whose behavior we observe what his (or her) intention is.[6] This person, in turn, could properly describe the situation in which he intends to do such-and-such with the words "I intend to do such-and-such". Using a technical label for this situation, then, we might say it is a situation in which someone has an "I-intention". I shall so use the phrase "I-intention" in what follows. It is fair to say that I-intentions are the focus of traditional action theory. This questions what I-intentions amount to, and how one is to understand, the relationship between I-intentions on the one hand, and the movements of a person's body on the other.

Among other things Searle argues, congenially, that in order to give an adequate account of collective intentional behavior one must go beyond I-intentions.[7] In other words, collective intentional behavior is not merely a matter of I-intentions, if it is a matter of I-intentions at all.[8] He also has a more positive doctrine about collective intentional behavior, a doctrine whose nature is less clear than it might be, and whose adequacy can be questioned. Or so I shall argue.

---

[5] For some discussions around the time of Searle's own see Gilbert (1989: Chap. 4), also Gilbert (1990), repr. in Gilbert (1996), and Bratman (1993).

[6] I continue now with "or she", and so on, understood when appropriate.

[7] "Congenially": to the present author. See Gilbert (1989: Chap. 4). In discussion of the nature of collective intentional behavior this introduces a notion of *joint commitment* that is not reducible to I-intentions or, more generally, I-commitments. See also Gilbert (1997).

[8] Searle would, I believe, say that I-intentions are involved (perhaps he would say that they have to be involved if a human being is to do something) but that, importantly, they are in some sense *derivative* in this case. See, e.g., his discussion of the members of the football team p. 403, penultimate paragraph.

3. **WE-INTENTIONS**

Before moving to Searle's discussion, I make a point that I take to be unproblematic pretheoretically. It concerns ordinary language or everyday speech. I also introduce a related technical term of my own.

The point is this: When there is collective intentional behavior each participant could properly describe the situation in terms of what *we,* as opposed to *I,* intend.[9] To take an example of Searle's, suppose that the members of a football team are engaged in the execution of a pass play.[10] As the sentence within quotation marks is normally understood, each of them could properly describe the situation as follows: "We intend to execute a pass play." The same goes for an orchestra. In the appropriate circumstances the members could properly describe the situation as follows: "We intend to play Beethoven's Fifth Symphony." And so on. So one way of posing the question "What is collective intentional behavior" is this: "What is it for us, as opposed to me, or you, to intend to do something?"[11] In technical terminology I now introduce, I rephrase the question as: what is a **we-intention**?[12]

This resembles one of the focal questions of traditional action theory, mentioned above: what does an I-intention amount to? That question might also be put: what is it for me to intend to do something? Or, in more general terms, what is it for an individual human being to intend to do something? The question just raised is, rather: what is it for two or more individual human beings collectively to intend to do something? What is it for an intention accurately to be ascribed to *them*?

3.1. Searle's Discussion

Searle's discussion takes off from the question (in my terms): what is a **we-intention**?[13] Thus early in his article he writes:

How, one wants to ask, could there be any group behavior that wasn't just the behavior of the members of the group? After all, there isn't anyone left to behave once all the

---

[9] That is not to say that he might not also properly say something about what he himself intends, see the previous note.

[10] Op. cit. p. 403.

[11] Cf. the title of Gilbert (1997): "What is it for *Us* to Intend?"

[12] The noun phrase "we-intention" is used in Tuomela (1984) and elsewhere. Searle uses both the noun phrase "we intend" (without a hyphen, as in "how could there be a "we intend" that . . .", p. 402) and the noun phrase "we-intention" (with a hyphen) as in "The thesis that we-intentions are a primitive form of intentionality" (p. 407). I discuss Searle's use of these phrases in the text below. I use boldface type to indicate that "we-intention" is being used in the sense indicated in the text.

[13] In his discussion of collective intentionality Searle allows that non-human animals have collective intentions. As said in note 1 above, I focus my discussion here on the case of human beings. Most of Searle's examples relate to that case. I do not take it to be a criterion of adequacy for a theory of this case that it applies without modification to the case of non-human animals. Cf. Gilbert (1989: 442–444).

members of the group have been accounted for. And how could there be any group mental phenomenon except what is in the brains of the members of the group? How could there be a "we intend" that wasn't entirely constituted by a series of "I intend"s?[14]

What does Searle have in mind when he writes of a "we intend"? Here it seems that he means to refer not to an element in a given individual's thinking, or, more broadly, some state of that individual's mind, but to whatever the phrase "a group mental phenomenon" refers to—if anything. In my terms, it seems that Searle's question could be phrased as follows: how could there be a **we-intention** that was not entirely constituted by a series of "I-intentions"?

As noted earlier, one point Searle wants to make is the following negative one: a given **we-intention** is not, or is not simply, constituted by a series of I-intentions. He argues this by reference to the following example. This also introduces his positive proposal, to which I turn shortly.

Imagine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. Each person has the intention expressed by the sentence "I am running to the shelter". But for each person, we may suppose that his or her intention is entirely independent of the intentions and behavior of others. In this case there is no collective behavior: there is just a sequence of individual acts that happen to converge on a common goal. Now imagine a case where a group of people in a park converge on a common point as a piece of collective behavior. Imagine that they are part of an outdoor ballet where the choreography calls for the entire corps de ballet to converge on a common point. We can even imagine that the external bodily movements are indistinguishable in the two cases. . . . Externally observed, the two cases are indistinguishable, but they are clearly different internally.

Searle considers the difference between the ballet case and the case of those running for shelter, and makes several points on this score. I shall focus on the first: in the case of those running for shelter

each person has *an intention that he could express without reference to the others*, even in a case where each has mutual knowledge of the intentions of the others.[15]

whereas in the ballet case

The individual "I intend"s are in a way we will need to explain, derivative from the "we intend"s.[16]

What does Searle have in mind by "the 'we intend's" here? In the previous place where he wrote of "a 'we intend'", he seemed to refer to a "group mental phenomenon", a **we-intention** in my terms. Now he is suggesting that, in the

[14] Searle, op. cit. p. 402.
[15] Searle, op. cit. p. 403, emphasis mine.
[16] Ibid.

ballet case, there are several "we intends" present.[17] In addition, he is suggesting that there are individual "I intends" in the picture that are derivative from these "we intends".

It seems, then, that he is talking about something other than **we-intentions** at this point. For presumably, there is but one relevant **we-intention** in the ballet case. As the dancers would put it: "We intend to converge on a common point. That is our intention." Just as in the individual case it makes little sense to say that, at some one point in time, I intend, many times over, to do such-and-such; so it makes little sense in the collective case to say that, at some point in time, we intend, many times over, to do such-and-such.

With respect to the supposedly derivative status of one or more "I intends", had Searle not written of "we intends", in the plural, one might think he is saying this: when there is a **we-intention**, individual "I-intentions" somehow derive from it.[18] That, however, is not what he seems to be saying.

According to Searle there are several "we intends" in the picture, each corresponding to one of the dancers. Each dancer, that is, has his own "we intend". I take Searle's suggestion to be that each one's "I intend" derives from his own "we intend". So one is led to ask, what is it for an individual to have a "we intend"?

What Searle goes on to say suggests roughly the following. To have one's own "we intend" or, as I shall henceforth put it, one's own we-intention (another phrase Searle uses in the same context) is to be in a special state, one corresponding in some way to sincere utterances of the form "We intend to perform act A", "We are doing A" and the like. Crucially, this state is a state of an individual human being.[19] It is expressed or given voice when one says, sincerely, "We intend. . . ."[20] Searle takes it, I should add, that one may be in this state

---

[17] I am assuming, evidently, that this is not a misprint. There are quite a few misprints in the original publication of this article, so this assumption is perhaps problematic. As Savas Tsohatzidis has pointed out to me, however, the pertinent elements in this sentence are not changed in the reprint of the article in Searle (2002: 92). And the reading that follows from it squares well with other things Searle says, as will soon be clear. It is true that Searle's next sentence begins with the words. "That is, . . ." and points in a different direction. Searle may, even so, have changed direction himself here. Suffice it to say that I am focusing on one of the things that seem to be going on in this complex paragraph, one which fits well with things Searle says later.

[18] Such a derivation could occur, presumably, by something like the following route. Each party understands that the **we-intention** in question exists, and understands that given this **we-intention**, it is appropriate for him to form certain I-intentions, which he does. To say this leaves open, of course, *how* the pertinent **we-intention** renders certain I-intentions appropriate.

[19] In the technical terminology that I am using here, then, individual human beings, (and, in Searle's view individuals of some other species) have we-intentions. In contrast, a **we-intention** is (in my technical terminology here) the intention of a group—whatever that may amount to. It is not something that a single individual can have.

[20] Cf. Searle, p. 403: "the individual intentionality, expressed by "I am doing act A," is derivative from the collective intentionality, "We are doing act A." Tuomela and Miller (1988: 370) refer to "a we-intention *expressible by* "We shall do X" (or "We will do X"). As they conceive of we-intentions, these are shared by those who do something together—"shared" in the sense that all have *qualitatively* identical we-intentions.

without expressing it in language. It is, however, expressible by a sincere "We intend. . . ."

I have said that, in my construal of Searle, we-intentions correspond to *sincere* utterances of the form "We intend. . . ." This is because I assume Searle would not say that a we-intention was present in the case where someone says "We intend . . ." with the intention to deceive another as to what is going on. This person is also ready to say "It is not the case that we intend to do A" to one he does not want to deceive. It is this kind of case that I am assuming should be ruled out.

What I have said so far about we-intentions as Searle understands these may provoke the following response. Should I not have said something like this: the special state in question corresponds in some way to sincere *and correct* utterances of the form "We intend . . ."? Should I not have said, more precisely, that a we-intention is the state someone is in *when there is a* **we-intention** and he is one of "us"—a state he could express by sincerely saying "We intend . . ."? It is clear from what follows in Searle's text that he does not want to include such a condition on we-intentions.

That said, one might immediately wonder if there is indeed a unique state, a particular mode of being, to which all sincere utterances of "We intend . . ." correspond. Surely some such utterances will be correct, justified, responses to the way things are. They will have a particular, sui generis, type of causal history. Others will not have that type of causal history. They may have a variety of widely different causal histories—they may be based on a variety of mistakes, dreams, hallucinations, and so on. Perhaps, though, there is a unique state common to all of those who do or could sincerely express themselves by saying something of the form "We intend . . ."

It may next be pointed out that if, for instance, I am the member of an orchestra in concert, and I sincerely say to my colleague (who seems to be under a misconception) "We are playing Beethoven's *Ninth*!", I would normally be understood to be referring neither to a state of mine nor a state of his, nor to a state of mine, on the one hand, and his, on the other, but rather to a state, or situation, of ours. I am saying precisely that *we* are doing something—all 30 of us, or however many we are.

I assume that Searle does not mean to deny this, but that he wishes to focus on the psychological state one is oneself in when one sincerely refers to what *we* are doing or to what *we* intend.

Though Searle does not, to my knowledge, emphasize this contrast, I assume he would insist that a we-intention is significantly different from what one might call (perhaps inappositely) a *they-intention*, which is (by definition) the psychological state one is in when one sincerely and warrantably refers to what *they* are doing or what *they* intend. This is interesting because, presumably, one and the same fact—the existence of a particular **we-intention**—is referred to when people say "We intend to converge on a common point" and "They intend to converge on a common point", provided that the same people are included

within the "We" and the "They". The fact that I am (as a matter of logic) included within the "we" is presumably the key to the difference.

Suppose the account that I have so far given of Searle's concept of a we-intention is correct as far as it goes. The following questions arise: what is a we-intention? What kind of state is it?

A natural response is that it is a kind of belief. Certainly, one might express one's belief that one's group had a certain intention by saying something of the form "We intend to. . . ." Another possibility, though perhaps a more difficult one to grasp, is that it is a kind of intention.

I say that the latter possibility is "perhaps a more difficult one to grasp" because one might think at first blush that a person could only express his own intention and his own belief. If I can't express your intention (as opposed to saying or claiming (and believing) that you have the intention), how can I express our intention (as opposed to saying or claiming (and believing) that we have it), if "we" includes you as well as me? I don't say this question cannot be given a satisfactory answer, but it is a question that is likely to arise.

Searle refers to a we-intention as involving a particular kind of "intentionality": "collective intentionality".[21] This is not conclusive, however, given Searle's broad construal of the term "intentionality". Nonetheless, it does seem that Searle is inclined to see a we-intention as a kind of intention or, to put it more circumspectly, as akin to an I-intention in important ways. This interpretation is born out by what Searle says later in the article.

3.2. Searle's Theses

I have now covered much of the discussion whose main points Searle summarizes in terms of three theses.[22] Thesis 1 is as follows: "There really is such a thing as collective intentional behavior that is not the same as the summation of individual intentional behavior."[23] This thesis, I take it, concerns **we-intentions**. That is evidenced by the reference to a "summation" of individual intentional behavior.

Thesis 2 runs thus: "we-intentions cannot be analyzed into sets of I-intentions, even I-intentions supplemented with beliefs, including mutual beliefs, about the intentions of other members of the group."[24] That this also concerns **we-intentions** is evidenced by the reference to "sets" of I-intentions. It is then much the same as Thesis 1, but amplified by reference to I-intentions that are supplemented by mutual beliefs about the intentions of other members.[25]

---

[21] Searle, p. 403.

[22] I shall not discuss a fourth and fifth theses that are stated in the last section of Searle's article.

[23] Searle, p. 402.

[24] Ibid., p. 404.

[25] This thesis is stated prior to the discussion rejecting the account of "we-intending" in Tuomela and Miller (1988). It is worth noting that Tuomela and Miller were not attempting to give an account of **we-intentions**, which is what Thesis 2 concerns. Tuomela has responded to Searle's criticisms in a number of places including Tuomela (1995).

Searle prefaces Thesis 3 with the following two points or, as he sees them, constraints on an acceptable account of—anything. First: *society* contains nothing but individuals, and *consciousness* is "in individual minds, in individual brains". Second, "all *intentionality*, whether collective or individual, could be had by a brain in a vat or by a set of brains in vats."[26]

Thesis 3 is: "The thesis that we-intentions are a primitive form of intentionality, not reducible to I-intentions plus mutual beliefs, is consistent with these two constraints."[27] Here the reference to I-intentions and *mutual* beliefs seems to hark back to Thesis 2, and the topic of **we-intentions**. Searle goes on, however, to focus on the we-intentions of particular individuals. It is possible, therefore, that his point would have been better captured with the omission of "mutual".

What he goes on to say is this:

> We simply have to recognize that there are intentions whose form is: We intend that we perform act A; and such an intention can exist in the mind of each individual agent who is acting as part of the collective. . . .[28]

In order that there be collective behavior, Searle says, each individual member of the supposed collective must have an appropriate we-intention "in his head".

If I have such a we-intention, Searle asserts, I take it that "my collective intentionality is in fact shared . . . I am not simply acting alone." I could, however, be mistaken. The "apparent presence and cooperation of other people" could be an illusion.

From what Searle says, he thinks that my mistake may be far more radical than a mistake about whether anyone is doing anything with me right now. I could also, it seems, be mistaken in thinking there were other people at all. I myself could be nothing more than a "brain in a vat".[29]

4. SOME OUTSTANDING ISSUES

How far has our understanding of **we-intentions** been advanced by Searle's discussion as that has so far been represented here? Searle's rejection of the idea that **we-intentions** are constituted by a set of I-intentions of the people in question is one I share, for reasons I have advanced elsewhere.[30] That he has an

---

[26] Searle, p. 406–407; emphasis mine.

[27] Ibid., p. 407.

[28] Loc. cit.

[29] Loc. cit., all quotations in this and the previous paragraph. Searle does not here spell out the "brain in the vat" idea. It is something like this: the said brain generates experiences qualitatively akin to those it would have if it were in contact with various other things and people—but it is not.

[30] See, e.g., Gilbert (1989, 1990). In these discussions I reject accounts of acting together in terms of personal goals (I-goals). The same points can be made, mutatis mutandis, of personal intentions. See Gilbert (1997).

acceptable positive account of what a **we-intention** is, however, is not obvious. As his discussion develops, indeed, it seems that his main interest is not so much in developing a complete account of **we-intentions**, but rather in emphasizing that the primary constituents of **we-intentions** are we-intentions, not I-intentions.

I shall focus on four main problems regarding what Searle says about **we-intentions**. The first two concern what he says about we-intentions.

4.1. The Spare Characterization of We-Intentions

One thing Searle suggests, I take it, is that an orchestra intends to perform Beethoven's Fifth Symphony only if the members of the orchestra individually we-intend to play that symphony. And this means that each of them is in an intention-like state expressible by the sincere avowal "We intend to perform Beethoven's Fifth Symphony".

Now something like this may be true. But we-intentions are so underdescribed that it is hard to know what to make of it. I don't say that it is "mysterious" in that pejorative sense that philosophers often use when they want to rule something out of court. It simply begs for more to be said.

Searle might at this point insist that there *is* no more to be said. He has shown, he might say, that we-intentions are a *primitive phenomenon*. One must then ask whether he has really shown this, and, indeed, what it is that he has tried to show.

At the outset of his paper Searle uses the phrase "primitive phenomenon" twice. He says, first:

> Collective intentional behavior is a primitive phenomenon that cannot be analyzed as just the summation of individual intentional behavior;

Now, to say that a phenomenon, or an idea, is primitive can mean that it cannot be analyzed at all. There's nothing to be said about it. A rose is a rose is a rose, so to speak. Searle might mean to say this about collective intentional behavior—that is, the phenomenon of **we-intention**—but it seems that he does not, since he does give something of an analysis of **we-intention**, in terms of we-intentions.

His real point in this quotation seems, indeed, not to be that collective intentional behavior cannot be analyzed, but that it cannot be analyzed *as just the summation of individual intentional behavior*. I take his next invocation of primitiveness, immediately following the other, to relate to we-intentions, which are my present concern. He says:

> and collective intentions expressed in the form "we intend to do such-and-such" or "we are doing such-and-such" are also primitive phenomena and cannot be analyzed in terms of individual intentions expressed in the form "I intend to do such-and-such" or "I am doing such-and-such."

If we take this as uttered in the same spirit as the previous invocation of primitiveness, we will take it as largely a negative point about how we-intentions are to be analyzed. They are *not* a matter of what I intend. They are not, in other words, I-intentions.[31]

Suppose, then, that Searle's references to primitiveness are essentially references to what **we-intentions**, and we-intentions, are not. In particular, they are not just a matter of I-intentions, as one might think. It is unclear, in both cases, whether and why we should allow ourselves to stop here. Why should we assume that no more can be said about either of these things?

As to we-intentions, these are the building blocks, for Searle, of **we-intentions**, and according to him, **we-intentions** and similar phenomena lie at the foundation of social institutions. It is obviously important for him to say as much as he can about we-intentions. I am concerned both that he has said very little, and that he has not shown that more cannot be said.[32]

He has, of course, said something. A we-intention is an intention-like state expressible by such a sincere avowal as "We intend to perform Beethoven's Fifth Symphony", and is not to be identified with an I-intention, whether an I-intention to perform Beethoven's Fifth Symphony (which is unlikely on practical grounds) or an I-intention to play a role in a performance of Beethoven's Fifth Symphony in which the relevant others take part.

Suppose that *is* all that can be said about we-intentions as such.[33] One might then ask: when and why would such a state be appropriate to the circumstances? This leads to my second concern.

4.2. The Need to Understand **We-Intentions**

Can we understand what a we-intention is without first understanding what a **we-intention** is? On the face of it, certainly, it would seem that only when there is a particular **we-intention** is a corresponding we-intention in Searle's sense *appropriate*. Someone might dream or hallucinate or simply mistakenly believe in the existence of a particular **we-intention**. He will then be in a position sincerely to express himself by saying "We intend. . . ." Unless, however, he does dream or hallucinate or mistakenly believe that there is a **we-intention** in the picture, he will not be able sincerely to express himself by saying "We intend. . . ." In short, he will not be in a position to we-intend.

---

[31] My interpretation here is supported by the following in Searle (1997: 427): "Collective intentionality is a primitive notion in the sense that it does not reduce to individual intentionality."

[32] Cf. Zaibert (2003: 221).

[33] Searle has written a great deal in other places about intentions and intentionality in general. Thus what he says in Searle (1990) can presumably be amplified so as to reflect his views on these matters. He would say, for instance, that an intention-like state is one with a world-to-mind direction of fit (e.g., Searle 2001: 38). Such points do indeed help to fill out the picture of **we-intentions**—to some extent. I shall not attempt to pursue such filling out here. I doubt that the particular issues I raise would be affected by its details.

It seems, then, that in order fully to understand what a we-intention is one needs to understand what a **we-intention** is.[34] For one needs to know what we-intentions assume or presuppose. In relation to we-intentions themselves, then, this question presses: has Searle provided us with an adequate account of **we-intentions**?

4.3. Searle on **We-Intentions**

Searle may believe that we-intentions are such that one could be a permanently isolated brain in a vat, and still have a we-intention. And there may be problems with this idea relating to the conditions under which one is capable of having certain kinds of thought.[35] It is not necessary to discuss this particular issue, however, in order to see that an account of **we-intentions** in terms of a correlated series of we-intentions is seriously lacking.[36] Yet Searle says little more about **we-intentions** than this.[37]

He does in fact suggest an addition to this picture. By a "correlated series" of we-intentions I mean a series whose elements fit together in the right way. Thus, if one member of the series is a we-intention of individual A that posits individual A and individual B, and only A and B, as the members of the "we", its correlate would be a we-intention of individual B that, also, posits A and B, and only A and B, as the members of the "we". Searle at one point suggests that when there is a **we-intention** each of those with a pertinent we-intention assumes that the other people he posits as members of the "we" also have pertinent we-intentions.[38] So, perhaps one should interpret him as saying more than that a **we-intention** involves a correlated series of we-intentions. Rather it involves a correlated series of we-intentions possessed by individuals each of whom assumes that there is indeed such a correlated series, and not only his own we-intention. Though one who makes such an assumption could, in the general case, be mistaken, in the case of an actual **we-intention**, he is not.

---

[34] Cf. Hornsby 1997: 430 f.

[35] Cf. Meijers (2001). Tollefsen (2004), citing Davidson, alludes to the possibility that a brain in a vat may lack any "we-concept" at all.

[36] See the following paragraph on what I mean by a "correlated series" here.

[37] A similar position may have been taken by Wilfrid Sellars. See, e.g., Sellars (1963) which at one point attempts to probe the "force of such phrases as '*we* intend. . . .'" (203). Though short, the discussion that follows is very suggestive and worth more scrutiny than I can offer here. Sellars appears to anticipate Searle's conception of a we-intention when he writes: ". . . when the concept of a group is 'internalized' as the concept of *us, it becomes a form of consciousness* and, in particular, a form of *intending*." (ibid.: Sellars's emphasis). I take it that this form of intending on the part of the named individuals is at issue when, earlier in the paragraph from which the previous quotation was taken, Sellars addresses **we-intentions**. There he says "We intend . . ." is clearly not the logical sum, so to speak, of "Tom intends . . .," "Dick intends . . .," "Harry intends . . .," etc. Yet "the fewer the people in the group who . . . intend that X do A, the less defensible becomes the statement that the group . . . intends that X do A" (ibid.).

[38] On p. 407 Searle writes, "Of course I take it . . . that my collective intentionality is in fact share; I take it in such cases that I am not simply acting alone."

Even given this amplification, there is still a problem with respect to the account. In order to show this I present some examples.

The first example involves brains in vats, but these are not permanently hallucinated brains. Suppose that we have three brains in three vats, each a long way from the other. They are the brains of three people who once roamed the earth, and knew each other, namely, Bob, Carol, and Ted. Let us suppose that Bob is now to be identified with his brain, Carol with hers, and Ted with his. Bob's brain is Bob as he now is, and so on.

Suppose that each of Bob, Carol, and Ted we-intends to meet the others in San Francisco in December. Bob, for instance, will then express himself thus "We—Carol, Ted, and myself—intend to meet in San Francisco in December." This would not merely be a case of purported reference to others, but rather a case of actual reference.[39] Let us suppose each one assumes the others also we-intend in an analogous way.

This case may already go beyond what Searle would take to be the minimal case of a **we-intention** comprised of correlated we-intentions and suppositions about such correlation. Be that as it may, the conditions so far posited do not suffice to allow one to say, in all seriousness, that Bob, Carol, and Ted collectively intend to meet in December. In other terms, one cannot say that *they* intend to meet in December; that this is *their* intention.

Perhaps the pertinent we-intention of each is a necessary condition or aspect of *their* intending to meet. Perhaps, too, the assumption of each of the pertinent **we-intention** of each is also a necessary condition. I shall leave those questions open. The point is that these things are not sufficient. In short, it is not the case that there is a **we-intention** to do something just in case each of us we-intends, according to Searle's understanding of this, to do that thing, and assumes that the others do also. Something crucial is lacking.

I say this on the basis of my pretheoretical or intuitive understanding of what it is for *us* to have an intention. Intuitively, we do not seem to have anything approaching a complete analysis of the intuitive idea of a **we-intention** here.

It is important to note that one can make the point without reference to the philosophical trope of a brain in a vat. I now turn to an example in which the people in question are understood to be embodied and to live in the world in the usual way.

Suppose Ben is currently thinking, with respect to himself and Elaine, "We intend to get married". Indeed, he expresses himself thus to his parents. Elaine is in a similar position. And each assumes the other would sincerely say the same thing if prompted to do so. If a **we-intention** was a series of correlated

[39] According to Searle, a we-intention need not involve successful as opposed to purported reference: "Collective intentionality in my head can make a purported reference to other members of a collective independently of the question whether or not there actually are such members" (p. 407).

we-intentions, and so on, then it would be the case that there was a **we-intention** to get married, the members of the "we" being Ben and Elaine. But surely the description of the situation so far is not enough to show that they do. If Ben's parents learn that Ben and Elaine have never discussed getting married with one another, they would surely judge Ben's announcement to be inaccurate.[40]

One might, of course, wonder how realistic this example is, irrespective of the embodiment of its participants. It may not be that likely, but it could happen. Ben might believe, on insufficient grounds, that Elaine had indicated to him her readiness to marry him, for instance, and that he had indicated the same to her, as she knew. Perhaps, he had a very vivid dream to that effect and he is unaware, now, that this was just a dream. Or perhaps he is simply making wild inferences from the data available to him, such as a brief smile from Elaine in his direction. Elaine might be in a similar position. Each, thus feels justified in saying what he or she says, but neither is so justified.

I suppose it is even possible that Ben and Elaine get married as a result of this situation. One day Ben meets Elaine and shyly says: "Shall we fix a date for our wedding?" Elaine is not fazed by this, given her mistaken view of the situation, and names a date that appeals to her. At this point they have agreed on a wedding date, and have, consequently, a genuine **we-intention**. As a result of this, we may suppose, they do get married.

The last case shows that even mistaken we-intentions can have significant consequences. Here is another example of that.

Consider again Searle's example of a case in which "a group of people in a park converge on a common point as a piece of collective behavior."[41] Just as a set of similar I-intentions can lead to behaviorally indistinguishable behavior on the part of the individuals involved, so can a set of correlated we-intentions coupled with the shared assumption that such we-intentions exist, when there is as yet no **we-intention** in the picture. There may have been no prior pact, discussion, or communication between the parties. Nor need any one be aware of what the others are doing. Each could be hallucinating in such a way that he is able to move forward without colliding with the others, doing his part, as he thinks, in a collective action. In that case, intuitively, *these people did not do anything together*. They simply acted in a coordinated fashion, on the basis of certain convenient hallucinations.

---

[40] Suppose Ben says to his parents, of himself and Elaine, "We are going to get married", as opposed to "We intend to get married". What Ben says could be taken in part to express an intention of his own, an I-intention, one also expressible by his saying "I intend to bring it about that we get married—and I am going to succeed." What Ben says could also be taken to express a we-intention. If all that is true is that Ben has the I-intention just mentioned, his we-intention will properly be judged to be unwarranted. Something like this happens in the film *The Graduate,* in which the graduate in question is named Ben and the woman he loves is Elaine. At first Ben's parents are overjoyed (thinking he has a warranted we-intention), then they are irritated (realizing that all he has is an I-intention).

[41] Searle, p. 403.

One general conclusion that suggests itself is this: **we-intentions** are not a purely mental phenomenon. They are not all about a mind, or minds. Something needs to happen in the world "outside" our individual minds for the core assumption behind our we-intentions to be true. There is nothing metaphysically suspect about such a statement. It sounds like plain common sense.

4.4. The Special Normativity of **We-Intentions**

Why do **we-intentions** matter? Why should we care to have them, rather than correlated we-intentions and correct assumptions about their existence? If we all mistakenly believe there is a **we-intention**, then presumably we will act and talk to one another accordingly and little will be lost. However, if the mistake is discovered something will be lost—unless the parties have come to the point of creating a **we-intention** as a result of their mistaken we-intentions—as happened in the last stage of the story of Ben and Elaine.

As I have argued elsewhere, when people understand that, as they would put it, *we* intend to do something, they understand themselves to be related to one another in particular ways. In particular, they understand themselves to have certain rights against one another and certain correlative obligations towards one another. These are rights to actions that help to satisfy the **we-intention**, and a consequent entitlement to rebuke one another for not acting appropriately, and to demand appropriate action when it does not look as if it will otherwise be forthcoming.[42]

I believe it to be intuitive that such rights and obligations somehow inhere in **we-intentions**. I have elsewhere offered an account of **we-intentions** that, as I have argued, shows how this is possible. For present purposes the point to be made is that it is hard to argue that such rights and obligations inhere in we-intentions and assumptions about them.

5. PACTS, AGREEMENTS, WE-INTENTIONS, AND **WE-INTENTIONS**

Though there is no space to go into this here, there is reason to think that **we-intentions** are always created by virtue of an agreement or something deeply akin to an agreement that would not count as an agreement proper as these things are normally reckoned.[43] The type of communication involved may be minimal, but it would be greater than is assumed to exist in any case involving brains in vats that are not in contact with one another.

Searle recognizes the pertinence of agreements to the generation of **we-intentions**. Thus consider two cases he describes in a central part of his discussion. In the first case

---

[42] Gilbert (1989, 1990, 1997, etc.), Meijers (2003) endorses the point and sees it as problematic for Searle.

[43] This was the gist of Gilbert (1989: Chap. 4). I argue there and elsewhere that both agreements and other forms of joint activity are a matter of joint commitment. For amplification as to the nature and genesis of joint commitments see, for instance, Gilbert (2005), Gilbert (2006b: Chap. 7).

[Some] business school graduates get together on graduation day and form a pact to the effect that they will all go out together and help humanity by way of each pursuing his own selfish interests.[44]

In the second case

. . . one of the members of a softball team loses his wallet at the game. . . . the members reason that the chances of finding it are best if they each act separately; and each searches for the wallet in his own say, ignoring the others. They then set out in a coordinated and cooperative way to search for the wallet by acting with complete lack of coordination and cooperation. . .

It is surely true, as Searle implies, that the business school graduates and the members of the softball team interact in such a way that a **we-intention** arises in each case. In the wake of their pact, each business school graduate can correctly refer to the whole bunch and say, "We plan to help humanity, by pursuing our selfish interests." And after they have together reasoned their way to a conclusion as to how to search for the wallet, every member of the softball team can say to people they come across such things as this: "My team members and I—we're searching for a lost wallet. The others are way over there."[45]

What is it to enter a pact or agreement? Whatever else it is, it surely involves some form of *communication* between the parties. Thus it is not plausible to argue that it is a matter merely of a confluence of matching psychological states of individuals, nor even of the common knowledge of such states. It may well be a matter of the mutual communication of certain states.[46] It is implausible, however, to conjecture that it is a matter of we-intentions. For if one thing is clear about we-intentions from what Searle says, it is that they are expressible by utterances of the form "We intend . . ." or "We are doing. . . ." A pact is entered into, rather, on the basis of such utterances as "Shall we . . .?" and "Yes, let's". That state, if any, that is expressible by "We intend . . ." comes *later*.

6. CONCLUDING REMARKS

In this article, I have reviewed some key points in Searle's classic article "Collective Intentions and Actions" and noted a number of issues that arise in relation to what he says. Among other things I have noted that if we look closely at the

---

[44] Searle, p. 403; the following quotation is also from here.

[45] I should emphasize that I have been taking Searle's reference to the way the members of the team "reason" to imply that they talk to one another about the missing wallet and arrive at something tantamount to an agreement to separate in order to find it. Without this assumption, or something close to it, the claim that there is a **we-intention** in the situation would be less clear. Indeed, the whole thing could be a matter of each member's reasoning in private and I-intending to go wherever he saw there was a way no one else had taken. That said, I take it, that in both of these cases a **we-intention** emerges from a pact or agreement or something much like one.

[46] See Gilbert (1989, 1990). For concordant discussion see Tuomela (1995), Meijers (2003). For my own view of agreements as joint decisions constituted by a joint commitment with a particular type of content see, for instance, Gilbert (1993).

situations in which people are comfortable saying or implying that there is a **we-intention** in the situation, we find that they associate **we-intentions** with special rights and obligations of the parties. This is evidence that something akin to an agreement lies at the foundation of **we-intentions** generally. It seems that it is best to try to understand this something as best we can before attempting to understand the mental states of the human participants in **we-intentions**, in particular those mental states involved or expressed when someone sincerely says something of the form "We are doing . . .", "We intend. . . ."

I propose, now, that if we automatically accept such initial constraints as Searle proposes, we stand to hamper our investigations. Consider his insistence that consciousness is "in individual minds, in individual brains". There may be a sense in which this is indubitable. But does it mean that an adequate account of the phenomenon people have in mind when they say things of the form "We intend . . ." must be couched in terms of the states of individual minds or brains? How could it mean that? The account sought is an account of the referent or intended referent of a certain form of words in English. Is it plausible to insist that one knows what this must be like in advance? One can say, perhaps, that since "I intend . . ." refers to a state of an individual's mind or brain, then "We intend . . ." may well do so, or it may well refer to a series of individual minds in certain states. It may, but it seems a mistake to assume a priori that it must.

It is in fact entirely plausible to expect that, when coupled with the first person plural pronoun, or indeed, the third person plural pronoun, the verb "intend" refers not just to facts about individual minds, if it refers to those at all, but to facts about the relationship of certain human beings one to another. Such relationships may involve their minds, but also their bodies, their perceptions or knowledge of each other, their communications, and so on. The same goes, of course, for statements of the form "We believe . . .", "We feel remorse over . . .", and so on.

Depending upon the outcome, it might in the end be said, I suppose, that our talk about what we intend, for instance, is not about *intention* at all, since it is not reducible to talk about individual minds and/or brains alone. Alternatively, our assumptions about what intention is, in the general case, may have to be altered.[47]

In sum, if Searle's overall social constructionist project is to have a firm foundation, his discussion of both we-intentions and **we-intentions** needs either to be amplified or amended. This is true, of course, of most discussions of important topics. All of those working in this area must be grateful to Searle for his lively, provocative discussion, and for the salience his work in the area has provided for the philosophy of social phenomena.[48]

---

[47] On this issue, in relation to collective belief, see Gilbert (2002).

[48] I thank Aaron James for discussion and Savas Tsohatzidis for comments on the penultimate draft of this paper.

REFERENCES

Bratman, M. (1993). Shared intention. *Ethics,* 104: 97–113.

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

Gilbert, M. (1990). Walking together: a paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15: 1–14. Reprinted in Gilbert (1996).

Gilbert, M. (1993). Is an agreement an exchange of promises? *Journal of Philosophy*, 90: 627–649. Reprinted in Gilbert (1996).

Gilbert, M. (1996). *Living Together: Rationality, Sociality, and Obligation*, Lanham, MD: Rowman & Littlefield.

Gilbert, M. (1997). What is it for *Us* to intend? In G. Holmstrom-Hintikka and R. Tuomela (eds.), *Contemporary Action Theory*, Vol. 2, Dordrecht, The Netherlands: Reidel. Reprinted in Gilbert (2000a).

Gilbert, M. (1998). In search of sociality. *Philosophical Explorations*, 1: 233–241. Reprinted in Gilbert (2000a).

Gilbert, M. (2000a). *Sociality and Responsibility: New Essays on Plural Subject Theory*, Lanham, MD: Rowman and Littlefield.

Gilbert, M. (2000b). Sociality, unity, objectivity. *Proceedings of the 1998 World Congress of Philosophy: Invited papers*, pp. 153–160.

Gilbert, M. (2001). Joint action. In N.J. Smelser and P.J. Baltes (eds.), *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam: Elsevier Science.

Gilbert, M. (2002). Belief and acceptance as features of groups. *Protosociology*, 16: 35–69.

Gilbert, M. (2005). Shared values, social unity, and liberty. *Public Affairs Quarterly*, 19: 25–29.

Gilbert, M. (2006a). Rationality in collective action. *Philosophy of the Social Sciences*, 36: 3–17.

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

Gilbert, M. (2007). Mutual recognition, common knowledge and joint attention. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson (eds.), *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, www.fil.lu.se/hommageawlodek.

Hornsby, J. (1997). Collectives and intentionality. *Philosophy and Phenomenological Research*, 57: 429–434.

Meijers, A. (2003). Can collective intentionality be individualized? *American Journal of Economics and Sociology*, 62: 167–193.

Rosenberg, J. (1980). *One World and Our Knowledge of It*. Dordrecht, The Netherlands: Reidel.

Searle, J.R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press. Reprinted in Searle (2002).

Searle, J.R. (1995). *The Construction of Social Reality*. New York: The Free Press.

Searle, J.R. (1997). Précis of *The Construction of Social Reality. Philosophy and Phenomenological Research*, 57: 427–428.

Searle, J.R. (2001). *Rationality in Action.* Cambridge, MA: MIT Press.

Searle, J.R. (2002). *Consciousness and Language*. Cambridge: Cambridge University Press.

Searle, J.R. (2006). Social ontology: some basic principles. *Anthropological Theory*, 6:12–29.

Sellars, W. (1963). Imperatives, intentions, and the logic of "ought". In G. Nakhnikian and H.-N. Casteneda (eds.), *Morality and the Language of Conduct*, Detroit, MI: Wayne State University Press.

Taylor, C. (1980). Critical notice of Jonathan Bennett's *Linguistic Behavior*. *Dialogue*, 19: 290–301.

Taylor, C. (1985). *Philosophical Papers*, Vol. I: *Human Agency and Language*. Cambridge: Cambridge University Press.

Tollefsen, D. (2004). Collective intentionality. *The Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/c/coll-int.htm.

Tuomela, R. (1984). *A Theory of Social Action*. Dordrecht, The Netherlands: Reidel.

Tuomela, R. (1995). *The Importance of Us*, Stanford, CA: Stanford University Press.
Tuomela, R. and Miller, K. (1988). We-intentions. *Philosophical Studies*, 53: 367–389.
Zaibert, L. (2003). Collective intentions and collective intentionality. *American Journal of Economics and Sociology*, 62: 209–232.

FOUNDATIONS OF SOCIAL REALITY IN COLLECTIVE INTENTIONAL BEHAVIOR

Kirk Ludwig

INTRODUCTION

The foundation of any account of social reality is an account of the nature of collective behavior, and in particular collective intentional behavior. Social practices and social interaction of any sort involve some form of collective intentional behavior essentially. If we do not understand the latter, we do not understand the former.

Collective intentional behavior is not just the aggregate or summation of individual intentional behavior. John Searle provides a vivid example of this in his paper "Collective Intentions and Actions" (1990; all citations to page numbers alone are to this article in the reprint in Searle 2002).[1]

> Imagine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. Each person has the intention expressed by the sentence "I am running to the shelter." But for each person, we may suppose that his or her intention is independent of the intentions and behavior of others. In this case there is no collective behavior; there is just a sequence of individual acts that happen to converge on a common goal. Now imagine a case where a group of people in a park converge on a common point as a piece of collective behavior. Imagine that they are part of an outdoor ballet where the choreographer calls for the entire corps de ballet to converge on a common point. We can even imagine that the external bodily movements are indistinguishable in the two cases; the people running for shelter make the same types of bodily movements as the ballet dancers. (p. 92)

The same movements, by the same people, may, on different occasions, be something they do intentionally as a group or only as individuals. The fundamental difference must lie with the differing intentions (and perhaps other attitudes) of the individuals on the two occasions, for everything else is constant.

To set out the basic issues that arise in the project of analyzing collective intentional behavior, let us begin by contrasting [1]–[3] with [4]–[6].

---

[1] This paper provides the fullest description of Searle's account of collective intentions. Later discussions, in (Searle 1995, pp. 23–26) and (Searle 1998, pp. 118–120), are shorter summaries of the argument of (Searle 1990).

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 49–71.
© 2007 Springer.

[1] I intend to sing the national anthem
[2] I sang the national anthem
[3] I sang the national anthem intentionally
[4] We intend to sing the national anthem
[5] We sang the national anthem
[6] We sang the national anthem intentionally

There is a reading of each of [4]–[6] on which they express merely that each member of the group picked out by 'we' intended to sing the national anthem, sang it, and sang it intentionally, that is, on which each member of the group could say truly [1]–[3]. But there is also another reading of each of [4]–[6] which implies that what we intend we intend jointly, and what we did and did intentionally we did together and did together intentionally. I will call the former reading of [4]–[6] the distributive reading and the latter the collective reading. To take [5] as an example, the distributive reading can be represented as in [5′].

[5′] Each of us sang the national anthem.

If [5] on its collective reading is true, then [5′] is true, but not vice versa. To understand collective intentional behavior, we need to understand the difference between what [1]–[3] express on the one hand, and what [4]–[6] express, on their collective reading, on the other. This can also be put as the problem of understanding the difference between the distributive and collective readings of [4]–[6].

There are two main problems in the philosophy of collective action.

First, what is the ontology of collective intentional action? In particular, must we admit the existence of irreducibly group or plural agents over and above individual agents? A powerful argument for the existence of group agents derives from the observation that the only difference between [1]–[3] and [4]–[6], on their collective reading, appears to be the presence of a plural subject term in [4]–[6] in place of the singular subject term in [1]–[3]. As the referent of the subject term in [1]–[3] is the intender and agent of the action in the individual case, so it seems, by analogy of form, we must accept that the referent of the subject term in [4]–[6] is the intender and agent in the collective case. Some philosophers have embraced this conclusion and sought to explain how we can make good sense of it (see, e.g., Copp 2006; Gilbert 2000; Tollefsen 2002; Stoutland 1997). With Searle, I think this conclusion should be avoided if at all possible. We start with individuals and their intentions. In certain circumstances, typically those in which they wish to cooperate with one another, individuals conform their intentions to a common purpose. In these circumstances (which remain to be precisely delineated) we say they did something together intentionally. The difficulty is to see how to understand sentences such as [4]–[6] on the collective reading compatibly with this individualistic perspective.

Second, what is the psychology of collective intentional action? There are two principal subquestions that arise here.

First, if there are irreducibly group or plural agents, then it seems we must attribute to them psychological states such as intention, and presumably belief and desire. How can we make sense of this? Along with Searle, I accept a form of the connection principle, according to which nothing is a mental state unless it is a conscious mental state or a disposition, *inter alia*, to produce a conscious mental state. While it seems conceptually possible that there should be a conscious mind supervening on certain groups of individuals in certain circumstances, I think that we have no reason to believe that anything of the sort occurs or that we suppose anything of the sort when we use collective action sentences. This provides one of the most powerful reasons against admitting irreducibly group or plural agents.

Second, even if we eschew collective agents, there is an important and difficult question about the psychology of collective intentional behavior. I will call the sort of intention that an individual has in performing a singular action an I-intention. [1] expresses an I-intention, for example. Whatever is involved in collective action, it involves among other things individuals doing things with intentions. I will call the intention an individual has in participating in a group's doing something intentionally a we-intention. Thus, a we-intention is not an intention of a group, but of an individual who is a member of a group. If [4] is true on its collective reading, then each of the members of the group has a we-intention directed toward singing the national anthem. We need this notion even if we think there are group agents with intentions, for, of course, no one thinks that collective intentional behavior occurs without individual members of the group having intentions connected with their actions which contribute to what the group does. Our second question then is: what is the difference between I-intentions and we-intentions? In particular, are we-intentions a distinctive sort of intention? Can we understand we-intentions in terms of the concepts which we already deploy in understanding individual intentional action?

I will argue that once we understand properly the logical form of plural action sentences on their collective reading, we will be able to see that they do not carry any commitment to irreducibly plural or group agents and that all the resources needed for understanding we-intentions are already available in our understanding of individual intentional action. This is not the same as saying that we can analyze we-intentions in terms of I-intentions—or even I-intentions with mutual beliefs about a common goal. It is to say, however, that no conceptual resources not already deployed in understanding I-intentions are required to understand we-intentions.

I begin by looking at Searle's original and distinctive account of collective intentional behavior. It was through Searle's seminars at Berkeley in the late 1980s that I was originally introduced to the problem of collective intentional action. There is much that I admire and agree with in Searle's account, and I have learned a lot from it. There are also, however, some aspects of it which are not completely clear to me. In addition, if I understand his account correctly, it locates the difference between I-intentions and we-intentions in the

psychological mode of these attitudes rather than in their contents. As I understand it, it is for this reason that Searle holds that collective intentionality is a "primitive phenomenon" (p. 90). I am more optimistic that we can capture what is distinctive about we-intentions by a difference in contents of a single psychological mode.

The paper divides into two parts. In the first part, I look at Searle's account and the motivations for it. I will aim to say how I understand it, and raise some questions about it. This will set the stage for the second part, in which I will present my positive proposal, for it will give us a set of constraints and problem cases which any adequate account should be able to meet. My method will be to take a close look at the proper way of projecting the standard account of the logical form of singular action sentences to plural action sentences in light of the ambiguity between collective and distributive readings of plural action sentences, and then extend this to the analysis of plural intention sentences, such as [4], on their collective reading, to arrive at an account of the structure and content of we-intentions. This will provide us with an account which shows what is distinctive about we-intentions, meets our constraints, takes care of problem cases, and deploys only notions already deployed in our understanding of individual intentional action.

1. SEARLE'S ACCOUNT OF WE-INTENTIONS

Searle places two constraints on any adequate account of we-intentions.

Constraint 1
It must be consistent with the fact that society consists of nothing but individuals. Since society consists entirely of individuals, there cannot be a group mind or group consciousness. All consciousness is in individual minds, in individual brains.

Constraint 2
It must be consistent with the fact that the structure of any individual's intentionality has to be independent of the fact of whether or not he is getting things right, whether or not he is radically mistaken about what is actually occurring. (p. 96)

I think both of these constraints are correct and important. I think the best defense of them against objectors is to show that we can understand collective intentional behavior compatibly with them, for no one would be inclined to deny either one except insofar as they were thought to be incompatible with facts about collective action.

Searle's account of collective intentional behavior aims to meet constraints 1 and 2 compatibly with the thesis that

We-intentions cannot be analyzed into sets of I-intentions, even I-intentions supplemented with beliefs, including mutual beliefs, about the intentions of other members of a group. (p. 93)

I will call this *the irreducibility thesis*.

A question of interpretation arises about the irreducibility thesis, connected with the interpretation of 'I-intentions'. I believe that my characterization of 'we-intentions' as intentions that individuals have in participating in group intentional behavior conforms to Searle's use. There may be a difference in our use of 'I-intentions', however. I have characterized I-intentions as intentions individuals have when involved in individual, as opposed to group, intentional behavior. In some places Searle uses 'I-intention' compatibly with the way in which I have characterized it. However, in some places there is the suggestion that Searle intends 'I-intention' to cover any intention expressed with a sentence of the form 'I intend. . . .' These are not equivalent because the former way of characterizing 'I-intentions' leaves it open that what distinguishes I-intentions from we-intentions has to do with what replaces '. . .' in sentences of the form 'I intend. . . .' My positive proposal in the second part of the paper takes this route. This is important because it determines whether I can accept the irreducibility thesis as Searle intends it. I am inclined, on the whole, to think that Searle intends the official characterization of 'I-intentions' to go by way of reference to the intentions individuals have when engaging in singular action, rather than by way of what sentences of a certain form express. This does not build into the categories in which the discussion is conducted a view about what distinguishes I-intentions from we-intentions. In this case, I can accept the irreducibility thesis as Searle intends it.

Searle's account of we-intentions is developed in two stages, first briefly and informally, and then more fully using the resources first made available in his *Intentionality* (1983).

The informal account is

> . . . there are intentions whose form is: We intend that we perform act *A*; and such an intention can exist in the mind of each individual agent who is acting as part of the collective. In cases like that of the football team each individual will have further intentional content, which in ordinary English he might express in the form "I am doing act *B* as part of our doing act *A*." For example, "I am blocking the defensive end as part of our executing a pass play." We need only note that all the intentionality needed for collective behavior can be possessed by individual agents even though the intentionality in question makes reference to the collective. (pp. 96–97)

On this account, as I understand it, there are two components to the intentionality of individuals participating in group intentional action. First, each individual who is a member of a group which does something intentional has an intention of the form 'We intend that we perform act *A*'. Second, each also has an intention he or she could express using the sentence 'I intend to perform act *B* as part of our performing act *A*' (plus appropriate beliefs). This is the "further intentional content."

The first part of this, however, is not fully clear to me. In saying that 'We intend that we perform act A' is the form of an intention each individual has, it is unclear what state is being attributed to each individual, none of whom

is identical to the group. Each individual could think *we intend that we perform act* A, and intend that he does his part in their performing act A. But this attributes only one *intention* to each of the agents together with a *belief* that they intend to do something. So it is not equivalent to what is said above. Since any intention that an individual has should be attributable to him in a sentence in which the subject term refers to him, I suggest, at least for the moment, that we take it that the intention that each individual has in virtue of participating in a group that does something intentionally could be expressed by him by a sentence of the form 'I intend that we perform act *A*', or, somewhat more generally, 'I intend that we *F*' where *F* is replaced by a verb phrase whose main verb is an action verb. I believe this to be compatible with Searle's detailed development of the account in the second stage. If we interpret the above passage in this way, then it is also compatible with reading the irreducibility thesis with my characterization of I-intentions and we-intentions.

The second stage of Searle's account aims to say in more detail what the form is of the intentions that individuals who participate in collective intentional action have. If what I have just said is right, then this is a matter of making clearer the form of the intention expressed by 'I intend that we F'. Searle's method, which seems to me to be the right one, is to start with the form of singular intentions and ask how it is to be extended to the collective case, compatibly with our two constraints. Searle focuses on intention-in-action, but I think the same issues arise for prior intention, so I will just use the verb 'intend' in what follows. Searle uses the notation 'S(*p*)', where S stands for the type of psychological state, and *p* for its content. This leaves it unclear whose psychological state we are talking about. In the case of collective intentional behavior, it will be helpful to make this explicit. I will therefore use the following notation, which is equivalent to Searle's, except for making explicit whose state we are talking about: *x* intends (*p*).

Let us take now Searle's example of someone who intends to fire a gun by pulling the trigger. In *Intentionality*, Searle introduces a principle for identifying the content of a propositional attitude with mind-to-world direction of fit (e.g., belief) or world-to-mind direction of fit (e.g., desire). The content of the attitude is given by the proposition which has to be true for the propositional attitude to be satisfied, where this is a general term for the positive valuation of a bivalent evaluation of the attitude. For example, a belief is satisfied if it is true, and an intention is satisfied if it is carried out. Then to identify the full content of an intention, we must ask under what conditions it is carried out or not. This is an important principle, which I will make use of in my positive proposal. I will call this *the satisfaction principle*. Searle argues that using this principle reveals an important component of the content of an intention which is not represented in the complement clause of reports of intentions. Consider 'John intends to fire the gun by pulling the trigger'. Since the attitude expressed has a bivalent evaluation condition, its content must be given by a full sentence. In this case, the infinitival phrase indicates that the subject of the complement clause is the same as the subject of

the main clause, namely 'John'. Representing this, we get, 'John intends (he fire the gun by pulling the trigger)', where the 'he' is understood as coreferring with 'John'. But it is not enough for John to carry out his intention that as a matter of fact he fires the gun by pulling the trigger. He may do that accidentally, for example, while reaching into a drawer to retrieve a pair of scissors. For him to carry out his intention, that very intention must play a causal role in what he does. If it is a prior intention, it must lead to what Searle calls an intention-in-action that is coeval with the action, directly about it, and which guides it while it is being carried out. That intention-in-action must itself have the same feature, i.e., it must, to be carried out, be causing the action which it is about.

Thus, we need to fill in the content of the intention above to accommodate this. As Searle articulates it, it goes as follows: John intends (this intention causes it to be the case that the trigger pulls, which causes it to be the case that the gun fires). I understand 'trigger pulls' and 'gun fires' here as indicating event types which could occur independently of any agents.

The question, then, in light of the method, is how this is to be extended to 'We intend to *F*'. As I understand Searle, this in turn is the question how this is to be extended to '*x* intends that we *F*' where *x* takes on as a value an individual, not a group.

Take now an example (Searle's example). Jones and Smith are preparing a hollandaise sauce. They cooperate. Jones stirs while Smith pours, and they intend thereby to mix the sauce. What are the intentions each has in doing this in virtue of which they are doing it together intentionally? As Searle notes, it is not enough that Jones intends (this intention causes it to be the case that the ingredients are stirred) and Smith intends (this intention causes it to be the case that the ingredients are poured), for this is compatible with, for example, Jones noticing that Smith is pouring the ingredients for hollandaise sauce and opportunistically stirring. This could occur without Smith and Jones intentionally making hollandaise sauce. Smith might care less or even fail to attend to what Jones is doing and so not realize they are making hollandaise sauce. It does not help if we add that Jones/Smith intends (this intention causes it to be the case that the ingredients are stirred/poured, which causes it to be the case that the sauce is mixed). For this could fit a scenario in which each thinks he is opportunistically taking advantage of what the other is doing, and, in that case, they would not be making hollandaise sauce together intentionally.

Searle suggests and rejects another proposal, on which there is "a special class of collective intentions and that is all that is needed" (p. 101). He continues, "On this account, from Jones's point of view the intentionality is this:

collective i.a. (this collective i.a. causes: ingredients are stirred,
causes: sauce is mixed)."

The content here is read: this collective intention-in-action causes it to be the case that the ingredients are stirred, which causes it to be the case that the sauce is mixed. Since on Searle's view only individuals have intentionality, I take it that we are to think that each of Jones and Smith has such an intention, so that we

could put it this way (using 'we-intend' in the place 'collective i.a.'): Jones/Smith we-intend (this we-intention causes it to be the case that ingredients are stirred/poured, which causes it to be the case that the sauce is mixed). The complaint Searle has with this is that "it leaves out the fact that Jones[/Smith] is making an individual contribution to a collective goal . . . this account leaves it as mysterious how the collective intentionality can move the body"; each has to "personally . . . intend to do something if the sauce is ever going to get mixed" (p. 101). Since these intentions are intentions of the individuals, it is not clear to me why *how it moves them to action* should be more mysterious than how singular intentions move them to action. However, perhaps the difficulty may be put in this way: intuitively, when Jones and Smith each participates in their mixing the sauce, each of them does have particular individual intentions directed at their parts of what they are doing together, and this is a general feature of group intentional behavior. So what is left out of this account is the role of these individual intentions to do their parts in the collective action.

If individual intentions (or singular intentions) to do their parts in a group action are always present when individuals act as parts of groups, then any we-intention they have (taking this now as a sui generis type) would be satisfied only if the group action is brought about by way of such singular intentions. This suggests then reading into the content of the we-intention that it bring about such singular intentions, and this is the next proposal that Searle considers: Jones/Smith we-intends (this we-intention causes it to be the case that Jones/Smith has an I-intention which causes it to be the case that the sauce is stirred/poured, which causes it to be the case that the sauce is mixed).

The difficulty with this, Searle suggests, is that it misrepresents the relation between the collective intention and the singular intention. If I understand, the basic charge is that intuitively when I intend, in the relevant sense, that we do something, it is not part of what I intend that I form further singular intentions, even if I do form further singular intentions as a result. This seems right. Further intentions arise as a result of our intentions through rational deliberation given that one has settled on the intended end, not as a result of being part of what is intended. So we have still not quite got hold of the right relation between the collective intention and what "I have personally to intend to do" (p. 101).

The next move that Searle makes is original and clever, but I also find it puzzling. Let me try to explain carefully how I think it goes.

We begin by rethinking the account of a singular intention that involves a by-means-of relation, such as the intention expressed by: x intends (this intention causes it to be the case that the trigger pulls, which causes it to be the case that the gun fires). Searle suggests that this does not fully represent the form of the intention. He suggests that this "isn't just any old type of i.a. [intention-in-action], it is an achieve-B-by-means-of-A type of i.a." and that we

> . . . might think of the notation that represents this type of i.a. as containing two free variables, "A" and "B"; and these variables are then bound by clauses inside the brackets that function as nouns. What we are trying to say is that I have an achieve-B-by-means-of-A sort

of intention whose content is that that-the-trigger-pulls-as-A causes it to be the case that-the-gun-fires-as-B. (p. 102)

Searle suggests we can represent this as

i.a. B by means of A (this i.a. causes: A trigger pulls, causes: B gun fires)

I wish to rewrite this to try to make more perspicuous what I take the suggestion to be.

(∃A)(∃B)(Jones/Smith intends B by means of A (this intention causes a trigger pulling A, which causes a gun firing B)

I am not sure this fully corresponds to Searle's intentions. My reasons for this representation are as follows. I have introduced explicit existential quantifiers with wide scope because we need to bind variables both inside and outside the content clause, and so the scope of the binding operators has to cover both. I have not included as part of the quantifiers any restriction on a type (e.g., '(∃A: A is a trigger pulling)'), since that A is a trigger pulling and B a gun firing is properly part of the content of the intention, which may misfire, after all. One might execute the intention but be opening a drawer by pulling on its handle instead of firing a gun by pulling its trigger. One could easily modify this if it were thought to matter. I will postpone further comment on this until after we have seen how it is to be used to help with the problem of we-intentions.

The suggestion based on this view of singular by-means-of intentions is that we think of the we-intention as a by-means-of intention as well, but one where the means is a singular action—not a singular intention—and the goal is the collective goal, and that we write in the nature of the actions as collective and singular in the 'by means of' portion of the verb phrase, as follows:

i.a. collective B by means of singular A (this i.a. causes: A stirred, causes: B mixed)

Rewriting this in accordance with my suggestion above, we have:

(∃A)(∃B)(Jones/Smith intends collective B by means of singular A (this intention causes a stirring/pouring A, which causes a mixing B)

This accomplishes two things: (i) It keeps the idea that the collective intention involves the idea that a collective action is accomplished by singular actions. (ii) It avoids putting this into the content of the intention and makes it rather a modification of the type of intention expressed. (i) takes care of the idea that the collective intention requires for its satisfaction the formation of an individual intention. It does this not by making explicit mention of a singular intention as a means, but of a singular action as a means, which in turn will require the formation of a singular intention. (ii) secures that the type of intention is distinct from singular intentions without relying on a difference in content. This looks desirable because if we put this into the content, as in

Jones/Smith intends (this intention causes it to be the case that there is singular stirring/pouring, which causes collective mixing),

it would again seem subject to counterexamples in which each individual takes opportunistic advantage of what the other is doing, without their intending in the appropriate sense to do something together.

I admire the creativity of Searle's suggestion and how he works his way to this proposal. Yet, I am not at all certain that I understand it. And I have a number of reservations about it. I will now explain what these are. This will motivate an additional constraint on a proper solution. And I will try to meet all the constraints in the second part of this paper.

A key feature of the proposal is to locate the difference between I-intentions and we-intentions not in their content but in the kind of state expressed. In considering this, I want to start first with a proposal Searle rejects, before considering his own proposal. I will raise some objections to this proposal different from those we considered above, which I think don't apply fully to Searle's own, but which will help to set the stage for its discussion.

The proposal that Jones, for example, we-intends (this intention causes it to be the case that the sauce is stirred, which causes it to be mixed) locates the difference with I-intentions not in the content of the intention but in the psychological mode. A first problem here is that this seems more to label a problem than to solve it, for we are left with the question what the difference is between such a we-intention—which is a psychological state had by an individual—and an I-intention, if it does not lie in the content. What account is to be given of this? A second problem is that intuitively the sort of state that individuals have when engaging in intentional cooperative behavior with others is not of a distinct type. Imagine that you are intentionally pushing a car, and then someone else comes and joins you, and you are now pushing the car together intentionally. What happens? There are some changes of course. You come to believe that the other person intends to push the car with you, you believe you are pushing it with him, you accept his help, and you continue pushing it as you were before, intending now to do it with him. But is it really plausible that there is some change in the psychological mode of the intention? If we can find any other solution, it would be preferable. A third problem is that if there were genuinely a distinct kind of intention involved, it would seem plausible that this would be reflected in a distinct verb form, for in other cases where we have differences in psychological modes, we also have a difference in vocabulary to express them. But this requires us to suppose that there is a kind of ambiguity in the verb 'intend' which is not made explicit syntactically, and not just in English. This is not a knockdown argument, but it does give us reason to pause, and to look for another solution. Fourth, from the standpoint of practical deliberation, reasoning about whether to join with others to do something looks to be the same sort of process as reasoning about what to do generally. It looks as if it ought to issue in an intention to do something of the same sort as well—with a content

appropriate for the practical problem and its solution—and then if others decide similarly, we have a group of people who intend to do something together. It looks, in other words, from the standpoint of practical deliberation, like a single type of intention is involved as the issue of practical reasoning about either individual or group action.

However, Searle's final proposal has some important advantages over this initial one. For one of the charges I have leveled above is that it makes it look as if the we-intention is sui generis and not continuous in any intelligible way with singular intentions, and therefore more a label for a problem than a solution to one. But the final proposal in fact represents the form of collective and singular intentions as the same except for a small point. It seeks to make intelligible what the difference is by appeal to a difference in one point in a common structure. Each has the form

x intends . . . by means of ___ (p)

The difference resides in what fills in the '. . .'. In the case of a means-end singular intention, it is 'B' for the first and 'A' for the second, or, we might even write, 'singular B' and 'singular A'. Then for the collective case, the only difference lies in replacing 'singular B' with 'collective B'. This makes the account of we-intentions continuous with the account of singular intentions, and represents sentences attributing we-intentions as sharing the same logical form as sentences attributing singular intentions.

Why am I dissatisfied then? First, I am doubtful of the motivation for the representation of the singular intention as involving a modification of the verb. An intention is a commitment to carrying out a plan of action. If the plan is to bring about B by doing A, to fire a gun by pulling its trigger, then we have said enough about what intention someone has when we represent him as intending (this intention causes it to be the case that a trigger moves which causes it to be the case that a gun fires). The type of intention is not different when the agent simply intends to pull the trigger but not to fire the gun by doing that. If we are not justified in representing the singular means-end intention in the special form Searle suggests, then the suggestion for the collective case is not a simple step from the individual case, and the objections canvassed above to the first proposal apply.

Second, the existential quantifiers that, if my representation of the proposal is correct, are required would make it impossible for someone to intend to do B by doing A without there being some actions corresponding to 'B' and 'A': but if we allow that a brain in a vat could have the same intentionality as any of us, this cannot be a requirement on our having the intentions we have. If I intend to fire a gun by pulling a trigger, and do so, then the values of 'B' and 'A' for me are the firing of the gun and the trigger pulling; but there are no such actions to appeal to for a brain in a vat, though it could have the same intentions I do. This worry extends to the collective intention case. It may be that I have read into Searle's notation something he does not intend. But in that case I am uncertain how to understand the proposal.

Third, setting aside these worries, it is not clear to me that the proposal solves the problem. Since the intentionality in collective intentional behavior consists entirely in states of individuals, the proposal is that when members of a group of people all have intentions with appropriately corresponding contents, they collectively intend to do something, and then if they all carry out their intentions, they do something together intentionally. The intentions, on this proposal, which they all have, are of the form

(∃A)(∃B)(x intends collective B by means of singular A (this intention causes a F-ing A, which causes a G-ing B)

where the same thing goes in for 'B' for each member of the group and different things typically for 'A'. The difficulty I have in mind is that this seems to allow that it could accidentally happen that a group of people have intentions of this form, and do something together, though none of them know about any of the others, and their reasons for thinking that they were doing something with others were ungrounded, confused, or mistaken. Let us suppose that, at three institutions for the criminally insane, in three widely separated locations, three different inmates, Jules, Jim, and Jill, each coincidentally dream of being visited by the angel of God and being told that he or she is to cooperate in a mission to blow up the Golden Gate Bridge. Jules believes he has been assigned the job of planting explosives at the base of the north tower; Jim believes that he has been assigned the job of planting explosives at the base of the south tower; Jill believes that she has been assigned the job of planting explosives halfway along the main span. They all believe that they are working with others, who are to remain unknown to them, to prevent any of them giving away the others if captured. They each make their escapes, make their way to San Francisco, set explosives, coincidentally on the same day, but without coming in contact with each other, and then set them off, again, coincidentally, at about 5 p.m. in the evening, blowing up the Golden Gate Bridge together. Yet, intuitively, they did not do this together intentionally. Something still is missing.

Fourth, is it even clear that the account handles the case of Jones and Smith each taking advantage, as each thinks, of what the other is doing, to mix the hollandaise sauce? The reason the question arises is that in this case each is undoubtedly aiming at their doing something together, i.e., at a collective action, only not one that is collectively intended. They each expect that they will mix the hollandaise sauce, but neither thinks they will do it together intentionally. So they each have a means-end intention to bring it about that they do something by his doing something. It looks then as if, in line with Searle's proposal for how to represent means-end intentions, they each have an intention of the form

i.a. collective B by means of singular A (this i.a., causes: stirring/pouring A, causes mixing B)

But they are not mixing the hollandaise sauce together intentionally.

I do not think that these concerns that I have raised are decisive. Some may rest on misunderstandings. Others may be met by modifying the account in

some way. But I am inclined to ask whether a different approach may prove more fruitful. One thing that is central to my concern about Searle's approach is that it works by way of introducing a distinctive collective mode of intending. Other things being equal, it seems to me that we should look for a way to understand collective intentional behavior without introducing a special mode of intending. The only alternative is to try to understand the difference in terms of the contents of the intentions—and perhaps beliefs—of individuals who are engaged in collective intentional behavior. I therefore propose a third constraint:

Constraint 3
An account of we-intentions must locate the difference between I-intentions and we-intentions in the contents of the intentions, not in their mode.

As in the case of the first two constraints, beyond its intuitive appeal, I think the best defense of this one is to show that it can be met, and I undertake that next.

2. THE LOGICAL FORM OF PLURAL ACTION SENTENCES[2]

Let us return to the contrast between [1]–[3] and [4]–[6] (repeated here) with which we began.

[1] I intend to sing the national anthem
[2] I sang the national anthem
[3] I sang the national anthem intentionally
[4] We intend to sing the national anthem
[5] We sang the national anthem
[6] We sang the national anthem intentionally

I follow Searle in thinking that the proper way to approach the question of the structure of plural action sentences is to determine the proper thing to say about singular action sentences and see what the implications of this are for the plural case. I also follow Searle in rejecting the attribution of psychological states to groups, and so of intentional actions to groups. Further, I accept with Searle the satisfaction principle, and so that the content of an intention is to be determined by what conditions must be met for it to be said properly to be carried out.

The account I will give rests on two things. The first is the event analysis of singular action sentences. The second is the observation made at the outset that plural action sentences are ambiguous between a distributive and collective reading.

I will begin with the analysis of [2] and how to project that to [5]. Then I will turn to [3] and [6], and finally to the vexed case of [1] and [4].

The event analysis of singular action sentences treats [2] as introducing an implicit existential quantifier over events. It was introduced to explain how adverbs contribute to determining the truth conditions of action sentences in Donald Davidson's 1967 paper "The Logical Form of Action Sentences"

[2] In the following I draw on (Ludwig, forthcoming).

(reprinted in (Davidson 2001)). The action verb is treated as having an extra argument place for an event which is bound by the existential quantifier, and the adverbs are treated as contributing predicates of the event variable bound by the quantifier. This explains their compositional structure in terms of standard logical devices. This is now the standard analysis of the logical form of action sentences, and, more generally, of sentences whose main verb is an event verb. In linguistics, it is standard to separate out different case or thematic roles for arguments as conjuncts in the scope of the existential quantifier, and I will follow that practice in the following. In addition, I will explicitly require in the case of singular action sentences that there be a unique agent of the event expressed by the action verb. To keep the formulations less complex, I will not in the following incorporate an analysis of the tense structure of the sentences under analysis. This can be added without affecting any of the conclusions reached below (see Lepore and Ludwig 2003). The resulting analysis of [2] is given in [2a].

[2a] $(\exists e)$(singing(e) and agent_{χ}(e, I) and (no y: y ≠ me)(agent_{χ}(e, y)) and of(e, the national anthem))

I have subscripted 'agent' because action verbs may require a particular form of agency. I take the agency relation to be a determinable, which has various determinate forms. The subscripted 'χ' represents then a parameter which is set by the action verb which determines the particular form of agency required. The need for this is brought out by the contrast between causing someone to die and killing him. If one kills someone, one causes him to die, but it does not follow from the fact that one caused someone to die that one killed him. If I hire an assassin to kill a rival, I cause my rival to die, but it is the assassin who kills him. In this case, the verb 'kill' imposes some additional constraints on the relation of the agent to the event (in this case a death) beyond his just causing the death: the causing must not go primarily by way of another's agency, i.e., it must be not be in this sense indirect. So we may say that 'kill' requires direct agency. There are at least these ways in which one can be an agent of an event: the event can be a primitive action of the agent's, something he does but not by doing anything else (raising one's arm); the event can be something which an action of the agent causes—of which a subvariety is direct causation (as in killing someone); the event can be one which something that the agent does constitutes in whole or in part (adjourning a meeting by saying 'This meeting is hereby adjourned'); the event can be one which is part of something the agent does (contracting one's forearm muscles by clenching one's fist). This suggests a further analysis of 'agent_{χ}(e, x)' as '$(\exists f)$(primitive-agent(f, x) and R_{χ}(f, e))', which we will see is particularly helpful for thinking about plural action sentences. However, for the most part I will elide this further articulation of the agency relation (see Pietroski 2000 for discussion).

If we replace 'I' with 'we' in [2a], we get [5a], which looks like the most straightforward application of this analysis to the case of plural action sentences (using 'Y' as a variable that takes as values whatever we take plural terms

to refer to—as we have to make a change in the quantifier to reflect a change in the sort of referring term that appears on the right of the identity sign).

[5a] (∃e)(singing(e) and $agent_{\chi}$(e, we) and (no Y: Y ≠ us)($agent_{\chi}$(e, Y)) and of(e, the national anthem))

This looks problematic because it seems to commit us to a group being an agent. But it is here that the observation that [5] is ambiguous plays a crucial role in arriving at the right analysis. On the distributive reading of [5], we understand it to be saying not that we sang together, but that, instead, *each of us* sang the national anthem. That is, we interpret 'we' as a restricted quantifier, a quantifier that ranges over members of the group. Thus, we can represent [5] on the distributive reading as

[Each x: x is one of us](x sang the national anthem),

where 'x' is a variable that takes individuals as values. 'x sang the national anthem' gets the analysis represented in [2a]. So this comes out as:

[5b] [Each x: x is one of us](∃e)(singing(e) and $agent_{\chi}$(e, x) and (no y: x ≠ y)($agent_{\chi}$(e, y)) and of(e, the national anthem))

[5b] has two quantifiers in it, and we know that in natural languages where scope relations are not explicitly represented in a sentence typically we can read two quantifiers in a sentence in either order, as in 'Everyone loves someone', which can be read as 'Someone is such that everyone loves him' or 'Everyone is such that there is someone he loves'. What then happens when we construe the 'we' in [5a] as '[each x: x is one of us]' and give the event quantifier wide scope? The idea here will be that in effect we are saying that there is an event such that each of us is an agent of it, and no one who is not one of us is, and that it is a singing of the national anthem. This is straightforward except for how to handle '[no Y: Y ≠ us](agent(e, Y))'. If we treat 'us' here as '[each x: x is one of us]', which binds the argument place occupied by 'us', then we need to also have a "step-down" quantifier from the plural variable 'Y' so that we have a singular quantifier binding the position occupied by 'Y'. Thus, we need to represent this as follows: '[no y: y is one of any Y & [for any x: x is one of us](y ≠ x)](agent(e, y))'. This says no member of any group which is not a member of us is an agent of e. Let us abbreviate this as '[only x ∈ us](agent(e, x))'. Similarly let us abbreviate '[each x: x is one of us]' as '[each x ∈ us]'. This gives us [5c].

[5c] (∃e)(singing(e) and [each x ∈ us]($agent_{\chi}$(e, x)) and [only x ∈ us]($agent_{\chi}$(e, x)) and of(e, the national anthem))

This says then: there is an event which is a singing of the national anthem of which each of us but no others are agents (in a particular way). And that is simply to say that it is something which we did together, i.e., this is the collective reading of [5]. This is motivated by two independently established semantic observations, one about the logical form of singular action sentences, and the

other about how we need to understand the contribution of 'we' to plural action sentences on the distributive reading. Putting these two things together, this account of the collective reading falls out. And it is evident that here we are not committed to any group agents at all. This shows that the argument from the similarity of surface form of [1]–[3] and [4]–[6] discussed in the introduction to plural agents is unsuccessful, for the surface form does not reveal fully the logical form; on the contrary, a proper analysis of the logical form of plural action sentences shows that we are not committed to any but individual agents.

So far so good. Let us turn now to [3] and [6]. In line with the motivation for the event analysis, it will be natural to take 'intentionally' in [3] to be introducing a predicate of the event variable introduced by the event quantifier, and the most straightforward approach is to treat 'intentionally' as introducing the verb 'intend' and the sentence it modifies as providing material which figures in the content of the intention, which will be an intention of the agent of the action at the time of the action. This yields:

[3a] $(\exists e)$(singing(e) and agent_χ(e, I) and (no y: y ≠ me)(agent_χ(e, y)) and of(e, the national anthem) and I intend-[singing(e) and agent_χ(e, I) and (no y: y ≠ me)(agent_χ(e, y)) and of(e, the national anthem)])

The event quantifier binding the variable inside the scope of 'intends' represents the intention as a *de re* intention about the action (i.e., directly about and not via a description), and this is intuitively right, for one could have an intention to do something even at the time one does it without one doing it with that intention, for it might be directed at a different coeval event. I used the '[. . .]' notation here rather than parentheses because I want to indicate that this is not yet a full analysis of the content of the intention. I will come back to this when we turn to [1] and [4]. A fuller analysis of [1] and [4] will then fill in what is left unarticulated in this.

The extension to [6] on the collective reading is straightforward:

[6a] $(\exists e)$[each x ∈ us](singing(e) and agent_χ(e, x) and [only x ∈ us](agent_χ(e, x)) and of(e, the national anthem) and x intends-[singing(e) and [each x ∈ us](agent_χ(e, x)) and [only x ∈ us](agent_χ(e, x)) and of(e, the national anthem)])

Here I have pulled the quantifier binding the second argument place in 'agent_χ(e, x)' out front to let it bind also the subject position of 'intends'. I have not bound the argument positions for agents inside the scope of 'intends' with this initial quantifier, however, for that would, incorrectly, require agents who do something together intentionally to have an intention that just he be an agent of it.

I turn now to the real crux of the matter, [1] and [4]. In the case of [1], the complement in logical form should be represented with a sentence with the same subject as the superordinate clause, and, likewise, [4]. We have rejected genuinely plural agents, and so likewise the idea that we attribute intentions to such agents.

This means that in [4] 'We' should be read distributively, as '[each x ∈ us]'. This gives us [1a] and [4a], where in [4a] 'us' and 'we' are understood to corefer.

[1a] I intend that [I sing the national anthem]
[4a] [Each x ∈ us](x intends that [we sing the national anthem])

The matrix in [4a] has a singular variable in the subject position. When we instantiate [4a] to a member of the group, we get a sentence that attributes to him—on one reading of it—a we-intention. The analysis of we-intentions then comes down to the analysis of the matrix in [4a] on its collective reading. The same verb is evidently used in both [1] and [4]. Therefore, the difference must lie in a difference in the complement. A first observation is that evidently we must not give the complement sentence in [4a] its distributive reading, but rather its collective reading. This gives us [4b].

[4b] [Each x ∈ us](x intends that [(∃e)(singing(e) and [each x ∈ us](agent$_{\chi}$(e, x)) and [only x ∈ us](agent$_{\chi}$(e, x)) and of(e, the national anthem))])

This is not our final analysis. But we can already observe how it handles some cases. This is enough to handle Bratman's (1992: 333) case of my intending that we go to New York, as he puts it, in the Mafia sense—by way of my kidnapping you and putting you in the trunk . For in that case, I do not intend that each of us be an agent of our going to New York, but only me. This also is enough to handle Searle's lovely case (p. 94) involving the graduates of a business school who have been taught Adam Smith's theory of the invisible hand, and who each intend to help humanity by pursuing his own self-interest "without reference to anyone else"—knowing that each of the others so intends and likewise knows that the others do. For in that case, none of them intend that each of them be an agent of the helping, for what they intend they intend without reference to anyone else.

But the analysis is still not adequate. One problem we have to face, a problem which Searle highlights, is how this we-intention is connected with what the agent who has it does. Let us turn our attention back to [1a]. Our analysis of this is incomplete. When we turn to an account of the conditions under which this intention is carried out, we must read more into its content, as we have noted. I am going to represent this slightly differently from the way Searle does, but I think equivalently. For my intention to sing the national anthem to be carried out, it must be the case that as a result of so-intending: I do something primitively which brings it about directly that a singing of the national anthem occurs in accordance with a plan I have or will develop (perhaps in the course of doing it). I add the requirement that it come about as planned to handle difficulties with deviant causal chains. I will suppose that in the case of a prior intention this includes that the action be brought about by the intention evolving into an intention-in-action which guides the action (as needed) while it occurs and which is *de re* about the action. Representing this, we have [1b].

[1b] I intend (as a result of my so intending: (∃e)(singing(e) and (∃f)(primitive-agent(f, I) and $\text{cause}_{(\text{directly})}$(f, e) in accordance with a plan I have or will develop) and (no y: y ≠ me)($\text{agent}_{(\text{directly})}$(e, y)) and of(e, the national anthem)))[3]

As an aside, it is worth noting that if we allow that 'intend' introduces a quantifier over states, as event verbs introduce a quantifier over events, we can make use of that to represent more explicitly what is meant by 'as a result of my so intending'. The idea would be roughly:

(∃*s*)(intention(*s*) & subject(*s*, I) & content(*s*, *s* brings it about directly that (∃e)(singing(e) and (∃f)(primitive-agent(f, I) and $\text{cause}_{(\text{directly})}$(f, e) in accordance with a plan I have or will develop) and (no y: y ≠ me)($\text{agent}_{(\text{directly})}$(e, y)) and of(e, the national anthem))))

In the interests of keeping the length of the formulations from getting out of hand, I will not use this elaboration below.

Now let us turn to [4b]. A difficulty in the case of [4b] is that the complement does not have a term that refers to the subject of the intention, and this signals that its form cannot be read straight off from [1b]. To see what to do, we should then turn our attention to singular intentions with complement clauses which do not refer to the subject of the attitude. Let us then apply the satisfaction principle to [7].

[7] I intend that Sue not drink too much at the party.

For this intention to be carried out, it is not enough that Sue not drink too much at the party. I may invite her, and intend that she not drink too much, but forget about it and do nothing toward that end. It may happen that she does not drink too much, perhaps because she came with that resolve, or left early for another appointment, or was dissuaded by someone else. My intention was not carried out. For my intention that Sue not drink too much to be carried out, I have to do something that has some positive effect on her not drinking too much. I have to do something that helps to secure (in the way planned) *that* she not drink too much. I have to, for example, hide the liquor bottles when I see her glass is empty, or ask her what her New Year's resolutions were (knowing one was not to drink), or leave in a place where she will see it a magazine with a cover story on alcoholism, expecting that these things, in the circumstances, will have the right result. Thus, in logical form, [7] should involve my doing something primitively to bring it about as a result of my so intending, in accordance with a plan I have or will develop, that Sue not drink too much at the party. We can represent this as in [7a].

[7a] I intend (as a result of my so intending: (∃f)(primitive-agent(f, I) and f brings it about in accordance with a plan I have or will develop that Sue not drink too much at the party))

---

[3] A quick note on '$\text{cause}_{(\text{directly})}$(f, e)': this should not be read as 'f is the one and only direct cause of e' but rather as 'f is a direct cause of e'. We need to allow that more than one agent can stand in this relation to an event, as when we lift a bench together.

Return to [4b], repeated here.

[4b] [Each x ∈ us](x intends that [(∃e)(singing(e) and [each x ∈ us](agent$_\chi$(e, x)) and [only x ∈ us](agent$_\chi$(e, x)) and of(e, the national anthem))])

How should we apply the pattern used in [7a] for [7] to [4b]? A first thought is that we should instantiate [4b] to a member of the group—let us say me—and then put the that-clause in the place of 'that Sue not drink too much at the party' in [7b]. This gives [4c].

[4c] I intend (as a result of my so intending: (∃f)(primitive-agent(f, I) and f brings it about in accordance with a plan I have or will develop that [(∃e)(singing(e) and [each x ∈ us](agent$_\chi$(e, x)) and [only x ∈ us](agent$_\chi$(e, x)) and of(e, the national anthem))]))

Unfortunately, this does not work. Seeing why, and where we went wrong, will show, I think, both what the key is to understanding the structure of we-intentions, and why it has been so elusive.

The difficulty is not that I do not intend to do my part. For I do intend that we do something, and that requires me to do my part in it, since I am part of the group. The difficulty rather is that my planning that we do something and your planning that we do something may not amount to our planning to do something. One case of this is that of Jones and Smith where each intends to take advantage, as he thinks, of the other's inattention to do something that will result in their mixing a hollandaise sauce. Another example is provided by two painters who think they differ on which is the best brand of paint to use, Sherman-Williams or Benjamin Moore, in painting a house. They each think the other prefers Benjamin Moore and will insist on using only Benjamin Moore for the project. But each intends that they should paint using only Sherman-Williams. To this end, each plans a deception. Each brings to the work site a Benjamin Moore can filled with Sherman-Williams. Each also brings in the trunk of his car another can of Sherman-Williams. Each then surreptitiously empties the other's Benjamin Moore can and replaces the paint in it with Sherman-Williams paint from the can he had hid in his trunk. They then proceed to paint using Sherman-Williams, each thinking he is fooling the other into doing so with him. They each have an intention which fits the pattern in [4c] with appropriate adjustments for content. But they did not have we-intentions with respect to painting with Sherman-Williams, and *they* did not paint using Sherman-Williams *intentionally*.[4] Intuitively, what is missing is that neither intended that they should be doing something in accordance with a shared plan, but rather each intended something that presupposed that they did not intend to act in accordance with a shared plan.

---

[4] I am not saying that they did not paint the house together intentionally, but that they did not paint it together using Sherman-Williams intentionally.

Where did we go wrong? Return to [4], which should pattern after [1], rather than [7], because it, like [1], has an infinitival complement. Thus, we should use the pattern in [1b], which gives us [4d].

[1] I intend to sing the national anthem
[1b] I intend (as a result of my so intending: (∃e)(singing(e) and (∃f)(primitive-agent(f, I) and $\text{cause}_{(\text{directly})}$(f, e) in accordance with a plan I have or will develop) and (no y: y ≠ me)($\text{agent}_{(\text{directly})}$(e, y)) and of(e, the national anthem)))
[4] We intend to sing the national anthem
[4d] We intend (as a result of our so intending: (∃e)(singing(e) and (∃f)(primitive-agent(f, we) and $\text{cause}_{(\text{directly})}$(f, e) in accordance with a plan we have or will develop) and (no Y: Y ≠ us)($\text{agent}_{(\text{directly})}$(e, Y)) and of(e, the national anthem)))

Here, of course, we want to read the content as involving a collective action, and so representing the first 'we' and the 'us' as quantifying over members of the group and taking narrow scope with respect to the quantifier over events. Likewise 'our so intending' should be read as 'each of us so intending'. In the case of 'a plan we have or will develop', we must read 'we' as it appears here as independent of the first, as distributive, but as taking narrow scope with respect to 'a plan', since the key idea is that there is a shared plan. The subject term of course is interpreted as a quantifier over members of the group. When we put these ideas together, we get:

[4e] [Each w ∈ us](w intends (as a result of each of us so intending: (∃e)(singing(e) and [each x ∈ us](∃f)(primitive-agent(f, x) and [∃p: plan(p) and [each y ∈ us](y has or will develop p)]($\text{cause}_{(\text{directly})}$(f, e) in accordance with p)) and [only z ∈ us]($\text{agent}_{(\text{directly})}$(e, z)) and of(e, the national anthem))))

This says that each of us intends *that* as a result of each of us so intending there is an event of singing of the national anthem of which each of us is a direct agent and there is a plan (the same one) which each of us has or will develop to do this and that no one who is not one of us is a direct agent of this singing.

We still though have to connect what is expressed by [4e] with actions of the individual. We have in [4e] an intention in which the complement would be expressed in English by a that-clause without the subject of it being identical to the subject of the superordinate clause. So we must apply at this point the pattern of analysis appropriate for [7], 'I intend that Sue not drink too much at the party'; that pattern is given in [7a]. Instantiate then [4e] to a member of the group—let it be me—and we have finally [4f].

[7a] I intend (as a result of my so intending: (∃f)(primitive-agent(f, I) and f brings it about in accordance with a plan I have or will develop that Sue not drink too much at the party))

[4f] I intend (as a result of my so intending: (∃f)(primitive-agent(f, I) and f brings it about in accordance with a plan I have or will develop that as a result of each

of us so intending: (∃e)(singing(e) and [each x ∈ us](∃f)(primitive-agent(f, x) and [∃p: plan(p) and [each y ∈ us](y has or will develop p)] (cause$_{(directly)}$(f, e) in accordance with p)) and [only z ∈ us](agent$_{(directly)}$(e, z)) and of(e, the national anthem))))

This is the final analysis of a we-intention.[5] A couple of quick remarks may help to ward off misunderstandings. First, to have such an intention does not require that one think that one has the power to determine what other people intend (see Baier 1997; Stoutland 1997; Velleman 1997). It is no more required in this case than it is in the case of my intending that Sue not drink too much at the party. What is required is that one think that what one can do will, given the rest of how one believes things are or will develop, which may include of course what one takes the dispositions of others to be, have a reasonable chance, weighted by the value of the outcome, of achieving its goal (see Bratman 1999 for discussion). Second, one might worry that the requirement of a shared plan is too strong, for surely when people do things together they often do so without having in full detail exactly the same plan in mind that they will follow. This is true enough, but the requirement is not that everyone share all the same plans in all their detail, but that there is a plan that they share, and at a suitable level of abstraction, e.g., reform campaign financing, this seems easy to meet.

3. CONCLUSION

By way of conclusion, let us ask how this account fares with our examples and with our constraints.

Does the analysis handle the cases that gave us trouble before? Yes. It handles the case of the business school graduates, as noted above, because none of them

[5] It is worth noting that this account of we-intentions includes no mutual belief requirement. That requirement attempts to secure a shared goal, but as Searle's example discussed above involving the business school graduates who have learned the theory of the hidden hand, a mutual belief requirement falls short of what is required. The shared goal is secured through the intentions of the participants in group intentional action. Further, it seems clear that mutual belief is not just too weak in one respect, but too strong in another. Suppose A and B are partners in crime who believe in honor among thieves, and they agree that if they are captured and interrogated by the police, neither will give up the other by confessing but rather stick to their mutually agreed upon cover story. Suppose they are captured and presented with a classic prisoner's dilemma. Each is told that if neither confesses, each will receive some jail time but not commensurate with the crime of which they are accused. If both confess, then each will receive a sentence at the lower end of the range appropriate for the crime. However, if one confesses and the other doesn't, then he will receive immunity while the other will receive a much longer jail sentence, while if he does not and the other does, then he will receive the longer jail sentence and the other will receive immunity. Suppose A and B under the stress of interrogation each come to believe that the other will most likely confess, given the structure of the payoffs. But suppose that each is nonetheless determined to observe the pact he made, hoping the other will as well, but expecting otherwise, and willing to accept the consequences. They both act in accordance with the plan which they had agreed on, and so carry it out together. Intuitively they have acted together intentionally to frustrate the police investigation by sticking to their cover story. But a mutual belief requirement to the effect that the others are doing their parts would rule this out, incorrectly, and so is too strong a requirement on having we-intentions.

intend anything with regard to their all doing something together—this is ruled out by the description. In the case of Jones and Smith taking advantage, as each thinks, of the other's inattention to mix hollandaise sauce, neither intends that they do what they do by way of a shared plan, for each thinks the other unaware of the other's plan that they mix hollandaise sauce. Similarly, in the case of the two painters, neither intends that they should be painting with Sherman-Williams by way of their sharing a plan to do so, since each thinks the other plans (and thinks they plan) something incompatible with that. What about the case of Jules, Jim, and Jill, who each break out of prison thinking they are blowing up the Golden Gate Bridge with other saboteurs and do so together? What they lack are intentions that are about the group that actually blows up the Golden Gate Bridge. None of them has any way of thinking of them as a group. Their dreams of being approached by the angel of God provide no route in thought that connects them with *any* other people engaged in a task with them. And so their intentions are not about any group and so not about *them*. Contrast this case with one in which someone approaches the three of them independently and gives each a coordinated role in a plan to blow up the Golden Gate Bridge, indicating that there are others who will participate but who must remain unknown. In this case, they do have a way of thinking about the group of which they are a part, namely, as the individuals the organizer has recruited for the job. In this latter case, if they blow up the bridge according to plan, they do it intentionally.

Does this analysis meet our three constraints? Yes. It is clearly consistent with the fact that society consists of nothing but individuals, and requires no psychological states except those of individuals. In the case of the requirement that it be consistent with an individual's intentions being independent of whether he is getting things right, the account is certainly consistent with no one else having the plans or intentions that he intends that they have. It will be consistent with no one else existing provided that we take the appearances of the plural referring terms in the representation of the content of the intentions to involve in the individual's thought some sort of mode of presentation of the members of the relevant group that makes no *de re* reference to anything external to the individual. And it is clearly consistent with the third constraint, for the distinction between I-intentions and we-intentions has been drawn by way of a difference in the contents of the relevant sorts of intentions.

Finally, if it is successful, it also shows that we-intentions can be understood without invoking any concepts which are not already available in our understanding of I-intentions, for all of the concepts expressed by terms used in the analysis of we-intentions are used in the analysis of I-intentions. Yet given how we-intentions and I-intentions are introduced, it does this without reducing we-intentions to I-intentions, or to I-intentions together with mutual beliefs.

REFERENCES

Baier, A. (1997). *The Commons of the Mind*. Chicago: Open Court.

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2): 327–341.

Bratman, M. (1999). I intend that we *J*. In his *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

Copp, D. (2006). On the agency of certain collective entities: an argument from 'normative autonomy'. *Midwest Studies in Philosophy*, 30: 194–221.

Davidson, D. (2001). The logical form of action sentences. In his *Essays on Actions and Events,* Second Edition. Oxford: Clarendon Press. Originally published in N. Rescher (ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press, 1967.

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

Ludwig, K. (forthcoming). Collective intentional behavior from the standpoint of semantics. *Noûs*.

Lepore, E., and Ludwig, K. (2003). Outline of a truth conditional semantics for tense. In Q. Smith and A. Jokic (eds.), *Tense, Time and Reference*. Cambridge, MA: MIT Press.

Pietroski, P. M. (2000). *Causing Actions*. Oxford: Oxford University Press.

Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, J. R. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication*. Cambridge, MA.: MIT Press. Reprinted in Searle 2002.

Searle, J. R. (1995). *The Construction of Social Reality*. New York: Free Press.

Searle, J. R. (1998). *Mind, Language, and Society: Philosophy in the Real World*. New York: Basic Books.

Searle, J. R. (2002). *Consciousness and Language*. Cambridge: Cambridge University Press.

Stoutland, F. (1997). Why are philosophers of action so anti-social? In L. Alanen, S. Heinamaa, and T. Wallgreen (eds.), *Commonality and Particularity in Ethics*. New York: St. Martin's Press.

Tollefsen, D. (2002). Collective intentionality and the social sciences. *Philosophy of the Social Sciences*, 32(1): 25–50.

Velleman, D. (1997). How to share an intention. *Philosophy and Phenomenological Research,* 57(1): 29–49.

JOINT ACTION: THE INDIVIDUAL STRIKES BACK

Seumas Miller

Joint actions are actions involving a number of agents performing interdependent actions in order to realise some common goals. Examples of joint action are: two people dancing together, a number of tradesmen building a house and a group of robbers burgling a house. Joint action is to be distinguished from individual action on the one hand, and from the 'actions' of corporate bodies on the other. Thus an individual walking down the road or shooting at a target are instances of individual action. A nation declaring war or a government taking legal action against a public company are instances of *corporate* action.

Over the last decade or two a number of analyses of joint action have emerged, of which John Searle's is one of the most important.[1] These analyses can be located on a spectrum at one end of which there is so-called (by Frederick Schmitt 2003) strict individualism, and at the other end of which there is so-called (again by Schmitt [ibid.]) supra-individualism. Roughly speaking, Searle is somewhere in the middle of this spectrum.

A number of these theorists have developed and applied their favoured basic accounts of joint action in order to account for a range of social phenomena, including conventions, social norms and social institutions. Moreover, a variety of arguments have been offered for and against many of these differing accounts. At the risk of oversimplification, I suggest that the tendency has been to eschew individualist accounts in favour of supra-individualist accounts, or at least in favour of the occupancy of hoped for anti-reductionist middle ground.

In this chapter I defend a relatively strong form of individualism, namely the Collective End Theory (CET)[2] against arguments emanating from the supra-individualists and from their fellow traveller anti-reductionists, such as Searle, who try to occupy middle ground.

Individualism, as I see it, is committed to an analysis of joint action such that ultimately a joint action consists of: (1) a number of singular actions; (2) relations between these singular actions. Moreover, the constitutive attitudes involved in joint actions are individual attitudes; there are no sui generis we-attitudes. Here, it is important to stress that individualism can be, and in the case of CET certainly

[1] See Tuomela and Miller (1988); Tuomela (1995, 2002); Gilbert (1989: especially Sect. IV); Searle (1990, 1995: especially Chaps. 1 and 2); Cohen and Levesque (1991); Bratman (1992, 1993); and Miller (1992a, 1995, 2001: Chap. 2).

[2] For a detailed elaboration of CET, see Miller (2001: Chap. 2; 1992a). For earlier applications of this account in relation to conventions, speech acts and social contracts, see Miller (1986a, 1986b, 1987, 1992b).

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 73–92.
© 2007 Springer.

is, a form of relationalism. It is relational in two senses. First, as mentioned earlier, singular actions often stand in relations to one another, e.g. two partners dancing, and the joint action in part consisting of the singular actions, also in part consists of the relations between the singular actions. Second, the agents who perform joint actions have intersubjective attitudes to one another, e.g. they look one another in the eye. In virtue of such intersubjective attitudes, they will also typically have interpersonal relations to one another. But intersubjectivity and interpersonal relations in this sense are not necessarily, or at least are not by definition, social or institutional. To suggest otherwise would be to beg the question against individualism in any interesting sense of the term.

By contrast, according to supra-individualists when a plurality of individual agents perform a joint action, the agents have the relevant propositional attitudes (beliefs, intentions, etc.) in an irreducible 'we-form' which is sui generis, and as such not analysable in terms of individual or I-attitudes. Moreover, the individual agents constitute a new entity, a supra-individual entity not reducible to the individual agents and the relations between them.

Individualism can be, and has been, targeted by supra-individualist or, at least, anti-reductionist style arguments of a number of kinds. Some of these arguments miss their mark by virtue of the fact that they ascribe claims or positions to individualism that it does not make or need not occupy.

For example, it is sometimes argued that individualism must hold that the attitudes involved in joint actions must not make any reference to irreducibly collective entities. However, this kind of argument confuses irreducibly collective *entities* with irreducibly collective *attitudes*. An individualist account of joint action does not need to hold the implausible view that there are no irreducibly collective entities, e.g. Australia, the USA, Finland and the European Union. Clearly, there are irreducibly collective entities and such entities can and do figure in the content of individual attitudes, e.g. John Searle believes that Finland but not Australia or the USA is a member of the European Union. However, the existence of collective entities does not entail the existence of collective *agents*, and specifically, it does not entail the existence of supra-individual or we-attitudes.

A second kind of argument pertains to normativity. It is argued that joint actions are inherently normative phenomenon and that individualists deny this. Here, there is often a confusion between types of normativity. On any account, individualist or otherwise, joint actions involve beliefs, intentions, ends and the like, i.e. attitudes that are normative in the weak sense that they are subject to some form of evaluation, e.g. truth and falsity, success and failure. So much individualists can and do acknowledge. Moreover, on any account, individualist or otherwise, many joint actions are *as a matter of contingent fact* going to be normative in some stronger sense, e.g. possess institutional properties. However, the critical question is whether or not joint actions are *necessarily* normative in some stronger sense that would be antithetical to individualism, e.g. *necessarily* have *institutional* properties. Individualists deny that joint actions necessarily have institutional properties, and have offered counterexamples by way of support. For example, two strangers from different cultures travelling in opposite directions

whose common pathway is blocked by a large fallen tree might spontaneously jointly act to remove the tree; but this might be a one-off joint action not governed by any institutional rule or regularity. I conclude that the charge against individualism that it has a normative deficit is at best not sustained.

Other arguments against individualist accounts of joint action get it more or less right about individualism per se, but get it wrong about the power and flexibility of individualism to accommodate a variety of features of social forms. Here the work of John Searle is salient. According to Searle, "In addition to individual intentionality there is collective intentionality. Obvious examples are cases where I am doing something only as part of *our* doing something. . . . In my view all these efforts to reduce collective intentionality to individual intentionality fail" (1995: 23f.). An important claim here is that individualism cannot deal with the fact that an individual agent sometimes acts qua member of a group. One of my aims in this chapter is to show that, contra Searle and others, CET can accommodate the concept of acting qua member of a group.

Still other arguments are directed at the details of the analysis of joint action provided by a given specific individualist account. In this chapter, I also consider the arguments propounded by Frederick Schmitt against the collective end theory of joint action (cf. Schmitt 2003.). Some of these arguments pertain to the interdependence of participant's actions in joint action. Schmitt argues that such interdependence is not a necessary condition for joint action. Other arguments are directed at my individualist notion of a collective end, e.g. Schmitt argues that collective ends are not a necessary condition for joint action. More importantly, given the individualist concerns in this chapter, Schmitt provides alleged counterexamples to the view that CET provides a sufficient condition for joint action. Here the argument against CET is as follows: Because CET refuses to acknowledge the defining non-individualist properties of joint action, e.g. Searle's sui generis property of collective intention, inevitably examples of non-joint individual action will be found—in this instance allegedly by Schmitt—which CET will wrongly identify as instances of joint action. The inference Schmitt draws from this is the same as that drawn by Searle in relation to individualist accounts Searle has considered: namely, that since individualism cannot adequately capture the concept of joint action then there is a need to abandon individualism in favour of some form of supra-individualism. I do not find Schmitt's arguments against the collective end theory compelling, and offer various counter-arguments in response. Before turning to the arguments of the anti-individualists, I need to outline CET.

1. THE COLLECTIVE END THEORY OF JOINT ACTION

Basically, CET is the theory that joint actions are actions directed to the realisation of a collective end.[3] However, this notion of a collective end is a construction out of the prior notion of an individual end. Roughly speaking, a collective end

[3] The notion of a collective end was introduced in Miller (1982: 6). See Miller (1986a: Sect. 6; 1986b: 133f.). See also Searle (1990) and Gilbert (1989: Sect. IV).

is an individual end more than one agent has, and which is such that, if it is realised, it is realised by all, or most, of the actions of the agents involved; the individual action of any given agent is only part of the means by which the end is realised. (See below for more detail on collective ends.)

So, a joint action simply consists of at least two individual actions directed to the realisation of a collective end. Accordingly, individual actions, x and y, performed by agents A and B (respectively) in situation s, constitute a joint action if and only if:

(1) A intentionally performs x in s and B intentionally performs y in s.
(2) A x-s in s if and only if (he believes[4]) B has y-ed, is y-ing or will y in s and B y-s in s if and only if (he believes) A x-s or is x-ing or will x in s.
(3) A has end e, and A x-s in s in order to realise e and B has end e, and B y-s in s in order to realise e.
(4) A and B each mutually truly believes that A has performed, is performing or will perform x in s and that B has performed, is performing or will perform y in s.
(5) Each agent mutually truly believes (2) and (3).

In respect of this account the following points need to be noted.

First, with respect to clause (2), the conditionality of the action is internal to the agent in the sense that if the agent has performed a conditional action then the agent has performed the action in the *belief* that the condition obtains. Moreover, the conditionality of the agent's action is relative to the collective end. A x-s *only if* B y-s *relative to the collective end e*. So B's y-ing is not a necessary condition for A's x-ing, *tout court*. For example, it might be that A has some other individual end e1, and that A's x-ing realises e1 (in addition to e). If so, A will x, even if B does not y. But it remains true that A x-s only if B y-s *relative to the collective end e*. Again, A x-s *if* B y-s *relative to the collective end, e*. So B's y-ing is not a sufficient condition for A's x-ing, *tout court*. If for some reason A abandons the end e, then even if B y-s, A will not x.

Second, the notion of a collective end is an individualist notion. The realisation of the collective end is the bringing into existence of a state of affairs. Each agent has this state of affairs as an individual end. (It is also a state of affairs aimed at under more or less the same description by each agent.) So a collective end is a species of individual end. Thus CET is to be distinguished from the accounts of theorists such as Searle who favour irreducibly collectivist notions of collective intentions.[5] A collective end is the same as an ordinary individual end in that qua end it exists only in the heads of individual agents. But it is

[4] Something weaker than belief is probably sufficient here; perhaps thinks it is likely, or even thinks it is quite possible. Ditto for the other beliefs—including the mutual true beliefs—involved in the definition. However, absence of any degree of belief would not be sufficient. For if the first agent did not think there was any chance the other agent would do his part, then how could the first agent be thought to have the end in question. After all, the end can only be achieved if both agents do their parts.

[5] See note 2.

different from an ordinary individual end in a number of respects. For one thing, it is a *shared* end. By shared end I do *not* mean a set of individual ends that would be realised by qualitatively identical, but *numerically different,* multiple states of affairs. Rather, I mean a set of individual ends that would be realised by *one single* state of affairs; so the coming into existence of the one single state of affairs in question constitutes the realisation of *all* of the individual ends that exist in the heads of each of the agents. For another thing, although collective ends are shared (individual) ends, they are ends that are *necessarily* shared by virtue of being interdependent; they are not ends that are shared only as a matter of *contingent* fact. Suppose you and I are soldiers being shot at by a single sniper. Suppose further that we both happen to see the sniper at the same time and both fire at the sniper in order to kill him. I have an individual end, and you have an individual end. Moreover, my individual end is one that I share with you; for the death of this one sniper will not only realise my individual end, it will simultaneously realise your individual end.[6] However, this is not yet a collective end; for the fact that I have as an end the death of sniper is not dependent on your having as an end the death of that very same sniper, and vice versa for you. A final point about collective ends is that they are 'open' to those that have them; collective ends are objects of mutual true belief. So if end e is a collective end, then the agents who possess e, mutually truly believe that they each have e.

Third, a collective end is not an intention. On CET, ends are distinguished from intentions. The notion of an end is coordinate with that of a means, but this is not so for intentions. To say that an action was intended is not in itself to say that the action is a means or that it is an end. Indeed, some intended actions do not involve means or ends. Consider a gratuitous act of arm raising. I intentionally raise my arm but for no purpose; I just do it, albeit intentionally. I have no *end* in so raising my arm. Moreover, I have no *means*; I do not raise my arm by means of, for example, a pulley. Rather, my arm is under my direct control, and intending to raise it is sufficient for it to be raised.

Fourth, CET relies on a fourfold distinction between ends, intentions, beliefs and desires, but makes use only of ends, intentions and beliefs.[7] So CET

---

[6] It might be argued that the two individual ends are not the same in that necessarily I have as an end not only that the sniper die, but that I kill him. (And likewise for your individual end.) (See Searle 1983, especially Chap. 4.) But my example is intended to be illustrative only; it serves this purpose even if it does not correctly describes reality. On the other hand, if this claim is true it will cause me to revise my account of collective ends; strictly speaking they will consist of a set of different individual ends to the effect that each have as an end not only that some state of affairs obtain, but that he contribute in part to the realisation of that state of affairs. However, this revision to the content of collective ends would not be fatal to my account. In any case I am already committed to the claim that ends bring with them means in the sense that an agent performs the means in the belief that it is the means, or part of the means, to his end.

[7] In fact, I distinguish between two sorts or 'intending', trying and intending proper. The latter but not the former entails a belief that one can succeed in the action 'intended'. This kind of distinction is a familiar one in the literature. See, for example, Bratman (1987: 136). It is of course also true that the end of some intentional action can itself become an intended action. See Searle (1983: Chap. 1).

does not have as constitutive elements affective attitudes such as desires or preferences. Having a desire (or preference) is not the same thing, on this view, as having an end. Ends, like intentions, are conative states. On the other hand, in distinguishing between ends and affective attitudes such as desires, the claim is not that there are ends that are neither directly nor indirectly the object of some motivating affective state such as a desire.

Fifth, CET, as outlined above, constitutes a core theory of joint action. But this core theory can be extended to cover a range of different cases of joint action in respect of which the following points can be made (cf. Miller 1992a): (i) There are many types of joint action in which it is not the case that the performance of a contributory action by any of the participating agents is dependent on the performance of contributory actions by *all* of the other participating agents. (ii) The actions x, y, etc. performed by participating individual agents might be of the same type. Each agent might have to perform the same kind of action. Further the 'action' might not even be a single discrete action. It might be a set of actions of the same or different types, or it might be an activity or even a whole enterprise. (iii) The collective end might not be a state of affairs but rather a process. And this end might relate to the actions performed to realise it as a consequence or a result or as a whole to its parts. (iv) The realisation of the collective end can itself be simply a means to some further end, or it can be an end in itself or it can be both means and end. (v) The collective end can be a lower order collective end or a higher order collective end, that is, a collective end with respect to other collective ends, e.g. the members of the executive of a political party might have as a collective end to reinforce the existing collective end of members of the party to build solidarity among the members. (vi) Higher order collective ends can be, what Schmidtz (1994: 228) refers to as maieutic ends: "ends that are achieved through a process of coming to have other ends". For example, the members of the board of directors of a company might have as a collective end to settle on the collective long term goals of the company. These ends to be chosen are final ends; they are not to be chosen for the sake of some other ends, but they are nevertheless, not yet settled on. (vii) The collective end can be one the participating agents need to keep in mind while they are performing the actions which are the means to that end; or it might be an end which the agents will only realise if they concentrate wholly on the means; or it might be an end of which the agents are not fully conscious at any time. An example of an unconscious collective end might be some end pursued by members of a particular social class, say, to exploit workers.[8] (viii) The actions constitutive of joint action can be performed at the same time and place, or at different places at the same time, or at the same place at different times, or at different places and times. Indeed, the actions of the individuals can be separated

[8] Unconscious ends are not necessarily 'open' in any stronger sense than that there is mutual belief with respect to them; and beliefs can be unconscious.

by thousands of miles and/or hundreds of years, e.g. building the Great Wall of China. (ix) The actions constitutive of joint action can be interactive—in the sense that in the performance of each individual action the agent adjusts his action to the action of the other e.g. dancing—or non-interactive (in this sense). (x) Joint action typically, though not necessarily, involves, either direct communication (of some sort or other) between each individual and every other individual, or indirect communication between the individuals—A communicates with B, and B with C, but A does not communicate with C.[9]

2. CONFUSING ATOMISTIC INDIVIDUALISM WITH RELATIONAL INDIVIDUALISM

One of the more common errors of those who attack individualist accounts of joint action is to confuse *atomistic* individualism with *relational* individualism. The former is open to well-known difficulties, but CET is a species of the latter. Margaret Gilbert and John Searle are guilty of this confusion, so let us consider some of their examples and accompanying arguments against individualist accounts of joint action.

Let us first consider Gilbert's mushroom picking example.[10] This example is put forward in the context of Gilbert's pursuit of an analysis of social action that helps itself only to individualist notions. The point of the example is to show that such a pursuit is hopeless; according to Gilbert, such analyses in terms of individualist notions will inevitably fail. Gilbert points out that a group of mushroom pickers who individually pick mushrooms are not engaged in social action. Moreover, they would not be engaged in social action, even if they were mutually aware of one another's mushroom picking activities. Nor would the fact that one mushroom picker had a brief conversation with another mushroom picker turn their mushroom picking into social action. Finally, even if all the mushroom pickers engaged in an occasional two-person conversation this would still not be sufficient for their *mushroom picking* to count as social action. I agree. More important, CET yields the same result. By the lights of CET, this mushroom picking activity is not social action, or even joint action; for the individual mushroom picking actions are not directed to a *collective* end. So, the mushroom picking example, does not show that an individualistic analysis cannot succeed. It merely shows that Gilbert did not have an adequate individualist account; she is working with a form of atomistic individualism.

Gilbert correctly claims that from the fact that agents A and B (successfully) perform a joint action properly describable as X-ing, it does not follow that A performed some individual action properly describable as X-ing. Thus A and B might be dancing together, but it does not follow that if A were doing the same movements alone, A would be dancing (Gilbert 1989: 165). But this is not

[9] See Cohen and Levesque for contrasting account (1991: 506).

[10] This discussion is taken from Miller (2001: 85–90).

required of *relational* individualist accounts, and certainly not of CET. CET in helping itself to the notion of a collective end enables a joint action description which outruns the description of any of the individual actions considered without reference to the end. Thus, Fred can dial a number and Bill can pick up a receiver and there is the joint action of communication. But this is only the case if each had as a (collective) end that there be communication. In the dancing example each has as an end that they dance. Here it is the bodily movement of each coupled with the *relations* between those individual bodily movements that is the end that each has. So this end is not adequately described by a description of the bodily movements of one of the partners.

Now consider John Searle's well-known example of would-be capitalists aiming at their own interest while having a collective intention to benefit all (Searle 1990: 404; cf. also, Hobbs 1990). Searle is correct in thinking that the collective intention is not reducible to individual intentions. But he is wrong to think that it is therefore an irreducibly collective intention. Rather it is a collective end of the kind deployed in CET. Let us look at his example.

On one way of taking the example, the agents' actions are overdetermined. They have both individual and collective ends. Each has the individual end to get rich for its own sake. In addition, each has the collective end of the common good which each believes will be realised by each pursuing his or her own interest. Qua actions directed at the collective end the agents are involved in joint action. Qua actions performed simply in the service of their individual ends, the agents are involved in individual action. On the other way of taking the example each has one ultimate end viz. the common good. However, paradoxically, each must aim at his or her own interest in order to achieve the common good. Here the individual end is not an end in itself, but simply a means to the further end of the common good. However, the individual actions do constitute a joint action; for they are actions that are performed, indeed *only* performed, in order to secure the common good.

3. ACTING QUA MEMBER OF A GROUP

Searle's example of the invisible hand mechanism conveniently introduces the alleged problem for individualist accounts of joint action of actions performed by individuals qua members of a group. This objection is put forward by Searle (1990), but also by Margaret Gilbert and, in its most sustained form, by Schmitt (2003: 147–150).

According to Gilbert, two or more agents engaged in joint action are, what she terms, "plural subjects" with a "shared intention" to perform a joint action. Here the shared intention is not reducible to individual intentions or ends, and the (joint) action is not reducible to individual actions (or individual actions in conjunction with individual intentions or ends and the relations between the individual actions). Indeed, Gilbert holds the view that plural subjects can be ascribed intentions, beliefs and moral responsibility, that are not possessed by

the individual human agents that (in some sense) constitute the plural subject. Gilbert has presented her views in a number of different places, but not always in a consistent or luminous manner.[11]

At any rate, she is committed to the following definition of the key notion of interest to us here, viz. shared intention: "Persons P1 and P2 have a shared intention to do A if and only if they are jointly committed to intending as a body to do A" (Gilbert 2000: 73). Here the notion of a joint commitment is itself an irreducibly collectivist notion: "So it is not the case that each of them is the sole author of part of the joint commitment, let alone of the joint commitment as a whole. Rather they comprise the author of the joint commitment together" (ibid.: 72). However, the notion of a commitment is itself somewhat unclear. At one place, she offers an apparently non-normative account: "If I have decided to do A, I am subject to a commitment to do A". At others commitment seems essentially normative: "the pre-theoretical judgment that shared intentions involve obligations with corresponding entitlements responds to the presence, given a shared intention, of a joint commitment" (ibid.: 76).

I have two responses to the above analysis of shared intention. First, the crucial notion of a commitment, including a joint commitment, conflates the normative notion of putting oneself under an obligation with the (non-normative) notion of non-reversibility. Conative states, such as intentions and decisions are (at least partially) non-reversible. A useful image here is that of jumping off a cliff; at some point the person jumping 'commits' himself, and so there is no going back. Similarly for intentions and decisions. However, this feature of non-reversibility has nothing whatsoever to do with normativity; in particular, it is an entirely different sense of commitment from that of putting oneself under an obligation, e.g. by making a promise. If commitment is taken to imply non-reversibility only, then the definition of shared intention in terms of joint commitments seems to be vacuous; shared intentions are (more or less[12]) joint intentions. If on the other hand, commitment is taken in the normative sense of putting oneself under an obligation, then the definition is false; intentions, including shared intentions, are not necessarily puttings of oneself under obligations (or a necessary result thereof).

Second, the other crucial notion of intending "as a body" to do A conflates two notions: (i) the notion of intending as the member of a group, and (ii) the notion of an individual intention, albeit an individual intention attaching to a single macro entity which has agents P1 and P2 as parts. If the latter notion is the one in question, then we do not have an intention, shared or otherwise, of P1 and of P2; rather we have the intention of the (alleged) single entity that has P1 and P2 as its parts. If the former, then there is no *apparent* barrier to the analysis

[11] For a paper that details different interpretations of Gilbert's account of collective responsibility see Miller and Makela (2005).

[12] The notion of a shared intention is arguably a somewhat weaker notion than that of a joint intention.

of the shared intentions in terms of the individual intention of P1 and the individual intention of P2. But is there in fact a barrier to the analysis of shared intentions in terms of individual intentions? In what remains of this section, I will address this issue.

The notion of acting qua member of a group is often quite straightforward since the group can be defined in terms of the collective end or ends which the group of individuals is pursuing. Consider a group of individuals building a house. Person A is building a wall, person B the roof, person C the foundations, and so on. To say of person A that he is acting qua member of this group is just to say that his action of building the wall is an action directed towards the collective end that he and the other members of the group are seeking to realise, namely a built house.

Notice that the same set of individuals could be engaged in different collective projects. Suppose persons A, B, C, etc. in our above example are not only engaged in building a house but also—during their holidays—in building a sailing boat. Assume that A is building the masts, B the cabin, C the bow, and so on. To say of A that he is acting qua member of this group is just to say that his action of building the masts is an action directed towards the collective end that he and the other members of the group are seeking to realise, namely a built boat. Accordingly, one and the same person, A, is acting both as a member of the 'house building group' (G1) and as a member of the 'boat building group' (G2). Indeed, since A, B, C, etc. are all and only the members of each of these two groups, the membership of G1 is identical with the membership of G2.

Moreover, when A is building the wall he is acting qua member of G1, and when he is building the mast he is acting qua member of G2. But this phenomenon of one agent acting as a member of different groups in no way undermines individualism. Indeed, CET is able to illuminate this phenomenon as follows: For A to be acting qua member of G1 is for A to be pursuing—jointly with B, C, etc.—the collective end of building the house; for A to be acting qua member of G2 is for A to be pursuing—jointly with B, C, etc.—the collective end of building the boat.

Further, let us suppose that G1 and G2 each have to create and comply with a budget; G1 has a budget for the house, and G2 has a budget for the boat. The members of G1 and G2 know that they must buy materials for the house and the boat (respectively) and do so within the respective budgets. Assume that A, B, C, etc. allocate $50,000 to pay for bricks for the house. This is a joint action. Moreover, this joint action is one that A, B, C, etc. have performed qua members of G1. G1 is individuated by recourse to the collective end of building the house, and the proximate (collective) end of buying bricks is tied to that group, G1, and to its ultimate end of building a house. Accordingly, A, B, C, etc. are not in buying the bricks acting qua members of G2. For G2 is individuated by the collective end of building a boat, and A, B, C, etc. do not qua members of G2 have any plans to build their boat from bricks!

Doubtless, many groups are unified in more complicated ways than in our above example. Consider the members of an organisation; each person occupies an organisational role. Elsewhere I have elaborated an account of the joint actions performed by such organisational role occupants (Miller 2001: Chap. 5). Specifically, I have introduced the concept of a *layered structure of joint actions*. Roughly speaking, a layered structure of joint actions is a set of joint actions each of which is directed to a further collective end; so it is a macrojoint action comprised of a set of constituent microjoint actions. I have also supplemented this account by recourse to concepts of conventions, social norms and the like, i.e. regularities in action to which members of organisations (and members of social groups) typically conform. At any rate, the point to be made here is that my account of the notion of acting qua member of a group in terms of acting in accordance with collective ends can be complicated and supplemented to accommodate various different kinds of acting qua member of a group. In the last section of this chapter, I return to this issue. In the absence of specific counterexamples, my account stands as an impediment to the claims of anti-reductionists such as Searle, Schmitt and Gilbert.

A favourite putative counterexample to individualist accounts that has been put forward by anti-reductionists is the members of a committee voting to, for example, elect one of their members as the chair of the committee. Here a problem allegedly arises by virtue of the fact that the committee is committed to whatever candidate received the largest number of votes, notwithstanding the fact that a number, perhaps even a majority, did not vote for that candidate. Assume that a majority of the members did not vote for candidate A. However, assume that candidate A received the largest number of votes. So the committee has voted in A and, therefore, each member of the committee has—qua member of the committee—endorsed A as the preferred candidate.

The fundamental notion required to understand these kinds of situation is the technical notion of what I have elsewhere termed, a joint mechanism (ibid.). What is a joint mechanism? Suppose we know that in some area there are bound to be disagreements not resolvable by recourse to rational argument to a point of consensus. In this area, we might decide to adopt procedures to resolve disputes. Voting and tossing a coin are procedures in this sense. The application of these kinds of procedures, even on a single occasion only, is action performed to realise a collective end. Let us call such procedures as these joint mechanisms.

Joint mechanisms consist of: (a) a complex of differentiated but interlocking 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. Thus a given agent might vote for a candidate. He will do so only if others also vote. But further to this, there is the action of the candidates, namely, that they present themselves as candidates. That they present themselves as candidates is (in part) constitutive of the input to the voting mechanism. Voters vote *for candidates*. So there is interlocking and differentiated action (the input). Further there is some result (as opposed to consequence) of the joint action; the joint

action consisting of the actions of putting oneself forward as a candidate and of the actions of voting. The result is that some candidate, say, Jones, is voted in (the output). That there is a result is (in part) constitutive of the mechanism. That receiving the largest number of votes is being voted in, is (in part) constitutive of the voting mechanism. That Jones is voted in is not a collective end of all the voters (although it is a collective end of those who voted for Jones.) However, that the one who gets the largest number of votes—whoever that happens to be—is voted in, is a collective end of all the voters.

The complex of joint action and its result is, qua application of the mechanism, in principle repeatable. But the complex of actions performed in any given application (and/or its result) must be able to differ from one occasion of application to the next. The action to be performed in the application of the mechanism is specified in such a way as to enable it to consist of different possible actions on different occasions of application. Or at least, the action is specified in such a way as to enable it to have a different result from one occasion to the next. I am speaking here of the result which is the output of the mechanism. Thus one must vote for a candidate, but which particular candidate is not specified. Or one must toss a coin, but it is not known in advance whether the result will be heads or tails.

Moreover this characteristic of delivering different actions or results on different occasions of application is a characteristic which can exist in virtue of different features of joint mechanisms. We have just seen in the case of the voting mechanism that one such feature is the capacity for differing actions to be performed on different occasions in conformity to the mechanism. We have also just seen, in the case of the mechanism of tossing the coin, that agents performing the same action on every occasion can deliver a different result from one occasion to the next, and do so in virtue of an internal property of the mechanism. (In the case of flipping the coin, the internal property is the equal probability of a tossed coin coming down heads or coming down tails.) Moreover, there are further kinds of case. One of these is the capacity of the mechanism to deliver a different result in virtue, not of different action input or of an internal property of the mechanism, but rather of some feature of the situation in which the mechanism is applied. Thus, an utterance of the term, "the bank" might produce a different result in one context (a conversation about money) from what it produces in another (a conversation about the Volga River). Moreover, it may be that there are special linguistic mechanisms which are such that the context *always* makes a difference to the result (whether by disambiguating or in some other way).

Another kind of case involves what I will call, following Grice (1989: 129), resultant actions. Thus a scrumhalf in the contact sport of rugby union may have the procedures of: tapping his left knee, if the lineout ball is to be thrown to the front of the lineout; tapping his right knee, if it is to be thrown to the back; tapping a knee once, if it is to be lobbed high; and tapping a knee twice, if it is to be thrown fast and low. The action of tapping his left knee twice means

that the ball is to be thrown fast and low, and to the front of the lineout. However, this is a resultant action, since it is derivable from the two basic procedures of tapping a knee twice and tapping the left knee.

Language in general appears to consist in part of joint mechanisms involving resultant actions.[13] Assume that there are the following joint procedures in a community: utter "Sydney" when you have as an end reference to Sydney; utter "Paris" when you have as an end reference to "Paris"; utter "is a city" when you have as an end ascription of the property of being a city; utter "is frequented by tourists" when you have as an end ascription of the property of being frequented by tourists. Then there might be the resultant joint action to utter, "Paris is a city", when you have as an end ascription to Paris of the property of being a city; and there might be the second, and different, resultant joint action to utter, "Sydney is frequented by tourists", when you have as an end ascription to Sydney of the property of being frequented by tourists. It is easy to see how by the inclusion of a conjunctive operation indicated by "and" additional linguistic combinations might yield multiple additional resultant actions, e.g. the communication effected by uttering "Paris is a city and Paris is frequented by tourists".

So joint mechanisms are such that their action inputs, and/or their results (outputs), may differ from one occasion of application to the next. This creates the possibility of individual choice, and indeed uniqueness, of resultant action. For example, the scrumhalf, although he is constrained in the sense that he has to follow the procedures, nevertheless, can choose between disjuncts of the procedures, and thus choose his resultant action. It also means that on occasion joint mechanisms, and joint procedural mechanisms, deliver outcomes other than those desired by at least some of the participants.

These two characteristics of joint mechanisms have the implication that in one sense the input of joint mechanisms is not a joint action, and the output is not a collective end. In particular, under the description, 'means to achieving my individual end', the action is not a joint action, and the end is not a collective end. Thus if when you flip the coin I call "Heads" in order to win the money, my action qua means to winning the money for myself is not joint, and my end is not collective. However, in another sense it is. Under the description, 'application of the mechanism', the individual actions of each of the individuals constitutes a joint action. Under the description, 'output of the application of the mechanism', the output is a collective end. Thus, if you flip the coin and I call "Heads" (or "Tails") and each of us does so in order to determine who is to win the money (the one who calls it correctly wins), then the individual actions of flipping the coin and calling "Heads" (or calling "Tails") constitute a joint action and my end is a collective end, as is yours. Similarly, when each of us

[13] Grice (1989: 129f.) developed his notion of a resultant procedure (as opposed to a resultant action) for precisely this purpose.

votes for possibly different candidates each of us acts in order to ensure that someone or other (perhaps my candidate, perhaps yours, i.e. whoever gets the most votes) is democratically elected and becomes leader, then we have performed a joint action in order to realise that collective end.

In short, many social groups deploy joint mechanisms, such as flipping coins or voting, and when they do they act as members of the group participating in those joint mechanisms. However, joint mechanisms are readily analysed in accordance with the basic precepts of CET, i.e. collective ends and associated attitudes. Accordingly, acting qua member of a group in this sense is entirely consistent with individualism, and with CET in particular.

4. SCHMITT'S OBJECTIONS TO THE SUFFICIENCY OF CET

As mentioned above, Fred Schmitt has put forward detailed arguments to refute CET. The arguments are made both to the necessity of CET and to the sufficiency of CET. However, it is the arguments in relation to the sufficiency of CET that are most germane to my purposes in this chapter. For, as indicated above, they purport to demonstrate the failure of individualism to identify the defining non-individualist properties of joint action and, therefore if sound, pave the way for some form of supra-individualism or (at least) anti-reductionism. The counterexamples to the sufficiency of CET offered by Schmitt are supposedly clear cases of *non-joint* purely individual action which CET would allegedly wrongly identify as instances of joint action. In doing so—the argument continues—CET would reveal that it has not identified key defining features of joint action because it has eschewed non-individualist properties, such as Searle's sui generis notion of collective intention.

By contrast, Schmitt's arguments in relation to the necessity of CET, even if sound, are consistent with an individualist account, albeit an individualist account that is a weaker form of individualism than CET. For Schmitt's alleged counterexamples to the necessity of CET consist of examples in which there is an outcome aimed at by each agent and brought about by the combined actions of the agents, but there is apparently no two-way interdependence of action and/or no collective end, and Schmitt invites us to view these as being, nevertheless, intuitively cases of joint action. Suffice it to say that not only do I reject his claim to have found counterexamples to the necessity of CET, but I also note that his examples do not seem to involve any properties not susceptible of individualist analysis.[14] Accordingly, if these examples were in fact instances of joint action then this would undermine supra-individualism and anti-reductionism,

[14] Schmitt's alleged counterexamples to the necessity of CET consist of scenarios such as the Alan/Betty duo who lift different sides of a sofa but in which each would do his or her part even if the other did not (because, for example, each feels bound by some prior agreement to do their part). If each believes that the other will not lift their side of the sofa then there is no interdependence of action. But now, contra Schmitt,

including Searle's version thereof. At any rate, let me then turn to Schmitt's alleged counterexamples to the sufficiency of CET.

Schmitt offers a number of alleged counterexamples to the sufficiency of CET. The first example is as follows. Two hermits, Sim and Tim, individually put food in the forest by way of paying tribute to the god, Chocolate. Thus far their actions are not interdependent, nor is there a collective end. However, Sim and Tim individually do not have enough food for tribute so they *each rely on the presence of one another's food to enable an adequate amount of tribute to be present*. Schmitt claims that the example is not joint action, but that according to CET it is joint action. Since each relies on the action of the other, there is interdependence of action. However, interdependence of action does not entail the existence of a collective end, and the example as described thus far does not seem to involve a collective end; therefore, the example, as described thus far, is not necessarily a counterexample to CET.

However, Schmitt posits that in this scenario Tim and Sim have the common end of *someone's* paying tribute to the god; so is it the case that someone's paying tribute to the god is a collective end in this scenario? Certainly, both Tim and Sim have individual ends. Tim has as an individual end that he, Tim, pay tribute; moreover, he has this end on the condition that Sim pay tribute. However, Tim does not have as an end that Sim pay tribute.

Let us consider the proposition that Tim has the end that *someone* pay tribute. This proposition is ambiguous. It could mean that Tim has as an end that *anyone* pay tribute. But Tim does not have this as an end. To see this, consider that Sim is someone, but Sim paying tribute will not satisfy Tim's end. The situation here is analogous to me looking for Mary, and only for Mary; I am not looking, for example, for Betty. If I am looking for Mary in this sense then I am looking for someone, but in the restricted sense of 'someone if that someone is Mary'. Accordingly, it does not follow that I am simply aiming to find someone in the sense of anyone. Indeed, it is false that I am looking for just anyone; this is why finding Betty, for example, would not realise my end.

---

the example is not an instance of joint action. On the other hand, if each lifts because each thinks the other might lift their side (even if not strongly believing that they will) then there will be interdependence of action and the example will now be by CET's lights, contra Schmitt, an instance of joint action. Schmitt also claims that "Each individual must act so that a joint action is performed, but the performance of a joint action does not by itself require that each individual acts in order that the joint action be performed". This is true, but CET is not committed to the proposition that "each individual acts in order that the joint action be performed. In Miller (2001: 82), a clear distinction is made between two sorts of joint action, one in which this claim is true and one in which it is not. For example, in the case of dancing in which the collective end is internal to that actions that realise it, two partners have as a collective end that they dance jointly. On the other hand, in my example of killing a tiger by throwing spears at it—a case in which the collective end is external to the actions that bring it about—the agents do not have as a collective end that they act jointly, but only that the state of affairs, a dead tiger, comes into existence.

Correspondingly, Sim has as an individual end that he, Sim, pay tribute, but does not have an end that Tim pay tribute. Once again, even if Sim has as an end that *someone* pay tribute, he does not have as an end that *anyone* pay tribute. Specifically, Sim only has this as an end: that someone pay tribute if that someone is Sim.

Accordingly, there is no *accurately described* state of affairs such that both Tim and Sim have that state of affairs as an end. So there is no collective end as required by CET. Accordingly, the example is not a counterexample to CET.

In response to this, it might be held that I have misdescribed the end of Tim and of Sim. Perhaps, contra what I have suggested above, the end of Tim (and of Sim) is that there be sufficient tribute to assuage Chocolate. If that is the end of each, and if each makes his contribution as a necessary (but not sufficient) condition for realising this end, then the end in question is, pace Schmitt, a collective end. However, thus redescribed the example is, I suggest, intuitively a species of joint action. Moreover, as CET requires, there is interdependence of action, and this interdependence of action is relative to the collective end each has of ensuring that sufficient tribute is paid to Chocolate.[15]

Schmitt offers another alleged counterexample. Gus and Heather are strangers and somewhat deaf and blind, and are each given a side of a board to hold by Ignatz. Neither Gus nor Heather is aware that the other is holding the other side of the board. Schmitt claims this is not joint action. I agree. There is no interdependence of action in the required sense; neither is holding a side of the board in the belief that the other is holding the other side. Now assume that Gus and Heather become mutually aware that each is holding a side of the board. Again, Schmitt claims this is not joint action. I agree; for neither has a collective end.

Schmitt now adds a further condition to his example and claims that in so doing, he has added a collective end. However, the scenario is evidently not an example of a joint action. The additional condition is that each embraces the end of acting to hold his or her part of the board. Accordingly, Gus' end is to hold *Gus'* side of the board, and Heather's end is to hold *Heather's* (different) side of the board. But here Gus' individual end is not Heather's individual end, and so there is no state of affairs that each has as an end, i.e. there is no collective end. So although Schmitt rightly claims that the scenario is not an instance of joint action, he is wrong to think that it involves a collective end. Accordingly, it is not a counterexample to CET.

---

[15] This version of the scenario is consistent with Tim and Sim not caring whether each contributes a necessary part of the tribute required, as long as sufficient tribute is paid. The point is that each will do whatever he needs to do and can do to ensure sufficient tribute is paid. So each has the collective end to ensure sufficient tribute is paid, and each does whatever is necessary *in the circumstances* to realise this end; depending on what the other does, one of the two might be able to get away with paying very little tribute.

5. CORPORATE ACTION

Earlier I invoked a distinction between joint action and corporate action. Corporate actions are the 'actions' of corporate entities such as nation-states, e.g. the USA defeating Japan in the Pacific campaign in World War II. Interestingly, Searle does not make this distinction.

My own position is that properly speaking there are no corporate actions. Putative corporate actions will either turn out to be corporate *doings* which are not also actions, or else turn out to be complex joint actions, such as those I earlier dubbed "layered structures of joint actions". A 'doing' is something that is done by something; however, there is no implication of an agent possessed of a mind. For example, the hurricane *did* something (in this sense); it destroyed the building. More specifically, I do not accept that macro-entities, such as nation states and corporations, have beliefs and intentions, and consequently are either rational or moral agents. Properly speaking, all social actions are performed by individual human beings and not by social entities.

It should be noted that in advocating this form of individualism I am not committed to any of the following claims: (1) Macro-entities such as social groups and organisations are reducible to the individuals who (at least in part) constitute them.[16] (2) Macro-entities such as social groups and organisations do not have causal powers. (3) Macro-entities such as social groups and organisations cannot be blamed in some non-moral sense of blame for causing undesirable states of affairs, such as pollution of the environment. (4) It is never appropriate for individuals, courts of law and so on, to treat organisations (in particular) as if they were rational or moral agents.

The idea that there are corporate actions in the strong sense of actions, and that collective entities have associated mental states, such as intentions and beliefs, is most famously advocated by Peter French. French defines corporations in terms of their CID or corporate internal decision structure (French 1984: 39, 41, 48f.). This structure consists of, firstly, hierarchically ordered positions with their attendant tasks (e.g. director, engineer, typist) secondly, procedures (e.g. the board of directors of Gulf has the procedure of voting on whether to join a cartel) and, thirdly, company policies (e.g. to maximise profit).

I take it that there is a strong presumption against the ascription of mental states such as intentions and beliefs, let alone full-blown moral agency, to institutions and organisations, including corporations.[17] Moreover, in doing so I am at one with Searle. Searle rejects the idea that, for example, his collective intentions could exist in the heads of anything but ordinary human persons—and certainly not in macro social entities such as corporations. But let me briefly sketch an argument against French.

---

[16] On the issue of reducibility of social entities see Ruben (1985).

[17] For criticisms see May (1983).

If we follow French in ascribing such sophisticated intentions to Gulf Oil as the intention to maximise profits, then we are apparently committed to ascribing to Gulf a whole network of sophisticated propositional attitudes concerning economic production, the workings of markets, and so on. For as recent work in the philosophy of mind has taught us, propositional attitudes, such as beliefs and intentions, cannot exist as stand alone entities, but only as interrelated components of networks of propositional attitudes. Thus a being which intends to maximise profits must have beliefs about what a profit is, and therefore what cost and revenue are (profits equal revenue minus costs), what a market is, what production, supply, demand and so on are. Moreover, these propositional attitudes are not rudimentary ones directly tied to specific forms of behaviour or external sensory stimuli. They are not, for example, simple beliefs of the kind a dog comes to possess when it discovers that an object in its sensory field is in fact a bone. Rather the relationship between these sophisticated beliefs is indirect, complex and traversable only by reflective reasoning. Indeed, a being with such a network of propositional attitudes would be capable of high level thought, and therefore be possessed of a language in which to do this thinking.

Moreover, this agent's thought processes would include planning for its future and doing so on the basis of its past mistakes, and the likely responses of other corporations. Thus French: "Simply to be a moral person is to be both an intentional actor and an entity with the capacity or ability to intentionally modify its behavioural patterns, habits, modus operandi after it has learned that untoward or valued events (defined in legal, moral or even prudential terms) were caused by its past unintentional behaviour" (French 1984: 165). Such a corporate agent is self-reflective; it not only distinguishes its present from both its past and its future, and itself from other corporations, it reflects on itself for the purpose of transforming itself. Such a being has higher order propositional attitudes, including beliefs about its own beliefs and intentions, and conceives of itself as a unitary whole existing over time. In short, it looks as though we now have a fully conscious, indeed self-conscious, being on our hands.

Nor do matters rest here. For if we are prepared to grant Gulf a mind, then why not all its subsidiaries, as well as all other companies and subsidiaries worldwide? Indeed, how can we stop at corporations? Surely governments, universities, schools, supermarkets, armies, banks, political parties, trade unions, English soccer teams' supporters' clubs (at least) now all have minds, albeit in some cases smaller minds (so to speak). Not only do we have a self-conscious mind, but we have an ever-expanding community of self-conscious minds. Surely this is preposterous.

Let me now return to Searle who agrees with me that there are no corporate actions in this sense. As already noted, according to Searle mental states, including his corporate intentions, exist only in the heads of individual human persons.

Earlier, I suggested that some putative corporate actions might turn out to be complex joint actions; specifically, layered structures of joint action. Consider an army winning a battle. At level one we have a number of joint

actions. The members of, say, the infantry jointly move forward on the ground (level one joint action). The members of, say, a mortar group jointly act to pound the enemy positions with mortar fire (level one joint action). Finally, air support is jointly provided by a squadron of fighter planes (level one joint action). So there is a series of level one joint actions. Now, each of these three (level one) joint actions is itself describable as an *individual* action performed (respectively) by the different groups, namely, the action of killing the enemy by mortar fire, protecting the advancing troops by providing air support, and moving forward to take and hold the ground occupied by the enemy. However, each of these 'individual' actions is a component element of a larger joint action directed to the collective end of winning the battle. For each of these individual attacks is part of a larger plan coordinated by the military leadership. So these 'individual' actions constitute a *level two* joint action directed to the collective end of winning the battle.

It seems that Searle could accept this kind of analysis of some putative corporate actions, i.e. he could accept the notion of layered structures of joint action. So wherein would lie the difference between Searle's anti-reductionism and CET in relation to these kinds of corporate action? The difference would lie in Searle's analysis of the constituent joint actions of the layered structure of joint action. Presumably, Searle would insist that contra CET, the analysis of joint actions, including in the context of layered structures of joint action, would necessarily involve collective intentionality—specifically, collective intentions—not reducible to individualist notions of intentions, ends and the like. At this point we have come full-circle; we have returned to the different respective analyses of joint action, viz. CET and Searle's analysis. In closing I make the following point. Searle's account posits sui generis collective intentions, and he argues that reductionist accounts such as CET do not adequately account for joint actions and associated collective intentionality. However, he nowhere offers a detailed (naturally, non-reductionist) explication of collective intentions; to this extent his account is under-theorised.

REFERENCES

Bratman, M.E. (1987). *Intentions, Plans and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M.E. (1992). Shared cooperative activity. *Philosophical Review*, 101(2): 327–341.

Bratman, M.E. (1993). Shared intention. *Ethics*, 104(1): 97–113.

Cohen, P.R. and Levesque, H.J. (1991). Teamwork. *Noûs*, 25(4): 487–512.

French, P. (1984). *Collective and Corporate Responsibility*. New York: Columbia University Press.

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

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

Grice, H.P. (1989). Utterer's meaning, sentence-meaning and word-meaning. In H.P. Grice, *Studies in the Way of Words*, Cambridge, MA.: Harvard University Press.

Hobbs, J.R. (1990). Artificial intelligence and collective intentionality: Comments on Searle and on Grosz and Sidner. In Philip R. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*, Cambridge, MA.: MIT Press.

May, L. (1983). Vicarious agency and corporate responsibility. *Philosophical Studies*, 43(1): 69–82.
Miller, S. (1982). David Lewis on conventions. *Philosophical Papers*, 11(2): 1–8.
Miller, S. (1986a). *Conventions and Speech Acts*. Ann Arbor, MI: University Microfilms International.
Miller, S. (1986b). Conventions, interdependence of action and collective ends. *Noûs*, 20(2): 117–142.
Miller, S. (1987). Conventions and social contracts. *Philosophical Papers*, 16(2): 85–106.
Miller, S. (1992a). Joint action. *Philosophical Papers*, 21(3): 1–23.
Miller, S. (1992b). On conventions. *Australasian Journal of Philosophy*, 70(4): 435–145.
Miller, S. (1995). Intentions, ends and joint action. *Philosophical Papers*, 24(1): 51–67.
Miller, S. (2001). *Social Action: A Teleological Account*, Cambridge: Cambridge University Press.
Miller, S. and Makela, P. (2005). The collectivist account of collective moral responsibility. *Metaphilosophy*, 36 (5): 634–651.
Ruben, D.-H. (1985). *The Metaphysics of the Social World*. London: Routledge.
Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
Searle, J.R. (1990). Collective intentions and actions. In Philip R. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in Communication*, Cambridge, MA.: MIT Press.
Searle, J.R. (1995). *The Construction of Social Reality*. London: Penguin.
Schmidtz, D. (1994). Choosing ends. *Ethics*, 104(2): 226–251.
Schmitt, F. (2003). Joint action: From individualism to supraindividualism. In Frederick Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Action*, Lanham, MD: Rowman & Littlefield.
Tuomela, R. (1995). *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford, CA: Stanford University Press.
Tuomela, R. (2002). *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.
Tuomela, R. and Miller, K. (1988). We-intentions. *Philosophical Studies*, 53(3): 367–389.

COLLECTIVE SPEECH ACTS

Anthonie Meijers

1. INTRODUCTION

From its early development in the 1960s, speech act theory always had an individualistic orientation. It focused exclusively on speech acts performed by individual agents. Paradigmatic examples are 'I promise that p', 'I order that p', and 'I declare that p'. There is a single speaker and a single hearer involved. In his book *Speech Acts,* for example, Searle's analysis of promising starts from the following description: "Given that a speaker S utters a sentence T in the presence of a hearer H, then, in the literal utterance of T, S sincerely and non-defectively promises that p to H if and only if the following conditions 1–9 obtain [etc.]" (Searle 1969: 57).Though this focus may initially have been due to Searle's methodological approach of starting his analysis of speech acts with clear-cut examples, it has led in the end to an unnecessary and undesirable bias in speech act theory.

In this paper I will extend the traditional analysis of speech acts by focusing on collective speech acts. These are acts performed by collective agents or addressed to collective agents. Our language is full of collective speech acts. Examples are: "We promise to be back in time", "We believe that this is the wrong approach", "The Security Council appeals to Israel and Hezbollah to stop fighting", "I want the orchestra to start with the first part", "We urge the government to reconsider its policies", and so on. Surprisingly, these collective speech acts have escaped the attention of almost all philosophers of language thus far.[1]

Collective speech acts are not just interesting in themselves because they are omnipresent in everyday language. They are also crucial for understanding an issue that belongs to the very core of Searle's philosophical project: the nature of social reality. In *The Construction of Social Reality* Searle defends the view that language is constitutive of social institutions. Institutional facts require symbolic representations that are publicly understandable. I will argue that this claim can be made more precise using the concept of a collective speech act.

[1] The only exception to my knowledge is Hughes (1984). I owe this reference to Margaret Gilbert. Hughes' paper has been ignored in the subsequent literature. Much earlier, at the beginning of the 20th century, Adolf Reinach developed in his phenomenology a proto-theory of speech acts for the legal domain, which included a discussion of speaking on behalf of a group. (See Crosby 1990.)

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 93–110.
© 2007 Springer.

2. A NOTE ON GROUPS

Groups do not form a single category. As a matter of fact, there are many types of groups: the fans in a soccer stadium, the team playing, the members of the club, the visitors older than 21 years, the executive committee of the club, a family attending the match, etc. There is not one way of classifying groups. This can be done in terms of the physical and/or the intentional relationships between the members of the group. A family is bound by biological kinship, fans by the physical boundaries of the stadium and by their desire that their team will win, a team by their common goal and the arrangements that have been made for the match (their strategy, division of labour), an executive committee by the institutional rules of the club, and so on.

Rom Harré (1997: 200) makes a useful distinction between three types of groups, which I will follow roughly here:

- *Taxonomic groups* or classes. An example is the visitors in the stadium older than 21. Members of the group only share a particular characteristic, but they do not have 'internal' relations. If one of them drops out, it will not affect the others.
- *Crowds*. For example, the crowd entering the stadium. Members share a common goal, but do not have beliefs about each other given that common goal, or rights and obligations.
- *Structured groups*. Members are either physically or intentionally related. An example of the former is a family, of the latter a team playing soccer. Team members share a common goal, they have beliefs about each other, and they have rights and obligations. The members of structured groups are internally related, i.e., the loss of one of the members will affect the others.

For the purpose of this paper I will restrict most of the analysis of collective speech acts to structured groups of the intentional type. The main reason is that speech acts, being *acts*, are performed by *agents* and if the notion of a collective agent makes sense it will refer to members of a group that are intentionally related. They share intentions, goals, beliefs, or desires. On the other hand, it is hard to understand how groups of the other types (family, crowd, taxonomic) could be characterized as collective agents at all, since they need not share any intentional states. Speech acts may be *addressed* to those crowds or taxonomic groups, an issue to which I will come back when I discuss the reducibility of collective speech acts.

3. SPEECH ACTS AS COMMUNICATIVE ACTS

There is a sense in which *all* speech acts are collective acts, and it is important to clarify this sense before starting the analysis of collective speech acts. Searle's speech act theory and most subsequent theories conceive of speech acts as communicative acts. They are the minimal units of communication (Searle 1969: 16). The model of communication on which Searle's theory is based is

Grice's model. According to Grice (1957, 1989), for a speaker to mean something by uttering X is for him to intend that the utterance of X will *produce some effect* (belief, desire) in the audience *by means of* the recognition of the speaker's intention. Unlike Grice, Searle defines this effect as the hearer's understanding of the utterance of the speaker. In doing so, he makes room for the well-known distinction between the illocutionary effect and the perlocutionary effect of speech acts.

The resulting account of communication shares with many other accounts—such as the traditional sender–receiver model, the information processing models in cognitive science, or the radical translation/interpretation model—the idea of communication as a *one-way transfer* between speaker and hearer. The speaker produces some effect in the hearer by means of his intentions or sends a message via a communication channel that needs to be decoded by the hearer, or utters a sentence that needs to be interpreted from the hearer's third person point of view. A dialogue or conversation, then, is accounted for by applying this one-way model reciprocally. It is a sequence of one-way transfers, where speaker and hearer change roles.

I have criticized this model for not being able to capture what is in my view essential for communication: that it is a form of *cooperation* (Meijers 1994, 2002a). Communication is not a series of monological speech acts, where speaker and hearer act independently of one another. There are forms of communication that are like that (e.g., advertisements). But even in these cases, the utterances make use of conventional means in order to be understood, where these conventions originate in prior cooperation or coordination. Communication in a full sense, as in dialogue, is different from monological utterances. A single speech act is embedded here in a larger conversational framework, where speaker and hearer cooperate in order to reach understanding with respect to the matter being communicated. This understanding cannot simply be analysed as the *hearer's* understanding of the speaker's utterance. Grice's theory is in that sense only part of the story. The result of dialogue is more properly described as *shared* understanding. If communication is successful, speaker and hearer share understanding with respect to the speaker's speech act. As Charles Taylor said, "Communication doesn't just transmit information (. . .). It brings about the acknowledgement that some matter is *entre nous,* is between us" (Taylor 1980: 295).

This is obviously a large issue that needs careful discussion (see again Meijers 1994, 2002b). But we need not go into the details of the analysis here. What is important for the discussion of collective speech acts is that there is a sense in which *all* speech acts are collective acts. As communicative acts they are performed together, and require collective intentionality. The intention to perform a promise, a request, or an order, is embedded in the collective intention '*We* intend to communicate'. In addition, a successful speech act requires 'uptake' or successful communication and the result, shared understanding, introduces again a collective dimension in the analysis.

In the discussion below I will take this collective dimension of speech acts for granted. It applies to *all* speech acts performed in dialogue, not to a particular type of speech act. My focus in this paper will instead be on a particular type of speech act, namely those speech acts which involve collective agents, either as speakers or as hearers. This type is different from speech acts which involve only individual agents, and which have been the exclusive focus of speech act theory thus far. I will call the latter *individual speech acts,* to distinguish them from *collective speech acts* which involve collective agents. There are three forms of collective speech acts:

– Those in which a collective speaker addresses an individual hearer (e.g., 'We expect you to be in time');
– Those in which an individual speaker addresses a collective hearer (e.g., when the coach says to the team 'Play your own game and not your opponent's');
– Those in which a collective speaker addresses a collective hearer (e.g., when the prime minister says to the secretary general of the UN: 'We accept resolution 1701 and will respect a ceasefire').

Collective speakers are restricted to structured groups of the intentional type. They have to be intentional agents. Collective hearers, on the other hand, can be either crowds or structured groups of the physical or intentional type. We may ask a crowd to disperse, tell the passengers waiting at an airport that all flights have been cancelled, or order an army to withdraw.

4. ARE COLLECTIVE SPEECH ACTS REDUCIBLE?

It is a well-known fact that in uttering a particular sentence, more than one speech act may be performed. The assertion "It will be a formal dinner", for example, may also be the directive speech act "Don't forget to dress properly". Searle's analysis of indirect speech acts has generated a lot of interest in this phenomenon. The fact that we can perform more than one speech act in uttering a single sentence, has relevance for the analysis of collective speech acts. It is an interesting question whether upon closer analysis collective speech acts are nothing but a collection of individual speech acts. The idea is that instead of performing *one* speech act, a collective speaker makes several speech acts, where the number will depend on the number of individuals making up the collective. The 'We' is conceived as a set of 'I's. Similarly, a speech act addressed to a collective hearer, for example a crowd, may be a set of speech acts, where the set is defined by the individuals making up the crowd. The obvious advantage of such a reductive strategy is that we don't have to introduce a new category of speech acts, since the existing types can do the job. This is an issue that every account of collective speech acts will have to address. Are collective speech acts reducible?

Let me start with an example. Imagine a case in which the management team of National Aeronautics and Space Administration (NASA) makes the following

speech acts at a press conference: "We believe that the shuttle is safe enough to launch", "We intend to launch the shuttle later today", and "We will do everything possible to make sure that the shuttle returns safely". A reductive approach will have to rewrite these collective speech acts as a combination of individual speech acts, where the 'We' is replaced by 'I's. One way to do this is to rewrite the speech act

(a) *We believe that the shuttle is safe enough to launch*
as a conjunction of the speech acts
(b) *I believe that the shuttle is safe enough to launch &*
(c) *I believe that the shuttle is safe enough to launch*
and so on, for all the members of the management team.

The question, then, is: Is this conjunction of individual speech acts equivalent to the original collective speech act?

To understand why they are not equivalent we have to analyse the conditions of success of speech acts. Searle (1969: 57ff., 64ff.) distinguishes four types of conditions of success: the propositional content condition, the preparatory condition, the sincerity condition, and the essential condition. The propositional content condition specifies conditions on the type of content that can be part of a particular type of speech act. As we will see, there are propositional content conditions that are specific to collective speech acts. This is a strong indication that the equivalence relation between collective speech acts and a conjunction of individual speech acts does not hold.

Imagine that (as has in fact happened in the past) there are dissenting opinions in the management team of NASA with respect to the safety of the shuttle. If the management team want to launch it, they will have to make up their mind with respect to the aircraft's safety. In such a situation it can very well be that the management team as a collective body takes the position that the shuttle is safe, while there are members who do not share this view. These members, then, are not willing to perform the individual speech act "I believe that the shuttle is safe", while they are willing to perform the collective speech act "We believe that the shuttle is safe". More generally, collective speech acts involve collective intentional states (beliefs, desires, intentions), whose content cannot be unconditionally attributed to the participants outside the context of their collective action. As independent individuals, they may believe, desire or intend differently.

A further indication that the equivalence relation does not hold is that in the example "We intend to launch the shuttle today", the reduction of the collective speech act to a conjunction of individual speech acts of the type "I intend to launch the shuttle today" does not make sense. The launching of the shuttle is a collective act which cannot be carried out individually and thus cannot be the content of an individual intention (similarly, I cannot have the individual intention to perform Monteverdi's *Vespers of St. Mary*, since it takes a choir and basso continuo to do that). Here the individualist may argue that there are ways

of rewriting the collective speech act that do not violate the propositional content condition. For example, one might rephrase the speech acts as a conjunction of individual speech acts of the type "I intend that we launch the shuttle today", or "I intend to do my part in launching the shuttle". Though this seems to avoid the problem in a number of cases, we will see later that such a reformulation violates other conditions of success of speech acts.

Both examples show that the propositional content condition makes a reduction of collective speech acts difficult and possibly restricted to cases where the propositional content of collective speech acts can satisfy the propositional content condition of individual speech acts. The other conditions for the successful performance of speech acts point in the same direction. Preparatory conditions are conditions that need to be fulfilled in order for the speech act to be performed properly. They concern, among other things, the physical, mental and social status of the speaker and the hearer. Collective speech acts obviously require a collective agent to be in place. A conjunction of individual speech acts, on the other hand, requires a number of individual agents to be in place. These requirements are the same *only if* a collective agent can be conceived of as a summation of individual agents. Or, put differently, only if a straightforward reduction of collective agency is possible. I will address this issue below.

Searle's third condition, the sincerity condition, specifies the psychological state expressed by a speech act. An assertion counts as the expression of the belief that the propositional content is true, a promise counts as the expression of the intention to do what is promised. The collective promise "We will do everything possible to make sure that the shuttle returns safely", thus counts as the expression of a collective intention to do everything possible to make sure that the shuttle returns safely. More generally, collective speech acts express collective intentional states, whereas individual speech acts express individual intentional states. The reductionist strategy works, then, *only if* collective intentional states can be reduced to (combinations of) individual intentional states (whatever their content is). I will argue later that such a reduction is not possible.

Finally, the essential condition specifies the illocutionary point of a speech act, i.e., what the speaker wants to accomplish in performing the act. The point of an order is to get the hearer to do what is ordered, the point of a promise is to place the speaker under the commitment to do what is promised. A collective speech act is in that sense different from a collection of individual speech acts. Take again the case of the collective promise "We will do everything possible to make sure that the shuttle returns safely". Here the management of NASA places itself under the commitment to do what is promised. They do this as a collective agent, even as an institutional body, and *not* as a collection of individual persons. The commitment to carry out the promised action thus remains in place even if individual members of the management are replaced by others (unless the promise is explicitly withdrawn).

The attempt to reduce a collective promise to a conjunction of individual promises can only succeed if a collective agent is nothing more than a summation

of individual agents and if a collective intention can be reduced to a collection of individual intentions. This was also the outcome of the discussion of the preparatory condition and of the sincerity condition. It is time now to address this issue directly: (i) Are collective agents nothing more than collections of individual agents and (ii) are collective intentional states nothing more than collections of individual intentional states? A negative answer will refute the reductive strategy, and will have the consequence that collective speech acts must be regarded as *sui generis* acts. These questions are obviously large questions which have been discussed extensively in the literature. I believe that the second question is the more fundamental one. Collective intentional states already have a 'We' as a subject, and the minimal form of a collective agent is two individuals sharing an intentional state (a physical or biological relation is not enough for agency, though it may result in a structured group).

That collective intentionality is an irreducible, primitive phenomenon has been argued by many philosophers, including Margaret Gilbert, Raimo Tuomela, and John Searle. In Searle's view, there is a "deep reason" why collective intentionality cannot be reduced to individual intentionality. "No set of 'I-Consciousnesses', even supplemented with [mutual] beliefs, adds up to a 'We-Consciousness'. The crucial element in collective intentionality is a sense of doing (wanting, believing, etc.) something *together* [my italics], and the individual intentionality that each person has is derived *from* the collective intentionality that they share" (Searle 1995: 25). Many species of animals, according to Searle, have this "biologically primitive" capacity for collective intentionality, i.e., the capacity for cooperation and *sharing* intentional states.

I believe that Searle is right. The awareness of doing something together cannot be 'synthesized' out of the awareness of individual agents doing something individually. Additional arguments can support the non-reductionist position. I have argued for the irreducibility of collective intentionality on the basis of an analysis of the commitments involved (Meijers 2003a). Take the example of the launching of the shuttle. Suppose that members of the management team agree to do this. Since launching the shuttle is a complicated affair, they will intend to do this *together*. Such an intention brings about commitments among the agents who make up the collective agent. There will be a division of labour in which they are committed to do their part. Not only that, they are also entitled to *claim* that others do their part as well. The collective intention thus creates inter-individual commitments and normative relationships among the members of the team. The crucial point for the present discussion is that no set of individual intentions can ever generate *inter-individual* commitments. An individual intention may eventually generate a commitment to oneself to do what one intends to do (but such a commitment may be nothing more than the original intention), but it can never generate commitments to others that I do that, let alone a claim by others that I do that. Collective intentions, on the other hand, create such inter-individual commitments and claims. From the fact that *we* intend to do something together, we are both committed towards each other

to do our part and I can claim that you do your part and you can claim that I do mine. Seen from this perspective, then, the irreducibility of collective intentionality stems from the impossibility to generate inter-individual commitments on the basis of individual intentionality.

There is an ontological side to this as well. Though I agree with Searle that collective intentionality is irreducible, I find his conception of collective intentionality problematic from an ontological point of view. Let me explain. Searle formulates two conditions of adequacy for an account of collective intentionality: (i) it must be consistent with the fact that society consists of nothing but individuals, and (ii) it must be consistent with the fact that all intentionality, whether individual or collective, could be had by a brain in a vat (Searle 1990: 407). These two conditions, which state the obvious for Searle since they are presented as "facts", result in an individualistic conception of collective intentionality. We-intentions are the intentions of *individual* agents. These agents are capable of individual *and* of collective intentionality. Ultimately, it is therefore possible that a *single* individual agent has a collective We-intention, for example the intention "We intend to launch the shuttle today".

This is a very unfortunate consequence of Searle's theory in my view. As we have seen, inter-individual commitments are part and parcel of collective intentionality. These commitments exist *between* individuals. They cannot be accounted for in terms of the intentional states of a *single* individual, because that would effectively amount to the reduction of relations to the intrinsic properties of one of the relata. I take it to be evident that this cannot be done. The question, then, is: Can Searle's theory be repaired to avoid this unacceptable consequence? I believe it cannot, because the consequence is strongly related to Searle's internalist conception of the mind. It would mean, among other things, giving up the second condition mentioned above. This condition, however, is not a peripheral item of the theory, but one of the key tenets of Searle's philosophy (see Searle 1983). As such it is a condition for *all* accounts of intentionality, including collective intentionality.

An externalist conception of collective intentionality, in which genuine relations *between* individuals play an important role, avoids these problems.[2] In this conception collective intentions need to have a foundation *in re*. There has to be another agent 'out there', so to speak, for collective intentionality to be possible. If no other agent is in fact 'out there', there is no collective intentionality. Of course, individual agents may have the false belief that their intentions are shared and they may act *as if* there is such collective intentionality. But the beliefs and the intentionality in question will then be just that: false beliefs and *as-if* intentionality.

Whatever the outcome of the debate on the ontology of collective intentionality is, the main result of our discussion for this paper is that collective

[2] See for a more elaborate account Meijers (2003b). See also the discussion on intentionality *de re* between McDowell (1991) and Searle (1991).

intentionality cannot be reduced to individual intentionality and that consequently *collective speech acts are sui generis acts*. They form a separate class of speech acts, and cannot be reduced to combinations of individual speech acts.[3]

5. THE ANALYSIS OF COLLECTIVE SPEECH ACTS CONTINUED

The previous sections indicated already a number of ways in which collective speech acts are different from individual speech acts. In this section I will explore more differences between individual and collective speech acts and in the next section I will discuss the consequences for Searle's taxonomy of speech acts.

As we have seen already, speech acts are communicative acts and as such they have certain characteristics. They are performed *in* the act of speaking, they are addressed to a hearer and they require uptake or understanding by the hearer in order to be successful. There is an important asymmetry between individual and collective speech acts in that a collective speaker (say the management of NASA) cannot literally utter sentences. Only individuals can do that. Collective speech acts are therefore performed by individual speakers on behalf of the group. The analysis of the intentions involved in collective speech acts has to reflect that.

In his pioneering article "Group Speech Acts", Hughes (1984: 388) analysed the intentions involved in collective speech acts and the conditions that have to be fulfilled for these speech acts to be successful:

> The conditions, then, of a group speech act are as follows [for a group G, speaker S and utterance X]:
>
> (1) There exists a group (G), this group has an illocutionary intention, and X conveys that illocutionary intention.
> (2) S (believes that he/she) knows the illocutionary intention of G and that X conveys this illocutionary intention.
> (3) G does not object to S uttering X on its behalf and if G intends for any specific individual(s) to utter X, it intends for S to utter X. S (believes that he/she) knows this.
> (4) #2 and #3 are (the) reasons S utters X.

These conditions are preparatory conditions for collective speech acts. They concern the individual speaker who will speak or write on behalf of the group.

---

[3] An exception may be those cases in which an individual speaker addresses a collective hearer, where the hearer is a crowd or a structured group of the physical type. Imagine that there is a fire in a building and that the porter orders the inhabitants via the loudspeakers in the rooms to leave the building as soon as possible. This collective speech act can probably be analysed as the conjunction of a number of individual speech acts, though even in this case there is the complicating factor that the speech act will become part of the public domain and that the hearers will have common knowledge about it. They know that the others know that they know that the others know [etc.] that the speech act has been performed.

The first condition is obvious and simply states that there has to be a group, that the group has to have an illocutionary intention, and that the utterance has to convey this intention. The second condition, however, introduces a new element in the analysis of speech acts. There is an interesting asymmetry between individuals and groups in an epistemic sense. Individuals usually don't have to make an extra effort to know their illocutionary intentions when performing a speech act. They just know it. Groups, on the other hand, do not have the type of epistemic access that individuals have. In order to know what a group's intention is, a conscious effort need to be made by the speaker who acts on behalf of the group. Usually he or she needs to consult other members and there has to be some procedure to decide on the group's intention. The condition states that the speaker knows or believes to know what the intention of the group is, and also that he or she knows that the utterance of X conveys this illocutionary intention. The third condition is about the delegation of the speech act. Notice that the actual speaker need not be a member of the group for the collective speech act to be successful. Hughes makes a useful distinction here between the 'intender' of the speech act (the group) and the actual speaker. In a collective speech act they may be different, while they are always the same in individual speech acts. The condition allows for the explicit delegation of the speech act, as well as for situations in which there is no such explicit delegation (for example, when all members can speak on behalf of the group). The final condition states that the reasons for performing a collective speech act should be internally related to the group's illocutionary intentions. If it is done for other reasons, for example, self-interest of the speaker, the speech act does not count as a *collective* speech act.

Hughes' conditions contribute to the analysis of the complex phenomenon of a collective speech act. Most of them just add to the list of preparatory conditions that Searle already distinguished. There is one point that deserves special attention. It concerns the way in which groups develop their collective aims, beliefs, desires, or intentions. For Hughes, "the ideal group is characterized by active, open, and free exchange among all members of the group. In this interchange the utterances of individuals are directed to all other persons of the group"; at the heart of such a group there is a "consensus mechanism" (1984: 385, 384). This mechanism takes care of the formation of the group's illocutionary intentions, or, more generally, its collective intentional states. Deviant cases are cases where the consensus mechanism is marginalized, usually because of urgent situations: an SOS message on behalf of the group without prior deliberation, or an offer for a cease-fire made by a gang leader during a fight between groups, without holding a referendum first. But even in these cases there may be a rudimentary consensus mechanism at work, for example, based on body language. But it appears, according to Hughes, "that [in these cases] we recognize group speech acts as such if we believe that there is some other basis by which the utterer 'knows' (is justified in believing he knows) the group's intention" (1984: 384).

Hughes's analysis suffers, in my view, from taking a particular type of group, an ideal democratic group, as the general paradigm for groups (*pars pro toto*), thereby

taking a particular type of mechanism for the formation of collective intentional states, the consensus mechanism, as a general mechanism. As a matter of fact, the situation of a really free exchange of ideas that leads to a consensus in which power relations among the members of the group do not play any role, seems to me rather exceptional. It is even doubtful whether this can be used as a paradigm on which real group processes can be modelled. As we have seen in Section 1, there are several types of groups, and probably several types of developing collective views. A family will be different in this respect from a company.

Groups in which agents are not only allowed to speak on behalf of the group, but are also authorized to act on behalf of the group, are especially interesting from the point of view of speech act theory. Imagine a structured group of the intentional type, in which the leader of the group has been authorized to do whatever he or she thinks is necessary in a specific situation, for example, a negotiation. The leader will utter speech acts such as "We believe the proposal is unacceptable", or "We will withdraw from the negotiations if the political situation does not improve". Or, to take another example, imagine that President George Bush utters the speech act "The US government believes that the axis of evil includes stem cell biologists", without consulting his ministers. His utterance is nevertheless a collective speech act made by the government of the US, because it is not an option for the members of his government qua members to deny that this speech act expresses a collective belief of the US Government to that effect. Why is this so?

Contrary to Hughes's idea, that in these cases we recognize group speech acts as such if we believe that there is some basis by which the utterer *knows* the group's intentional states about the subject, I believe that we encounter here a phenomenon that is similar to the one Searle (1989) described in his analysis of performatives. In a performative, for example, "I hereby promise you to come tomorrow", the speaker makes it the case that he or she is promising by *declaring* that he or she is promising. Similarly, an authorized speaker can make it the case that his utterance expresses a group belief by *declaring* that it expresses that belief. When Bush says that the US government believes that the axis of evil includes stem cell biologists, this belief has *de facto* become the collective belief of the US government. In that respect these collective speech acts are declarations, just as performatives are declarations. They are declarations in addition to being assertives, directives, commissives, or expressives.

6. CONSEQUENCES FOR SEARLE'S TAXONOMY OF SPEECH ACTS

Speech acts differ in many respects.[4] The style of the performance, the force of the illocutionary point, the extralinguistic institutions needed for the speech act, or the relation to the rest of the discourse, are among the many properties of speech acts that can be used to bring order in the apparent chaos. Searle's taxonomy is constructed on the basis of three main qualities of speech acts:

[4] In his paper "A Taxonomy of Illocutionary Acts" (Searle 1979: Chap. 1), Searle lists at least 12 respects in which speech acts can differ.

their illocutionary point, their direction of fit, and their expressed psychological state. The resulting categories are substantially different from each other and very useful in practice. Searle distinguished five types of speech acts that are well-known ever since: assertives, directives, commissives, expressives, and declarations. A key question for this paper is: Do collective speech acts fit Searle's taxonomy, i.e., can we distinguish the same types of collective speech acts on the basis of these qualities? If so, that would show that collective speech acts are not only different from individual speech acts, but also similar in important respects.

Assertives are the first category of speech acts. Their illocutionary point is to commit the speaker in varying degrees to the truth of the propositional content of the speech act; the expressed psychological state is a belief in that content; and the direction of fit is 'words-to-world', i.e., what is asserted in the speech act needs to fit the state of affairs in the world. For example, in my assertion "It is safe to launch the shuttle", I commit myself to the truth of the proposition that it is safe to launch the shuttle, the speech act expresses my belief to that effect, and the speech act is supposed to reflect this state of affairs in the world. Assertives can be performed by collective agents. In such a speech act a collective speaker is committed to the truth of the propositional content, the speech act expresses the collective belief in the content and the direction of fit is words-to-world.

There is, however, a complication with respect to the notion of a collective belief. In what sense can a collective agent be said to have a genuine belief and what would that add to the beliefs which the individuals making up the collective already have? For this discussion it is important to make a distinction between *belief* and *acceptance*.[5] To believe a proposition is to be committed to the truth of that proposition, where only epistemic reasons count as valid reasons for an agent to believe that proposition. Acceptance, on the other hand, is a different cognitive attitude. To accept a proposition is to adopt that proposition as a valid premise in one's practical reasoning about future actions, where in addition to epistemic reasons practical considerations play a role. For example, for reasons of prudence an agent may accept in one's practical reasoning the proposition that the Dow Jones Index will go down. In such a situation the agent may even believe in an epistemic sense that the Index will go up, but plays it safe and acts under the assumption that the Index will go down.

Given this distinction, there are two readings of a collective belief: We may interpret it as a genuine belief, or we may take it as the acceptance of a proposition. I have argued for the second view.[6] Let us review the options before us. If we take it as a genuine belief, only epistemic considerations can play a role; in matters of truth, practical considerations do not matter. This means that if the management of NASA asserts: "It is safe to launch the shuttle", the members

[5] The distinction goes back to Cohen (1992) and Bratman (1993).

[6] See Meijers (2002b) and (2003b), Gilbert (2002) and Wray (2001). The view has become known as 'rejectionism' with respect to collective beliefs in the literature.

of the management *qua members* are committed to the truth of that proposition. And if they are committed qua members, they are also committed as individuals. Otherwise we would have a contradiction: you cannot in one role believe that X is true and in another role believe that X is false. A collective belief in this reading, then, is nothing more than the conjunction of the individual beliefs of the members of the group. Consequently, if one of the members changes view and starts believing that the shuttle is *not* safe to launch, the earlier collective belief ceases to exist. The question, then, is: What does the notion of a collective belief add to our analysis here? I am tempted to say that it adds nothing. We can account for such a collective belief in a reductionistic way, because we can rewrite this belief as a conjunction of individual beliefs. The group's commitment to the truth of the proposition, however, does add something to the individual commitments to its truth.

In the second reading, a collective belief is interpreted as the collective acceptance of a proposition. Let us take again the situation in which there are dissenting opinions within NASA's management about the shuttle launch, but that they agree as a group that it is safe to launch the shuttle. At a press conference the chairman will then say: "We believe that it is safe to launch the shuttle." Though this is literally said to be a belief, the expressed psychological state is really the acceptance of a view, for the decision taken is based on epistemic *and* practical considerations. For example, they may have reasoned that they cannot postpone the launch indefinitely, that further upgrading of the shuttle will only marginally improve the situation, that funding may be in danger if they will not launch the shuttle shortly, and that therefore the majority view in the management will be followed. In this situation members *qua members* of the management will accept as their joint view that the shuttle is safe to launch, while some members will individually have reservations and believe differently. They accept it, and in their acceptance non-epistemic reasons play a role. My conclusion is that either we have a reductive notion of collective belief as the conjunction of individual beliefs, or we have an irreducible notion of collective belief which turns out to be the acceptance of a view upon closer analysis.[7] The expressed psychological state of a genuine collective assertion should be described as the collective acceptance of a proposition rather than as a collective belief.

Directives are Searle's second type of speech acts. Their illocutionary point is to get the hearer to do what is specified in the propositional content of the speech act. The expressed psychological state is the speaker's wanting that the hearer does this, and the direction of fit is world-to-words. The speech act is satisfied if the hearer does what the speaker wants him to do. Collective agents can perform such directive speech acts. For example, when the parliament asks a minister to be present at a meeting. The illocutionary point here is to get the minister to be present, and the expressed psychological state is a collective desire

[7] See also the discussion in the literature mentioned in the previous footnote. Volumes 16 (2002) and 18 (2003) of the journal *ProtoSociology* are partly devoted to this issue.

to that effect. Things are different when a collective agent is involved as the *hearer* of the speech act and the speaker is an individual agent. For example, when a coach urges his team to play defensively. The expressed psychological state then is an individual desire.

Commissives form the third category of speech acts and their illocutionary point is to commit the speaker to some future course of action. The expressed psychological state is the intention to do what is promised, and the direction of fit is world-to-words. Collective agents can perform collective promises. For example, when the government promises to the parliament to improve airport security. In doing so, the government places itself under the commitment to improve airport security, the expressed psychological state is the collective intention to do what is promised, and the speech act is satisfied if the world changes in a way that the speech act specifies. Similar to the case of directives, if the speaker of a collective promise is an individual agent and the hearer a collective, the expressed mental state is an individual intention.

Expressives are a type of speech act whose illocutionary point it is to express the psychological state specified in the propositional content. The sincerity condition or expressed psychological state is again the one mentioned in the propositional content, while expressives do not have a direction of fit. Collective agents are able to perform expressive speech acts. For example, when the government formally expresses regret for an action by one of its ministers. The point of such a speech act is to express this feeling of regret, and the expressed psychological state is the government's collective feeling of regret with respect to the action of the minister. Notice that agents need not actually have these (collective) psychological states, the speech act *counts as* the expression of these states. This is a general point that applies to all types of speech acts.

Declarations are the fifth and last type of speech acts Searle distinguishes. Their illocutionary point is to bring about a correspondence between the propositional content of the speech act and reality, i.e., to make true in the world what is said in the content of the act. The expressed psychological state is twofold, namely a belief and a desire, and the direction of fit is both words-to-world and world-to-words. Declarations can be performed by collective agents. For example, when the jury utters the verdict: "Guilty", the illocutionary point of the jury is to declare the subject to be guilty. Due to this act, he or she changes status from suspect to guilty person. Or, more precisely, he or she is both made a guilty person by this act and described as such. The expressed psychological states are the collective acceptance of the proposition that the suspect is guilty and the desire to change the suspect's status into that of a guilty person. As we have seen with the other types of speech acts, collective declarations can have collective agents as speakers and as hearers. An example of the latter is when a soccer team is declared to be the best team of the year by a panel of journalists.

The picture that emerges from our discussion is that collective speech acts roughly fit Searle's taxonomy of individual speech acts. The same types of speech acts can be distinguished using the three qualities of illocutionary point, expressed psychological state, and direction of fit. I said 'roughly' because we

encountered a problem with respect to the expressed psychological state in the case of assertives. For an individual assertive this may be a belief or an acceptance, for a collective assertive this is always the acceptance of a proposition. The problem can be solved by describing these two psychological states on a higher level of abstraction in the taxonomy. For example, both can be described as the 'assent to a proposition'. By definition, then, this would cover the individual speaker's commitment to the truth of the proposition, and the collective speaker's commitment to the practical value of a proposition in practical reasoning. In my view, such a solution is, to a large extent, a cosmetic, because it conceals rather than solves the initial ambiguity.

We are confronted now with a puzzle. On the one hand, collective speech acts are largely similar to individual speech acts in terms of the three main qualities of speech acts and the taxonomy based on them. On the other hand, we have found that collective speech acts are really different from and irreducible to individual speech acts (see Section 4). They are *sui generis* acts. As we have seen the main reason for their irreducibility is that collective agents and collective intentionality cannot be reduced to individual agents and individual intentionality. The resulting puzzle shows, in my view, that an important quality of speech acts is missing, which should have been included in the construction of Searle's taxonomy. It is the *kind of agency* involved in speech acts. This quality is orthogonal to the illocutionary point and the expressed psychological state of a speech act, in that these two qualities allow for an individual and a collective reading. It is no surprise, then, that the resulting taxonomy of speech acts allows for individual as well as for collective agents, both as speaker and as hearer.

The addition of an extra quality or dimension to the basis on which the taxonomy of speech acts will be constructed has a rather fundamental impact. It converts Searle's two-dimensional matrix into a three-dimensional one, as shown below:

Type of speech act	Illocutionary point	Direction of fit	Expressed psychological state
Assertive	commits the speaker to the truth/acceptance of a proposition	words-to-world	belief/acceptance
Directive	attempts to get the hearer to do something	world-to-words	desire
Commissive	commits the speaker to some future action	world-to-words	intention
Expressive	expresses a psychological state about a propositional content	none	varies with illocutionary point
Declaration	creates a correspondence between propositional content and reality	world-to-words and words-to-world	belief/acceptance and desire

The third dimension of 'kind of agency involved' can have a number of values, given the typology of groups I have discussed in Section 2. At one extreme there is only an individual agent, at the other there is a structured group of the intentional type, with several other types of groups in between.

One may wonder whether Searle's taxonomy already includes elements of collective agency, and thus a hidden third dimension after all. A case in point is Searle's category of declarations. Though there may be exceptional cases of individual declarations—for example, when a speaker defines a term—most declarations require extralinguistic institutions and thus collective intentionality and collective agency. And in many cases a speaker acts as the authorized delegate of an institution. Take, for example, the standard example of declaring somebody husband and wife. Though an individual civil servant performs this declaration, the official acts and can only act on behalf of an institution in such a case. It seems therefore more appropriate to say that it is the state that has declared them husband and wife, rather than the civil servant. That makes the act a collective speech act. Or, to give another example, if a head of state declares war to another nation, it is not that particular individual who declares war, but the state one represents. Again, this seems to be clearly a collective speech act.

7. THE CONSTRUCTION OF SOCIAL REALITY

On Searle's view (Searle 1995: Chap. 1), there are three building blocks of social reality: the imposition of function on entities that do not have that function prior to the imposition, collective intentionality and the distinction between constitutive and regulative rules. In a trivial (because definitional) sense every collective speech act creates a social fact, then, since it essentially involves collective intentionality

Social reality consists to a large extent of institutions and institutional facts. Institutional facts, according to Searle, are created by the collective imposition of status functions on entities according to the constitutive rule "X counts as Y in context C" (Searle 1995: Chap. 2 and 3). These status functions exist only because they are collectively imposed and cannot be achieved solely in virtue of the physics of these entities. Language is fundamental in this analysis in two ways: (i) there has to be a *symbolic representation* of the status function Y, since this function goes beyond the physics of the X; this symbolic representation has to be conventional and publicly understandable; (ii) the imposition is collective; this means, among other things, that communication will be involved in the process of the collective imposition of status functions.

Searle's claim that language is "essentially constitutive of institutional reality" (1995: 59) can also be phrased in terms of the previous analysis of collective speech acts. The collective imposition of function obviously requires a collective subject, the rule "X counts as Y in context C" has the form of a declaration, and since the Y term needs to be symbolically represented in a publicly understandable way, it will usually involve a linguistic act. Collective declarations have exactly

the requisite properties. When the chairman of a jury says: "The suspect is guilty", an institutional fact is created by a collective declaration and new deontic phenomena such as powers, rights, obligations, and duties will have been created. Collective declarations, then, are essentially constitutive of institutional reality.

8. CONCLUSION

Up to now, speech act theory has had a blind spot for those speech acts that are performed by collective agents. The paradigm case has been an individual speaker addressing an individual hearer. I have shown in my paper that speech acts performed by collective speakers are *sui generis* acts. They cannot be reduced to individual speech acts. The analysis of their propositional content condition, their preparatory conditions, their sincerity condition, and their essential condition has shown important differences with individual speech acts. Collective speech acts also involve preparatory conditions that are specifically related to their delegation and authorization. The consequence of my analysis is that Searle's taxonomy of speech acts has to be modified because an important classifying principle, the type of agency involved, is missing. Searle's claim that language is essentially constitutive of institutional reality can be rephrased on the basis of the previous analysis: collective declarations are the stuff institutional reality is made of.[8]

REFERENCES

Bratman, M. (1993). Practical reasoning and acceptance in a context. *Mind*, 102: 1–15.

Cohen, J.L. (1992). *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.

Crosby, J. (1990). Speech act theory and phenomenology. In A. Burkhardt (ed.) *Speech Acts, Meaning, and Intentions: Critical Approaches to the Philosophy of John R. Searle*. Berlin: De Gruyter.

Gilbert, M. (2002). Belief and acceptance as features of groups. *Protosociology*, 16: 35–69.

Grice, H.P. (1957). Meaning. *Philosophical Review*, 66: 377–388.

Grice, H.P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Harré, R. (1997). Crews, clubs, crowds, and classes. In J.D. Greenwood (ed.), *The Mark of the Social*. Lanham, MD: Rowman & Littlefield.

Hughes, J. (1984). Group speech acts. *Linguistics and Philosophy*, 7: 379–395.

McDowell, J. (1991). Intentionality *De Re*. In E. Lepore and R. van Gulick (eds.), *John Searle and His Critics*. Oxford: Blackwell.

Meijers, A. (1994). *Speech Acts, Communication and Collective Intentionality: Beyond Searle's Individualism*. Leiden, The Netherlands: Rijks Universiteit.

Meijers, A. (2002a). Dialogue, understanding and collective intentionality. G. Meggle (ed.), *Social Facts and Collective Intentionality*. Frankfurt: Hänsel.

Meijers, A. (2002b). Collective agents and cognitive attitudes. *Protosociology*, 16: 70–85.

Meijers, A. (2003a). Why accept collective beliefs? Reply to Gilbert. *Protosociology*, 18: 377–388.

Meijers, A. (2003b). Can collective intentionality be individualized? *American Journal of Economics and Sociology*, 62: 167–183.

[8] I would like to thank Savas Tsohatzidis for stimulating comments on an earlier version of this paper.

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
Searle, J.R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.
Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
Serale, J.R. (1989). How performatives work. *Linguistics and Philosophy*, 12: 535–558.
Searle, J.R. (1990). Collective intentions and actions. In P.R. Cohen, J.L. Morgan and M.E. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press.
Searle, J.R. (1991). Reference and intentionality. In E. Lepore and R. van Gulick (eds.), *John Searle and His Critics*. Oxford: Blackwell.
Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.
Taylor, C. (1980). Critical notice on *Linguistic Behaviour* by Jonathan Bennett. *Dialogue*, 19: 290–301.
Wray, K.B. (2001). Collective beliefs and acceptance. *Synthese*, 129: 319–333.

Part II
From Intentions to Institutions: Development and Evolution

THE ONTOGENY OF SOCIAL ONTOLOGY: STEPS TO SHARED INTENTIONALITY AND STATUS FUNCTIONS

Hannes Rakoczy and Michael Tomasello

Searle's (1995) book *The Construction of Social Reality* is one of the great works of late 20th-century philosophy. Like all great philosophical works, it provides fresh new insights and perspectives on phenomena that lie right before our eyes every minute of every day. After Searle's work, in our view, no right-minded thinker can seriously maintain a narrowly reductionist view of the nature of human experience. After *The Construction of Social Reality*, in our view, no right-minded thinker need take seriously philosophical theories that ignore the fundamentally social nature of many of the most important aspects of our daily lives.

The basic proposal is that humans live as much in a social and institutional world as they live in a physical one. In a football game, the teams and the score are as real as the grass and the dirt on which the game is played. Searle thus develops an ontological framework for the description of basically everything. We live in one world which at bottom consists of brute, observer-independent facts. At the very bottom of the brute facts are the physical facts, a level up the chemical, the biological, and among the latter the intentional facts.[1] Humans and other animals are intentional beings. That Jack believes Mary loves him, and that she hopes that he loves her are objective intentional facts about them. They are brute in the sense of holding independently of any particular observers. But humans (and according to Searle many other animals) not only are individual intentional beings, but they also enter into irreducibly collective we-intentionality with others. When Jack and Mary go dancing together, each intends not only to dance but that "we dance together". Where collective intentionality is involved, we deal with social facts, and again, social facts are brute, objective facts about groups of people. In addition, there is a specific class of collective intentionality that involves the collective application of constitutive rules and the assignment of some kind of deontic status to actions and objects—and here we leave the realm of brute facts and enter the realm of institutional reality, which is not observer-independent. That a group of people is a football team and that a certain move is offside are facts only within

[1] This prima facie looks like the usual levels picture invoking the notion of supervenience or a similar determination relation. Searle's formulation of this picture, though, is a bit idiosyncratic in that he talks of causation between lower and higher levels (e.g., 1983, 1992, 1995) which is inconsistent with a supervenience reading and most levels conceptions (see, e.g., Kim 1995). We will ignore these details in the following.

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 113–137.
© 2007 Springer.

the collectively intentional practice of playing football. Take away the practice, for example, in a culture that does not play football, and there is no team and no offside left (though the grass and the dirt, and the people and their intentionality remain). Human social life, from language to money, marriage, and government, is mainly a matter of such institutional reality.

In developing his theory Searle in many places makes empirical claims. Specifically, he claims that of the main components in the construction of human social reality, social animals such as hyenas and chimpanzees possess the most basic, collective intentionality, and that they lack only the more normatively loaded status functions and constitutive rules—and the deontic powers these create. But we think that Searle has his facts about the animals wrong, and that this has some unfortunate consequences for his analysis of the origins of human social reality. The empirical evidence suggests that other animals, including our nearest primate relatives, do not really have shared or collective intentionality of the human kind—using either Searle's own definitions and concepts explicitly, or reasonable extensions or operationalizations of them.

Obviously, empirical facts do not determine conceptual distinctions, but more the other way around, and so being wrong about the facts is in most cases not lethal to a philosophical analysis. But in this case Searle's promiscuous attributions of collective intentionality belie a lack of appreciation, in our opinion, of the complexities and power of this very rare form of social interaction. In this chapter we attempt to spell out these complexities and this power by comparing the social interactions of human children to those of our nearest primate relatives, chimpanzees, with regard to social interactions involving: (1) collaboration, (2) communication, and (3) social learning. The resulting analysis of the skills and motivations which constitute collective intentionality also enables us to propose some plausible hypotheses about how human children move developmentally from fairly local social interactions involving shared intentionality to participation in full-blown institutional reality in the form of constitutive rules, status functions, and the deontology these create.

1. CHIMPANZEES AND HUMAN CHILDREN

Human children as well as many other animals engage in complex, socially coordinated behavior. Here we compare children to our closest primate relatives, chimpanzees, with an eye on the question of how to distinguish social coordination from collective intentionality[2] proper.

[2] Previously we (Tomasello and Rakoczy 2003) distinguished between 'shared intentionality' and 'collective intentionality', the former referring to here-and-now collaboration with specific others, the latter referring to broader, normatively structured societal interactions with others representing generalized others sensu Mead. This distinction is close to Searle's distinction between social and institutional reality. To avoid terminological confusion, though, in the present chapter we follow Searle in using 'collective intentionality' and 'shared intentionality' interchangeably, in the broad sense of we-intentionality creating social reality, with institutional reality being a specific subclass of social reality.

1.1. Collaboration

Humans share with many species the ability to act intentionally. Piaget (1936) demonstrated that from 8–9 months human infants engage in coordinated means-ends behavior, discriminating between success and failure (failure often leading to the choice of different means). Köhler (1925) and subsequent research has demonstrated that chimpanzees and many other animals act intentionally in this same way. Searle (1994, 2001) has long emphasized the continuity of human and other animal individual intentionality, defending our common sense practice of ascribing intentions to cats, dogs, and Köhler's apes against overly linguistic philosophical theories of what intentionality requires (such as Davidson's), theories that "overintellectualize the mind" (Hurley, 2001). At the same time he has stressed the specifically human ways of thinking that are linguistically and culturally constituted. Searle thus advocates a differentiated approach to the individual intentionality of humans and other animals, attempting to identify both similarities and differences.

When it comes to collective intentionality, however, we think Searle is too quick in stressing commonalities between humans and other species: "Humans share with many species of animals the capacity for collective intentionality" (2005: 6f.); "Animals hunting together, birds cooperating in building a nest, and presumably the so-called social insects, such as ants and bees, manifest collective intentionality" (1999: 121). Differences between humans and other species come only with regard to constitutive rules and status functions. But based on empirical research, we would draw a line already at collective intentionality, specifically, we would draw a line between coordinated social activities and collective intentionality.

By collective intentionality in these animal examples Searle means shared we-intentions of the form "we hunt together", "we build a nest together", etc., where each participant "is acting, and doing what he or she does, only as part of a collective action" (2005: 6). In contrast to reductively individualist accounts, Searle does not attempt to conceptually reduce we-intentions to complex individual intentionality of the form "I intend to X", "You intend to X", "I know that you intend to X and know that I intend to. . .". In fact, we-intentions according to Searle are both conceptually irreducible and biologically primitive.

But what are the criteria for ascribing such we-intentions, particularly in the case of prelinguistic and nonlinguistic creatures? More specifically: what counts as evidence for distinguishing such we-intentional collaboration from mere social coordination? Searle (1990, 1995) gives very illuminating analyses for the case of adult cooperators who can be asked about their individual and collective intentions, analyses that—in contrast to some individualist proposals—allow us to distinguish cooperation proper from complex coordinated social behavior plus mutual knowledge. Unfortunately, however, these analyses are unhelpful in the absence of language.

Bratman (1992) identified a trio of characteristic features of shared collaborative activities: (i) mutual responsiveness (to each other's behavior), (ii) commitment

to the joint activity (including coordinating and meshing sub-plans necessary for joint success), and (iii) commitment to mutual support of each other in the respective roles. Mere social coordination exhibits (i), but falls short of (ii) and (iii). Tomasello et al. (2005) recently adapted this analysis for interpreting preverbal collaboration, attempting to operationalize Bratman's criteria. First, if in collaboration the participants pursue a shared we-intention, then if one participant simply stops doing her part, the other(s) should be displeased with this and attempt to reengage her in the joint project. Second, if in pursuing their shared goal the participants grasp the roles that each participant plays (including themselves), then they should be able over repeated enactments of the joint activity to reciprocate and reverse roles as needed.

The question thus arises whether chimpanzees engage in these kinds of shared collaborative activities in the wild. Chimpanzees join one another in agonistic interactions within the group (so-called coalitions and alliances), and they act together to defend the group from predators and other chimpanzee groups. But in these interactions each individual does basically the same thing; they just do it at the same time and with responsiveness to one another's behavior. The most complex coordinated activity of chimpanzees is group hunting, in which two or more males seem to play different roles in corralling a monkey (Boesch and Boesch 1989). But in a reinterpretation of participant behavior over time in these hunts, some observers have characterized this activity as essentially identical to the group hunting of other social mammals such as lions and wolves (Cheney and Seyfarth 1990; Tomasello and Call 1997). Although it is a complex social activity, as it develops over time each individual simply assesses the state of the chase at each moment and decides what is best for it to do. There is nothing that would be called collaboration in the narrow sense of joint intentions based on coordinated plans. Rather, group hunting plausibly is an instance of complex social coordination only.

Natural observations of human children, in contrast, suggest that during the second year they begin to engage in collaborative games with complementary roles and turn-taking structure, and in collaborative instrumental activities with clearly differentiated roles (Brownell and Carriger 1990; Eckerman and Didow 1996; Ratner and Bruner 1978; Ross and Lollis 1987). Natural observations on their own, however, are hard to interpret and need to be complemented by experimental data.

In older experimental studies with apes (e.g., Crawford 1937; Chalmeau 1994), the most complex behavior that could be extracted was something like two chimpanzees pulling a heavy object in parallel, and during this activity almost no communication among partners was observed (Povinelli and O'Neill 2000). More recently, Warneken et al. (2006) directly compared children (18 and 24 months old) and human-raised chimpanzees on experimental tasks that required collaboration with an experimenter. Some of these tasks had an instrumental problem-solving structure (toys for the children/food for the chimps could be obtained through an apparatus that had to be operated together), others

were ends-in-themselves, collaborative games. During the collaborative acts, the experimenter was programmed to interrupt the shared activity at a predetermined point (by ceasing to act in his role) to see if the subjects would attempt to reengage his participation.

Children engaged proficiently in both kinds of collaborative tasks, and during the interruption periods they communicated appropriately with the experimenter (e.g., pointed to his side of the apparatus) and tried to reengage him in the collaborative activity. Chimpanzees, in contrast, engaged with the experimenter in a coordinated way only in the instrumental tasks where there was a food reward to be gained, but not in the noninstrumental game acts. Furthermore, even in the instrumental tasks, when the partner interrupted his engagement, they never communicated with him or tried to reengage him in the collaborative activity (they rather tried on their own or disengaged).

With specific regard to the grasp of the role structure of joint acts, Carpenter et al. (2005) recently placed 1.5-year-old children in the context of novel collaborative games with an experimenter, where the child first learned to play one role and the experimenter the other complementary one. These children demonstrated their ability for role-reversal imitation by spontaneously acting out the other role when appropriate. Human-raised chimpanzees, in contrast, though they had acquired some imitative abilities through enculturation, revealed little competence for such role-reversal imitation (Tomasello and Carpenter 2005).

These results suggest that even very young children pursue shared we-intentions with others. In the experiments they were committed to continuing the joint activity, as indicated by their reengagement attempts, and they understood the roles each participant played and how they were interconnected, as indicated by their pointing to the other's part and by reversing roles. Chimpanzees' behavior, in contrast, is more plausibly interpreted as social coordination with others without the pursuit of a shared we-intention. These findings thus suggest that humans develop a nascent ability to engage in collaboration as the basic form of collective intentionality in the second year of life, before language acquisition has begun in earnest, and this development seems to be a uniquely human achievement.

1.2. Communication

Human communication, needless to say in a volume on John Searle, is a cooperative enterprise. And obviously humans engage in a very specific form of cooperative communication, namely language, which no other animal does. But though language is special, it is not a primitive. Rather it has its ontogenetic roots in pre-linguistic social and communicative activities, as pragmatist theories of language acquisition have stressed (e.g., Bruner 1983; Tomasello 1999, 2003). These activities themselves, we would now like to argue, are characterized by nascent collective intentionality, and are uniquely human.

It is true that chimpanzees, like many species, engage in varied and complex communicative behavior. Particularly, they gesture to one another regularly.

While some of their gestures are relatively inflexible displays invariably elicited by particular environmental events, an important subset are learned by individuals and used flexibly—such things as 'arm-raise' to elicit play or 'touch-side' to request nursing. We know that such gestures are learned because in many cases only some individuals use them, and indeed several observers have noted the existence of idiosyncratic gestures used by only single individuals (Goodall 1986). And their flexible use has been repeatedly documented in the sense of a single gesture being used for multiple communicative ends, and the same communicative end being served by multiple gestures (Tomasello et al. 1985, 1989). Flexible use is also evident in the fact that apes only use their visually based gestures such as 'arm-raise' when the recipient is already visually oriented toward them—so-called audience effects (Tomasello et al. 1994, 1997; Kaminski et al. 2004).

Though these gestures are flexible, intentional acts, they are all tied to concrete action contexts and center around moving others to behave in certain ways. While human infants use these kinds of gestures as well, from around 1 year they engage in communication which goes beyond such proto-imperative acts: they show objects to others, with no obvious aim apart from looking at them together; they point out things to other people simply to share attention and interest with them[3]; and they point for ignorant others to inform them about states of affairs (proto-assertive or proto-informative, Liszkowski et al. 2006).

Apes, in contrast, do not show objects to conspecifics, nor do they point for others. They do not do these things even though they are capable of gesturing flexibly, as indicated above, and though they can track what other individuals do and do not see. Thus, they follow the gaze direction of conspecifics to relatively distal locations (Tomasello et al. 1998); they follow another's gaze direction around and behind barriers to locate specific targets (Tomasello et al. 1999); and they even seem to know the content of what another sees, as individuals act differently depending on whether a competitor does or does not see a potentially contestable food item (Hare et al. 2000, 2001). So the question arises: If chimpanzees have the ability to gesture flexibly and they also know something about what others do and do not see—and there are certainly occasions in their lives when making someone see something would be useful—why do they not sometimes attempt to direct another's attention to something the other does not see by means of a pointing gesture or something equivalent? Some might object that

[3] For example: Child and another person sit in a room and something interesting happens which both see, e.g., lightning outside. The child points to the event, looking back and forth between the other person and the event. Recent experimental evidence shows that 1-year-olds want the other person not only to look at the event or at them, but to attend to the referent and share their attitude about it—if the other does not do so, infants keep on pointing (Liszkowski et al. 2004). Misleadingly, this form of pointing was originally called 'proto-declarative' in a seminal work by Bates et al. (1975). Following Searle's taxonomy of illocutionary act types, it is not so clear what this kind of pointing is proto- for. It most closely resembles expressives in that emotions are expressed and shared, and there is no obvious direction of fit.

they do do this on occasion in some experimental settings, but this in fact only deepens the mystery. The observation is that captive chimpanzees will often "point" (whole arm with open hand) to food so that humans will give it to them (Leavens and Hopkins 1998) or also, in the case of human-raised apes, to currently inaccessible locations they want access to (Savage-Rumbaugh 1990). This means that apes can, in unnatural circumstances with members of the human species, learn to do something in some ways equivalent to pointing (in one of its functions). And yet there is not a single reliable observation, by any scientist anywhere, of one ape pointing for another.

The reason why chimpanzees do not point and human infants do, we think, is that chimpanzee communication lacks the nascent collective intentionality we find in infant communication. Infants' communication—both production and comprehension—is embedded in frames of joint attention and action ("we are looking at this together", "we are doing that together") which make meaning and interpretation possible. Consider this very simple scenario, often called the object choice task. Tomasello et al. (1997) had one person, called the "hider", hide food or a toy from the subject in one of three distinctive containers. Later, a second person, called the "helper", showed the subject where it was by tilting the appropriate container toward them, so that they could see the prize, just before their attempt to find it. After this warm-up period in which he defined his role, the helper began helping not by showing the food or toy but by giving signs, one of which was pointing (with gaze alternation between subject and bucket as an additional cue to his intentions). The apes as a group were very poor (at chance) in comprehending the meaning of the pointing gesture, even though they were attentive and motivated on virtually every trial. (Itakura, et al. 1999, used a trained chimpanzee conspecific to give a similar cue, but still found negative results.) Human 2-year-old children, in contrast, performed very skillfully in this so-called object choice task. Subsequent studies have shown that apes are also generally unable to use other kinds of communicative cues (see Call and Tomasello 2005, for a review), and that even prelinguistic human infants of 14 months of age can comprehend the meaning of the pointing gesture in this situation (Behne et al. 2005).

It is important to recall that apes are very good at following gaze direction in general (including of humans), and so their struggles in this kind of task do not emanate from an inability to follow the directionality of the pointing/gazing cue. Rather, it seems that they do not understand the meaning of this cue—they do not understand either that the human is directing their attention in this direction intentionally or why she is doing so. As evidence for this interpretation, Hare and Tomasello (2004) compared this pointing gesture to a similar but different cue. Specifically, in the key experimental condition they had the experimenter first establish a competitive relationship with the ape, and then subsequently reach unsuccessfully in the direction of the baited bucket (because the hole through which he reached would not enable her arm to go far enough). In this situation, with an extended arm that resembled in many ways a pointing gesture

(but with thwarted effort and without gaze alternation), apes suddenly became successful. One interpretation is that in this situation apes understood the human's simple individual intention to get into the bucket, and from this inferred the presence of food there (and other research has shown their strong skills for making inferences of this type; Call 2004).

But understanding individual instrumental intentions is not the same thing as understanding communicative intentions. Nor is following gaze the same thing as understanding communicative intentions. In simple behavior reading or gaze following, the individual just gathers information from another individual in whatever way it can—by observing behavior and other happenings in the immediate surroundings and making inferences from them. The object choice task, however, is a cooperative and communicative situation in which the subject must understand the joint project, and the experimenter's communicative intentions, in its context. That is, she must understand that the pointing behavior of the human is located in the course of "our project" (see Wittgenstein's construction workers); it is not just an individual instrumental act, but rather something done "for me" and so is relevant in some way for the search task I am facing. Take away the joint project, and there is no communicative meaning left. For example, when in a control condition of the Behne et al. (2005) study the experimenter retreated from the joint project, performed some individual act (e.g. inspected her wristwatch) and in this course distractedly extended her index finger towards one of the boxes as in the pointing condition, children did not interpret this as a communicative act (i.e., did not choose this box).

It is thus the joint attentional frame, or common communicative ground, emanating from the joint project which gives the pointing gesture its meaning in specific contexts (Clark 1996). Thus, if you encounter me on the street and I simply point to the side of a building, the appropriate response would be "Huh?" But if we both know together that you are searching for your new dentist's office, then the point is immediately meaningful. In the object choice task, human infants seem to establish with the experimenter a joint attentional frame that "what we are doing" is playing a game in which I search for the toy and you help me. Upon seeing the point, then, the infant thus asks herself how it is relevant to the joint game (Sperber and Wilson 1986), which leads her to take the point as a proto-assertion as to where the toy is located.

As sophisticated as chimpanzee communication is in its own right, it remains on the level of individualistic intentionality, lacking the structure of we-intentionality in the form of joint attentional frames that characterizes infant communication from the second year on.

1.3. Social Learning

All primates learn from observing others do things. But the human version is distinctively different. Whereas nonhuman primates mostly focus on changes in the environment created by behaving others—so-called emulation learning—humans

focus also on the behavioral means used to create those changes. For example, if a demonstrator opens a nut in some clever way, chimpanzees learn that the nut opens, whereas humans learn in addition the clever strategy used (see Tomasello 1996, for a review). This may be because humans imitate others not just to learn effective ways of dealing with the environment instrumentally, but also to identify with others, to be like them—the so-called social function of imitation (Carpenter 2006). Finally, as noted above, humans also engage in role-reversal imitation in which they do to the other what the other has done to them, or they understand and reproduce the other's role in a collaborative activity merely by observing. The outcome is that human imitative learning enables a much more powerful form of cultural transmission than the weaker, and less social, social learning skills of nonhuman primates, resulting in a kind of cultural ratchet leading to cumulative culture evolution (Tomasello et al. 1993).

With specific reference to shared intentionality, humans in addition employ their skills of imitation in various kinds of interactive contexts involving we-intentionality. Quite often infant imitation takes place in a collective we-context set up by adults and involving informal adult instruction: adults show children what to do with objects, how "we" act with them, supplying feedback and assistance. Tomasello et al. (1993) called this, from the child's point of view, instructed learning, as the adult directions/instructions are learned and internalized along with some instrumental change in the world. No other species on the planet has reliably been observed to do such things in flexible ways across contexts (Kruger and Tomasello 1996). But infants by no means are just passively placed into such collective frameworks and there only perform their individual imitative acts. Rather, from early on they actively participate in collectively intentional forms of imitation of the form "I do what we do" that transcend the purely individualistic "I do what you did".

One such form involves the imitation of conventional objects, particularly artifacts, with *causal usage functions* à la Searle (1995). Not only do infants learn by imitation how to handle objects in the pursuit of idiosyncratic ends, but through imitation they also learn how "we" use scissors, pencils, and chairs, what they are "for". Preschool children reveal rich knowledge about teleological ascription of functions to artifacts when talking about them (e.g., German and Johnson 2002; Kelemen 1999), but how do we know whether younger children really imitate what "we do" with an artifact rather than just "what you did"? One source of indirect evidence comes from "functional fixedness" (Duncker 1945), the phenomenon that we find it hard to use artifacts in ways deviant from their normal use. Functional fixedness can be interpreted as an index of understanding the conventional usage of artifacts, of the ways "we use them". And recent experimental studies have revealed that children from 2 years after imitating novel acts with artifacts show just this kind of fixedness. Another source of evidence comes from naturalistic observations: children in their second year begin to use teleological normative vocabulary in describing artifacts as "broken" (Kagan 1981).

The normative dimension and the collective intentionality of imitation can be seen even more clearly in the case of early imitation of actions that involve some simple *status function* assignment à la Searle (1995). Of course children from 1 year learn a language, a system of status functions. And 2-year-olds have been observed to engage in seemingly normative behavior in response to other's linguistic 'mistakes' by correcting false statements (Pea 1980, 1982). The case of language is difficult to interpret in this context, however, because it is not clear whether children—before they acquire metalinguistic awareness from age 4–5—have to see an utterance in a dual perspective as sound event X which counts as speech act Y in context C, or whether they rather just see through language, so to speak (a point we will return to below). Clearer cases are games involving objects which children know and which get additional status in the context of a game. Simple rule games fall under this heading. A particularly interesting case is pretend play. Children imitatively learn to engage in pretend play (Rakoczy et al. 2005a), and from their second year engage in simple joint pretense. For example, they do such things as pretend with others that a stone is an apple. In the context of this pretense, the stone counts as an 'apple'. In recent studies we found that children from 2 years not only imitatively learn such game actions and engage in joint games, but indicate an awareness of the normative dimensions of games with status: when others do not respect the status function of the objects in the relevant game context, they protest (Rakoczy et al. in press. Rakoczy, submitted). Furthermore, after this protest, children often go on to supplement the critique with some explanation as to how to do the act correctly. That is, these children not only participate in collective instructed learning as pupils, they themselves have begun their career as instructors.

So, human infants become imitative learners from 1 year of age, their imitative abilities revealing remarkable systematicity. In contrast, we do not have good evidence for comparable imitative capacities in great apes. Imitation itself might be a cradle of cooperation in infants, and its absence in apes partly explanatory for their lack of cooperation proper. From the second year on, human imitation not only applies to individual acts, but involves a dimension of collective intentionality, in the form of instructed learning, and even normativity: Children learn how "we" treat objects and assign them functions, and what normative implications this carries. They even—in simple ways—instruct others to do so. Great apes, in contrast, do not reveal any such collective intentionality in their social learning, either in the form of instructed learning or in the form of normativity.

1.4. Summary

Nascent collective intentionality, involving the corresponding normative aspects, develops in the second year of human ontogeny in the areas of collaboration, communication, and social learning. In contrast, closer empirical investigations of animal social life do not reveal any convincing evidence for

collective we-intentionality in nonhuman species. Searle is too promiscuous in granting collective intentionality widely to many animals. This overattribution, we think, rests on two mistakes: First, Searle's notion of collective intentionality overemphasizes its primitiveness and neglects social-cognitive requirements on the part of the individuals participating in cooperation. Many species, such as ants, do not fulfil these requirements—they do not understand each other as subjects—and so cannot be proper cooperators. Second, as just argued at length, even for apes, who do fulfil these requirements, stricter criteria for ascribing collective intentionality proper are needed.

As to the first point, in some sense we agree with Searle that we-intentionality is conceptually not reducible to simpler forms of individual intentionality.[4] We agree that for collective "we"-intentionality there are no necessary and sufficient conditions in terms of individual "I"-intentionality. However, in his argument against reductive individualistic accounts, Searle might throw out the baby with the bathwater, might make we-intentions too primitive in the sense that he does not sufficiently consider individual cognitive prerequisites that are necessary though not sufficient for collective intentionality. Specifically, we-intentionality plausibly builds on, but is not reducible to individual social cognitive abilities to interpret each other. Only if I have some grasp on your acting intentionally, can I enter into a shared we-intentional action with you (see Tomasello and Rakoczy 2003).

At one point Searle does consider individual prerequisites for collective intentionality quite along the lines we do here: "Collective intentionality presupposes a Background sense of the other as a candidate for cooperative agency; that is, it presupposes a sense of others as more than mere conscious agents, indeed as actual or potential members of a cooperative activity" (1990: 414). But then he goes on to stress that these prerequisites are not cognitive ones, but rather belong to the non-intentional skills of the Background. If that simply means that we do not abstractly believe that others are persons (as we do not abstractly believe that there are objects in the world) Searle's proposal sounds fine. However, isn't Searle's general "sense", at least partly, a disposition to have certain beliefs about specific others in specific situations? For example, that Peter here and now can (consciously) see X, but not Y, that Mary now would like to do Z, etc.—and that in collectively intentional activities with Peter or Mary their perceptual and intentional situation would have to be coordinated with one's own?

While these might be to some degree just terminological matters, there is a more substantial issue. Social insect behavior is an instance of collective intentionality, says Searle, and collective intentionality presupposes a sense of others

[4] When Searle talks about individual and collective intentionality, he mostly talks about intentional attitudes of an individual in both cases, but in the former case the attitudes have the form "I intend", "I desire", etc., whereas in the latter they have an irreducible "we"–form ("We intend to take a walk together".). To avoid confusion, we will follow this terminology here.

as intentional agents and potential cooperators. So social insects understand each other as intentional subjects. But what kind of independent evidence do we have for such a claim—independent, that is, from the social coordination?

None, we think. In fact, whereas human infants develop this kind of understanding from 1 year, almost no other species (apart from great apes to some degree) does. What are the indicators? As intentional action (on the world-to-mind fit side) and perception (on the mind-to-world fit side) are the biologically and ontogenetically most basic forms of intentionality (Searle 1983), the most basic cases of understanding one another as intentional subjects should be the cases of perception and of intentional action. Roughly, infants from 1 year of age begin to manifest a bundle of new social behaviors such as following and directing the attention of others (on the perceptual side) and imitating others' acts (on the side of intentional action) that are triadic in structure (there is a triangle child–other–object/situation in the world) and that are best interpreted as reflecting a nascent understanding of other and self as equivalent perceiving and acting intentional subjects (Carpenter et al. 1998b). While many species engage in dyadic interactions with objects on the one hand, and with conspecifics on the other hand, triadic phenomena are markedly absent from most species except humans and to some degree other primates.

So Searle claims that collective intentionality is widespread in the animal kingdom and, by implication, that many species have a sense of each other as intentional agents. In contrast, we think there is no good evidence for collective intentionality proper in nonhuman species. And this is not surprising for most non-primate species given that we do not have good independent evidence for the requisite social cognitive abilities in these species.

Turning to the second point, we should note that the case of the great apes (and perhaps other primates) is quite different in that they do reveal some basic abilities to interpret each other. In their case we do have independent evidence that they understand each other as perceiving subjects (Hare et al. 2000, 2001) and as intentional agents (Call et al. 2004). Both infants and chimps thus go some way together from being intentional agents to understanding one another as such in simple ways. But whereas human infants move further from there to sharing into collective we-intentionality, we do not have good evidence for such a move in the case of the apes when we look at the phenomena more closely and with stricter criteria (as argued in Section 1.1–1.3.). And here it is plausible to see this as rooted in the fact that only humans develop the requisite "biological primitive sense of the other as a candidate for *shared* intentionality" (Searle 1990: 415; our emphasis).

Avoiding these two mistakes of overattributing collective intentionality not only helps us to appreciate the uniquely human skills of creating social reality, but also helps to isolate the even more sophisticated skills required for understanding constitutive rules and status functions, which lead into full-blown human institutional reality. This seems to happen in human ontogeny one or more years after infants' initial skills of shared intentionality have emerged.

Understanding the transition to these more sophisticated skills would thus provide a fuller and more differentiated account of collective intentionality. It is to that task that we now turn.

2. THE TRANSITION TO INSTITUTIONAL REALITY

Collective intentionality constitutes social facts. Many species, according to Searle, live in a world of social facts. Beyond this, however, institutional reality is constituted by special forms of collective intentionality involving the assignment of status functions and constitutive rules. And institutional reality is the heart of uniquely human forms of life, says Searle:

> The institutional ontology of human civilization, the special way in which human institutional reality differs from the social structures and behavior of other animals, is a matter of status functions imposed according to constitutive rules and procedures. Status functions are the glue that hold human societies together. (2005: 9)

While Searle is surely right about institutional ontology being a uniquely human phenomenon, we think overattributing collective intentionality in its simpler forms to many animals prevents one plausible explanation of why no other species on the planet has money, marriages, or property: namely, because no other species has the requisite social ontology (i.e., collective intentionality) on which to build an institutional ontology.

"X counts as Y in context C" is the formula for assigning status functions through constitutive rules, and therefore the logical basis of institutional life. For example, "This piece of wood counts as a queen in the context of the game of chess". Now, many species are individual intentional beings, capable of perceiving and classifying objects, as pieces of wood for example. So clearly the "X" is not the crucial part for explaining why only humans are institutional beings. In a recent paper (Searle, in press) Searle explores the idea that what is crucial is the "X counts as Y" part. This, so Searle claims, requires the ability to think about one object on two levels at once, and in the paradigmatic cases this in turn requires the ability to symbolically represent the X as being a Y. And nonhuman animals lack these abilities. We agree that no other species counts an object as having a collectively defined status beyond its brute identity, and that this is an important part of the story. But it might be that great apes, and perhaps other species are capable of privately perceiving objects on different levels. For example, apes seem capable of using objects such as replicas as scale models for states of affairs in the world in the way human children do (Kuhlmeier and Boysen 2002). True, this probably only involves appreciating natural meaning (Grice 1957), and falls short of understanding collective assignment of anything. But it can be seen as a simple form of treating objects on different levels at once. Thinking about objects on two levels is surely necessary for participating in status assignment, but not sufficient. So what is missing in the apes?

One possibility is that the "C" term already carries a lot of baggage, that the context C presents the background which has to be in place before the "X counts as Y" can be grounded in the first place. The context in question already has to be one of collectively intentional practices, and apes do not have such practices, and therefore their rudimentary capacities (if there are such) to think about objects at different levels privately (involving natural meaning) cannot be made good for creating status (involving nonnatural meaning). Searle seems to miss this possibility due to his overattribution of collective intentionality to large parts of the animal kingdom.

Taking this possibility seriously, the following picture of the ontogeny from individual intentionality to collective intentionality to institutional reality in humans emerges. Children in their first year and chimps go a long way together: they share simple forms of individual intentionality, and some basic forms of understanding intentionality in others and themselves. From the second year of life, however, the ways depart: infants enter into collective we-intentionality in the form of collaboration, preverbal communication and social learning. Early collaboration is confined to instrumental and simple play actions, early communication to gestures (such as pointing), and early social learning to the imitation of nonsymbolic acts. In other words, there is collective intentionality, there are shared practices, in some cases (imitation of acts with artifacts) even involving the collective assignment of causal usage functions—but there is no institutional structure yet. Context C is in place but still awaits status creation. How does this develop? Searle's (1995) fairy-tale about the historical development of borders probably is a good analogy. Imagine two tribes building a stone wall to separate their territories. In the course of the collaborative project of building it, the wall is assigned the causal usage function of separating the territories in virtue of its causal properties. Now over the years the wall decays, and after a while only some stones remain which could not physically fulfill the function of separating the territories. However, in the course of time the practice of not crossing the border has become so entrenched that the stones, even while not functioning as a wall anymore, have come to count as a line not to be transgressed: they have acquired the status of a border. And, we should like to add, they have acquired this status against the background of the context C, which is constituted by the tribes' collective practice of trying to get along with each other etc.

This is a good analogy for ontogeny in one respect: what today is a collective practice for children, yet without status, tomorrow becomes an institutional affair. In another respect, however, ontogeny is obviously different from history: while the historical story is supposed to show how status can arise in a whole community, children grow into an already existing institutional reality, and the question is how they develop the ability to participate in it. As status arises out of the tribes' collectively intentional practice, their collaboration and communication, so children's participation arises out of their engagement in collective intentionality (though with a central role for scaffolding by adult members of the

community). In contrast, as apes do not engage in collective intentionality proper, there is no context C rich enough to ground the emergence of institutional life.

Institutional reality, as Searle and many others have stressed, is a holistic affair: most status is constituted in a network of other status (think of the interrelated matters of money, exchange, private property, etc.), most constitutive rules are not understandable without understanding many others. But how can children then ever enter into this net? Two entering gates, two cradles for becoming an institutional being, we would like to suggest, are language (obviously), and playing games, particularly games of make-believe.

Pre-linguistic communication, as we have argued above, already is a collectively intentional activity. Based on that activity, infants in their second year acquire language as a public, conventional system of constitutive rules and status functions. Language is thus in one sense the continuation of communication with other means. But on the other hand, the means are so special in that they involve conventional meaning, constitutive rules, and thereby normative dimensions that they constitute more than just a continuation.

So children in their second year acquire the ability to correctly perform simple speech acts, at first often in conjunction with deictic gestures (the gesture typically handling reference, plus a linguistic predication), subsequently more independent from gesturing (Tomasello 2003). But not only do they speak correctly, *act in accordance* with the rules themselves—they show an awareness of *following the rules*. First, as mentioned above, 2-year-olds actively correct others' false statements and false labeling (Pea 1980, 1982). Second, soon after children have mastered some basic word use, they begin to play with language, mislabeling objects or making false statements in an obviously intentional and amused way (Johnson and Mervis 1997)—which can be interpreted as an index of their awareness of the normative structure that is violated.

Language as an activity governed by constitutive rules and the corresponding status functions has a double role, as Searle has stressed: being a system of status functions itself, it is also the logical foundation for most other status functions. That some X counts as a Y usually requires some way of symbolizing it as being a Y, and the default way to do this is using language to declare it a Y (on pain of well-known rule regresses this obviously does not apply, however, to status assignment in language itself).

However, although children do acquire language in the second year of life, and indicate some awareness of its normative structure, it is an interesting question whether they understand the status structure of this activity: whether they understand the dual nature of utterances as sound events that on a second level have the status of speech acts (see Kalish 2005). Probably young children do not understand this dual structure, but rather just implicitly master the linguistic practice until they develop a metalinguistic awareness some years later at 4–5 (e.g., Doherty and Perner 1998). They can use language, but see through it, as we use our eyes to see without seeing them (see Wittgenstein 1961, § 5.633). Language is thus the first status involving activity in which children learn to

participate. But their participation is such that they do not have to appreciate the dual structure of status function assignment, do not have to see something as an X while counting it as a Y.

This is different in the case of games, particularly games of pretend play with objects. Games are among the standard textbook examples of institutional activities ("This piece of wood counts as a queen in chess"). Games of pretend play are a special and developmentally very interesting case: there is a similar dual structure of status assignment, such that, e.g., a wooden block counts as something else in the course of the game. But interestingly, this something else, the "Y", is not, as in the case of rule games, only understandable through the game. Imagine an adult and a child 'cutting', 'peeling', and 'eating' a wooden block. "This wooden block counts as an 'apple' in our pretense game" is the form of the status function assignment (which is practically established in the standard case by saying things like "This is our apple now, okay?", see Walton 1990). This status assignment brings with it the usual normative structure of institutional practices: in the context of the game, the wooden block, counting as an 'apple', has to be treated appropriately, can be peeled, eaten, found disgusting, etc., but cannot be drunk or driven.

If pretend play was confined to simple pretense acts like pretending to sleep or pretending to fly, it might be objected that the same concerns as in the case of language would apply: that children do not necessarily have to see an act of pretending as X (flapping one's arms) which additionally counts as Y ('flying'). Perhaps this concern would hold, but in fact early pretense does involve treating objects ("props" as Walton 1990 calls them) as if they were something else, and the concern does not hold for these cases. Particularly, when there are familiar objects used in joint pretending, having a pretense status of being something else, the child has to, in some sense, be aware of the dual structure—otherwise we would not talk of pretense but of delusions (the child losing a tooth in trying to take a bite out of the wooden block does not count as pretending). In contrast to the case of language, the child thus has to see the object as an X that in the context of the game counts as a 'Y'.

Due to this logical structure of pretend play, we think games of pretense are one plausible cradle for the child to enter more broadly and actively into the world of status functions and institutional facts, more actively than in the case of language which does not require the dual view on Xs as Ys (Rakoczy 2006, 2007).

Children begin to engage in pretend play from 1.5 years, imitatively learning it from others and mainly practicing it in joint games of pretending (Rakoczy et al. 2005a,b). By 2 years, their pretending with others has become systematic: they respect the stipulations of the pretense scenario ("this wooden block is our apple, and this pen is our knife")—the status function assignment, that is—and produce normatively appropriate inferential acts. For example, they pretend to cut the wooden block with the pen, handle the pen 'carefully' because it is 'sharp', etc. (Harris and Kavanaugh 1993; Rakoczy et al. 2004; Rakoczy and Tomasello, 2006). And not only do they act appropriately themselves, but also

indicate an awareness of the normative structure of such practice more actively: when a third person joins the game, but makes a "mistake", i.e., does not respect the pretense status of an object, they protest and criticize her (Rakoczy, submitted). So by the third year of life, children have entered into the basics of this remarkable practice of games of pretending, collectively treat objects they know to be Xs as Ys, follow and respect the implications of the proto-constitutive rules of the game, and normatively criticize deviations from the rules.

In embryonic and isolated form we thus have here the basic structure of institutional reality in the games of 2-year-olds. Of course this is a long way from money, marriage, and universities, but the seeds are there, and so joint pretending quite plausibly can be considered the central cradle for, and the entering gate into institutional life. This proposal, we think, gains plausibility from the following lines of reasoning: First, all status function assignment is symbolic, or linguistic in a broad sense: "The sense in which symbolization in [a] broad linguistic sense is essential to all institutional facts is that the move from X to Y in the formula 'X counts as Y in C' is already a symbolic move" (Searle 1999: 155). This, we have already seen, is why language is in some sense the basis for all institutional reality. And, true, in the case of pretending, the standard way to set the scene is to declare things like "This is an apple now". But in many cases one can do without such linguistic (in the narrow sense) declarations. Rather, the introductory acts of the pretense can set the scene (think of putting on a Batman mask to make clear who is going to be in the game; or in our previous example, one could pretend to take the block from a tree, pretend to peel it, etc.). They are then symbolizing or linguistic acts in the broad sense Searle alludes to. The pretense acts make the required symbolizing move and the object itself is in some sense the symbol. This fits nicely with a remark by Searle: "In the limiting case, we can use the object itself to represent the Y status function" (1999: 155). Pretend play, we think, is just such a limiting case, and probably the ontogenetically primary one.

Second, let us come back to the question how, given that institutional reality is holistically structured, one can ever enter into this net. This, of course, is reminiscent of the question how one can ever learn a language. And the related question how one can ever come to interpret others as intentional beings. Regarding the latter question, we have already seen that Searle's claim that perception on the one side (mind-to-world fit) and intentional action on the other (world-to-mind fit) are the most basic forms of intentionality is shared by young children: they begin, at 1 year, to interpret others and themselves as perceiving and intentional subjects first, and then work their way inwards into the intentional net, so to speak, until they acquire a concept of belief, for example, at age 4 (e.g., Perner 1991). Now, in the case of institutional reality, are there "outer" areas of the net which can supply an entry analogous to observation sentences (or something close in case there are no such things) in the case of language, and to intentional action and perception in the case of understanding intentionality? One interesting possibility here is that some institutional activities are less bound up

in the net than others, and therefore more accessible. Games come to mind as a plausible candidate. In fact, Searle at one point explores such a possibility but then, we think, dismisses it too quickly. After having noted the principle of the holism of institutional reality he says:

> It might seem that games are a counterexample to this general principle because, of course, games are designed precisely with the idea that they should be forms of activity that do not connect with the rest of our lives in a way that social facts generally do. Today's baseball game need have no consequences for tomorrow in the way that today's wars, revolutions, buyings, and sellings are intended precisely to have consequences for tomorrow and into the indefinite future. Nonetheless, even in the case of games, there are systematic dependencies on other forms of social facts. (1991: 343)

True, there are still dependencies. But we think, generally not as intricately as in other areas of institutional life. For ontogenetic purposes, the interesting limiting cases are now those where there is little dependency, or at least those which can be practiced by a child without having to be aware of much dependency, and which therefore supply the possibility to enter into the institutional net. And pretend play, it seems, is just such a case. One aspect of the relative independence of pretense games from other institutional affairs is how the context C is specified. Usually, joint pretense is created on the spot in dyads or bigger groups playing together. That is, the context is paradigmatically defined as "our game here and now". Thus, both regarding the persons involved and temporally, this context is much more restricted, and therefore probably much more manageable as an entry than contexts which refer to whole societies or quantify over types of events, etc.

In sum, joint games of pretending are one important cradle for children's entering into institutional reality, indeed, the key cradle beside language which is of course in some sense the fundamental one. But it is one step closer to the bed of institutional reality than language in that pretending (in contrast to early discourse) requires children to take the double-view of status assignment on objects. And once they have mastered this double-view at the very outer edge of institutional reality, in "nonserious", short-lived, relatively isolated and personally restricted contexts, they can begin to work their way inward the institutional network towards ever more general, serious, and temporarily extended practices.

One possible way into the institutional net is via the development of performative speech acts. These speech acts, when uttered in the right context (such as "Your name is Peter" at a baptism) create a state of affairs just by being uttered (Searle 1969). And many of these states of affairs are institutional facts: "Your name is Peter"; "You are married"; "I hereby declare you guilty", etc. And many institutional facts would be unthinkable without the corresponding performative speech acts (Searle 1995). Now, as we have seen earlier, the typical way to set the scene in joint games of make-believe is to say things like "This is our apple now, and I'm now Adam and you're Eve". The striking structural analogy between such declarations and performative speech acts is by no means a coincidence, we think. Rather, such stage-setting in joint pretending should be seen as proto-performative speech acts. And these lay the basis for participation in the performative (in the

more narrow sense) making of institutional facts in a wider range of domains developing subsequently. What is "I'm Adam now, you're Eve" today in a game, tomorrow becomes "I hereby declare you husband and wife" in earnest.

Pretend and other games are probably the first areas where children participate in status assignment, involving a dual view on objects. But of course this is still some way from mature institutional life. Empirically, research on children's understanding of more "serious" status in such areas as money, private property or the nature of naming has typically found that children do not show much understanding in these areas before the school years (for an overview, see Kalish 2005). In a recent study by Kalish et al. (2000), for example, children's status understanding in pretense ("This is now our apple") was directly compared to their understanding of status in serious collective decisions about naming ("This horse is now called Peter") and property issues ("This is now yours"). While 3-year-old children understood "nonserious" status in the pretense context, only much older children (at 7 years) showed competence regarding serious collective decisions.

Such findings suggest that early participation in status assignment in pretense still lacks important building blocks of institutional life. In particular, though young children see a wooden block as such and as counting as 'apple' in a pretend game in some sense, they probably do not yet have much explicit understanding of the logical structure of this: of the fact that "this is a wooden block" is a brute fact, observer-independently true, whereas "This is an 'apple'" is an observer-dependent fact, true in the context of the game because people act according to it in some way. Grasping these logical structures requires mastery of basic epistemological notions. In particular, what is required is a grasp of the logical peculiarities of much status creation: such acts of creation—typically through performatives—have both a mind-to-world direction of fit (they aim at truth), and a world-to-mind direction of fit (they change the world).

The basic epistemological prerequisites are gradually acquired by children from the preschool years, with a major transition at age 4 when children master the basics of the concept of belief (for a review, see Wellman et al. 2001). Speaking a language and participating in specific forms of perspective-shifting discourse, and in discourse with propositional attitude constructions, has been found to be crucial for the development of the concept of belief and related epistemological notions (e.g., Lohmann and Tomasello 2003; see Astington and Baird 2005 for an overview). So language, itself of course a system of status functions, in a dialectical way, helps lay the foundations—via the development of epistemological notions—for developing more sophisticated forms of status assignment and institutional reality.

In sum, the general picture that emerges is this: Language and playing games, particularly pretense games, are the two central cradles for children's growing into institutional life which emerge in the second year of life, based on children's participation in simpler we-intentionality. Language is primary in the sense that most other status depends on language. But young children when they acquire language probably do not yet view speech acts in a dual way as sound that carries

meaning, but rather they see through language. In games of pretense, in contrast, young children not only participate in status assigning practices, they have to implicitly see the normative status structure of the enterprise, have to take a dual perspective on the objects involved. Being an activity rather isolated from the rest of societal life, being a concrete activity where the context C in which status holds is here and now, and involving proto-performatives ("This is our 'apple' now") structurally analogous to real, status-creating performatives, playing pretense games is an important cradle and foundation for children's entry into institutional reality more generally. Developing language and discourse skills enables the acquisition of central epistemological notions, and armed with such notions children can move on from their nascent skills in the area of games to a more general and explicit understanding of status and institutional reality.

3. CONCLUSION

Searle's program over the last decades is a heroic attempt to supply a general ontological framework for describing everything. Particularly his work in the ontology of intentional, social, institutional, and linguistic reality is philosophy at its best: conceptually enlightening and empirically useful for researchers in the cognitive and social sciences.

While applauding Searle's general approach of nonreductive naturalism, we have focused here on some shortcomings in his theory of collective intentionality.[5] Searle starts from the insight that reductive individualist accounts are fatally flawed in their attempt to find necessary and sufficient conditions for we-intentionality in terms of I-intentionality. But Searle, we think, overreacts by declaring we-intentionality too primitive. In particular, the necessary but not in themselves sufficient preconditions in terms of individual social cognitive abilities are neglected. On Searle's account we-intentionality is so primitive that reasonable criteria and restrictions for its ascription are missing, promiscuous attribution being the consequence. True, much socially coordinated behavior of many species prima facie looks like human cooperation. But on a closer look and with more stringent criteria it turns out we do not have good evidence to consider this behavior more than just coordination. And it turns out that for most species we do not have good evidence that they might even come close to fulfilling the individual social cognitive requirements for being cooperators in the first place.

[5] As empirical researchers we have not touched on the more purely philosophical controversies (though these also have implications for describing the relevant empirical phenomena, of course). For example, controversies regarding Searle's radical internalism (brains in vats can have we-intentions), the problem of "free-floating Y terms" (Smith 2003), the question how sharp the distinctions are between regulative and constitutive rules (how many rules of chess can you change without making it a different game?), and between usage and status functions (chairs could be—causally—used for many purposes, so isn't it like a status that they are for sitting?).

By not appreciating the complexities of collective intentionality proper (as empirical research documents it in early human ontogeny), by promiscuously granting it too widely, Searle neglects important and fundamental differences between the social ontology of humans and the social worlds of other animals. Consequently, the emergence of human institutional reality on top of social reality and the question why this is uniquely human is left mysterious: If apes have collective intentionality and social ontology (as Searle thinks), and if they can treat objects on two levels at once (as recent research might suggest), why don't they get their institutions going? In contrast to Searle, we think there is no collective intentionality proper in other animals, in most species because they do not even have the requisite social cognitive abilities, in apes because they do not make it from "I" and "You" to "We", from a simple understanding of each other to sharing with each other. And so there is no foundation for institutional life in the nonhuman animal kingdom. Human uniqueness begins with collective intentionality and social ontology, not with institutional reality.

True, in the current account we still have plenty of mysteries. But we hope we have located them better in the phenomena of the empirical world. Future empirical research into these phenomena would profit from philosophy supplying more refined analyses of collective intentionality, analyses filling the right middle ground between reductive individualism and just-so-primitivism, analyses that also make room for describing different forms of collective intentionality growing in complexity from child pretense to adult theatre.

Acknowledgements. This work was supported by a "Dilthey Fellowship" of the Volkswagen Foundation and the Fritz Thyssen Foundation to Hannes Rakoczy.

REFERENCES

Astington, J.W. and Baird, J.A. (2005). Introduction: why language matters. In J.W. Astington and J.A. Baird (eds.), *Why Language Matters for Theory of Mind*. Oxford: Oxford University Press.

Bates, E., Camaioni, L., and Volterra, V. (1975). Acquisition of performatives prior to speech. *Merrill-Palmer Quarterly,* 21(3), 205–226.

Behne, T., Carpenter, M., and Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science,* 8(6): 492–499.

Boesch, C. and Boesch, H. (1989). Hunting behavior of wild chimpanzees in the Tai-National-Park. *American Journal of Physical Anthropology,* 78(4): 547–573.

Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review,* 101(2): 327–341.

Brownell, C. and Carriger, M.S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Development,* 61: 1164–1174.

Bruner, J. (1983). *Child's Talk*. Oxford: Oxford University Press.

Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla,* and *Pongo pygmaeus*). *Journal of Comparative Psychology*, 118: 232–241.

Call, J., Hare, B., Carpenter, M., and Tomasello, M. (2004). 'Unwilling' versus 'unable': chimpanzees' understanding of human intentional action. *Developmental Science,* 7(4): 488–498.

Call, J. and Tomasello, M. (2005). What do chimpanzees know about seeing revisited: An explanation of the third kind. In N. Eilan, C. Hoerl, T. McCormack, and J. Roessler (eds.), *Joint Attention: Communication and Other Minds*. Oxford: Clarendon Press.

Carpenter, M. (2006). Instrumental, social, and shared goals and intentions in imitation. In S. J. Rogers and J. Williams (eds.), *Imitation and the Development of the Social Mind: Lessons from Typical Development and Autism*. New York: Guilford.

Carpenter, M., Nagell, K., and Tomasello, M. (1998b). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4): 176.

Carpenter, M., Tomasello, M., and Striano, T. (2005). Role reversal imitation and language in typically developing infants and children with autism. *Infancy*, 8(3): 253–278.

Chalmeau, R. (1994). Do chimpanzees cooperate in a learning task? *Primates*, 35: 385–392.

Cheney, D.L. and Seyfarth, R.M. (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago, IL: University of Chicago Press.

Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Crawford, M.P. (1937). The cooperative solving of problems by young chimpanzees. *Comparative Psychology Monographs*, 14: 1–88.

Doherty, M. and Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, 13: 279–305.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(5).

Eckerman, C.O. and Didow, S.M. (1996). Nonverbal imitation and toddlers' mastery of verbal means of achieving coordinated action. *Developmental Psychology*, 32: 141–152.

German, T.P. and Johnson, S.C. (2002). Function and the origins of the design stance. *Journal of Cognition & Development,* 3(3): 279–300.

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, MA: Harvard University Press.

Grice, H.P. (1957). Meaning. *The Philosophical Review*, 64: 377–388.

Hare, B., Call, J., Agnetta, B., and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59(4): 771–785.

Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61(1): 139–151.

Hare, B. and Tomasello, M. (2004). Chimpanzees are more skillful in competitive than cooperative cognitive tasks. *Animal Behaviour*, 68(3): 571–581.

Harris, P. L. and Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, 58(1)[231]: v–92.

Hurley, S. (2001). Overintellectualizing the mind. *Philosophy and Phenomenological Research*, 63(2): 423–431.

Itakura, S., Agnetta, B., Hare, B., and Tomasello, M. (1999). Chimpanzee use of human and conspecific social cues to locate hidden food. *Developmental Science*, 2(4): 448–456.

Johnson, K.E. and Mervis, C.B. (1997). First steps in the emergence of verbal humor: a case study. *Infant Behavior & Development,* 20(2): 187–196.

Kagan, J. (1981). *The Second Year: The Emergence of Self-Awareness*. Cambridge, MA: Harvard University Press.

Kalish, C. (2005). Becoming status conscious: children's appreciation of social reality. *Philosophical Explorations*, 8(3): 245–263.

Kalish, C., Weissman, M., and Bernstein, D. (2000). Taking decisions seriously: young children's understanding of conventional truth. *Child Development*, 71(5): 1289–1308.

Kaminski, J., Call, J., and Tomasello, M. (2004). Body orientation and face orientation: two factors controlling apes' begging behavior from humans. *Animal Cognition*, 7(4): 216–223.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition,* 70: 241–272.

Kim, J. (1995). Mental causation in Searle's 'biological naturalism'. *Philosophy and Phenomenological Research,* 55: 189–194.

Köhler, W. (1925). *The Mentality of Apes*. New York: Harcourt Brace.

Kruger, A.C. and Tomasello, M. (1996). Cultural learning and learning culture. In D.R. Olson and N. Torrance (eds.), *The Handbook of Education and Human Development. New Models of Learning, Teaching and Schooling*. Oxford: Blackwell, pp. 369–387.

Kuhlmeier, V.A. and Boysen, S.T. (2002). Chimpanzees (*Pan troglodytes*) recognize spatial and object correspondences between a scale model and its referent. *Psychological Science,* 13(1): 60–63.

Leavens, D. and Hopkins, J. (1998). Intentional communication by chimpanzees: a cross-sectional study of the use of referential gestures. *Developmental Psychology,* 34(5): 813–822.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., and Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science,* 7(3): 297–307.

Liszkowski, U., Carpenter, M., Striano, T., and Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development,* 7(2): 173–187.

Lohmann, H. and Tomasello, M. (2003). The role of language in the development of false belief understanding: a training study. *Child Development,* 74: 1130–1144.

Pea, R.D. (1980). The development of negation in early child language. In D.R. Olson (ed.), *The Social Foundation of Language and Insight: Essays in Honor of Jerome Bruner*. New York: W.W. Norton.

Pea, R.D. (1982). Origins of verbal logic: spontaneous denials by two- and three-year-olds. *Journal of Child Language,* 9(3): 597–626.

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Piaget, J. (1936). *The Origins of Intelligence in Children.* New York: W.W. Norton.

Povinelli, D. J. and O'Neill, D. K. (2000). Do chimpanzees use their gestures to instruct each other? In S. Baron-Cohen, F. Tager, H. Cohen, and D.J. Cohen (eds.), *Understanding Other Minds*, 2nd edn. Oxford: Oxford University Press.

Rakoczy, H. (2006). Pretend play and the development of collective intentionality. *Cognitive Systems Research,* 7: 113–127.

Rakoczy, H. (2007). Play, games, and the development of collective intentionality. In C. Kalish & M. Sabbagh (eds.), *Conventionality in cognitive development: How children acquire representations in language, thought and action. New Directions in Child and Adolescent Development. No. 115* (pp. 53–67). San Francisco: Jossey-Bass.

Rakoczy, H. (submitted). Young children understand the normative structure of joint pretend games.

Rakoczy, H. and Tomasello, M. (2006). Two-year-olds grasp the intentional structure of pretense acts. *Developmental Science*, 9(6): 558–565.

Rakoczy, H., Tomasello, M., and Striano, T. (2004). Young children know that trying is not pretending—a test of the "behaving-as-if" contrual of children's early concept of "pretense". *Developmental Psychology,* 40(3): 388–399.

Rakoczy, H., Tomasello, M., and Striano, T. (2005a). On tools and toys: How children learn to act on and pretend with 'virigin' objects. *Developmental Science,* 8(1): 57–73.

Rakoczy, H., Tomasello, M., and Striano, T. (2005b). How children turn objects into symbols: A cultural learning account. In L. Namy (ed.), *Symbol Use and Symbol Representation*. New York: Erlbaum.

Rakoczy, H., Warneken, F., and Tomasello, M. (in press). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology.*

Ratner, N. and Bruner, J. (1978). Games, social exchange and the acquisition of language. *Journal of Child Language,* 5: 391–401.

Ross, H.S. and Lollis, S.P. (1987). Communication within infant social games. *Developmental Psychology,* 23(2): 241–248.

Savage-Rumbaugh, S. (1990). Language as a cause-effect communication system. *Philosophical Psychology,* 3: 55–76.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, MA: Cambridge University Press.
Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
Searle, J.R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, and M. Pollack (eds.), *Intentions in Communication*. Cambridge, MA: MIT Press.
Searle, J.R. (1991). Intentionalistic explanation in the social sciences. *Philosophy of Social Sciences,* 21(3): 332–344.
Searle, J.R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Searle, J. (1994). Animal minds. *Midwest Studies in Philosophy,* 14: 206–219.
Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.
Searle, J.R. (1999). *Mind, Language, and Society. Philosophy in the Real World*. London: Weidenfeld & Nicolson.
Searle, J.R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.
Searle, J.R. (2005). What is an institution. *Journal of Institutional Economics,* 1(1): 1–22.
Searle, J.R. (in press). What is language: some preliminary remarks. In S. L. Tsohatzidis (ed.), *John Searle's Philosophy of Language: Force, Meaning, and Mind*. Cambridge: Cambridge University Press.
Smith, B. (2003). The ontology of social reality. *American Journal of Economics and Sociology,* 62(1): 285–299.
Sperber, D. and Wilson, D. (1986). *Relevance*. Cambride, MA: Harvard University Press.
Tomasello, M. (1996). *Do apes ape?* In C. M. Heyes and B. G. Galef (eds.), *Social Learning and the Roots of Culture*. New York: Academic Press.
Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, M. and Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.
Tomasello, M., Gust, D., and Frost, G.T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates*, 30(1): 35–50.
Tomasello, M., Call, J., and Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development,* 68(6): 1067–1080.
Tomasello, M., Call, J., and Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour,* 55(4): 1063–1069.
Tomasello, M. and Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development,* 70(1): 1–136.
Tomasello, M. and Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind and Language,* 18(2): 121–147.
Tomasello, M., Kruger, A.C., and Ratner, H.H. (1993). Cultural learning. *Behavioral and Brain Sciences,* 16: 495–552.
Tomasello, M., Call, J., Nagell, K., Olguin, R., et al. (1994). The learning and use of gestural signals by young chimpanzees: a trans-generational study. *Primates,* 35(2): 137–154.
Tomasello, M., Hare, B., and Agnetta, B. (1999). Chimpanzees, *Pan troglodytes,* follow gaze direction geometrically. *Animal Behaviour,* 58(4): 769–777.
Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences,* 28: 675–691.
Tomasello, M., George, B., Kruger, A. C., Farrar, M. J., and Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution,* 14: 175–186.
Walton, K. L. (1990). *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.

Warneken, F., Chen, F., and Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development,* 77(3): 640–663.

Wellman, H.M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development,* 72(3): 655–684.

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul.

SOCIAL REALITY AND INSTITUTIONAL FACTS: SOCIALITY WITHIN AND WITHOUT INTENTIONALITY

Robert A. Wilson

1. INTRODUCTION

John Searle's *The Construction of Social Reality* (1995) offers an account of the nature of social reality that complements and builds on the views of language and mind that Searle has developed in his earlier books (Searle 1969, 1983, 1992). It shares with those books a combination of a high level of both philosophical rigor and accessibility, and takes the reader down a persuasive path from the basic questions "What is social reality?" and "What are institutional facts?" to Searle's detailed answer to these questions.

My twofold aim in this paper will be to provide reasons for questioning Searle's answer, and to sketch an alternative way of thinking about the relationships between intentionality and "social facts" or "social reality"—both expressions that Searle uses freely, and what I would prefer to call, in parallel with intentionality, *sociality*. As the title of Searle's book on social reality suggests, his aim is to provide an account of sociality that shows how sociality can be both a construction and a part of reality, how there can be objective facts that we nonetheless play a role in constructing. Institutional facts, which are a focus of his work here and more recently (e.g., Searle 2003), are paradigms of such facts.

I want to suggest that that focus, and perhaps Searle's broader concern to address social constructivism and attacks on realism in epistemology that frames his discussion in *Construction*, results in a view of sociality that is misleading in several important ways, including in how we should view certain forms of nonhuman cognition and in how we should think about the relationship between intentionality and sociality. My argument will turn on the innocuous-sounding point that the two questions listed above—about social reality and about institutional facts—require importantly different answers, and that by focusing primarily on the latter question, the one about institutional facts, Searle presents a skewed answer to the former question, the one about social reality. Let me begin with a brief sketch of Searle's view of institutional facts.

2. SEARLE'S ACCOUNT OF INSTITUTIONAL FACTS IN *CONSTRUCTION*

Searle begins with the contrast between what G.E.M Anscombe (1958) has called "brute facts", facts that are, in Searle's words, "totally independent of any human opinions", and institutional facts, "so called because they require human

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 139–153.
© 2007 Springer.

institutions for their existence" (Searle 1995, p. 2; in what follows, all references by page numbers alone are to Searle 1995). Searle aims to understand how facts of this latter kind are possible by providing an account of their intrinsic nature and by explaining their relationship to other kinds of facts and, ultimately, to brute facts. Searle holds, surely correctly, that these two enterprises concerning institutional facts—the "what" and the "where"—are intimately related. Answering the Kantian question about social or institutional facts for Searle is constrained by the following picture of our ontological situation:

> We live in a world made up entirely of physical particles in fields of force. Some of these are organized into systems. Some of these systems are living systems and some of these living systems have evolved consciousness. With consciousness comes intentionality, the capacity of the organism to represent objects and states of affairs in the world to itself. Now the question is, how can we account for the existence of social facts within that ontology? (p. 7)

While the details of Searle's answer are (as details tend to be) complicated, its basic conceptual contours are not.

According to Searle, institutional facts involve (indeed, presuppose, as the quotation above suggests) intentionality. In particular, they involve not simply the intentionality of individuals but a special form that this intentionality takes, *collective intentionality*. In recognizing some forms of intentionality as collective, Searle is not positing any form of "group mind", or a kind of intentionality that can be found outside of the heads of individuals (cf. Pettit 2003, Wilson 2004: Chaps. 11–12, 2005b). But neither does Searle take his view to be reductive in that it eliminates collective intentionality in favor of individual intentionality. Searle takes himself to walk a middle path here insofar as he implies that certain intentional states that individuals have take a *first person plural* form: they are not simply of the form "I intend that p" or "I believe that p" but "*We* intend that p" or "*We* believe that p".

Such collective intentionality is critical to understanding institutional facts, claims Searle, since it plays a crucial role in assigning what he calls *status functions* to particular brute facts. These are functions assigned to brute facts that those facts have and can perform *only* because we have collectively assigned them that function. To use the example that Searle provides in introducing the idea of a status function (pp. 39–41), consider a wall that begins by marking a physical boundary but that comes, over time, to be physically eroded down to some base stones and to serve a symbolic function: marking the boundary of a certain kind of territory. While the wall may have the function of keeping out intruders, as Searle says, "in virtue of sheer physics" (p. 39), once it decays into a line of stones it can maintain this function only by some kind of collective imposition of function. Here we come to recognize mere stones arranged in a certain way as having a particular status vis-à-vis how we might act, and it is our collective recognition of that status function that is the *sine qua non* for those stones so arranged to have that function. Should our collective intentionality shift over time, such that the stones come to be viewed as mere stones again, the

institutional facts about their status as boundary markers will likewise disappear, becoming the basis for historical facts about a small part of previous human culture.

One way in which Searle sometimes expresses this point about the collective ascription of status functions is by saying that we use *constitutive rules* of the form "X counts as Y" in order to assign a status function to some piece of the physical world and so create an institutional fact. Constitutive rules are the rules that do not simply regulate a preexisting activity but create the possibility of those activities. Institutional facts and social reality are constructed through our adoption of many constitutive rules that assign status functions to parts of the physical world.

The assignment of status functions provides the key to understanding Searle's account of institutional facts and social reality, for it explains how certain kinds of facts are generated through our collective actions, and so how those facts are ontologically related to facts whose existence we can take for granted. By two or more individuals sharing first person plural intentional states, whether they are intentions, beliefs, or desires, that assign status functions, those individuals are able to create social or institutional facts. Such facts form clusters or networks, have a normative dimension to them (either in terms of what they require or in terms of what they allow), and can come to guide the actions of both those sharing the first person plural intentional states that generate them as well as those who simply grasp the social reality that such facts constitute.

After showing how this idea applies to the institution of money and going on, in Chapter 3, to articulate the key role that language plays in creating and sustaining institutional facts, Searle continues in Chapter 4, with a generalization of his basic analysis to institutional facts more generally, listing as amongst the social phenomena to which the account applies "marriage, property, hiring, firing, war, revolutions, cocktail parties, governments, meetings, unions, parliaments, corporations, laws, restaurants, vacations, lawyers, professors, doctors, medieval knights, and taxes" (p. 79). Further interesting details follow in later chapters, but as I have said it is the basic picture of social reality and institutional facts that I'm chiefly interested in here. In that picture, as Searle says (in typically evocative language):

> The central span on the bridge from physics to society is collective intentionality, and the decisive movement on that bridge in the creation of social reality is the collective intentional imposition of function on entities that cannot perform those functions without that imposition (p. 41).

With this central span in mind, let us now consider the two questions with which we began—"What is social reality?" and "What are institutional facts?"

3. COLLECTIVE INTENTIONALITY AND INSTITUTIONAL FACTS

Searle is no doubt correct in highlighting the central role that intentionality plays in the creation and maintenance of many kinds of (if not all) institutional facts, including the full range of institutional contexts indicated by the list of

examples above. And it is not simply intentionality but a fairly sophisticated form of intentionality—both higher-order and shared intentionality—that is necessary for there to be social facts of the kind on which Searle focuses, institutional facts. That we are the only species on the planet which has created institutional facts through such forms of intentionality is striking, and Searle has provided a framework for understanding many aspects of distinctively human cultural institutions and culture more generally. But there are reasons to question whether there is enough in Searle's account to adequately answer the question of what institutional facts are and how they are possible, and whether what he does say here answers the broader question of what social reality is.

The first reason concerns the issue of our species' uniqueness in the world as creators of institutional facts. Although our uniqueness in the world is not in doubt, our uniqueness as bearers of the kind of collective intentionality that Searle rightly holds to be necessary for such facts is more questionable. While Searle does not come right out and deny that nonhuman animals have the capacity for the appropriate form of collective intentionality, one that assigns status functions to brute facts, the comments that he does make regarding nonhuman intentionality are skeptical and deflationary about their abilities (e.g., pp. 38–40, 63, 70–71). The question, in a nutshell, is whether nonhuman animals can assign status functions to brute facts. If not, why not? If so, why has this not resulted in the generation of institutional facts?

This is clearly, in large part, an empirical issue. But in determining how to assess the empirical evidence here, we need to be clear about just what form collective intentionality must take in order to be of the sort necessary for the creation of institutional facts. Suppose that we follow Searle in thinking of collective intentionality as first person plural intentionality, "we" (rather than "me") intentionality. The feature of such intentionality that makes for a kind of mental Rubicon between human and nonhuman intentionality, claims Searle, is the ability to

> impose functions on phenomena where the function cannot be achieved solely in virtue of physics and chemistry but requires continued . . . cooperation in the specific forms of recognition, acceptance, and acknowledgment of a new *status* to which a *function* is assigned (p. 40).

As becomes clearer in Chapter 3, where Searle is articulating the role of language in creating social reality, the "specific forms of recognition, acceptance, and acknowledgment" that he seems to have in mind here are those that generate *deontic commitments* in a sphere of rights and obligations (see esp. p. 70–71). Language, or something like language, is necessary to operate in such a deontic domain, claims Searle.

If Searle is correct about this, then the reason that nonhuman animals cannot assign status functions to brute facts is that they lack language or some language-like form of expression and communication. This entails that they cannot have the kinds of individual thoughts, and so the kind of collective intentionality,

both with a deontic component, necessary for the creation of institutional facts. It will thus be true that their behaviors and actions are not governed by constitutive rules of the form "X counts as Y", but this is a consequence of their lacking the right kind of prior intentional capacities.

I think that this reason for denying nonhuman animals the capacity to ascribe status functions to brute facts is not a good one, and that Searle is in fact mistaken to think that nonhuman animals lack this capacity. Let me pick up on what I see as a dialectical weakness in Searle's position here, and use that to introduce a puzzle about this way of articulating the relationship between collective intentionality and institutional facts that, in turn, helps to identify a problem for Searle's broader view of institutional facts and social reality. I begin with an analogy, one not all that removed from our topic.

Consider the issue of whether nonhuman animals have the social capacity *to play*, as many ethologists and behavioral ecologists claim they do (Bekoff and Byers 1998). Arguably, social play shares key features with behaviors that might be considered putative examples in which status functions are ascribed: it is a joint action that arises from the individuals involved recognizing and accepting that action as an instance of play, and would not count as play were such recognition and acceptance absent (Bekoff and Allen 1998). But suppose now that, faced with putative examples of social play in nonhuman animals that satisfy these criteria, one were to insist that the psychological capacity necessary for *genuine play* needs to include a capacity to generate deontic commitments, things that can, cannot, and must be done as part of the action's counting as play, and that are (or can be) articulated as constitutive rules governing play. One could then claim that this allows us to demarcate human play from nonhuman "play", since even if there is something that looks very much like genuine play between many nonhuman animals, it lacks a critical part of the overall intentional profile that genuine play has.

The question we would surely want to pose here is what the basis is for insisting on such a psychologically enriched notion of social play. In the context of a dialectic with someone who is perfectly happy to defer to ethologists, behavioral ecologists, and others who study animal behavior in talking of animal play, it will not do to respond by simply appealing to the alleged nature of play. That would clearly beg the question at issue. Nor would the situation be changed significantly by pointing to uncontroversial features of *human* play that perhaps require the capacity for deontic commitments, such as deferral to a referee or umpire, or the idea of fairness in play. For clearly the question at issue is not whether nonhuman play is as elaborate or as sophisticated as human play—I do not know of anyone who would claim that it is—but whether there is a core behavioral capacity or a capacity for action that is shared by humans and at least some nonhuman animals that is called "play" or "social play". Even in the human case, we require a conception of play that applies paradigmatically to those whose lives are replete with play—children—and much the same problem arises here as with the case of non-human animals.

This brings us to the role of language both in generating deontic commitments and in the social phenomena of interest. In the case of human social play, language clearly does play a role, not simply because lots of our play involves literally playing with language (jokes, irony, wordplay), but also because we typically initiate and terminate play with linguistic utterances ("Let's play!", "Enough!", "I give up!") and use language as a means of maintaining and directing the play action. Insofar as there are forms of human play that essentially involve language, these will simply be forms of play that are unavailable to nonlinguistic creatures. But is language crucial to social play *per se*? Play is not simply behavior, but behavior generated by a kind of shared intentionality. In us, that shared intentionality leads to or includes certain articulable norms of play, and these regulate both play behaviors and what we might think of as institutions of play—explicit games being the most obvious example. Perhaps language plays a crucial role in the creation and maintenance of such institutions, as Searle suggests, and perhaps it plays such a role in the explicit articulation of the norms of play. But it is difficult to see what further, crucial role language could have in social play *per se*, or in the shared intentionality that produces it.

The same is true of examples such as that of territory marking, which Searle discusses cursorily. While territory marking operates via olfaction in many species, it is a far cry from operating "in virtue of sheer physics", as erecting a wall around an area functions to mark a boundary and keep out intruders. In some species of animal olfactory cues are tied directly to particular behaviors in stimulus-response fashion, the olfactory detection of oleic acid by worker ants and the subsequent ejection from the nest of any ant so scented being one example sufficiently well-known to have become a mainstay of the philosophical literature (e.g., Allen and Hauser 1991; Sterelny 1995). But this very much *isn't* how territory marking works in psychologically more sophisticated animals, such as canids. (If it were, as Colin Allen pointed out to me, territory *defense* would seem puzzling, since territory *marking* would suffice to repel intruders.) Particular scents are used for different purposes by canids, and the patterns that these form depend on further contextual variables, such as whether the animal is dominant, whether it forms part of a mating pair, and where the animal is in the territory, such as at a junction or a periphery (Allen, Bekoff, and Crabtree 1999). Whether scented objects serve to mark a territory for an individual wolf or coyote, both for an individual marking the territory and an individual observing a territory boundary so marked, depends on a range of social variables, including on its being treated as a territory by others. When there is evidence that scents are so treated, individual species members treat them as such, and when there is not, they are mere scents. Just as stones that appear in certain kinds of patterns are typically signs of past human presence and intention (and those in wall-like patterns more specific signs), so too are similar scents detected on spatially separate but related objects typically signs of past animal presence and intention. But whether those signs become the basis for the recognition of something more specific, such as a territorial boundary,

depends in part on the collective recognition of that boundary. There is nothing about certain scents themselves that means "Beware! Territory here!"; rather, they come to take on that kind of significance or status in virtue of species-specific collective recognition. Territory marking and defense in at least some animals seems to be, *contra* Searle, a fairly clear example of the ascription of status functions in the absence of human institutions.

Reflecting on these two examples also highlights the tight coupling between the character of the collective intentionality that Searle thinks is required for there to be institutional facts, and those institutional facts themselves. On Searle's view, there is a relatively narrow gap between the individual psychological capacities necessary for participation in a world of institution facts and the creation of those facts themselves, and this provides the basis for explaining our uniqueness as creators of institutional facts. Nonhuman animals fail to create such facts because they lack some crucial aspect of those capacities, claims Searle. In terms of Searle's notion of status functions, non-human animals cannot assign status functions to brute facts because they lack a crucial shared psychological capacity for doing so. I shall move on now to probe this putative relationship between collective intentionality and institutional facts, at least insofar as those facts are representative of a larger class of social facts or social reality.

4. WHAT IS SOCIAL REALITY?

So far, I have used the analogy to social play to question the appropriateness of a view, such as Searle's, that insists on a psychologically-enriched profile of the participants in a social action that (we agree) requires shared intentionality and, in some sense, norms governing play action, but for which Searle seems to require something like explicit norms that can be articulated in a language by the participants in constitutive rules. That additional requirement seems inappropriately stringent in the case of social play, something that not only non-human animals but that infants and very young children appear to engage in, and the suggestion is that it is likewise too strong a constraint to impose on the capacity to assign status functions as well. But I want to use the example of social play in another way now: to question whether a view that offers what I have called a tight coupling between psychological capacities and institutional facts is likely to have the resources for answering both of our questions, "What are institutional facts?" and "What is social reality?" This introduces a second reason to question the adequacy of Searle's account of both institutional facts and social reality.

Whatever we say about the examples of social play and territory marking in at least some animals, it seems hard to deny that such play is social in a sense relevant to the question "What is social reality?". Play is interpersonal, it involves participants who have at least second-order intentionality, and it requires at least some level of shared intentionality, we-intentionality. (That is

no doubt part of what makes it *fun*.) Interactions having these features are typical of many behaviors that we find in human and nonhuman communities, including both cooperative and competitive interactions, behaviors that are aggressive or defensive, and those that involve resource acquisition or protection or by contrast that seem frivolous and unconnected to the struggle for existence. Some of these, such as grooming and alliances in primates, social play and territory marking in canids, and collective food sharing in bats, have been studied in detail. The question is where they fit in Searle's framework, given that they do not involve *institutional* facts, and that they may or may not involve the psychological capacities that Searle thinks necessary for the creation of institutional facts.

One option is simply to deny that such behaviors are, in the relevant sense, parts of social reality at all. In terms of Searle's initial dichotomy between brute facts and institutional facts, this would be to recognize them as brute facts. Some facts, one might say, are brute facts about the social world, such as who interacts with whom, what percentage of a population engages in a certain behavior, or what an individual's ranking is amongst his or her peers. None of these need involve human institutions and so need not be institutional facts, and need not presuppose the kind of collective intentionality necessary for there to be such facts. Thus, the Kantian question of how what we might call *brute social facts* are possible does not arise, or at least does not fall within the range of Searle's account of social reality, and how it is possible.

At several points Searle himself suggests this kind of response when he *stipulates* that the expression "social fact" refers to "any fact involving collective intentionality" (p. 26, cf. also p. 122, and his 2003, p. 198). All of the social phenomena that I have identified above do, I think, involve collective intentionality, at least conceived simply as first person plural intentionality, although they may not involve the more complicated forms that such collective intentionality can take and that, let us suppose, is necessary for the creation and recognition of institutional facts. Searle could look to exclude them by denying either that they involve collective intentionality or collective intentionality of the right kind, making them social phenomena that are interesting in their own terms, perhaps, or to someone else, but (apparently) not within Searle's own purview.

This option, whether or not it is one that Searle himself means to adopt, is ill-advised. What it does, in effect, is to construe social reality narrowly and in its most extreme form collapse the question "What is social reality?" into the question "What are institutional facts?". But for anyone working seriously in the social or behavioral sciences, even those who are focused exclusively on institutional facts, the institutional domain is a proper subset of the social domain and not coextensive with it, something that Searle himself recognizes (e.g., pp. 26, 121). This is not simply because of a concern with understanding nonhuman social behavior (although that is a concern of some, even in the social sciences), but because even in the human realm the kinds of interactive behaviors of which social play is typical can and do take place both inside and outside of institutional contexts. Examples that are species-typical for *Homo sapiens* include

social play and other forms of joint attention, cooperative endeavours that range from sharing one another's company (say, on a walk) to undertaking long-term projects that require a division of labor (such as raising children together), and telling one another stories or our sharing our narrative histories with one another.

Since members of our species now live in a world full of institutional facts (in Searle's sense), including those involving cultural artifacts, religious practices, formal rituals, and technological innovations, many of these interactive behaviors—perhaps even most of them—are mediated by institutional facts. We can jointly attend to a movie rather than play together, can share each other's company not only during a walk but also on a cruise or at dinner, can undertake joint projects that draw on technological wonders ranging from the pen to the computer, and can tell one another stories not simply face-to-face but at a distance through mobile phones and the internet. Our social lives are permeated with institutional facts, true, but my point is that this is an institutional overlay to a preexisting social realm, an extension of capacities and abilities that are presupposed by, rather than exhausted by, the institutional forms in which they are manifest. And some of these capacities and abilities are *social* (or psychosocial), not simply psychological, in nature. They are interpersonal, other-directed abilities that are manifest in our direct interactions with one another, even in a world crammed with institutional facts.

So simply denying this range of examples as part of social reality is not a real option. A second and more plausible option is available, however, for someone taking Searle's general approach to social reality and institutional facts faced with examples of non-institutional sociality. Rather than simply denying that such examples are social by stipulation, one could maintain that our best bet for coming to understand them is by focusing on their manifestation in behaviors and actions that involve institutional facts. Institutional facts exemplify a broader realm of social facts, and by understanding them—how they are possible—we will understand that broader realm, and how it is possible. We might think of institutional facts as paradigmatic social facts, or as representing in a particularly perspicuous way what is problematic or puzzling about social facts. Either way, there is an *epistemic* justification for the focus on institutional facts. Since Searle himself conducts much of his general discussion using the terms "social reality", "social phenomena", and "social facts", but the details of his account are focused on institutional facts, this option may be one that Searle himself finds attractive.

Yet the perspective that I have sketched also indicates something unsatisfactory about this second response to the examples I have presented, and I think this can be revealed in terms of Searle's own explicit framework. As we have seen, Searle is clear that institutional facts require a collective psychology of a certain kind, and goes so far as to identify social reality as a kind of (part of, subset of) what we might call *intentional reality*, the part of reality that involves intentional facts. But while we can concede that this is true of institutional reality,

it is not true of social reality more generally. Even on Searle's own hierarchical representation of where institutional reality fits in the overall ontology of common sense (p. 121), it is clear that there are social but non-institutional facts, and that there are several features distinctive of institutional facts that are not shared by intentional and social facts more generally: they involve the ascription of status functions and are often linguistically mediated. Thus, while a focus on institutional facts may help us to understand some aspects of both intentionality and sociality, an exclusive focus on them would be like trying to understand mammals but focusing on monotremates, or birds by focusing on penguins. In particular, while concentrating on institutional facts may allow us to understand the sense in which some forms of sociality are constructed (in the sense that preoccupies Searle), it also obscures the equally important sense in which there are forms of sociality or social reality that are *not* constructed. It is with these in mind that I want to conclude with some thoughts about the place of sociality vis-à-vis both intentionality and institutional reality.

5. THE DOMAINS OF THE INTENTIONAL, THE SOCIAL, AND THE INSTITUTIONAL

Like Searle, I am interested in the question of how institutional facts are possible. But unlike Searle, I do not think that we best answer this question by emphasizing psychological capacities of a certain kind that individuals possess and bracketing or ignoring the broader social or interpersonal domain of which the institutional is a part. Rather, we would do better in answering *both* the question of what institutional facts are and that of what social reality is by trying to understand the relationships between not just the intentional and the institutional, but between the intentional, the social, and the institutional. Pulling apart the social and institutional domains, and so treating the question "What are institutional facts?" separately from the question "What is social reality?", makes for a more complicated story, but one that seems required by the richness of the phenomena in both the institutional and social domains.

One of the complications that arises in treating the two questions separately lies in the relationship between intentionality and sociality. Despite Searle's stipulation that he uses the expressions "social facts" and "collective intentional facts" so as to be coextensive (p. 122), our previous discussion should lead us to wonder about where aspects of the social world that do not involve collective intentionality fit into Searle's ontology. For not only are there social behaviors that require less than the full intentional profile that Searle views as necessary for the creation of institutional reality, but there seem to be such behaviors that require very little by way of intentionality at all, let alone shared, collective intentionality. For example, species of bird flock and species of fish school, and both are regarded as social behaviors; insects that live in large colonies with a division of reproductive labor are called "social insects"; and again whatever we think of whether the territoriality of animals is established by the ascription of status functions, that behavior is treated as a paradigm social behavior by those

who study it. In short, while there are forms that sociality takes that require intentionality, especially collective intentionality, there are others that do not. For this reason, while institutional facts presuppose intentional facts, social facts in general do not.

As the first indented quote of Searle's that I provided in the paper indicates, Searle himself adopts a fairly traditional, hierarchical, levels-based view of reality, according to which there are distinct physical, living (or biological), intentional (or psychological), and social levels, with entities at "higher" levels being organized systems of entities at immediately "lower" levels. It is testimony to the power of this idea of a hierarchical nesting of levels of reality (or things, facts, processes, etc.) that there has never been a serious competitor to that idea, at least amongst those who accept some measure of realism about the entities identified at the physical level. Recent alternatives seem to reduce to variants (e.g., distinguishing levels from orders, Kim 1998), or simply to denials of the view, together perhaps with a rejection of the conception of an integrated ontology (e.g., pluralism, Dupré 1993).

While I think that the levels-view is misleading in various ways, I am also skeptical of there being a global alternative to it. The best we can do, I suspect, is to identify particular problems introduced by (or that cannot be solved by invoking) this metaphor for the relationship between the many things there are in the world, and then to propose local alternatives to it. Elsewhere (Wilson 2003, 2005a: Chaps. 9–10) I have argued that a levels-view provides a view of natural selection that is misleading in several respects, suggesting that we think of certain kinds of properties of genes, individuals, and groups, and the processes that they feature in, as *entwined* or *fused*, rather than simply present "at different levels of selection". A variant of that idea may be applicable in thinking about the relationship between the psychological, the social, and the institutional, but we need a better sense of what the limits are to the levels metaphor as Searle and others have used it to make sense of social reality.

We have already met one obvious problem for the idea that social facts are a special kind of intentional fact: social facts can be found both "higher" than intentional facts and "lower" than them. There is an evolutionary reason for this, one that may help us to reflect on the relationships between sociality and intentionality in a way that sheds some light on the two questions with which we began, "What are institutional facts?" and "What is social reality?".

The social realm straddles both sides of the psychological or intentional domain because social aggregation is a more pervasive fact about the living world, especially the mobile, multicellular living world, than is intentionality. The fact of social aggregation, a direct product of how most organisms reproduce, means that at least some of one's conspecifics are typically a prominent feature of mobile, multicellular organisms. Thus, finding ways to interact with them constitutes an adaptive problem that such species typically face. Of course, one way to solve that problem is to interact with conspecifics as little as possible, and while some species have found that solution, they are very much a tiny minority.

Even in species that are sometimes called "solitary", such as orangutans or most species of spiders, offspring are relatively helpless and it is in their evolutionary interest typically to know how to gain protection and food from others, typically their parents, for at least some period of time. Sexual reproduction introduces another basis for sociality amongst the mobile and multicellular, since as crude as some of the strategies for mating are in other species by our own lights, they nonetheless are social in nature, even when they are coercively so.

So sociality is a pervasive feature of the animal world, some forms of which pre-date intentionality and others of which coevolved with particular forms of intentionality. More sophisticated forms of intentionality are layered on more basic forms that coexist with social dispositions and behaviors, and that in turn give rise to more sophisticated forms of sociality. Searle is right, I think, to take some of these forms—those involving institutional facts—to require not simply shared intentionality but shared intentionality of a special sort. Yet such institutional facts and the social reality they constitute are very much the tip of the social iceberg, the part that we can see from our own institution-laden point of view. Such facts arise when organisms with intentional capacities come to collectively represent in certain ways, and they can do that only against a pre-existing background that includes other social phenomena. This is not so much to imply that sociality should be conceptualized at a "lower level" than intentionality as to suggest that we replace such hierarchical thinking with something more like a cycle of social-intentional-institutional facts that dynamically build on one another over time.

While I have introduced this idea in a sketchy evolutionary scenario, the same general conception can be adopted in thinking about ontogenetic development, about everyday human life, and about cultural change in human societies. Consider human development. We are each born with a cluster of innate capacities—physical, intentional, and social in nature—and come to acquire others through these and the experience we have of the world. We do aim to understand what the physical basis is for both the initial capacities and those that we come to acquire over time, as well as the broader relationships that hold between particular physical, intentional, and social capacities. But we are born with sociality, both in terms of our innate orientation towards others and in terms of finding ourselves born into a social world, and the complex interplay between the physical, the intentional, and the social changes our standing with respect to each over time. Human sociality is not a kind of intentionality, although there are particular forms of sociality (such as friendship) that require particular forms of intentionality (e.g., a theory of mind), and that in turn may require specific physical abilities. Conceptualizing these as part of a layered reality can be an impediment to understanding the relationship between "the intentional" and "the social", as I think it is in evolutionary terms.

Searle would be right to point out that none of this—the appeal to non-institutional forms of sociality, the questioning of the levels metaphor, or pointing

to the interplay between the intentional and the social on both evolutionary and developmental timescales—sheds light on the basic problem with which he has grappled. That is a problem about the construction of social facts, a problem that Searle addresses by showing the connection between such facts and collective intentionality of a particular kind. What I am suggesting, however, is that Searle's preoccupation with that problem, a problem that at its core is one about the construction of institutional facts, has left him blind to the greater part of sociality or social reality, and especially its relationship to other aspects of reality, including intentionality.

6. CONCLUSION

I have argued for three chief, related conclusions in this paper. First, in Section 3, I argued that although Searle is right to highlight human uniqueness in the world as creators of certain kinds of social facts, he is wrong to likewise see us as the only creatures with the kind of first person plural intentionality necessary to create such facts. In part this is because Searle builds in more than is strictly necessary to collective intentionality (as the example of play aimed to show), and in part it is because Searle has either not explored or not taken seriously enough some of the relevant literature on nonhuman cognition (as the example of territory marking aimed to show). From Searle's own discussion of status functions, it seems clear to me that a range of nonhuman animals have the intentional capacity to ascribe such functions. Since Searle thinks that they do not, his views direct us to provide an intentional account of their lack. By contrast, if I am correct that they do have this capacity, at least in a rudimentary form, then the question of why only we create institutional reality, or at least the kind of elaborate institutional reality that typifies *Homo sapiens*, needs a different kind of answer. Expressing this in terms that Searle himself occasionally does—in terms of the notion of *culture*—perhaps helps to locate what might sound like an arcane debate between philosophers within a framework that those in the social sciences will recognize. To understand ourselves as distinctively cultural animals, we need to look beyond individual and collective intentionality.

Second, in Section 4, I argued that the view mediating this inference from human uniqueness in one domain—that of social facts—to another—that of collective intentionality—is mistaken. This is the view that the domains of sociality and collective intentionality are tightly coupled. Searle holds this view in part because he requires more than mere first person plural intentionality for institutional facts, namely a form of such collective intentionality that requires the explicit recognition of norms and even language. Perhaps this is required for institutional reality, but this brings out the point that institutional reality is a special part of sociality, not simply the whole of it or even a representative part of it. Not only does the ascription of status functions require a less psychologically enriched set of capacities than Searle thinks it does, but there are large tracts of sociality that seem to fall outside of the purview of Searle's view of the

"construction of social reality". Precisely because of its emphasis on institutional facts, status functions, and a particularly enriched form of collective intentionality, Searle's account has little to say about the forms of sociality that are not constructed in the way in which institutional facts surely are. While Searle himself recognizes the distinction between mere sociality and institutional reality, his bracketing of the former skews the account he gives of the latter.

My final point, articulated in Section 5, is that we need, collectively, to do some serious thinking about the broader relationships between sociality, intentionality, and institutional facts. The questions "What are institutional facts?" and "What is social reality?" are not simply different but different *kinds* of question. The first is what we might call a "puzzle question", a question that might be rephrased as the question of how institutional facts are possible and that is premised on a prior puzzle about how some parts of reality might be constructed yet objective. The second is not a puzzle question at all but a question about a range of phenomena—those I gather under the heading "sociality"—and how they are to be understood. Some of these phenomena, those concerning institutional facts, are puzzling for the reasons that Searle articulates, and there are many sensible and plausible things that he says about them in responding to the puzzle question. But the question about sociality is much broader, and to answer it we need a more sophisticated understanding of the relationship between intentionality and sociality, one that I suspect will take us beyond the levels metaphor that has served as a crutch in this and other areas of philosophy for too long.

REFERENCES

Allen, C. and Hauser, M. (1991). Concept attribution in non-human animals: Theoretical and methodological problems in ascribing complex mental processes. *Philosophy of Science*, 58: 221–240.

Allen, J.J., Bekoff, M. and Crabtree, R.L. (1999). An observational study of coyote (Canis latrans) scent-marking and territoriality in Yellowstone National Park. *Ethology*, 105: 289–302.

Anscombe, G.E.M. (1958). On brute facts. *Analysis*, 18: 69–72.

Bekoff, M. and Allen, C. (1998). Intentional communication and social play: How and why animals negotiate and agree to play. In M. Bekoff and J.A. Byers (eds.), *Animal Play: Evolutionary, Comparative, and Ecological Perspectives*. Cambridge: Cambridge University Press.

Bekoff, M. and Byers, J.A. (eds.) (1998). *Animal Play: Evolutionary, Comparative, and Ecological Perspectives*. Cambridge: Cambridge University Press.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Pettit, P. (2003). Groups with minds of their own. In F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield.

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, J.R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.
Searle, J.R. (2003). Social ontology and political power. In F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MA: Rowman & Littlefield.
Sterelny, K. (1995). Basic minds. *Philosophical Perspectives*, 9: 251–270. Reprinted in K. Sterelny, *The Evolution of Agency and Other Essays*. Cambridge: Cambridge University Press.
Wilson, R.A. (2003). Pluralism, entwinement, and the levels of selection. *Philosophy of Science*, 70: 531–552.
Wilson, R.A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge: Cambridge University Press.
Wilson, R.A. (2005a). *Genes and the Agents of Life: The Individual in the Fragile Sciences: Biology*. Cambridge: Cambridge University Press.
Wilson, R.A. (2005b). Collective memory, group minds, and the extended mind thesis. *Cognitive Processing*, 6: 227–236.

Part III
Aspects of Institutional Reality

THE VARIETIES OF NORMATIVITY: AN ESSAY ON SOCIAL ONTOLOGY

*Leo Zaibert and Barry Smith**

For much of the first fifty years of its existence, analytic philosophy shunned discussions of normativity and ethics. Ethical statements were considered as pseudo-propositions, or as expressions of pro- or con-attitudes of minor theoretical significance.[1] Nowadays, in contrast, there are prominent analytic philosophers who pay close attention to normative problems and important books written by such philosophers on topics in law and social justice and on social and institutional ontology. Here we focus our attention on the work of Searle, at the same time drawing out an important connection between Searle's work and that of two other seminal figures in this development: H.L.A. Hart and John Rawls.

Hart was, within the context of recent analytic philosophy, the most important philosopher of law and Rawls the most important political philosopher. Still a child of the 20th century, as we shall see, Searle tends to assume that there is but one type of normativity within the realm of social institutions. Like Hart and Rawls, he thereby neglects features which are of crucial significance for an adequate understanding of social reality.[2] Our main goals are twofold. On the one hand we wish to expose how this neglect constitutes a shortcoming of Searle's ontology of social reality.[3] On the other hand, our attention to the ways in which this neglect plays out in the normative philosophy of other luminaries of 20th century analytic philosophy should help us to identify an entrenched trend, and also thereby contribute to its reversal.

* The authors thank the Alexander von Humboldt Foundation, Ingvar Johansson, Kevin Mulligan, Mariam Thalos, and Raimo Tuomela. This article draws on discussions in Zaibert 2005 and in Smith 2003.

[1] See, e.g., Hare 1963, Ayer 1952, Wittgenstein 1965.

[2] Among the most important contributions to the field of social ontology we find Bratman (1999), Gilbert (1989), Searle (1995, 2001), Tuomela (1995, 2002).

[3] We shall follow Searle's lead in accepting a distinction between social and institutional reality. According to Searle institutional facts are a subset of social facts, characterized by the fact that they involve the transmission, cessation, or creation of power, whereas social facts do not directly relate to power. See Searle 1995: 38, 79ff. Searle rejects, however, the idea of a sharp distinction between these two classes (1995: 88ff.).

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 157–173.
© 2007 Springer.

1. HART AND SOFT POSITIVISM

We first set the stage for our more detailed treatment of Searle's views on normativity with a discussion of Hart and Rawls. The work of Hart, especially, forms part of a famous debate between natural law theorists and legal positivists, a debate which reveals that the question of the ontological status of laws has historically been linked to normative issues of moral philosophy. Natural law theorists affirm that immoral law is not law; that is, they believe that the ontological status of laws is determined by their relation to morality, in accordance with the famous motto: "*Non videtur esse lex quae justa non fuerit*". Legal positivists, on the other hand, insist that law is law independently of whether or not it is moral. According to the classical legal positivism of John Austin, for example, the issue of the legal status of law is an entirely empirical affair, to be established primarily through the determination of pedigree and enforceability. Was the entity or institution created and maintained in existence in accordance with the right sorts of rules? Is the entity such that the state can coerce people into complying with it?

According to Austin, we are to understand the nature of a legal system by starting out from the case of someone forcing someone else at gunpoint to hand over his wallet. The normativity of the law differs from the normativity of the highwayman only in this: that the law normally functions on the basis of threats alone; only in extreme circumstances it is necessary to bring guns into play.

In *The Concept of Law* (first published in 1961, references below are to the second edition, Hart 1994), Hart deploys a sustained attack on traditional legal positivism. His criticism of Austin is both elegant and persuasive. Hart himself still defends a positivistic conception of the ontological status of the law, but he rejects traditional positivism, above all because of its superficial treatment of rules. The rules the gunman imposes upon his victim—"Hand over your wallet", "Don't do anything stupid"—are all of the same type: they demand certain sorts of conduct. The law, however, operates on the basis of two types of rules, which Hart calls primary and secondary. Primary rules are duty-imposing; they demand conduct in just the way in which the gunman's actions do. Secondary rules are power-conferring; they make certain sorts of situations possible—they are rules about rules. A rule that states that a judge is entitled to decide how to interpret a primary rule is a secondary rule; it gives the judge the power to settle disputes by establishing what the correct interpretation of a law is.

It is possible, perhaps, to imagine an entire society in which there existed only primary rules. But such a society would be profoundly inept when it comes to resolving controversies about the laws themselves or about their interpretation. A situation, on the other hand, in which secondary rules would arise in relation to highwaymen robbing stagecoaches belongs, at best, to the world of Monty Python.

With only one kind of rule in its conceptual armoury, Hart argues, traditional positivism is unable to distinguish between two crucially distinct phenomena: (1) being *de facto* obliged and (2) having a genuinely normative obligation. If a gunman puts a gun to your head, you might indeed be, as a matter of empirical fact, obliged to hand over the money. For you to have a normative obligation, in

contrast, it is necessary that you accept not only the empirical fact of your being obliged but also the rightness of the *system* which makes this so (even if you do not accept specific rules in this system). You accept that to do this or that is your *duty*; that it is *the right thing to do*. This notion finds no purchase in the realm of actions performed in response to gunmen's threats.

Hart refers to this dimension of acceptance as the "internal aspect" of obligations, to which he opposes an "external aspect"—the only one that traditional positivism is capable of explaining. He asks us to imagine someone describing the functioning of a street light in a busy intersection in the following way: when the street light becomes red in the direction of the cars, the likelihood that cars will stop, and that pedestrians will cross the street is very high; when the street light becomes green in the direction of the cars, the likelihood that cars will move forward and pedestrians will stay put increases. Obviously, Hart points out, such a description fails to mention a fundamental element of what is really going on. The red light is not merely a *sign* that allows us to predict that drivers and pedestrians will behave in this or that way; rather it is a *reason* which gives rise to this or that behavior. The red light indicates not simply *that* I stop, but that I *ought* to stop. This notion of a *reason* is not available to traditional legal positivism.

Since Hart is himself a positivist, it might look as if by introducing normative elements into his determination of the ontological status of laws he concedes too much to natural law theory. After all, for Hart as for natural law theorists, whether a given entity is or is not law depends on normative factors. He insists, however, that he has carved out an intermediate theoretical space between natural law and traditional positivism, which he calls "soft positivism" (Hart 1994: 250ff.).

Hart's strategy—though he does not himself admit it—is to distinguish between two types of normativity. On the one hand is the robust normativity of the natural law theorist, illustrated for example by the Ten Commandments. On the other hand is Hart's own brand of normativity—what we might call soft normativity—which is what is necessary, in his view, for the existence of laws. Soft normativity is the sort of normativity that flows logically from the very nature of secondary rules. Secondary rules create institutions, and these institutions in turn create the very possibility of certain sorts of acts. Hart himself appeals to the example of games in order to illustrate this point (Hart 1994: 56ff., 141ff.). A group of people can play football without requiring the presence of a referee of any sort. But when a referee is present and disputes arise, then the referee will have the last word in resolving such disputes. His appointment is however possible only insofar as the players accept the secondary rules that make the institution of refereeing possible. That the referee has the last word is part of the content of the corresponding secondary rule, and it is this same rule which gives rise to the normative component in the referee's decisions. When a referee declares "penalty kick", for example, he is not merely providing an indication of what is likely to happen next (any more than a traffic light is providing an indication of likely traffic flows). Rather, his declaration is the very reason which *explains* what happens next, because it explains what *ought* to be done.

But there is a problem with Hart's approach. The sense of 'ought' as expressing soft normativity, the sense of ought that is involved in rules of games like chess or football, is radically different from the sense of 'ought' that is involved, for example, when someone says that we ought to treat other human beings with respect, or that we ought not to gratuitously harm them. We believe that any ontology of legal institutions that does not do justice to the distinction between these types of normativity is doomed to fail.

2. RAWLS AND RULE-UTILITARIANISM

In 1955, at the beginning of his career, Rawls published an important article called "Two Concepts of Rules" (Rawls 1955; references below are to the reprint in Rawls 1999b), a work which was unfortunately overshadowed by his later *A Theory of Justice* (Rawls 1999a) but which has nonetheless exerted some considerable influence along the way. It has been translated into numerous languages and it is a mainstay in anthologies dealing with moral philosophy. It consists of an attempt to defend utilitarianism against certain traditional objections relating to the alleged incapacity of utilitarians to deal with the institutions of promise and of punishment, and to the widespread supposition that utilitarians must perforce allow on felicific grounds the occasional breaking of promises and the punishing of innocents.

Rawls's defense of utilitarianism, which has become a commonplace in many philosophical circles, goes roughly as follows: utilitarianism should not be seen as a theory that seeks to maximize general welfare in every instance. Rather, it is a theory that seeks to devise general rules of behavior of a sort that would *tend* to maximize welfare. The idea is that, once the rules have been established, then they must be followed, even if violating rules on this or that occasion yielded a net increase in general welfare. It is then unlikely that human beings would ever endorse on felicific grounds *rules* that would authorize the breaking of promises or the punishment of innocents.

In this way Rawls draws the nowadays familiar distinction between act- and rule-utilitarianism, and this constitutes the first half of his article. It is however the somewhat neglected second half which is important for our purposes. Indeed, the distinction which occupies him in the first half Rawls himself considers to be rather obvious (Rawls 1999b: 33). What he considers not obvious is the existence of a certain ambiguity regarding the notion of a rule, as between what he calls summary rules and practice rules.

A summary rule is simply a guide for action, formulated on the basis of experience. For example, if upon encountering caustic persons in the past one has established that the best course of action has been to keep a low profile, one might decide on encountering a caustic person now that it is best to do the same. Summary rules are inductive. The decisions they are based upon are logically prior to the rules themselves.

Rawls's practice rules, in contrast, are not inductive; they are not the result of such recollection of past events, and they are logically prior to the cases in which

they are applied. An example of a practice rule would be the rules involved in games like baseball. Here the rules precede the game. What counts as a 'run' in baseball is not the result of looking back at what things have counted as 'runs' in past baseball games and then concluding: "well, this must also be a 'run'". Practice rules, rather, give rise to the very possibility that the cases in which they are applied can indeed occur. Thus they are not mere generalizations from past behavior. Practice rules define the very behavior which they at the same time permit. In chess, bishops move diagonally; the issue as to whether or not to move a bishop diagonally is not a genuine dilemma within the context of playing chess. If someone were to insist on moving his bishop non-diagonally, then he would *eo ipso* no longer be playing chess.

According to Rawls the rules of rule-utilitarianism are precisely practice rules. They are rules which define the very institutions they regulate. The normativity of rule utilitarianism, as Rawls conceives it, is the logical normativity of the system of propositions which describe institutions that rule-utilitarianism itself creates, such as promising and punishment. The State, for example, does not really have the option of whether or not to punish an innocent person, for punishing the innocent is logically forbidden by the very practice rule which sets up the institution of punishment itself (Zaibert 2003). Deciding to punish an innocent person is analogous to deciding to move a bishop non-diagonally in chess. As Rawls would have it: "To engage in a practice, to perform those actions specified by a practice, means to follow the appropriate rules" (Rawls 1999b: 37).

On Rawls's interpretation, then, the main difference between act- and rule-utilitarianism is not merely related to the issue of where to apply the welfare-maximizing measure (namely, to rules concerned with act-types rather than with act-tokens). Rather, rule-utilitarianism differs from act-utilitarianism in that it is a *logical* theory. In defending himself against charges that his view might be too conservative (insofar as he may be taken to blindly endorse the status quo of existing social institutions), Rawls states: "The point I have been making is rather a logical point", and then he continues: "where a form of action is specified by a practice there is no justification possible of the particular action of a particular person save by reference to the practice" (Rawls 1999b: 42). Utilitarianism in the hands of Bentham and Mill is a moral theory concerned with the same substantial normative issues as are addressed by natural law theorists; Rawls transforms it into a logical doctrine.

Whereas in "Two Concepts of Rules" Rawls seeks to defend utilitarianism, in *A Theory of Justice* and other later works he seeks to develop a neo-Kantian theory of the justice of social institutions that is opposed to utilitarianism. Yet there is nonetheless a certain connecting thread between the two works, which is the importance Rawls gives to the logical structure of institutions. The emphasis on procedural and formal justice in *A Theory of Justice* (Rawls 1999a: 74ff. and passim) can be seen as a reflection of the logicist leanings found in his early defense of utilitarianism. Focusing on Rawls' concern with the logic of institutions allows us to see the two works within a single context, and it allows us also to see the challenge which Rawls faces: in transforming normativity as traditionally

conceived into a matter of the logical consequences of rules of a certain type, rules which we adopt when we choose to engage in certain practices, Rawls (like Hart) makes questions like: "Why should we keep promises?" or "Why should we endorse a social order based on these or those principles?" of a piece with the question "Why should we play the game of chess rather than some other, slightly different game?" Let us turn now to Searle, and see how he seems to face a similar fate.

3. SEARLE AND OBLIGATIONS

In one of his earliest articles, "How to Derive 'Ought' From 'Is'" (Searle 1964), Searle claims that he found a way of showing that from purely descriptive premises we can derive normative conclusions. In other words, that he has shown how to bridge the gap between "is" and "ought", between matters of fact and judgments of value.

The best place to begin our discussion is Searle's analysis in *Speech Acts* of what he calls "The Naturalistic Fallacy Fallacy": "the fallacy of supposing that it is logically impossible for any set of statements of the kind usually called descriptive to entail a statement of the kind usually called evaluative" (Searle 1969: 132). The view that Searle wishes to defend is, in his own words, that:

> the view that descriptive statements cannot entail evaluative statements, though relevant to ethics, is not a specifically ethical theory; it is a general theory about the illocutionary force of utterances of which ethical utterances are only a special case.

How can I become obliged by merely uttering certain words, say, "I promise to mow your lawn"? Searle's gambit, in embryo, is as follows. He wants us to see the traditional problem of the naturalistic fallacy as a particular case of a putatively more general problem in speech act theory. It is then this latter problem, of the normativity associated with speech acts, which Searle sets out to solve—not, as many authors have too quickly assumed, the traditional problem of moral normativity.

Searle himself is emphatic about the fact that whatever relevance his views might have regarding moral normativity would be a mere side effect of his concern with a *logical* problem about the illocutionary force of certain utterances. As a propaedeutic warning, he tells us that we must avoid "lapsing into talk about ethics or morals. We are concerned with 'ought' not 'morally ought'" (1969: 176). And again: "Let us remind ourselves at the outset that 'ought' is a humble English auxiliary, 'is' an English copula; and the question whether 'ought' can be derived from 'is' is as humble as the words themselves" (1969: 176). The humble sense of 'ought' with which Searle is concerned is the same sense as that in which, when playing chess, you *ought* to move your bishop diagonally. We note in passing that this sense of 'ought', interesting as it might be, is at best of indirect significance for moral philosophy.

Searle's treatment of the humble sense of 'ought' is reminiscent of another treatment of these matters in the writings of A.N. Prior, who noted that, from

the premise that "Tea drinking is common in England", one could validly infer that "either tea drinking is common in England or all New Zealanders ought to be shot" (Prior 1949).[4] Of course, this inference constitutes no contribution whatsoever to the solution of the metaethical problem regarding the nature of moral propositions.

To be sure, Searle's derivation of an 'ought' from an 'is' is not as vacuous as Prior's *reductio*. But it is similarly irrelevant to ethics. For it merely tells us something about the *meaning* of the word 'promise'. Promising *means* undertaking an obligation, and undertaking an obligation *means* that one ought to do whatever one has obliged oneself to do. The problem is that this sense of obligation falls short of capturing the sort of obligation that involves moral normativity. As Searle admits, "whether the entire institution of promising is good or evil, and whether the obligations undertaken in promising are overridden by other outside considerations are questions which are external to the institution itself" (1969: 189). Such external considerations are very often precisely of a *moral* nature.

There is something odd, then, about Searle's attempt to examine what he describes as the *general* problem of the naturalistic fallacy, for the classical interest of philosophers in this fallacy has been focused precisely on its properly ethical dimension. So it was for Hume (1888), for Moore (1959), and for Popper (1966). These authors leave no doubt that they are dealing with an ethical problem.

The problem with Searle's treatment of the naturalistic fallacy is brought out nicely by D.D. Raphael in the context of writing on the justification about political obligations. Why does the citizen have a duty to obey the laws about the State? Raphael points out that there is an answer to this question which is "simple and obvious": "It follows logically that *if the State is authoritative*, i.e., has the right to issue orders to its citizens and the right to receive obedience from them, the citizens are obliged to obey those orders" (Raphael 1990: 175, emphasis added). Raphael rubs home the downright platitudinous character of this sort of answer: "the citizen is legally obliged to obey the law because the law *is* that which imposes legal obligations" (1990: 175). And then he compares this sort of answer with the passage in which Hamlet is asked by Polonius, "What do you read my lord?" and Hamlet replies, "Words, words, words". Though both answers are "formally correct", as Raphael puts it, they tell us "virtually nothing" (1990: 175). Something similar happens with Searle's derivation of 'ought' from 'is'. The very meaning of promising is that one ought to do what one has promised to do. But this sense of 'ought' is indeed humble, and it is dramatically different from the sense of 'ought' that has preoccupied moral philosophers throughout the ages.

In spite of his reminding us of the humble nature of the problem he seeks to solve, toward the end of his derivation of 'ought' from 'is', Searle asks: "what

[4] See also David Brink's discussion of naturalism in Brink 1989: 150ff.

bearing does all this have on moral philosophy?" His answer deserves to be quoted in full, with emphasis added:

> At least this much: It is often claimed that no ethical statement can ever follow from a set of statements of fact. The reason for this, it is alleged, is that ethical statements are a sub-class of evaluative statements, and no evaluative statements can ever follow from a set of statements of fact. The naturalistic fallacy as applied to ethics is just a special case of the general naturalistic fallacy. I have argued that the general claim that one cannot derive evaluative from descriptive statements is false. *I have not argued, or even considered, that specifically ethical or moral statements cannot be derived from statements of fact*. (Searle 1969: 187)

Clever as Searle's manoeuvre is, it nonetheless misrepresents the case that has traditionally been made by those who believe that there is an is/ought gap. Classical moral philosophers have not subsumed the ethical problem under the general speech act problem in order then to show that, since there is a gap concerning that general problem, the gap must extend to the particular ethical version of the problem. It has been enough to point out that there is no way to bridge the gap in the particular case of morality. Searle is rather alone in his interest in the *general* naturalistic fallacy.

In his famous article Searle states that he is going to show that the venerable view to the effect that 'ought' cannot be derived from 'is' is flawed by presenting a counterexample to this view. He then says:

> It is not of course to be supposed that a single counter-example can refute a philosophical thesis, but in the present instance if we can present a plausible counter-example and can in addition give some account or explanation of how and why it is a counter-example, and if we can further offer a theory to back up our counter-example—a theory which will generate an indefinite number of counter-examples—we may at least cast considerable light on the original thesis. (Searle 1964: 43)

The needed theory has been long in the making. *Speech Acts*, in which "How to Derive 'Ought' From 'Is'" was reprinted with minor modifications, was indeed the first step; but it is only with the publication of his two most recent major works—*The Construction of Social Reality* (Searle 1995) and *Rationality in Action* (Searle 2001)—that we have Searle's views on the ways in which speech acts contribute to the construction of social institutions. Indeed, Searle's philosophy has gained in depth and in comprehensiveness with these recent works—but then for this very reason the neglect of morality within his total system is all the more striking.

The world Searle investigates in these two books includes "the world of Supreme Court decisions and of the collapse of communism" (Searle 1995: 120); it includes marriages, money, government, and property rights, and discussions about altruism and egoism. And Searle expressly claims to be interested in the "basic ontology of social institutions"—of *all* social institutions. Yet still he avoids tackling head on the problem of the normativity of social institutions. In these recent works he has emphasized above all the importance of

promising. Promises, he tells us, are present in "all" or "virtually all" speech acts.[5] Marriages, money, property rights, and contracts all contain promises. And promises create obligations. But how?

Searle's answer is elegant and complex. As in Hart and Rawls, it revolves around a distinction between two types of rules, which in terms coined by Searle already in *Speech Acts*, are called 'regulative' and 'constitutive'. Regulative rules regulate forms of behavior that exist independently and antecedently (Searle 1969: 33). Constitutive rules—like Hart's secondary rules and Rawls's practice rules—create or define new forms of behaviour (Searle 1969: 34ff.). Thus when someone violates a constitutive rule, he *eo ipso* places himself outside of the institution to which the form of behavior defined by the rule belongs. Violating a regulative rule, in contrast, may give the violator a reputation for bad manners or reckless driving, but does not *ipso facto* place him outside of any institutions.

Rules of etiquette are regulative. It is perfectly intelligible to say that someone acted in ways that satisfy such rules even if that someone is unaware of the fact that he was satisfying such rules. Contrast this case with a community in which a group of 22 people gather together and move about while kicking a ball in more or less the same way as would a group of people playing football; they would not really be *playing football* unless a set of rules defining football was already in existence, and unless they knew about these rules. The latter constitute the very possibility of the activity of playing football.

4. ON SOFT NORMATIVITY

We wish to refer back to the notion of 'soft' normativity we introduced above, and suggest that it is an accurate corollary to Searle's (and Hart's and Rawls's) views that they characterize the normativity of social institutions as *soft*—of a piece with the normativity we find in games.[6] The constitutive-rules-based 'oughts' of games are, however, defeasible to a very high degree. Certainly when playing baseball one *ought* to go to first base after four bad pitches are thrown, but no one *ought* (in any interesting sense) to play baseball in the first place: any baseball player can walk off the field, can abandon the game, any time he wishes—though of course if a professional baseball player were to do this, he would probably lose his job.

This last remark reminds us that there are other types of oughts in games, in addition to those based in constitutive rules. For example: one ought to remain in the baseball field even after humiliating oneself by missing an easy catch. Players in a game of basketball can 'foul' their opponents several times in order

[5] Searle 2001: 181. In the Spanish version of *Rationality in Action*, which was published earlier than the English version, and which is virtually identical to the latter, it is stated that "all" (not merely "virtually all") speech acts contain an element of promising.

[6] See, for example, the section entitled "Games and Institutional Reality" in Searle 1995: 66–71.

to prevent them from scoring, but they ought not to stab or shoot their opponents. One might try to explain the latter sorts of normativity by appealing to the fact that, for example, by embarking on a game of baseball one has in a sense promised not to leave the playing field after making silly mistakes, or that all human beings have in a sense promised not to kill in general and that this promise covers also one's behavior when playing basketball. This strategy, however, robs terms like 'promise' and 'contract' of their customary meanings. Moreover, at least some of the mentioned obligations seem not to be obligations of the sort which one could acquire by means of promises or contracts.

Legal and sociopolitical institutions, similarly, give rise to obligations not only of the constitutive-rule-based sort but also of other sorts. According to Hart, for example, Nazi laws are genuine laws in the constitutive-rule-based sense—but they are at the same time laws that one *should not follow*. Famously, Hart charged that Gustav Radbruch's abandonment of positivism in the post-Nazi era was the result of his "half-digested" understanding of "the spiritual message of liberalism" (Hart 1997: 31), whereby Radbruch had failed to see that even the staunchest positivists share the "conviction that if laws reached a certain degree of iniquity then there would be a plain moral obligation to resist them and withhold obedience" (Hart 1997: 30). Presumably, Hart would agree that this "plain moral obligation" is not a game-related obligation. Significantly however he does not discuss what type of obligation it might in fact be, and this is the sort of discussion that Searle avoids as well.

To see that something is wrong with the identification of all normativity with the normativity of games, we can appeal to Wittgenstein's remarks on the nature of games in the context of his treatment of the notion of family resemblance in the *Philosophical Investigations* (Wittgenstein 1953: § 66ff.). According to Wittgenstein, no definition formulated in terms of necessary and sufficient conditions can apply to all games. In light of Searle's views on what we have called soft-normativity, however, it is tempting to suggest that *being created by a set of constitutive rules* would amount, precisely, to the sought-for definition. Whenever you are in the presence of an entity which exists in virtue of constitutive rules, you are *eo ipso* in the presence of a game, and *vice versa*.

This resolution of Wittgenstein's puzzle comes at a price, however—for it forces an over-large scope upon the notion of game, which now turns out to include sociopolitical institutions like promising, punishment, marriage, and government. Note that if Searle and our authors are correct, then this would in no way count against it. For whenever Searle, Hart, and Rawls wish to explain the normativity of such institutions they do indeed invariably end up talking about the way in which swinging at the third strike entails that you ought to leave the baseball field. This move, if we are right, is not a matter of happenstance. Rather, it reveals that Searle and our other authors have maneuvered themselves into a position where they do not have the tools to draw the distinction between games and sociopolitical institutions.

Part of the compelling force of the "why should I play this game anyway?" objection to the thesis that all normativity is soft normativity turns on the *conventional* character of games. For even if there existed something like a game of life (Ralls 1966), the skeptic could still ask a reformulated question: "Why should we not alter its rules?" Constitutive rules, after all, are not merely to a high degree defeasible, they are also easy to change: at some point in their history virtually all games had rules different from those they have today.

Hospitals are, by definition, places where physicians and nurses ought to care for patients. If there were a hospital in which nurses and physicians systematically harmed their patients, then we would not be content simply to claim that this institution is, by definition, no longer a hospital and leave it at that. Obviously, we would claim that the physicians and nurses *ought* to care for their patients, and that this obligation is not merely the result of the constitutive rules governing hospitals and medical professions.

Other sorts of normative claims: that murder is wrong, or that it is appropriate for wrongdoers to apologize, that purely accidental (nonnegligent) wrongdoing is not blameworthy, etc., are not only not easily defeated, they are also—and even more conspicuously—not *easily changed*. Whereas the number of fouls a basketball player can 'legally' make in the course of a game can at any time be changed, the prohibition against stabbing his opponents is not likely to change at all.

5. ON ROBUST NORMATIVITY

In spite of the fact that Hart cares about legal institutions, that Rawls cares about political institutions, and that Searle cares about social institutions, they, and the legions who have followed in their footsteps, have all avoided addressing the challenge encapsulated in Raphael's charge of triviality—the challenge that their respective logical analyses tell us "virtually nothing" about the normativity that is interwoven in the fabric of institutions of the various non-game-related types referred to in the foregoing. For aside from the sorts of normative demands to which secondary rules, practice rules, and constitutive rules give rise, there exist in law, politics and society other types of demands which are similarly non-conventional. Each of us believes that he has an obligation to respect other human beings; each of us believes that he has an obligation to apologize to those we have wronged. These beliefs do not depend for their existence on any promises we have made, and neither do the associated obligations. Each of us believes, similarly, that intentional wrongdoing ought to be blamed more severely than unintentional wrongdoing; each of us believes that wrongdoers ought to be blamed. These views, again, are clearly normative, and they do not depend for their existence on any promises or contracts.

We believe (with Searle) that a minimum dose of realism is necessary for any sane philosophy (Searle 1955: xiii and Chap. 7 passim). Moreover (also with Searle) we understand realism as an ontological thesis: "realism . . . is not a theory

of truth, it is not a theory of knowledge, and it is not a theory of language", and Searle himself has recently admitted that "if one insists on a pigeonhole, one could say that realism is an ontological theory: it says that there exists a reality totally independent of our representations" (Searle 1995: 155). Yet Searle avoids the discussion of realism as pertains to the dimension of moral normativity. Indeed at crucial junctures Searle shuns ontology entirely. Thus, in the introduction to *Intentionality*, Searle praises the methodological advantages of approaching the analysis of mental phenomena from the perspective of intentionality in the following terms: "one advantage to this approach, by no means a minor one, is that it enables us to distinguish clearly between the logical properties of Intentional states and their ontological status; indeed, on this account, the question concerning the logical nature of Intentionality is not an ontological problem at all" (Searle 1983: 14). Searle believes that a logical approach to intentional phenomena can allow him to repeat the success of his logical analyses of obligation in "How to Derive 'Ought' From 'Is'". In *Rationality in Action*, still more recently, Searle has suggested that we can enjoy some of these same benefits by providing a logical account of notions such as self, freedom, and responsibility; that is, that we can talk about these notions without having to deal with the embarrassing ontological questions that had affected their treatment in earlier times.

In tandem with the shunning of realism as it pertains to ethics, of course, goes the shunning of precisely those types of normativity which are not soft. We can morally criticize Nazi institutions; we can accept that promises do not obligate if what is promised is itself immoral, we can expect—and sometimes accept—apologies when we are wronged. But to tackle theoretically these genuinely moral dimensions of social institutions we need to go beyond merely tracing the *logical* paths connecting speech acts, institutions, and consequent obligations.

To see what more is needed, let us pay closer attention to the normativity that is associated with our intentional states. This is, we suggest, more fundamental than the normativity associated with speech acts. Paradoxically, perhaps, we find some support for this thesis in Searle's own philosophy, above all at the outset of *Intentionality*, where he writes:

> A basic assumption behind my approach to problems of language is that the philosophy of language is a branch of the philosophy of mind. The capacity of speech acts to represent objects and states of affairs in the world is an extension of the more biologically fundamental capacities of the mind (or brain) to relate the organism to the world by way of such mental states as belief and desire, and especially through action and perception. (Searle 1983: vii)

We fully agree with Searle's assumption regarding the priority, biological and otherwise, of intentional states over speech acts, though we wish he had done more to exploit this insight in his recent work on social reality. We say this not because we deny the general value of speech act theory. Our claim is, precisely, that its value should not be overestimated, and that in particular the concern

with practice, secondary or constitutive rules which we find in Hart, Rawls, and Searle has already yielded all the fruits that it is worth collecting. Constitutive rules do give rise to claims which exhibit some sort of normative force, but they are not nearly the end of the story of normativity.

A no less vital chapter in this story deals with a different sort of normative force—that which derives from intentional states. What happens if we focus not on speech acts in giving an account of legal and sociopolitical institutions, but rather on the intentional states which underlie them? Speech acts are in their entirety contingent, first in the sense that one can choose to perform them or not, and secondly in the sense that they need not have existed at all. It is indeed hard to imagine a society in which something resembling promising did not exist, but given Searle's analysis of speech acts as products of constitutive rules such a society is not impossible. Some intentional states are not contingent in either of these two senses.

By Searle's own admission, the intentional state of intending is crucially important for promising: if you promise to X then you must intend to X. But where the skeptic can raise the concern as to why he should play the "promising game", there is no parallel concern in relation to the phenomenon of intending. This is because what happens when one intends is not the result of applying human conventions. And while it is hard to imagine a society which did not develop a practice more or less identical to promising as we know it, it is downright impossible to think of human beings who do not intend.

6. THE NORMATIVITY OF INTENTIONS

The structure of intending is rather complicated, and it is the subject of a very extensive debate. Virtually all participants to this debate, however, agree that intending is subject to more stringent rational considerations than are other intentional states. We could have contradictory desires without thereby being irrational, but for one who has contradictory intentions, i.e., one who intends to have a meal and not to have a meal simultaneously, a charge of irrationality will almost always succeed.

Intentions are connected to actions in ways in which mere desires are not. You can only intend to do things that you believe are up to you, and when you intend to do X, then if your intention is to be fulfilled X must come about "in the right way", i.e., in the way the intending agent foresees that X should come about. These two features of intentions not only distinguish them from related phenomena like desires or wishes, but also explain why intending to do X *commits* us in certain ways. If you form an intention today to visit friends tomorrow, forming that intention somehow settles your deliberative process; you are now committed to visit your friends tomorrow. This does not mean that you cannot possibly change your mind: the commitments that arise from intentions are defeasible, just like those that arise from promising. But there is nonetheless a stark contrast between the way commitments arise from intentions and the way

they arise out of speech acts. Forming intentions is itself optional, but once they are formed, the commitments which follow from them do not arise in virtue of constitutive rules imposed, as it were, from without; rather, they arise solely in virtue of the intrinsic nature of the intentions themselves.

Imagine that you communicate to your students your intention to tidy up your office. Month after month, indeed semester after semester, students visit you and see that you have done nothing of the sort: your office is ever messier. Regularly they ask: "What about your intention to tidy up your office?", to which you reply: "It is still there". Nothing has prevented you from carrying out your intention; you simply have not done so. After some time your students will be justified in believing that either you do not have the intention to clean your office at all (that you have been lying, or confused as to what it is to have an intention), or that, if you do have the intention, then you are somehow irrational.

If, in contrast, you had merely wished or desired to tidy up your office, then your inaction would be evidence neither of irrationality nor of dissimulation or confusion. This is not to say that there are no constraints on what we can desire. Your desire that a fairy godmother should materialize and tidy up your office would properly be counted as a sign of irrationality, just as would the corresponding cognitive state of believing that a fairy godmother is on her way to do the job. Such constraints are, however, more stringent in the case of intentions than in the case of other mental states.

What does this tight connection between intentions and rationality tell us about normativity? We note, first, that acting goes hand in hand with the possibility of blame. Acting *intendedly* means acting in such a way that one is committed to acting in precisely the way one acts. If, therefore, what one does intendedly is a bad thing, then one is clearly at least not less blameworthy for doing it than if one had done it unintendedly. This normative principle, namely that intended wrongdoing ought to be blamed more severely than unintended wrongdoing, is rooted in the intrinsic nature of the phenomenon of intending, and not related to conventional constitutive rules.

The same reasoning explains why doing bad things on the basis of a commitment to those bad things is evaluated differently from doing those same bad things in the absence of such commitment. Regardless of whatever general character traits one possesses, being committed to a bad thing makes one, *ceteris paribus*, no less blameworthy than if one does this bad thing without being so committed. This normative principle follows, again, from the intrinsic nature of intentions, and it is quite unlike those normative claims that follow from conventional constitutive rules.

In order to drill home this point it is profitable to take a look at Christine Korsgaard's *Locke Lectures* which open with a statement to the effect that "Human beings are condemned to choice and action".[7] This statement is part of

[7] Korsgaard's lectures are available at: http://www.people.fas.harvard.edu/~korsgaar/#Locke%-20Lectures, from which all quotations here are taken.

Korsgaard's ambitious project of showing how "we human beings constitute our own personal or practical identities—and at the same time our own agency—through action itself. We make ourselves the authors of our actions, by the way that we act". Clearly, when Korsgaard says "through action" she means "through *intentional* action". Indeed, she points out that

> to call a movement a twitch, or a slip, is at once to deny that it is an action and to assign it to some part of you that is less than the whole: the twitch to your eyebrow, or the slip, more problematically, to your tongue. For a movement to be my action, for it to be expressive of *myself* in the way that an action must be, it must result from my entire nature working as an integrated whole.

Twitches are not actions because they do not express our selfhood in any meaningful way. Slips are more problematic precisely because slips of the tongue can in some cases be actions, though except in rare and contrived cases, unintended actions. It is however precisely intentions which constitute our selfhood; and it is intentions, too, which constitute the principal grounds for blameworthiness of our actions.

According to Korsgaard "there is no *you* prior to your choices and actions, because your identity is in a quite literal way *constituted* by your choices and actions". And then Korsgaard adds:

> The identity of a person, of an agent, is not the same as the identity of the human animal on which the person normally supervenes. Human beings differ from the other animals in an important way. Because we are self-conscious, and choose our actions deliberately, we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical identity—that the other animals lack. It is this sort of identity that makes sense of our practice of holding people responsible, and of the kinds of personal relationships that depend on that practice.

What distinguishes our identity from that of animals is, in other words, our capacity to act intentionally; our capacity to act intentionally is of course wholly dependent upon our more fundamental capacity to form intentions. And, ultimately, it is these capacities to form intentions and to carry them through which make sense, not only of the practice of "holding people responsible", but of other normative phenomena such as the apportioning of praise and blame.

The intrinsic nature of intentions gives rise in this way to important normative principles. As R. Jay Wallace puts it, the intentional actions, and ultimately the intentions, of morally responsible people "are thought to reflect specially on them as agents, opening them to a kind of moral appraisal that does more than record a causal connection between them and the consequences of their actions." (Wallace 1996: 52) In order for agents to be the subjects of judgments of praise and blame it is necessary that agents be autonomous beings, and the role played by intended action in the constitution of this autonomy is a rich and still untapped source of insight.

7. CONCLUSION

The normative principle to the effect that *to bring about an evil outcome intentionally is never less blameworthy than to bring it about unintentionally* is in no sense analogous to the principle that in chess bishops move diagonally. For the rules of chess are open to deliberation. One could choose or invite others not to play chess; one could propose that chess be played differently; and thus one could affect the way in which the soft normativity of chess plays itself out in reality. One cannot, in contrast, refuse to accept or propose adjustments in the normativity of intending.

There is, then, normativity in intentional states themselves, before they give rise to speech acts. But this is still not the end of the story as concerns the manifold varieties of normativity. Thus we still cannot explain why murder is wrong. And we still do not yet have the means to do justice to those features of normativity turning on virtue, character traits, and like phenomena, which are the fare of neo-Aristotelian ethics. Our discussion of intentions is meant simply to establish that there are provinces in the kingdom of normativity that have nothing to do with conventional rules. Surely some of these provinces affect the structure of social ontology: it is rather hard to accept that all social reality is a matter of soft normativity, yet it is a view of this sort with which Searle's otherwise groundbreaking work on social ontology is still stuck.

REFERENCES

Ayer, A.J. (1952). *Language, Truth and Logic*. New York: Dover.
Bratman, M. (1999). *Faces of Intention*. Cambridge: Cambridge University Press.
Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
Gilbert, M. (1989). *On Social Facts*. London: Routledge.
Hare, R.M. (1963). *Freedom and Reason*. Oxford: Oxford University Press.
Hart, H.L.A. (1994). *The Concept of Law*. Second Edition. Oxford: Oxford University Press.
Hart, H.L.A. (1997). Positivism and the separation of law and morals. In R. Dworkin (ed.), *The Philosophy of Law*. Oxford: Oxford University Press.
Hume, D. (1888). *A Treatise Concerning Human Understanding*. Ed. L.A. Selby-Bigge. Oxford: Clarendon Press.
Moore, G.E. (1959). *Principia Ethica*. Cambridge: Cambridge University Press.
Popper, K. (1966). *The Open Society and its Enemies*, Vol 1, rev. edn. Princeton, NJ: Princeton University Press.
Prior, A.N. (1949). *Logic and the Basis of Ethics*. Oxford: Oxford University Press.
Ralls, A. (1966). The game of life. *Philosophical Quarterly*, 62: 23–34.
Raphael, D.D. (1990). *Problems of Political Philosophy*. Second Edition. Atlantic Highlands, NJ: Humanities Press.
Rawls, J. (1955). Two concepts of rules. *The Philosophical Review*, 64: 3–32.
Rawls, J. (1999a). *A Theory of Justice*. rev. edn. Cambridge, MA: Harvard University Press.
Rawls, J. (1999b). *Collected Papers*. Ed. S. Freeman. Cambridge, MA: Harvard University Press.
Searle, J.R. (1964). How to derive 'ought' from 'is'. *The Philosophical Review*, 73: 43–58.
Searle, J.R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.
Searle, J.R. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.
Searle, J.R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.
Smith, B. (2003). The ontology of social reality. *American Journal of Economics and Sociology*, 62: 285–299.
Tuomela, R. (1995). *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford, CA: Stanford University Press.
Tuomela, R. (2002). *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge: Cambridge University Press.
Wallace, R.J. (1996). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
Wittgenstein, L. (1953). *Philosophical Investigations*, transl. G.E.M. Anscombe. Oxford: Blackwell.
Wittgenstein, L. (1965). A lecture on ethics. *The Philosophical Review*, 74: 3–12.
Zaibert, L. (2003). Punishment, justifications, and institutions. *Studies in Law, Politics, and Society*, 30: 51–83.
Zaibert, L. (2005). *Five Ways Patricia Can Kill Her Husband: A Theory of Intentionality and Blame*. Chicago: Open Court.

A BEHAVIOURAL CRITIQUE OF SEARLE'S THEORY OF INSTITUTIONS

Ignacio Sánchez-Cuenca

1. INTRODUCTION

In *The Construction of Social Reality* (Searle 1995) and in several subsequent works, John Searle has developed an extremely elegant and original theory about institutions. Moreover, that book has reinvigorated an old discussion on the ontology of the social world that languished in obscure sociological treatises.

Searle likes to present his theory of institutional facts as a purely analytical inquiry about the logical structure of institutional reality. Precisely because his task is philosophical in nature, he feels that he can proceed without paying much attention to positive theories in the social sciences about the emergence and functioning of institutions. Actual institutional facts (the existence of money, people getting married, etc.) appear as illustrations of a highly abstract theory, not as objects of empirical scrutiny. His goal is elucidatory, not explanatory. He wants to make clear how institutions can be constructed, how institutional reality is possible within the physical world, not under what conditions institutions survive, or why different societies have different institutional orders, or how institutions evolve and change.

However, this sharp separation between philosophical and positive theories about institutions is not always possible or desirable. I want to argue in this chapter that a closer look at positive theories about institutions and rules shows that Searle's philosophical analysis is not necessarily wrong, but one-sided and incomplete. This is due, in my opinion, to the fact that Searle neglects certain kinds of institutional facts (those that have distributional consequences and entail conflicts) and privileges some others (such as conventions that serve as coordinating devices). It appears sometimes as if Searle were trying to accommodate all institutional reality into the categories that are useful to understand one fundamental but very peculiar institution, language. As I explain later, it is not by chance that Searle's favourite example of an institutional fact is the existence of money. Money, as a mean of exchange, shares with language some crucial features that can by no means be extrapolated to institutions in general.

Before going into details, I want to present a sketch of the argument. I start with two separate claims. The first one is about the distinction between regulative

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 175–189.
© 2007 Springer.

and constitutive rules. The logical formulae that Searle proposes in order to analyse constitutive rules do not ensure that the activities that the rules mention are actually constituted by them. I present several examples in which it becomes obvious that the dependence of the activity on the rule is a matter of degree. Some activities are fully dependent on the rule, but some others are only partially so. The second claim is that attitudes of endorsement are not necessary for constitutive rules to work. I show that constitutive rules can be imposed under the threat of violence, and I focus on the case of dictatorships to make it clear that their survival has little to do with the attitudes of the population.

These two claims can be combined. I argue next that in order to understand how constitutive rules work in practice, we need to understand why people comply (that is, act in accordance) with them. Compliance is a problem, in my view, when the dependence of the activity on the rule is partial (claim 1), that is, when there are institutional and non-institutional ways to achieve similar outcomes. And, compliance can be secured regardless of the attitudes of the people complying (claim 2). The problem of compliance does not arise for certain kinds of constitutive rules such as those that define the grammar of a language or those that define money. These are examples where the activity is fully constituted by the rule.

People may comply with constitutive rules for many different reasons: out of prudence, out of habit, or simply because of the acceptance of, or agreement with, the rule. The important point here is that the existence of attitudes of a certain kind is not a necessary condition for compliance. We can abstract the various reasons behind compliance and establish simply that constitutive rules work when they are in equilibrium, that is, when no one has reasons not to comply with them. In the final part of the chapter, I explore the idea of institutions as equilibria, and how it compares with Searle's theory of institutions.

Ultimately, Searle's philosophical analysis is not empirically neutral. The whole project of a purely philosophical analysis of institutions is misconceived. It cannot avoid some implications about the beliefs and motivations of those who have to make choices given the rules. It seems to me that attitudes and collective intentionality are not necessary to understand how constitutive rules are created or why people comply with them. My general point is that to get a right picture of institutions we need some combination of philosophical insight and social scientific knowledge about how institutions work.

2. HOW MUCH IS CONSTITUTED BY CONSTITUTIVE RULES?

When Searle presented his distinction between regulative and constitutive rules in his classical article "How to Derive 'Ought' From 'Is'" (Searle 1964), he illustrated it with two extreme examples: the rules that regulate polite table manners and the rules that establish the game of chess. The former regulate eating, the latter create the possibility of playing chess. The activity of playing chess, like the activity of producing grammatical sentences, is made possible by constitutive rules (the rules of chess, the rules of grammar). In the realm of politics, similar examples can be

found: no one can be a member of a democratic parliament without an electoral system (a set of constitutive rules that transforms votes into seats).

It is worth noting that the two logical formulae that according to Searle reflect the structure of constitutive rules have only an indirect bearing on the issue of whether the activity is made possible by the rule or not. In the more general formulation, a constitutive rule has this structure:

X counts as Y in C.

X is the entity on which the status-function, Y, is imposed. Searle offers a more precise formula that makes the structure of the status-function clearer. The term Y establishes a set of rights and obligations, following this structure:

We accept (S has power (S does A)).

The two formulae are equivalent. It may be said that the person who obtains more votes (X) counts as the president of the United States (Y) in the presidential elections (C), or alternatively, that the president of the United States has the power, for instance, to veto laws passed by Congress. The set of rights and obligations of the president is an exhaustive description of the status of being president of the United States (of what 'counting as Y' means).

In my view, neither of these logical expressions guarantees that the activity is made possible by the rule. This can be shown through some examples employed by Searle himself. Let us take first the case of war. Searle points out that wars have an institutional side, in virtue of which the parties involved in the war undertake some obligations and rights about the treatment of prisoners, the protection of civilians, and so on and so forth. In the United States, a war is fully institutional when it is legally declared. An armed attack (X) counts as war (Y) in an international conflict (C) when the countries declare war.

Searle openly admits that war quite often proceeds non-institutionally, that is, without a formal declaration, and without paying much respect for the obligations and rights that derive from international treaties on the conduct of war. The conclusion in the *Construction of Social Reality* is worth quoting: "'War' thus oscillates between naming a type of large-scale social fact and a type of institutional fact" (Searle 1995: 89). Yet, if the same phenomenon is sometimes a social (non-institutional) fact and sometimes an institutional fact, it follows that the constitutive rules that make war an institutional fact do not make the activity as such possible. The constitutive rules create in this case some rights and obligations, allowing us to distinguish between legal (or institutional) and illegal (or non-institutional) wars, but this is not sufficient to conclude that the constitutive rule makes war possible. Unless, of course, one is willing to say that the constitutive rule makes legal wars (as opposed to non-legal wars) possible. This would be true merely by definition: every constitutive rule would make possible the activity as defined by the rule. But this is an interpretation that Searle rules out in *Speech Acts* (1969: 35), for good reasons. This understanding of constitutive rules has little to do with the sense in which chess rules make possible an activity that depends entirely on these rules.

Let us take now the case of marriage. Marriage is a form of partnership. But the partnership formed by a couple who want to live together can be achieved non-institutionally, through cohabitation, out of a personal commitment (out of love, or out of interest and convenience). Marriage, as such, does not make living together possible. Yet, marriage institutionalizes cohabitation, and as Searle (1995: 81) points out, it makes it possible that people can remain married even if they do not live together. More generally, it creates new rights (for instance, about adoption, about property arrangements). Part of the activity that falls under the institution of marriage is constituted by the institution, but another part can be carried out non-institutionally.

Marriage is a particular case of a contract. The point can be extended to contracts more generally. Contracts are constitutive rules that define the rights and obligations of the parties who enter into some exchange (a worker and a firm, a tenant and a landlord, etc.) Yet, the exchange is not made possible by the contract. The exchange could take place informally, based on mutual trust among the parties. Contracts exist to reinforce trust thanks to the potential intervention of a third party, but they do not make the exchange itself possible. They make exchanges easier through the creation of sets of rights and obligations that can be enforced by a third party, but they do not constitute the exchange as such.

What the war and marriage examples have in common is that in both cases the constitutive rules impose status functions that generate what Searle calls deontic powers (institutional rights and obligations). But it is one thing to say that constitutive rules generate deontic powers and a different thing to conclude that constitutive rules make possible an activity that without the rules would not be possible. In fact, the formula "We accept (S has power (S does A))" seems closer to the idea of deontic powers than to the idea of constituting anything.

The original distinction between regulative and constitutive rules becomes somewhat fuzzy now. If we put aside momentarily legal claims, we could say that rules about war behaviour are regulative rather than constitutive, since they regulate warfare behaviour (e.g. civilians are not to be killed, prisoners are not to be tortured). These rules make a distinction between civilized and uncivilized wars, just as table-regulative rules distinguish between polite and impolite eating habits. If we say that rules of polite table behaviour regulate eating, but do not make eating possible, we seem forced to conclude that rules about war establish how wars should be conducted, but do not make war possible.

The difference between rules of polite table behaviour and war rules is that the latter are part of the international legal system and therefore give rise to legal claims about the rights and obligations of the parties. Rules of polite table manners do not bring about deontic powers simply because they are not part of a legal system. But suppose that a legal system covered table manners. There would then be some rules that would define the institution of polite table behaviour,

so that accepting to eat in a table with someone else would entail undertaking some rights and obligations.[1]

If the incorporation of rules into the legal system is enough to make them constitutive, then constitutive rules cannot be defined in terms of making certain kinds of behaviour possible. What makes legal rules different from purely regulative rules is simply that they create rights and obligations. These rights and obligations are sometimes, but not always, enabling, in the sense that chess rules are enabling with regard to playing chess.

To put it briefly: certain activities are made possible by rules, but neither Searle's examples nor Searle's analysis of the logical structure of constitutive rules capture the constitutiveness of these rules. Rather, his analysis of constitutive rules is ultimately about deontic powers.

One might think that this amendment to Searle's theory is inconsequential, since after all what he is interested in providing is an analysis of how deontic powers are created. He could therefore keep everything intact except the original intuition he used to justify the distinction between the two kinds of rules. However, I am going to argue that this lack of clarity about the extent to which the activity is constituted by the rule has serious consequences for the whole analysis of deontic powers.

3. ATTITUDES AND DEONTIC POWERS

The formula "We accept (S has power (S does A))" is an abbreviation of a more general one (Searle 1995: 111):

> We collectively accept, acknowledge, recognize, go along with, etc. that (S has power (S does A)).

Searle, by using the "etcetera" rider, does not fully specify the set of attitudes that may enter into the formula, but the idea is that collective attitudes of endorsement are crucial here. Something similar holds for the "X counts as Y in C" formula:

> the 'counts as' formula captures the essential feature of status functions, namely, that the function cannot be performed without a certain set of attitudes on the part of the participants in the institution. (Searle 2003: 27)

These attitudes are required because Searle argues that constitutive rules are generated through collective intentionality. Collective intentionality is required

---

[1] In *Speech Acts*, Searle warns against distorted conceptions of regulative rules. For instance, one could feel tempted to say that "Non-wearing of ties at dinner counts as wrong office behaviour" (Searle 1969: 36). However, he blocks this move by saying that in this example the "counts as" is a term of appraisal ("wrong"), not a specification of what something is counted as. On the other hand, in *The Construction of Social Reality* (1995: 50), he explicitly says that the imposition of legal sanctions for the breakdown of regulative rules generates constitutive rules (the imposition of a status function to the person who violates the regulative rule).

to impose a status function on physical entities, so that these physical entities can perform the functions we assign to them simply because we agree on those functions. The recurrent example is money: certain bits of paper count as money because we impose a status function on them (the function of serving as a means of exchange). This is why Searle takes acceptance and "collective agreement" to be equivalent. An institution works when people agree with the assignment of functions contained in the institutional rules the institution is made of.

Searle reserves a large role for collective intentionality in his theory. There is collective intentionality when the agent does something as part of a collective endeavour in which the participation of others is necessary. According to Searle, even two prizefighters are engaging in collective intentionality: they understand that the activity of each is part of a collective contest. Searle equates "social facts" with collective intentionality, institutional facts being a subclass of social facts. Only collective endeavours are social facts. Non-organized crime, for example, is excluded. Sociologists would not be happy with this way of talking, but my main concern here is not terminological. The issue I want to discuss now is rather whether institutional facts always require the kinds of attitudes that Searle associates with collective intentionality, from "going along with" to "full endorsement". Let me be precise: I do not question that Searle's description in terms of such attitudes fits many relevant cases; my point is that there are many other equally relevant cases that cannot be reduced to this description.

To begin with, there is some ambiguity about who belongs to the "we" of the formula, as some of Searle's commentators have already noted. Barry Smith (2003: 27) mentions contested borders, where each national group has different beliefs about what the border separating two countries is. Likewise, Leo Zaibert (2003: 61) mentions the possibility of contested government, where some group in society (e.g. republican Catholics in Northern Ireland) does not recognize the institutionally constituted authority (e.g. the British government).

Searle is fully aware of these conflicts, but he thinks that they do not pose a threat to his theory. In his view, these are problems that have to be solved empirically, but do not represent a challenge to the philosophical analysis of institutional facts (Searle 2003: 307). Yet, the way in which these disputes are solved could have theoretical significance. If some group imposes its point of view on the other group through the sheer use of force, the resulting deontic powers will exist despite the opposition by one of the groups. It is hard to discern the role of attitudes of collective agreement under these circumstances.

For the sake of simplicity, think of this imaginary case: there is a society formed by three people, A, B and C. Their economy is very simple: they survive cultivating land. A is much stronger than B and C, to the point that B and C are certain that they would lose their life in a fight with A. One day A communicates this message to B and C: "From this moment on, I declare that I am the ruler and that you, B and C, are the ruled. The ruler has the right to impose a tax on the ruled. By virtue of this power, I decree that you, the ruled, will give half of your crops to me, the ruler." B and C do not participate at all in this declaration.

They do not recognize any authority or legitimacy to A, and they consider the announced tax to be mere confiscation, since they do not receive anything in return. Yet, both B and C deliver half of their crops to A simply out of fear. They do not want to fight with the "ruler" and therefore they comply.

Given the previous definitions, there seems to be no collective intentionality here. Therefore, we cannot refer to this outcome as an institutional or social fact. But, on the other hand, there is a constitutive rule, and there is compliance with this rule. The only element that is missing is any collective intentionality in virtue of which B and C "accept" or "agree on" the deontic powers that A has created for himself. Searle insists that acceptance does not imply approval. Might we say then that B and C accept A's deontic powers? Perhaps "acceptance" is too strong. However, "going along with" could serve. B and C go along with A's deontic powers because they are afraid of the consequences of not complying. What this means in the present context is merely that B and C understand A's deontic powers and comply out of prudence. To put it another way, B and C are aware that A's deontic powers are backed by force and conclude that compliance pays off. I do not see any role for shared attitudes in this story.

The story of this three-person society is not very realistic, but it draws our attention to the case of dictatorships in which a minority imposes constitutive rules defining deontic powers on the majority of the population. Although it is true that those in the minority engage in some form of collective intentionality when they agree on the constitutive rules that they are going to impose on the whole society, the fact is that those in the majority who fully reject these rules may comply with them without having any kind of collective attitude or intentionality. The attitudes of the majority are quite irrelevant. The only requirement for the rules to work is that those in the oppressed majority are aware of these rules and of the consequences of not complying with them. But being aware of the rules is not an adoption of an attitude towards them, let alone a form of acceptance.

Searle (1995: 91) rightly points out that police forces and armies are also systems of status-functions. However, this observation does not invalidate my claim. On the one hand, armies, despite being institutions, can act against other institutions (e.g. when armies participate in coups). On the other hand, once the police or the army possesses weapons and are willing to use them, they possess some brute power or brute force that is independent of the rest of society: what matters is only that members of the police or the army obey the commands they receive from within.

Besides, what matters regarding the threat power of police or of the army is not the attitudes of the population about who has the right to use force, but the expectations of the population about the possible results of violence. If people expect that the rulers are going to use the police or the armed forces to repress the population, it is irrelevant that the repressed population accepts or goes along with the institutional organization of brute force. It is expectations, not attitudes, what we have to analyse here.

What is ironic about the example of dictatorship is that the imposed constitutive rules work best precisely when there is no collective intentionality in the oppressed majority. Dictatorships survive when they disorganize society, when individuals feel like social atoms: that is, when people make the decision whether to comply or not individually, without much information about what others think. Searle's comments on the collapse of the Communist countries are instructive here. He takes this historical experience to prove his point about the importance of attitudes for the maintenance of the institutional system. Suddenly and unexpectedly, Communist regimes "collapsed when the system of status-functions was no longer accepted" (Searle 1995: 92). But this is not entirely accurate. A couple of comments are in order here.

On the one hand, it is not clear what Searle means when he says that people no longer accepted status-functions. In the same page of the previous quotation, he writes, referring to the pre-1989 period, that "most people did not think that the system of status-functions was morally acceptable, much less socially desirable." If people complied with the regime, it was, he admits, because of police action backed by the Soviet threat of invasion. The only difference between the pre-1989 disapproval and the 1989 disapproval is that in 1989, people rebelled against the regime due to the removal of the Soviet threat. The attitudes can be held constant, there is something else that changes. The key factor here is not people's attitudes, but the momentous decision of not complying with the rules.[2]

Research on the fall of the German Democratic Republic has shown that rebellion was not due to a change in attitudes. What the literature reveals is that the maintenance of the social and institutional order was not based on individual attitudes, but on the absence of collective action. Both Timor Kuran (1991) and Susanne Lohmann (1995) have argued that the key variable here was the degree to which citizens were aware of each others' beliefs. Everyone was fed up with the system, but each was uncertain about how extended this feeling was in society, precisely due to the social and political isolation of individuals under communist rule. Both Kuran and Lohmann assume that there was a generalized falsification of preferences in the public realm that averted the emergence of collective action. The first manifestations of resistance produced a snowball effect. They made it clear to some people that there were others willing to change things. Collective action revealed how widespread discontent with the regime was.[3]

The lesson to be drawn from the eastern European revolutions is not that attitudes matter to explain the endurance of constitutive rules (or their sudden collapse). Rather, what really matters is the capacity of those who want to

[2] For an ingenious formal model on the role of the Soviet threat of invasion, see Zielinsky (1995).

[3] The differences between Kuran's and Lohmann's models are small: whereas Kuran assumes that the protests of radicals triggered the snowball effect, Lohmann argues that people had discounted the rebellion of the radicals: what ignited the movement were the first signals that moderates were also participating in the movement.

replace these rules to organize collectively. It is a matter of collective action, of forging a movement of people against the rules. In the absence of collective action, constitutive rules backed by a repressive system can survive independently of the attitudes of the people.

4. COMPLIANCE

In the previous section, I have argued that shared attitudes are not always necessary for the creation and enforcement of constitutive rules. Constitutive rules can be imposed on people and people may comply with them out of fear. Although Searle says very little about compliance in his writings, it is a key issue if we want to understand how institutions are created and maintained.

The reason Searle has so little to say about compliance is that he tends to approach social reality through the lens of language and games, but these are two very special institutions, if only because people do not have reasons not to comply in these cases. Language is a coordinating device and no one has incentives to deviate in the use of grammar rules. If someone decides to break the rules of language, that person will soon find out that communication with others becomes impossible.

Games are not simply coordinating schemes, since they generate winners and losers. But games are, so to speak, isolated spheres of activity: winners and losers are such only with regard to the game, without further consequences. If somebody is unhappy with the prospect of losing the game, she can abandon it. What she cannot do is break the rules to win the game.[4]

It is not by chance that Searle has privileged the example of money. Money has many features in common with language. Both are forms of coordination in which individual deviations do not make sense. Searle has projected the ontology of language into social and institutional reality without giving due attention to the significance of compliance. In fact, most political activity bears little resemblance to games, language or money. A person's quality of life may depend crucially on whether the person is a winner or a loser in the political realm. Because of the consequences of politics in terms of welfare, the issue of compliance with the rules is not irrelevant. In politics, therefore, we observe breakdowns of the institutional order that are absent in the cases of conventions such as language. There are coups, revolutions, secessions, invasions and institutional crises that are not solved according to the rules.

What is the difference between language on the one hand and politics on the other? In order to answer this question, we have to bring back constitutive rules and the degree of dependence of the activity on the rule. In Section 1, we saw

[4] Note that the possibility of cheating does not invalidate my claim. When someone cheats in a game, she is deceiving the other players, but she is not really breaking the rules, for the game proceeds according to the rules.

some extreme cases. Some constitutive rules have a constitutive power close to zero: wars can occur regardless of the constitutive rules that define the rights and obligations of the parties. Some other rules have full constitutive power: chess can only be played according to chess rules. In between we have many intermediate cases.

Suppose that the losing candidate in a presidential election, unhappy with the electoral results, launches a coup and with the support of the army becomes the new president. Of course, this is not a thought experiment. It has occurred quite frequently in history. The new ruler is not an institutional one. He is, strictly speaking, an invalid president, for he has not been chosen according to what the rules establish. But thanks to the threats of the army, people do not react against the coup and this person becomes the new ruler. He calls himself the president and people refer to him as the president, even if the constitution establishes that someone counts as the president only if he obtains more votes than any other candidate in the presidential election. Once in power, the new, non-institutional, president rules. He does not rule like the former democratic (institutional) president, but he rules nonetheless.

What this illustration shows is that there are institutional and non-institutional ways of being the ruler of a country. Although the activities of institutional and non-institutional rulers are not exactly the same, they are sufficiently similar to warrant the conclusion that the invalid president has broken the rules of the game. The invalid president has not complied with the rules. Elsewhere (Sánchez-Cuenca 2003) I have analysed this problem in full detail, arguing that the generic problem of compliance arises when outcomes that are achieved institutionally (winning elections) can also be achieved non-institutionally (mounting a coup), that is, when the activity is *partially* dependent on the rule.

Note, by contrast, that the compliance problem does not exist in the case of chess. It is impossible to win without following the constitutive rules of chess. If some player physically removes the piece that counts as the king of the rival, she does not become the winner. A player only wins when she checkmates the opponent. On the other hand, it is possible to rule regardless of electoral results. The point of playing chess is to win, but the point of the "political game" is not to win elections for the sake of winning: winning is instrumental to ruling and to the exercise of power, and this can be done institutionally or non-institutionally.

The compliance problem brings together the two parts of the argument developed so far. On the one hand, there is a problem of compliance when the degree of dependence of the activity on the rule is not total (Section 1). On the other hand, if there is a compliance problem, it is not obvious that the deontic powers created by constitutive rules rest on shared attitudes, for people may be coerced to comply as happens for instance in dictatorial regimes. The rule can exist and be enforced without the people who comply with it holding the right attitudes (Section 2).

There are several possible explanations of why people comply with constitutive rules. First of all, people may comply because they accept, agree or go along

with the rule. This is the possibility explored in *The Construction of Social Reality* and more fully developed in Searle's later book, *Rationality in Action* (2001). According to Searle, constitutive rules generate obligations, or what he calls desire-independent reasons for action. The paradigmatic case is that of promises. To make a promise is to place oneself under an obligation. Thus, the person who makes a promise puts himself or herself under the obligation of keeping the promise: the person ought to keep the promise (this is how the derivation of "ought" from "is" works). "Recognitional" rationality (Searle's expression) makes the obligation that follows from the promise binding for its author.

The problem, as Searle emphasizes, is that these desire-independent reasons coexist with desire-dependent reasons, and the latter may be more powerful than the former. Therefore, although the agent has a reason to keep the promise, prudential reasons may prevent him or her from doing so.

It is worth comparing Searle's approach to promises with the game theoretic one. As Thomas Schelling showed in his classic book, *The Strategy of Conflict* (Schelling 1960), rational people do not believe promises and threats issued by agents who do not have incentives to carry them out. This idea was later on formalized in game theory: a promise is credible when it is part of a subgame perfect equilibrium (or when it is temporally consistent). Since then, the idea has given rise to a rich body of theoretical and empirical literature. In economics, as Barro and Gordon's seminal work has shown (Barro and Gordon 1983), the independence of central banks is grounded on the lack of credibility of governments' promises not to introduce inflation by surprise. In political science, the game theoretic analysis of the credibility of promises has been employed to explain the Serb-Croat war (Fearon 1998), the Glorious revolution (North and Weingast 1989), or the American civil war (Weingast 1998). Precisely because desire-independent reasons generated by promises rarely triumph over prudential reasons, societies have developed systems of third-party enforcement to make promises credible. One of the functions of legal systems is to induce people to keep promises, given that desire-independent reasons are no guarantee that promises will be kept.

Searle does not pay much attention to economic theory or game theory. In his opinion, economic models of human behaviour are dead wrong. This is so obvious to him that he does not spend much time developing his objections, some of which are rather sketchy. Take this example: "if I value my life and I value twenty-five cents (. . .), there must be some odds at which I would bet my life against a quarter" (Searle 2001: 6). Searle claims that this result does not make any sense and considers that to be a fatal objection. The objection could be understood as a challenge to what in choice theory is known as the continuity axiom.

The continuity axiom is formally necessary to prevent infinitely valued goods (or, more technically, lexicographic preferences). It simply states that if the agent prefers A over B over C, it is always possible to find a probabilistic lottery combining A and C such that the agent still prefers that lottery to B. This can

be guaranteed with a sufficiently low probability of C. Searle seems to assume that one's life is among one's infinitely valued goods. But we could think of the case in this way: an agent prefers "crossing the street and buying the newspaper" (A) to "staying at home without buying the newspaper" (B) to "being run over by a car" (C). What the axiom implies is that there is a sufficiently low probability α of being run over by a car such that the agent considers that $(1 - \alpha)A + \alpha C \geq B$. In words, despite the high value of life, a very small probability of being run over by a car does not deter the agent from crossing the street to buy the newspaper. The fact that Searle does not accept this result is interesting in itself, but it is not really an objection to the theory.

Searle has a better point when he insists that in most decision problems economic theory is of little help, since the real problem of practical reason is not to make a choice given a preference order, as assumed in economics, but to order the many conflicting preferences the agent has (Searle 2001: 30). The deliberative phase that precedes choice consists precisely in the ordering of preferences. This objection, unlike the previous one, does not concern the consistency of the theory, but rather its domain of applicability, and should be taken very seriously. If economic theory works better in the economic sphere than in other contexts, it is precisely because it is easier to order the preferences when the agent is involved in economic exchanges with other agents.

Despite, however, Searle's misgivings about economic theory, it is worth trying an economic approach to constitutive rules. In *The Construction of Social Reality* Searle affirms that "institutions survive on acceptance" (1995: 118). We have seen above that if we interpret "acceptance" in attitudinal terms, we are led to some strange results about the "we" of the "we accept . . ." formula and about the attitudes of those who reject the rules that are imposed on them coercively. But if we get rid of attitudes and collective intentionality, we can provide a more realistic description of how constitutive rules operate. More concretely, we can follow the economic approach and interpret "acceptance" behaviourally, in terms of compliance. This is what I discuss in the next section.

5. INSTITUTIONS AS EQUILIBRIA

Institutions survive when the social groups that could destroy them do not have reasons to do so. It is not a matter of attitudes. Attitudes can be consistent with very different courses of action. Now, what kinds of reasons am I talking about? The reasons for compliance may be both desire-independent (external) and desire-dependent (internal) reasons. The fact that economists tend to rule out desire-independent reasons is partly due to the fact that they want to prove that there can be compliance even when people are moved exclusively by desire-dependent reasons. Admittedly, this is not a very realistic assumption, but it can be justified as a limiting case: if an institution can survive when motivations are exclusively based on self-interest, it is easier still if people act out of desire-independent reasons for action.

I take here an eclectic view, in the sense that the reasons for complying with the rules can be some sort of "everything considered" reasons that weigh both internal and external reasons. When, everything considered, people do not have reasons not to comply with the rules, we have generalized compliance. In game theoretic terminology, this amounts to saying that the rules (or the institution they create) are an equilibrium. An equilibrium is a combination of strategies such that no one has incentives to deviate from them. A rule survives when the equilibrium strategies of the players are those of complying with the rule.

The concept of equilibrium is much more general than the concept of acceptance. A given set of rules can be in equilibrium for a wide variety of reasons: because people agree with the rules, or because people, though disagreeing with the rules, do not find ways to organize collectively against them, or because the cost of replacing the existing rules with some new ones is too high, or even because of habit (because people used to comply with the rules and continue to do so almost through inertia).

The concept of equilibrium can be equally applied to regulative and constitutive rules. There is nothing particularly interesting about why people comply with regulative rules. It may be that the regulative rule is backed by sanctions, or that there is some consensus in society about the importance of complying with the rule, or that complying with the rule is costless. The real challenge is to understand how compliance with constitutive rules can be an equilibrium.

Constitutive rules generate deontic powers. These powers may or may not have distributional consequences. For instance, language and money do not produce winners and losers. Everyone gains from the possibility of speaking a language or of having a means of exchange. When the deontic powers associated with some constitutive rule make everyone better off, no one will have reasons not to comply with the rule.

If deontic powers do have distributional consequences, so that some people gain more than others thanks to the existence of the rule, then the losers may have little reason to comply with the rules. Here we have to distinguish two possibilities. In the first, deontic powers merely reflect the distribution of power in society. Very roughly, we can understand this social or brute power as the capacity for collective action of the different groups or classes that compose the society. If the institutional system simply represents the relative power of these groups, these groups will not have reasons not to comply, since given their social power they cannot rationally expect to obtain a greater lot than the one obtained institutionally. Rules or institutions would embody the bargaining solutions among the various social groups with regard to different social conflicts.

This, I think, is quite an unlikely possibility, for deontic or institutional powers rarely match social power.[5] And this is what makes compliance puzzling.

---

[5] In his reconstruction of historical materialism, Gerald Cohen argues, very abstractly, that for each ownership right (i.e. a deontic power) it is possible to find a non-institutional matching power. From this point of view, compliance is not problematic, since constitutive rules simply reflect the economic powers of the different groups (see Cohen 1978: 220ff.)

If deontic powers go beyond social powers, if there is not a perfect correspondence between the two, then why do the social groups that could be better off without the constitutive rule comply with it? If deontic powers defined by constitutive rules alter the distribution of social power, why do the groups negatively affected by the rules comply with them?

It is hard to provide convincing answers at this level of abstraction. Instead of trying a general argument about compliance, it may be more fruitful to focus on some concrete example. The institution of democracy is an excellent case in point, and one that has been extensively analysed in terms of constitutive rules and compliance. Basically, democracy is a system in which rulers are selected in contested elections. This, obviously, requires a complex set of constitutive rules defining the powers of the ruler, the procedures of aggregating votes, etc. What makes democracy so unique is the inherent role that uncertainty plays in its functioning. Election, a mechanism with a strong component of randomness, decides who the winners and the losers are. The dilemma of compliance is therefore faced by the various political forces in its crudest form: although initially all agree to run for the election, once one group discovers that it is the loser, it may decide not to accept the result of the election. In democracy, the winner is the political force that wins more votes. But having more votes does not necessarily mean having more social power. Perhaps other political forces, despite being less numerous, have greater military power, or simply greater capacity for collective action. Thus, the results of elections cannot be equated with any kind of bargaining solution that reflects the relative brute power of each group.

In his "Democracy as an equilibrium", Adam Przeworski (2005) has shown why compliance with the rules of democracy is so rare in poor countries and why compliance becomes almost universal above a certain income level. To put it in a nutshell: suppose that the loss of freedom associated to dictatorship is abhorred by everyone, independently of income; suppose also the marginal utility of income is decreasing. In poor countries, the overthrow of democracy may generate gains in income that compensate for the loss of freedom in dictatorship. In rich countries, by contrast, the marginal utility of income is so low compared to the cost of dictatorship that political forces are better off complying, even if they are permanent losers in the electoral contest.

This simple mechanism fits the observed association between the survival of democracy and income. Poor democracies have very little chances of surviving, for there are parties with incentives not to comply in case they lose elections, whereas rich democracies survive regardless of electoral results. The theory can be enriched and developed in several directions: for instance, Przeworski shows that if brute (military) power is balanced, or if society is sufficiently egalitarian, there can be compliance even in poor countries.

Note how far we have moved away from Searle's original framework. We still have constitutive rules, but attitudes have disappeared. Constitutive rules may have different degrees of constitutiveness. When there is no full constitutiveness by the rule, complex problems of compliance with the rule emerge. If there is

compliance, the rule is in equilibrium. If there is no compliance, the rule collapses. The problem of institutional order is basically the problem of why people comply with institutional arrangements that create deontic powers which alter the underlying distribution of social power. In other words, what we have to explain is under what conditions do institutions become equilibria.[6]

REFERENCES

Barro, R.J. and Gordon, D.B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, 12: 101–121.

Cohen, G.A. (1978). *Karl Marx's Theory of History: A Defence*. Princeton, NJ: Princeton University Press.

Fearon, J. (1998). Commitment problems and the spread of ethnic conflict. In D.A. Lake and D. Rotchild (eds.), *The International Spread of Ethnic Conflict*. Princeton, NJ: Princeton University Press.

Kuran, T. (1991). Now out of never: The element of surprise in the East European Revolution of 1989. *World Politics*, 44: 7–48.

Lohmann, S. (1995). The dynamics of informational cascades. The Monday demonstrations in Leipzig, East Germany, 1989–91. *World Politics*, 47: 42–101.

North, D. and Weingast, B.R. (1989). Constitutions and commitment. *The Journal of Economic History*, 49: 803–832.

Przeworski, A. (2005). Democracy as an equilibrium. *Public Choice*, 123: 253–273.

Sánchez-Cuenca, I. (2003). Power, rules, and compliance. In J.M. Maravall and A. Przeworski (eds.), *Democracy and the Rule of Law*. Cambridge: Cambridge University Press.

Schelling, T.C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Searle, J.R. (1964). How to derive 'ought' from 'is'. *The Philosophical Review*, 73: 43–58.

Searle, J.R. (1969). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J.R. (1995). *The Construction of Social Reality*. New York: Free Press.

Searle, J.R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.

Searle, J.R. (2003). Reply to Barry Smith. *American Journal of Economics and Sociology*, 62: 299–309.

Smith, B. (2003). John Searle: from speech acts to social reality. In B. Smith (ed.), *John Searle*. Cambridge: Cambridge University Press.

Weingast, B.R. (1998). Political stability and civil war: Institutions, commitment, and American democracy. In R.H. Bates et al. (eds.), *Analytic Narratives*. Princeton, NJ: Princeton University Press.

Zaibert, L. (2003). Intentions, promises, and obligations. In B. Smith (ed.), *John Searle*. Cambridge: Cambridge University Press.

Zielinsky, J. (1995). The Polish transition to democracy: a game-theoretic approach. *Archives Européennes de Sociologie*, 36: 135–158.

[6] I am grateful to Ricardo Parellada and Adam Przeworski for their useful comments.

SEARLE VERSUS DURKHEIM

Steven Lukes

In *The Construction of Social Reality* (Searle 1995) and subsequent writings that develop its argument (notably Searle 2003 and his contributions to D'Andrade 2006), Searle proposes a general account of social reality, of what constitutes 'social facts.' His purpose is to give an account of social reality in general and of institutional reality in particular. On his account, animals—including human animals—are social, exhibiting collective intentionality, but human animals have institutions, created and sustained entirely in individual minds by collective intentionality (which cannot be reduced to individual intentionality) and enabling them, through the constitutive power of language, to act on desire-independent reasons. Several commentators (notably Gross 2006) have been struck by what they see as a convergence between this account and that of Emile Durkheim, who, one century earlier, set out his account of social reality in his *The Rules of Sociological Method* (Durkheim 1982), whose first chapter asks 'What is a Social Fact?', and in subsequent writings, including his essay on 'Individual and Collective Representations' (Durkheim 1953).

At first blush, there would seem to be a case for this convergence thesis. At the very end of his life Durkheim wrote this:

> The great difference between animal societies and human societies is that in the former, the individual is governed exclusively from *within itself,* by the instincts (except for a limited degree of individual learning, which itself depends upon instinct). On the other hand, human societies present a new phenomenon of a special nature, which consists in the fact that certain ways of acting are imposed, or at least suggested *from outside* the individual and are added on to his own nature: such is the character of the 'institutions' (in the broad sense of the word) which the existence of language makes possible, and of which language is itself an example. They take on substance as individuals succeed each other without this succession destroying their continuity; their presence is the distinctive characteristic of human societies, and the proper subject of sociology. (Durkheim 1982: 248)

Searle allows that animals can be social but proposes that what distinguishes human social reality is that it consists in 'institutional facts' and Durkheim writes that 'one may term an *institution* all the beliefs and modes of behavior instituted by the collectivity; sociology can then be defined as the science of institutions, their genesis, and their functioning' (Durkheim 1982: 45). Both see institutions as 'language-dependent' (Searle 1995: 64), and as examples of institutions both cite legal and moral rules and financial systems. Like Durkheim, Searle takes language itself to be an institution: 'linguistic facts are

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 191–202.
© 2007 Springer.

also institutional facts' (Searle 1995: 72). Both Searle and Durkheim write of individual and collective 'representations' and for both the notion of 'collective representations' plays a central role in accounting for social facts. For Searle, 'all social and institutional reality exists in individual minds together, of course, with all of the material objects and other features of the context in which those minds operate' (Searle 2006b: 59); for Durkheim social life is 'made up entirely of representations' (Durkheim 1982: 34). Both thinkers affirm a belief in the unity of nature (bodily and mental, material and social or cultural). Thus Searle writes that 'we live in exactly one world' and Durkheim that 'there is no realm of nature that is not bound to others' (Durkheim 1953: 23). Both seek to reconcile this with an account of what renders human social reality distinctive. Both firmly maintain that we can attain objective knowledge of social facts. Both adhere to 'realism' applied to both nonhuman and human domains and to a straightforward version of the correspondence theory of truth which for Searle means that 'statements are made true by how things are in the world that is independent of the statement' (Searle 1995: 219) and for Durkheim that 'how the facts are classified does not depend on [the sociologist], or on his own particular cast of mind, but on the nature of things' (Durkheim 1982: 76). And both are engaged in combating what they see as the irrational currents of their times—postmodernism and relativism for Searle, 'resurgent mysticism' for Durkheim (Durkheim 1982: 33).

Nevertheless, Searle roundly rejects the idea that their views converge. Faced with claims that his work exhibits 'unacknowledged and unreconstructed Durkheimism,' that he is 'a Durkheimian at heart' and that his account reexpresses 'what are essentially Durkheimian ideas in an analytic idiom' (Gross 2006: 46, 51), Searle counters that 'the fundamental ontology in Durkeim is mistaken' (Searle 2006b: 62), that Durkheim had 'an inadequate conception of social facts' (Searle 2006b: 57), and that his view 'is not remotely like my view and it cannot be made consistent with what we know about how the world works anyway' (Searle 2006b: 63). Indeed, provoked to take a closer look at Durkheim's work, he writes that the situation 'is much worse than I had originally thought. His conception of social ontology is not only inconsistent with mine, but it is flawed in ways I did not originally realize' (Searle 2006b: 58).

The gist of Searle's critique is that Durkheim

> fails to distinguish social facts in general from the special subclass of institutional facts, he thinks that social facts are essentially coercive, and he thinks that they exist outside of individual minds. He says that the essence of social facts is that they are 'external to the individual and endowed with a power of coercion, by reason of which they control him.' This is exactly the opposite of my view. I think we need to start by distinguishing social facts in general from institutional facts, that human institutions are in large part enabling and empowering and that the collective intentionality which creates and sustains them is entirely in individual minds. (Searle 2006b: 57)

Searle writes that his most important point is that Durkheim has no account of deontology, of the deontic powers that are 'the glue that holds society together'

(Searle 2006b: 64). These are created by the assigning of status functions; these consist in 'rights, duties, obligations, responsibilities and so on' (Searle 2006a: 20), which in turn create desire-independent reasons for action. Durkheim, Searle claims, 'gives no evidence of understanding this point': namely, 'that there is a class of functions that can *only be performed in virtue of the collective acceptance of a deontic status,* and not in virtue of physical structure alone' (Searle 2006b: 64). Moreover, despite his recognition of the importance of the social character of language, Durkheim (alongside recent sociological theorists, notably Habermas, Bourdieu, and Foucault) failed to see how institutional facts 'require linguistic or symbolic modes of representation or they cannot exist' (Searle 2006b: 65).

Searle spells all this out at length. He examines, in particular, Durkheim's use of analogies to establish the independence and exteriority of social facts—the *sui generis* level of the social—and his claim that this derives from 'the association of minds' (Durkheim 1953: 26). He argues that the analogies Durkheim adduces to support this claim fail—the living cell's relation to chemical particles; the hardness of bronze lying in neither copper, tin nor lead; the liquidity and other properties of water being absent from its component gases; and the independence of mental states from their neuronal substratum of brain cells. The picture, says Searle, is inadequate because

> the relation of individual minds to collective intentionality is *constitutive and not causal*. The collective intentionality is constituted by what is in individual minds. It could not work otherwise. The explanation of human behavior requires intentional causation, but intentional causation can only work if it is internalized in individual minds. Collective intentionality can only function causally if it is constituted by what is in the minds of individuals. For Durkheim it is impossible to explain the mechanisms by which social facts 'constrain' individuals, because the collective representations are not in individual minds. (Searle 2006b: 62)

And he argues that the constraint, or coercive power, that Durkheim sees as distinctive of social facts 'covers over too many distinctions,' failing to distinguish between regulative and constitutive rules, between deontic desire-independent reasons for action and social forms of coercion, and between, for instance, feeling an obligation and being subject to the pressure of a crowd emotion.[1]

Searle's response to Durkheim's work raises a series of intriguing questions. First, there are exegetical issues. Is Searle in any respects *unfair* to Durkheim? Does he mischaracterize what Durkheim wrote? How divergent *are* their views? And are there plausible ways of reading Durkheim that would render them less so? And second, there are substantive issues. How convincing is Searle's critique? How deep does it go? And can we construct a convincing Durkheimian response?

---

[1] These last points are explored in the introduction to Lukes 1973.

I

In all this, is Searle unfair? In two cases, he plainly is. In the first place, he writes that Durkheim 'fails to distinguish social facts in general from the special subclass of institutional facts' (Searle 2006b: 57) but, as we have seen, Durkheim did draw just such a distinction, and he does so in several places (including the preface to the second edition of *The Rules*). Second, Searle criticizes Durkheim's famous maxim that social facts should be considered as things, arguing that in doing so, he contrasts 'things' with 'ideas' and that he fails to see that '*we can have an epistemically objective science of a domain that is ontologically subjective*' (Searle 2006b: 63). But Durkheim explicitly states—repeatedly and ever more explicitly throughout his career—that social facts consist in 'representations and actions' (Durkheim 1982: 52); and in 1908 wrote that 'in social life, everything consists of representations, ideas and sentiments' and that 'all sociology is a psychology, but a psychology *sui generis*' (Durkheim 1982: 247). He does not contrast 'things' with 'ideas', but rather with *preconceived* ideas, deriving from prejudices or intuitions or ideological doctrines, in advance of empirical research. He writes that to 'treat phenomena as things is to treat them as *data*' (Durkheim 1982: 69). Chapter 2 of *The Rules* sets out some sensible procedures for achieving such objectivity. Moreover, the maxim should be understood in its historical context. Impressed by the successes of German social science, Durkheim was urging that education in Third Republic France should be *à l'école des choses*, to induce in students a proper sense of the complexity of their social and political world, the better to predict and control it.

A third case where unfairness can be alleged is less obvious. Searle writes: 'At no point does he say, as I do, that all collective representations exist only in individual minds' (Searle 2006b: 59), and yet, in all fairness, he also cites Durkheim writing, in *The Division of Labor* that the collective or common *conscience* is 'only realized in individuals' (Searle 2006b: 61). He might also have cited Durkheim asserting that 'society comprises only individuals' (modified in a footnote to asserting that they are 'the only active elements in it') and then adding that therefore 'social life can have no other substratum than the individual consciousness' (Durkheim 1982: 39, 46). As we can already see, the issue to which these quotations allude is central to Searle's critique of Durkheim and much hinges on what Searle means by existing 'in individual minds' and Durkheim by 'substratum,' but what does seem obvious is that Durkheim, while repeatedly insisting that social facts are external to individuals, also clearly recognizes that they are also internal to, or internalized by them.

How important is it that Searle is mistaken in claiming that Durkheim does not acknowledge the subclass of institutional facts? That depends on how Durkheim understood 'institutions.' How different is his conception from Searle's? If different enough, then the convergence thesis is refuted. And the same goes for Durkheim's conception of 'representation.' Searle comments that Durkheim's diverges decisively from his own, because Durkheim's

'collective representations' cannot exist in the minds of individuals. In my view, all collective intentionality exists only in the minds of individuals. And it is not enough for a theorist to use the word 'representation', he has to specify exactly the propositional contents and various other features of the logical structure. (Searle 2006b: 66)

What these last two instances show is that we must be alive to the possibility, indeed likelihood, that Durkheim and Searle diverge by virtue of using the same terms in significantly different ways. And this is all the more likely, and significant, for two reasons.

The first is that the two authors have quite distinct purposes and projects in mind. Searle's concern is, by developing his theory of speech acts and theory of mind, to identify, as a philosophical enterprise, the distinguishing features of human social reality in order to show how this fits into the one world that 'consists entirely of physical particles in fields of force' (Searle 1995: xi). This leads him to reject theories which postulate further realities. On the other hand, he aims to answer 'the challenge of sociobiology' with its 'implicit message . . . that human beings are not different from other social animals and that the terms in which we need to understand human social behavior are essentially biological and above all evolutionary' (Searle 2006a: 6). And he also wants to insist, against those who want to make sociology seem like a natural science,

how different it is from the natural sciences. Sociology and the other social sciences are largely concerned with human behavior. The behavior in question consists in large part of actions done for reasons, and often the actions occur within humanly created institutions. But these facts, that actions are done for reasons, and thus subject to constraints of rationality, and that sometimes the actions are done within human institutions, together imply that the behavior in question is performed under the presupposition of free will, and empowered by systems of humanly created deontologies. (Searle 2006c: 124)

Durkheim's concern, by contrast, was indeed to establish sociology's credentials as a natural science. *The Rules*, intended to provide that rationale and guiding principles for future research, was directed both at the hostile and the skeptical and at fellow scholars and future collaborators. These would engage in specialist studies in a way that would ultimately transform the various social sciences into the systematically organized branches of a unified social science. They would be ever more deeply penetrated by 'the sociological idea', which implied that 'social facts are solidly linked to each other and above all must be treated as natural phenomena, subject to necessary laws' (Durkheim 1982: 195). Durkheim also wrote that 'social phenomena are only distinguishable from other phenomena by virtue of their greater complexity' (Durkheim 1982: 148). On the other hand, as we have seen, he assumed that sociology must have a subject matter peculiarly its own that would not be 'confused with that of biology and psychology' (Durkheim 1982: 50) and he was ever more insistent that its domain consists in 'representations and actions' (Durkheim 1982: 52). Furthermore, he stressed that his purpose was not 'to anticipate the conclusions of the discipline by stating a philosophical view' of the nature of social reality but

merely to indicate how, by outward signs, it is possible to identify the facts that the science must deal with, so that the social scientist may learn how to pick out their location and not to confuse them with other things. (Durkheim 1982: 43)

A second reason why the two authors are likely to diverge even in their understandings of the same terms is, simply, the time gap of one hundred years that separates them. For, as Searle repeatedly points out, Durkheim 'happened to live in the prehistory of this study of intentionality' (Searle 2006b: 62), and 'lacked the resources of contemporary philosophy' (Searle 2006b: 58). He lacked 'the conceptual resources even to state' the difference between being constrained by crowd emotions and being obliged by a promise (Searle 2006b: 65). So he failed to grasp other crucial distinctions, notably those between 'observer-relative and observer-independent phenomena' and between 'the ontological sense of the objective/subjective distinction and the epistemic sense' (Searle 2006b: 63).

So is Searle's reading of Durkheim unfair? With the not unimportant exceptions noted earlier, and given its admitted restriction to Chapter 1 and some other sections of *The Rules* and a subsequent 'closer look,' Searle's reading is fair enough. But the indicated restriction raises a most interesting question. Is there a significant gap between what we may call Durkheim's 'official' position, as set out in *The Rules* and other methodological writings and his 'unofficial' position, as exemplified in his sociological practice? And does that gap, if it exists, bear upon the important substantive issues raised by Searle's critique? I shall now suggest that there is indeed such a gap.

If we examine Durkheim's actual explanations in the light of this question it becomes clear that both official and unofficial positions are in evidence there. Consider his three best-known theories: the development of the division of labor as demographic volume and moral density increase along with the growth of markets and cities; the explanation of rising suicide rates by 'the social causes' of suicide, namely insufficient normative regulation (anomie) and/or insufficient social integration (egoism); and the generation and sustenance of religious sentiments and beliefs in collective experiences and their reiteration through ritual and symbolism.

In each of these three cases the official doctrine frames the explanations. Thus, in *The Division of Labor in Society*, he writes that 'everything occurs mechanically' and even describes his hypothesis as stating a 'law of gravitation of the social world,' in which

the partitions that shut off the different parts of society are more and more disappearing by the very nature of things, as a result of a kind of natural erosion, whose effect can even be reinforced by the action of violent causes. Population shifts thus become more numerous and more rapid, and migratory paths are channeled out as these movements occur. These form the network of communications. Such movements are more especially active when several of these paths intersect: these are the towns. Thus social density increases. As for the growth in volume, it is due to causes of the same kind. The barriers the divide peoples are similar to those that separate the various alveola within the same society and disappear in like manner. (Durkheim 1984: 288)

The social factors he adduces as causes are ranged along the continuum he identifies in the first chapter of *The Rules*, from the most 'crystallized' or 'morphological' features of social structure through legal and moral norms to states of the *conscience collective* and currents of opinion. In *Suicide*, after dismissing 'organico-psychic' and physical factors as explanatory, he argues that the book vindicates social realism, proving the existence of 'realities external to the individual' that are 'as definite and substantial as those of the psychologist or biologist' (Durkheim 1951: 37–38, 39). It was, he claimed, important to recognize their reality and conceive them as 'a set of forces which determine our behavior from without, just like the physico-chemical forces to whose influence we are subject.' Indeed, so truly are they phenomena (*choses*) *sui generis*, and not verbal entities, that one can measure them and compare their relative magnitude (Durkheim 1951: 309–310). And in *The Elementary Forms of Religious Life*, he likewise aims to offer an explanation that accounts for religious representations by reference to social causes. Thus he maintains that 'collective representations that refer to sacred beings' are plunged into the very source of religious life—the 'collective effervescences' that periodically occur when 'the individuals are assembled and in direct relations with one another, at the moment when everyone communes in the same idea or emotion' (Durkheim 1995: 350, 349). And he seeks to establish causal connections between structural or 'morphological' social facts and the content of religious belief and ritual and, most audaciously, with the fundamental categories of the understanding and modes of classification.

In all this talk of social causation, in the three cases cited and elsewhere, Durkheim continually resorts to the use of metaphors and analogies taken from the natural sciences of his time. In *The Division of Labor* it is the organic analogy that predominates, but thereafter the prevalent imagery comes from 19th-century physics, notably thermodynamics. In *Suicide* he writes that for each people there is 'a collective force of a determinate amount of energy, impelling men to self-destruction' and that the strength of such forces can be measured 'as one does the strength of electric currents or sources of light' (Durkheim 1951: 309–310). He uses this language very widely to describe the genesis and operation of collective ideas and sentiments, and the phenomenon of sacredness. So, for instance, 'the extreme ease with which religious forces radiate and diffuse' is compared with the way in which 'the heat or electricity that any object has received from outside can be transmitted to the surrounding milieu, and the mind readily accepts the possibility of that transmission' (Durkheim 1995: 327), and religion in general is seen as consisting in 'religious forces, human forces . . . moral forces' (Durkheim 1995: 421). Indeed, 'behind [religious] beliefs there are forces' and a 'theory of religion must first of all show what these forces are, of what they are made and what their origins are' (Durkheim 1913: 66). And he often described morality as 'a system of forces' which draw all their power of action from representations (see Lukes 1973: 36).

All the foregoing is, obviously enough, evidence of Durkheim's official position. But once we pay close attention to the content of his explanations, another

picture comes into view—a picture peopled by recognizably human actors. It is a picture that is somewhat closer to that sketched by Searle in discussing what distinguishes the social from the natural sciences, namely one in which human actions are typically done for reasons, under rational constraints, often within human institutions, often presupposing free will, and empowered by deontologies.

Why, as populations grow and grow denser and cities and markets develop, do people specialize, in ways that, according to Durkheim, generate social differentiation and the growth of 'organic solidarity? Only because people respond to the pressures of constraints by seeking out opportunities—acting rationally and making choices, using resources (such as money) and in pursuit of objectives whose very existence presupposes normative structures (such as markets).

Why does the suicide rate rise in the presence of anomie or egoism? Only because vulnerable, suicide-prone people, when norms and stable expectations break down or are lacking, or when they lack significant social attachments, are *deprived* of the sorts of reasons for acting and living in ways that protect them from the desperate extreme of killing themselves (and others from other, less extreme reactions). In short, you could say that Durkheim's explanation of suicide requires the counterfactual presupposition of informal 'deontological' frameworks specifying rights, obligations, responsibilities, and the like, that enable people to live normal, 'healthy' lives and that his theory of suicide is a theory of 'disablement' that occurs when those frameworks break down or weaken.

And, finally, how are we to explain the fact that in all known societies, from the most 'elementary' to the most advanced, people direct their thoughts, their feelings and their ritual practices toward sacred beings (from totems to the gods to God) set apart from everyday life and experience, inducing awe, reverence, and fear and in whose reality they believe? Durkheim's answer is, of course, complex, but it comes down to considering the faithful as reasoning creatures living within normative structures (from clans to churches) and possessing language, and who have distinctive emotional and cognitive needs. They seek and find strength, exaltation, and moral remaking through identifying with their 'society,' rallying around the same sign or emblem (for 'the totem is the flag of the clan'); they seek to render their world, natural and social, intelligible with the materials at hand, so that religion offered 'a first representation of what the relations of kinship between things might be' and 'a first explanation of the world' (Durkheim 1995: 222, 239); and they seek to represent that world, thus understood, imaginatively and aesthetically, in the form of stories and symbols (see Lukes 1973: 462–470).

II

Searle's critique of Durkheim, then, is directed at the official doctrine—or, to be more precise, at Searle's construal of the official doctrine. As we have seen, this is, in certain respects, unfair, in particular in its claim that, since Durkheim

thinks that social facts, and collective representations in particular, exist outside of individual minds, he therefore thinks that they do not exist in individual minds. But Durkheim clearly believed that they do both—that they are both external and internalized. His focus on the exteriority of social facts was a focus on the obvious but sociologically central truth that any given individual is surrounded, preceded, and outlasted by innumerable, intermeshed expectations, and requirements. In their investigations of this truth's manifold implications sociologists and anthropologists use such terms as socialization, norms, roles, cultural imperatives, organizational and institutional logics, and the like.

Of course, Searle recognizes this truth, writing that much 'collective intentionality is simply inherited, and many, perhaps most, of the institutions we grow up in—families, money, private property, the stock market, and summer vacations—are simply accepted uncritically'(Searle 2006b: 67). He writes that his attempt is to 'describe exactly what is accepted and rejected'; he is not concerned with 'the analysis of particular patterns of acceptance and rejection' (Searle 2006c: 67). But such an analysis would precisely have led him to focus on how the external becomes internal. Searle does not address this, which, he concedes, is 'an interesting area for sociological investigation' (Searle 2006b: 67). This is, I think, because, in his concern to show that collective intentionality is 'not mysterious or inexplicable' (Searle 2006a: 16), he thinks that this requires saying, as he does over and over again, that 'all collective intentionality exists only in the minds of individuals,' that 'social reality exists entirely [*sic*] in individual minds' (Searle 2006b: 66, 59).

He reads Durkheim as claiming that it 'surpasses its substratum and exists in part outside the substratum' and that it is '*the association of individual minds* that results in collective intentionality' (Searle 2006b: 59, 61). To read him thus is entirely understandable, since that, after all, is what he *wrote*. But it is a philosopher's reading. Recall that Durkheim was not doing social ontology: he was setting out methodological rules for sociologists and, in this capacity, using, or rather misusing, analogies to direct their attention to two incontrovertible facts: that 'certain ways of acting are imposed, or at least suggested *from outside* the individual and are added on to his own nature' and that institutions 'take on substance as individuals succeed each other without this succession destroying their continuity.'

Searle is, of course, absolutely right to draw our attention to Durkheim's misuse of his analogies. His explanation of why it is a misuse—that collective intentionality is constituted by what is in individual minds—is innovative and profound. So are the other implications that he has derived from his theory of speech acts and the essentially constitutive role of language, in particular his reflections on the enabling role of social institutions, through the creation and sustenance of deontic powers. He is certainly right that there is no attention paid to all this in Durkheim's methodological writings and thus no theorization of what is involved. It is true that Chapter 1 of *The Rules* is all about the exteriority and the generality-plus-independence of social facts. It is also true that it

contains a confusing and confused discussion of external constraint that plays on quite distinct ways in which individuals can be constrained. And it is true that there is no attempt, no hint of an attempt, to account for the central paradox that has preoccupied Searle: that institutional facts are real but they only exist because people think they exist. Searle writes that he does not 'see that Durkheim resolved the paradox or even showed an awareness that it was a paradox that he had made apparently inconsistent remarks' (Searle 2006c: 123). I agree that Durkheim, unlike Searle, did not resolve the paradox, but I do think that his writings abundantly express it and I disagree that the components of his view of social facts as collective representations are inconsistent. On the reading I have suggested, they are not only consistent but sociologically sound.

Searle resolves the paradox by arguing that institutional social facts, while epistemically objective, contain ontologically subjective elements that are essential to their existence: that, in short, 'these objective facts only exist in virtue of collective acceptance or recognition or acknowledgement' (Searle 2006a: 15, 13). He contends that Durkheim did not grasp the distinctions needed to see this. I maintain, on the contrary, that throughout Durkheim's work we can find highly suggestive answers to a question that is, in turn, raised by Searle's resolution. This is a question left entirely unaddressed by Searle himself. He asks: 'What does "collective acceptance or recognition or acknowledgment" amount to?,' but he does not ask how it is achieved.

Searle remarks that by 'collective acceptance' he does not mean to imply approval, that it is intended to capture 'a continuum that goes all the way from enthusiastic support to grudging submission' and that 'perhaps a less misleading term would have been "recognition"'[2] (Searle 2006c: 123). But this is already to suggest that people may be less or more submissive to the internalization of the shared understandings or collective representations that make the institutional social facts (and hence their deontic powers—rights, obligations, responsibilities, and so on) possible. To embrace that suggestion is surely to accept the case for inquiring into the transmission and inculcation of ideas, from the most concrete and specific to the basic categories of thought, into the mechanisms of socialization, into the conditions under which the recognition of shared norms and integration into shared values break down or are weakened, and into the role of ritual and symbolism in generating and regenerating 'enthusiastic support' for the collective representations that unite communities. These are precisely the inquiries that Durkheim pursued and encouraged others to pursue. In this way, Durkheim can be seen to have at hand a robust response to Searle's critique.

And perhaps that response can be rendered even more robust by paying attention to what Searle has to say about what he calls 'the background'—the 'capac-

[2] Searle's usage recalls H.L.A. Hart's notion of the secondary rules, or 'rules of recognition' in *The Concept of Law* (Hart 1961). Hart displays a similar lack of interest in how their internalization is achieved.

ities and abilities' that render one 'at home in his society', and '*chez lui* in the social institutions of the society', that render one 'sensitive to the rules', the 'motivational dispositions' that 'give sense' to people's beliefs and desires (Searle 1995: 147, 136). Searle has little to tell us about what the mechanisms are that generate and sustain these dispositions, other than that they have 'evolved'. My final suggestion is that here too the Durkheimian move is to explore whether and to what extent these are socially inculcated and reinforced. Just this is suggested by two of the recent sociological thinkers whom Searle criticizes alongside Durkheim, as deprived of modern philosophical sophistication concerning language, namely Bourdieu and Foucault. Both in different ways suggest that socially significant dispositions are inculcated, or as they both say 'inscribed', within human bodies. Thus Bourdieu writes of social agents as 'endowed with habitus, inscribed in their bodies by past experience': social norms and conventions are 'incorporated' in their bodies, thereby generating 'a permanent disposition, a durable way of standing, speaking, walking, and thereby of feeling and thinking' (Bourdieu 1990: 70). For Bourdieu the dispositions that constitute *habitus* are 'spontaneously attuned' to the social order, perceived as self-evident and natural (Bourdieu 2000: 138–139).[3] There are, of course, huge, unanswered questions and serious objections that are raised by this picture, not least the lack of any explanatory account of how the social structure gets inscribed into bodies (briefly discussed in Lukes 2005 and at length in Lahire 1998). My point is simply that here too, in its approach to what Searle calls 'the background,' a Durkheimian approach does not stand refuted by Searle's critique but rather offers a response to the challenge it raises.

REFERENCES

Bourdieu, P. (1990). *The Logic of Practice*. Translated by Richard Nice. Stanford, CA: Stanford University Press. (Original French edition 1980.)

Bourdieu, P. (2000). *Pascalian Meditations*. Translated by Richard Nice. Stanford, CA: Stanford University Press. (Original French edition 1997.)

D'Andrade, R. (ed.) (2006). Searle on institutions: symposium of articles. *Anthropological Theory,* 6(1): 5–125.

Durkheim, E. (1913). Contribution to discussion of 'Le Problème religieux et la dualité de la nature humaine', *Bulletin de la Société française de philosophie,* 13: 63–75, 80–87, 90–100, 108–111.

Durkheim, E. (1951). *Suicide: A Study in Sociology*. Translated by J. A. Spaulding and G. Simpson and edited with an Introduction by G. Simpson. London: Routledge and Kegan Paul. (Original French publication 1897.)

Durkheim, E. (1953). *Sociology and Philosophy*. Translated by D. F. Pocock with an Introduction by J. G. Peristiany. London: Cohen & West.

Durkheim, E. (1982). *The Rules of Sociological Method and Selected Writings on Sociology and its Method*. Edited with an Introduction by Steven Lukes. Translated by W. D. Halls. London and New York: Macmillan. (Original French publication 1893, second edition 1901.)

[3] Searle himself comments on the parallel between his conception of the Background and 'Bourdieu's important work on the "habitus"' (Searle 1995: 132).

Durkheim, E. (1984). *The Division of Labor in Society*. Translated by W.D. Halls with an Introduction by Lewis Coser. New York: Free Press. (Original French publication 1893, second edition 1902.)
Durkheim, E. (1995). *The Elementary Forms of Religious Life*. Translated with an Introduction by Karen E. Fields. New York: Free Press (Original French publication 1912.)
Gross, N. (2006). Comment on Searle. *Anthropological Theory,* 6(1): 45–56.
Hart, H. L. A. (1961). *The Concept of Law*. Oxford: Clarendon Press.
Lahire, B. (1998). *L'homme pluriel: les ressorts de l'action*. Paris: Nathan.
Lukes, S. (1973). *Emile Durkheim: His Life and Work*. New York: Harper & Row.
Lukes, S. (2005). *Power: A Radical View*. Second Edition. London and New York: Palgrave Macmillan.
Searle, J.R. (1995). *The Construction of Social Reality*. New York: Simon and Schuster.
Searle, J.R. (2003). Social ontology and political power. In F.F. Schmitt (ed.) *Socializing Metaphysics: The Nature of Social Reality*. Lanham, MD: Rowman & Littlefield.
Searle, J.R. (2006a). Social ontology: some basic principles. *Anthropological Theory,* 6(1): 12–29.
Searle, J.R. (2006b). Searle versus Durkheim and the waves of thought: reply to Gross. *Anthropological Theory,* 6(1): 57–69.
Searle, J.R. (2006c). Lukes and "substantial social scientific work". *Anthropological Theory,* 6(1): 122–125.

SEARLE'S DERIVATION OF PROMISSORY OBLIGATION

Savas L. Tsohatzidis

1. INTRODUCTION

In "How to derive 'ought' from 'is'" (Searle 1964), perhaps the most famous among his early articles, John Searle set out to show that what is sometimes called "the naturalistic fallacy"—the fallacy that is allegedly committed by those who affirm that it is possible to deduce evaluative conclusions from wholly non-evaluative ('descriptive') premises—is not at all a fallacy. The reason that it is not, Searle claimed, is that there are certain kinds of clearly evaluative statements (in particular, statements about what a particular person *ought* to do or has the *obligation* to do) that *are* logically deducible from certain sets of wholly non-evaluative ('descriptive') statements (specifically, from sets of statements that include a statement about what that person has *promised* to do). That being so, Searle contended, not only is there no such thing as "the naturalistic fallacy," but those claiming that there is commit themselves to the denial of the validity of a series of logically impeccable inferences (and so become victims of what might be called "the naturalistic fallacy fallacy").

Searle presented an improved version of his proposed derivation of 'ought' from 'is' (in effect, of statements about obligations from statements about promises) in the final chapter of his first major book, *Speech Acts* (Searle 1969), arguing that all criticisms of the derivation that in the meantime had been produced have been unsuccessful, and claiming that the derivation is one among the many philosophically interesting by-products of the general account of meaning and speech acts that that book proposes. In outline, Searle's presentation of the improved version of his derivation proceeds as follows. From a *descriptive* statement to the effect that a speaker X has uttered, at time *t*, a sentence of the from "I (hereby) promise to p", where p refers to a future action of X, we can infer, Searle says, the equally *descriptive* statement that X has promised, at *t*, to p, simply by making the empirical assumption that X's utterance satisfies the conditions that the theory of speech acts specifies as necessary and sufficient for a speaker's utterance to constitute a promise on his part to perform a future action. But once these two purely *descriptive* statements are in place, Searle contends, they logically necessitate, in conjunction with three analytic (and so, non-evaluative) statements, a series of clearly *evaluative* statements, and so contradict the thesis this is impossible to logically deduce evaluative conclusions from

S.L. Tsohatzidis, (ed.), Intentional Acts and Institutional Facts, 203–217.
© 2007 Springer.

non-evaluative premises. In particular: (a) in conjunction with the analytically true statement that *whoever promises, at a given time, to do something, undertakes, at that time, the obligation to do it*, the statement that X promised, at *t*, to p entails the statement that X undertook, at *t*, the obligation to p (e.g., the statement that Tom promised, at *t*, to give Mary five dollars entails the statement that Tom undertook, at *t*, the obligation to give Mary five dollars); (b) in conjunction with the analytically true statement that *whoever undertakes, at a given time, the obligation to do something, has, at that time, the obligation to do it*, the statement that X undertook, at *t*, the obligation to p entails the statement that X had, at *t*, the obligation to p (e.g., the statement that Tom undertook, at *t*, the obligation to give Mary five dollars entails the statement that Tom had, at *t*, the obligation to give Mary five dollars); and finally, (c) in conjunction with the analytically true statement that *whenever one has the obligation to do something, one ought, as regards that obligation, to do it*, the statement that it was true of X at *t* that he had the obligation to p entails the statement that it was true of X at *t* that he ought, as regards that obligation, to p (e.g., the statement that it was true of Tom at *t* that he had the obligation to give Mary five dollars entails the statement that it was true of Tom at *t* that he ought, as regards that obligation, to give Mary five dollars). But since statements about what a particular person at a particular time *ought* to do, or has the *obligation* to do, are clearly evaluative statements, it follows, Searle concludes, that such statements *can* be logically deduced from appropriate sets of wholly non-evaluative statements, contrary to the claim that it is never possible to logically deduce evaluative statements from wholly non-evaluative ones.

Searle's early thesis about the deducibility of statements about obligations from statements about promises deserves attention not only on account of its intrinsic interest, but also on account of the special significance it has acquired in the context of the highly ambitious philosophical projects that he has launched in *The Construction of Social Reality* (Searle 1995) and in *Rationality in Action* (Searle 2001). One of the main aims of *The Construction of Social Reality* is to explain, compatibly with a naturalistic world-view, how institutional facts are possible. A key element of the proposed explanation is that institutional facts are facts that only exist because groups of agents collectively acknowledge their existence by virtue of accepting constitutive rules of the form "X counts as Y in context C", whereby status functions are assigned to entities. And since, according to Searle, the assignment of a status function to an entity amounts to the creation of a *deontic power* (that is, of a right or of an obligation, or, more generally, of an entitlement or of a requirement) involving that entity (as Searle succinctly puts it (1997: 451), "the imposition of a status-function is the imposition of a deontology"), the fundamental kinds of things that, in the domain of institutional reality, agents bring into existence simply by acknowledging their existence are the deontic powers (e.g., the rights or the obligations) that follow from the collective assignment of status functions to entities.

But since, according to Searle, deontic powers (as opposed to brute powers) are not determinable on the basis of the physical constitution of the entities to which status functions are assigned, it is necessary, in order for a deontic power to be brought into existence, to be publicly and conventionally *represented* as being brought into existence. And since the primary means of *public* and *conventional* representation is language, language is, according to Searle, a necessary precondition for the existence of institutional facts (which explains, in his view, why humans, who are unique among animals in possessing linguistic abilities, are also unique among animals in creating institutional structures). Now, if it is true that deontic powers, whose creation is the defining feature of institutional structures, are only possible thanks to the existence of language, it should not be surprising to find that an obligation, which is a particular kind of deontic power, is brought into existence by a promise, which is a particular kind of linguistic act. Viewed in the context of *The Construction of Social Reality*, therefore, Searle's early derivation of statements about obligations from statements about promises does not merely represent, if correct, an isolated instance of the dependence of an extra-linguistic institutional fact on a linguistic fact, but rather a paradigmatically clear, though by no means unique, instance of the quite pervasive dependence that, according to Searle, obtains between facts of the former kind and facts of the latter kind.

Searle's early derivation has a related, and equally important, connection with another major theme of his recent work, namely, the critique of the classical model of practical reason that he develops in *Rationality in Action*.[1] One of the fundamental defects of the classical model of practical reason, Searle contends, is that, according to it, the only reasons for action that agents could possibly recognize and have are their desires or other states essentially dependent on their desires. And the fundamental move that, according to Searle, is required in order for that defect to be overcome, consists in acknowledging that, among the reasons that can motivate agents to act, there are not only the *desire-dependent reasons* that agents may recognize, but also the *desire-independent reasons* that they may recognize, and especially those desire-independent reasons that they cannot *fail* to recognize because they intentionally *create* those reasons for themselves. But understanding the significance of such desire-independent reasons for action amounts, in Searle's view, to understanding the significance of linguistically generated deontic powers, and in particular the significance of promissory obligations. For, according to Searle, the prototypical instance of an intentionally created desire-independent reason for action is the *obligation* that an agent creates for himself when he *promises* to perform some future action: that obligation is clearly a reason that the agent has to perform the future

[1] See also Searle (1999) for a preliminary statement of that critique, explicitly connected with the topic of Searle's early derivation.

action; and it is, furthermore, a reason that, on the one hand, holds quite independently of the agent's desires (even if an agent does not *want* to do what he has promised to do, the fact that, through his promise, he has placed himself under the *obligation* to do it, provides him with a *reason* for doing it) and, on the other hand, cannot fail to be recognized by the agent, since it was not dictated by others but was purposively brought into existence by the agent himself through his act of promising. Indeed, not only are the obligations created by acts of promising paradigm cases of intentionally created desire-independent reasons for action, but it is also the case, Searle suggests, that virtually all other linguistic acts create, in ways that are perhaps less obvious but no less certain, many other obligations of a similar sort (e.g., an assertion creates for its author the obligation to provide, if challenged, evidence for the truth of its content, a request creates for its author the obligation not to obstruct the realization, by its addressee, of the state of affairs that the addressee has been requested to realize, etc.); and since each of these obligations constitutes a reason for action that holds quite independently of the agent's desires, it follows that virtually every normal use of language constitutes a counterexample to the classical model's assumption that the only reasons for action that an agent might have are, or are essentially dependent on, the agent's desires. Obviously, Searle's early thesis that statements about promises logically entail statements about obligations acquires deep significance in this context, and that makes it understandable why it is frequently invoked by him throughout *Rationality in Action*. For although, as Searle now stresses, promises are not, in his view, unique among speech acts in creating obligations (and therefore desire-independent reasons for action), their conceptual ties with the creation of obligations are so strong and evident that they provide a model on which all linguistically generated obligations (and, therefore, all linguistically generated desire-independent reasons for action) should be analysed.

The main purpose of this chapter is to argue that Searle's thesis that promises necessarily create obligations is open to a family of decisive counterexamples. The reason why these counterexamples have so far remained unnoticed is, in my view, that most of the critical literature that the thesis has provoked (and which I do not propose to review here) does not concern itself with Searle's central contention that statements about promises entail statements about obligations (does not address, that is, what I will henceforth call the *deducibility claim* that Searle makes), but rather effectively grants to Searle that statements of the former kind entail statements of the latter kind and only disputes Searle's *separate* claim that statements of the former kind are purely 'descriptive' whereas statements of the latter kind are purely 'evaluative' (it only addresses, that is, what I shall henceforth call the *categorization claim* that Searle also makes). My counterexamples concern specifically the deducibility claim to which Searle is centrally committed—the claim, that is, that, necessarily, whenever a statement of the form 'X promised to p' is true, a corresponding statement of the form 'X ought to p' is also true—and, if valid, they make it strictly speaking unnecessary

to further discuss the categorization claim: if, contrary to Searle's contention, 'X promised to p' does *not* entail 'X ought to p', then Searle would presumably not expect any momentous philosophical conclusions to follow *merely* from the assumption that the former of these statements is 'descriptive' and the latter 'evaluative', even if that assumption turned out to be correct. I believe, however, that it is itself doubtful whether Searle would currently be entitled to that last assumption, given some independent theses that he has recently propounded; and I consequently believe that it is doubtful whether he would *now* be entitled to all the conclusions he was originally aiming to draw from his discussion of the 'is'/'ought' question, even assuming, contrary to fact, that statements of the form 'X promised to p' do entail statements of the 'X ought to p'. I will therefore use the present occasion to first offer some thoughts on the status of categorization claim, and will then proceed to the primary task of determining the truth or falsity of the deducibility claim.

2. THE INSTABILITY OF THE CATEGORIZATION CLAIM

Assume for the sake of argument that statements of the form "X promised to p" do entail corresponding statements of the form "X ought to p". As Searle recognizes, in order for the existence of such entailments to imply anything determinate about the "naturalistic fallacy," it should be pre-theoretically quite obvious that statements about promises cannot be anything else than purely 'descriptive' and that statements about obligations cannot be anything else than purely 'evaluative'. And Searle, in all presentations of his derivation, does take both of these things to be pre-theoretically quite obvious—so obvious, in fact, that he spends no time at all in *arguing* in their favour. One might doubt, however, that pre-theoretical obviousness could be reasonably assumed in that area (after all, 'descriptive' and 'evaluative' are not terms like 'red' and 'blue', or even like 'polysyllabic' and 'monosyllabic'), and one might accordingly expect that, given different theoretical frameworks within which the labels 'descriptive' and 'evaluative' are being deployed, different decisions as to how these labels should be applied to statements about promises or to statements about obligations would be called for. I suspect that Searle would eagerly deny both that the conditions of application of the labels 'descriptive' and 'evaluative' are theory-sensitive in that sense, and that the labels themselves lack the degree of stability that his categorization claim presupposes. I would like to suggest in this section, however, that such a denial would be unwarranted, by showing that some of Searle's *own* recent theoretical positions commit him to the thesis that statements about promises are no less *evaluative* than statements about obligations, contrary to his explicit early claim that such statements, unlike statement about obligations, are clearly not evaluative.

The source of the instability is Searle's thesis, frequently appearing in *The Construction of Social Reality* and in subsequent works, that *linguistic facts are institutional facts*—a thesis that he also expresses by saying that language is an

institutional phenomenon. One apparently troublesome consequence of that thesis that is not directly relevant to the present discussion and need only be briefly addressed for the purposes of clarification, is that, in conjunction with another thesis of *The Construction of Social Reality* to which I have already referred—namely, the thesis that language is a precondition of the existence of institutional phenomena—it seems to create a problem of either circularity or infinite regress at the foundations of Searle's account of institutions: if language is a precondition of *every* institutional phenomenon, and if it is *itself* an institutional phenomenon, then it seems to follow that language is a precondition of language, which would appear to be a rather unattractive proposition on which to build a theory of institutions. Now, Searle is fully aware of that particular problem, and devotes a significant part (1995: 72–76) of a chapter of *The Construction of Social Reality* to working out a solution to it—a solution that, for the purposes of the present discussion, I shall deem to have been successful (even though I have doubts about its cogency). His solution, in a nutshell, is to hold that the sense that the term "language" has when it occurs in the thesis that language is a precondition of all institutions is the sense in which it refers to a *general capacity for symbolization* that members of the human species possess as a part of their biological endowment, whereas the sense that the term "language" has when it occurs in the thesis that language is a social institution is the sense in which it refers to the set comprising each one of the *particular* systems of symbolizing conventions that *particular* human communities create by exercising that common biological capacity. According to Searle, language in the first of these senses is clearly a non-institutional (or, as he sometimes prefers to call it, a pre-institutional) phenomenon, whereas language in the second of these senses is clearly an institutional phenomenon (or, more precisely, a collection of institutional phenomena): thus, the capacity to use sounds in order to represent objects is a pre-institutional capacity that every normal member of the human species has; but the fact that the very same sequence of sounds is a *word* in one human community and not a word in another human community, or the fact that what are phonetically the same words in two human communities may have totally different *semantic contents* in these two communities are clearly, according to Searle, institutional phenomena (as a later passage in *The Construction of Social Reality* explains (1995: 99), such terms as "word", "sentence" or "semantic content" do not refer to intrinsic properties of the sound structures that recur in the language of a human community, but rather to "statuses" that have been collectively "imposed" on those sound structures by members of that community). By distinguishing, therefore, between these two senses that the term "language" can have, one can avoid, Searle proposes, both the threat of circularity and the threat of infinite regress that may appear to be generated by the simultaneous adoption of the theses that language (in the first of the indicated senses) is a precondition of all institutional phenomena and that language (in the second of the indicated senses) is itself an institutional phenomenon.

The problem that is of primary concern to us here can now be stated clearly, and is a consequence of Searle's thesis that any *particular* human language (that is, any *particular* system of symbolizing conventions that a human community has devised) is an institutional phenomenon. It is part of that thesis—and is, of course, explicitly acknowledged by Searle—that the fact that a sequence of sounds is the utterance of a *sentence* of a particular human language, or the fact that a sequence of sounds can constitute the performance of a particular *illocutionary act* in a particular human language, are *institutional* facts, since they are not facts that are due to the intrinsic physical properties of the sounds concerned, but are rather facts that only exist because they are collectively acknowledged to exist by the members of a particular human population—in other words, because members of a particular human population *accept* (or, as Searle likes to put it, *count*) a sequence of sounds as having the *status* of a sentence, or the *status* of an illocutionary act. Thus, in giving what he takes to be especially clear examples of institutional facts that are created out of brute facts through the imposition of "status functions on entities" in accordance with "the general formula 'X counts as Y in C'", Searle writes: "I make noises through my mouth. So far, that is a brute fact: there is nothing institutional about it. But, as I am a speaker of English addressing other English speakers, those noises *count as* the utterance of an English sentence; they are instances of the formula 'X counts as Y in C'. But now, in an utterance of that English sentence, the Y term from the previous level now functions as the X term at the next level. The utterance of that English sentence with those intentions in that context *counts as*, for example, making a promise" (Searle 1998: 128–129). So, both the fact that a sequence of sounds is the utterance of a *sentence* of a particular language and the fact that the utterance of that sentence is the performance of a particular *illocutionary act* in that language are, for Searle, institutional facts—indeed, *paradigmatic* examples of such facts.

But now recall that, according to Searle, *whenever* an institutional fact is brought into existence through the assignment of a status-function to an entity certain *deontic powers* are necessarily created—that is, certain rights and obligations (or more generally, certain entitlements and requirements) that did not previously exist come to exist within the human group that collectively acknowledges the fact's existence. If that is so (if, to put it in Searle's own words, "everything turns out to be deontic" (1995: 109) in the domain of institutional reality), then it follows that the *mere* fact that a sequence of sounds is the utterance of a *sentence* of a particular language, or the *mere* fact that a sequence of sounds is the performance of a particular *illocutionary act* in a particular language, are already *deontic* facts of particular kinds, each associated with its proprietary set of entitlements and requirements. But if that is true, and if it is also true, as Searle is independently committed to holding, that *every* statement of a deontic fact is an *evaluative* statement, then it follows that the mere statement that a sequence of sounds is the utterance of a sentence of a particular language, or the mere statement that a sequence of sounds is the

performance of an illocutionary act in a particular language, are already *evaluative* statements of a special kind.

The upshot of all this is, of course, that *every* singular statement figuring in the derivation of 'ought' from 'is' that Searle had originally proposed turns out to be, by his own stipulations, and contrary to his original intentions, ultimately evaluative. Searle's derivation, it will be recalled, was supposed to show that from certain purely *non-evaluative* singular statements such as (1) and (2)

(1) Tom uttered the sentence "I hereby promise to give Mary five dollars"
(2) Tom promised to give Mary five dollars

one can validly deduce, through the mediation of appropriate analytic truths, certain deontic—and therefore, according to him, unquestionably *evaluative*—singular statements such as those in (3) and (4):

(3) Tom has the obligation to give Mary five dollars.
(4) Tom ought to give Mary five dollars.

But, in view of Searle's conception of linguistic reality as a special kind of institutional reality, it now turns out that (1) and (2) are *themselves*, in their own subtle way, no less evaluative than (3) and (4) are, since (1) and (2) report particular *institutional* facts of the linguistic variety, since *every* institutional fact is ultimately, according to Searle, a *deontic* fact, and since every statement of a deontic fact is, according to Searle, an *evaluative* statement: As it should be clear by now, a sequence of sounds cannot, for Searle, be Tom's utterance of *the sentence* "I hereby promise to give Mary five dollars" unless Tom is a member of a linguistic population where that sequence of sounds is collectively *accepted* (or *counted*) as the utterance of a sentence; and a sequence of sounds cannot, for Searle, be Tom's *promise* to give Mary five dollars unless Tom is a member of a linguistic population where that sequence of sounds is collectively *accepted* (or *counted*) as the making of a promise. But if (1) and (2) report facts that are institutional, and therefore, in their own subtle way, no less deontic than those reported in (3) and (4), then, given Searle's non-negotiable assumption that deontic statements are necessarily evaluative, it follows that (1) and (2) are no less *evaluative* statements (though, of course, less *blatantly* evaluative statements) than (3) and (4) are. Consequently, even if it should turn out, as Searle claims, that, from (1) and (2), one can validly deduce (3) and (4), it would *not* follow, as Searle *also* wanted to claim, that this shows that certain evaluative statements can be deduced from certain *non*-evaluative ones. In short, Searle's early claim that "uttering certain words in certain conditions just *is* promising, *and the description of these conditions needs no evaluative element*" (Searle 1964: 50; second emphasis added) is simply inconsistent with his current view that the conditions under which a series of sounds constitutes a promise are conditions whose satisfaction transforms the occurrence of that series of sounds into an institutional fact by

imposing on it a particular *deontology*, and are, therefore, as full of 'evaluative elements' as any deontological condition is.[2]

Of course, it would be open to Searle to try to avoid this result by radically modifying his theory of institutional facts in a way that would allow him to withdraw his current commitment to regard statements like (1) and (2) as subtly evaluative, and so to restore his early claim that such statements are definitely not evaluative. But not only do I not see any basis on which such a move could be motivated within Searle's theory as it is now developed, but it is clear to me that, if attempted, it would only reinforce the point with which the discussion in this section has begun, namely, that the question of the applicability of the labels 'evaluative' or 'non-evaluative' to many kinds of statement cannot be answered in the theory-neutral way that Searle was taking to be possible when he originally proposed the derivation of statements about obligations from statements about promises. If that is so, the only sufficiently well-defined issue that remains to be discussed with respect to Searle's derivation does not concern the categorization claim that it embodies (nor, consequently, the implications or non-implications of that claim on the "naturalistic fallacy"), but rather the deducibility claim that it embodies. It is to the discussion of that claim that I now turn.

3. THE FALSITY OF THE DEDUCIBILITY CLAIM

In order to show that, contrary to Searle's deducibility claim, statements about promises do not entail statements about obligations, it would be sufficient to show that there exist conditions under which it is possible for a statement of the form "X promised to p" to be true of an individual at a given time but impossible for corresponding statements of the forms "X has the obligation to p" or "X ought to p" to be true of that individual at that time.[3]

Such conditions clearly exist. Suppose that Tom has just heard a story about Satan, but ignores that, unlike himself, Satan is merely a creature of human imagination. If, outraged by one of Satan's alleged deeds, Tom says, at *t*, "I hereby promise to kill Satan", the fact that Satan does not exist has no power to make it false that Tom has *promised*, at *t*, to kill Satan. But the fact that Satan does not exist certainly makes it false that it is Tom's *obligation* to kill Satan, or that Tom *ought* to kill Satan. An imaginary thing cannot be a thing that one *ought* to kill, though it *is* a thing that one can, in one's ignorance, *promise* to kill (and, incidentally, it is also a thing that one can, in one's ignorance, *believe* to be worth killing). For that reason, a statement like (5) is not contradictory, whereas a statement like (6) certainly is:

---

[2] It is worth noting that, as the second sentence in the above quotation from Searle makes clear, Searle's original position was not merely that "promise" is not an 'evaluative' word, but that none of *the conditions under which* a series of sounds constitutes a promise involves any evaluative element.

[3] I will be assuming with Searle that 'ought'-statements have at least one sense in which they clearly ascribe obligations, and that this is the sense of their occurrence in Searle's derivation.

(5) Tom promised to kill Satan, but (as everyone except Tom knows) Satan does not exist.
(6) Tom ought to kill Satan, but (as everyone except Tom knows) Satan does not exist.

And the fact that a statement like (6) is contradictory whereas a statement like (5) is not contradictory suffices to show that, contrary to Searle's deducibility claim, "Tom promised to kill Satan" does not entail "Tom ought to kill Satan".

Vacuous names (that is, names without referents, such as "Satan") present an obvious problem for Searle's derivation. Since Atlantis is a non-existent city, one can be under no *obligation* to liberate Atlantis; but that does not make it impossible for one to *promise* to liberate Atlantis, if one mistakenly believes both that such a city exists and that it is unjustly occupied. Since Sherlock Holmes is only a fictional detective, one can be under no *obligation* to hire Sherlock Holmes for the investigation of a robbery; but that does not make it impossible for someone to *promise* to the victims of a robbery to hire Sherlock Holmes for the investigation of that robbery, if one mistakenly believes, or simply wants to lead *others* to mistakenly believe, that Sherlock Holmes is a real detective. Beliefs in the existence of non-existent entities may sometimes be sufficient to make a *promise* involving those entities possible, but they are never sufficient to make an *obligation* involving those entities possible.

Vacuous names, however, are not the only, or the commonest, kinds of expression that show Searle's deducibility claim to be untenable. What I shall here call *vacuous descriptions* (that is, definite descriptions that, as a matter of contingent fact, no object satisfies) give rise to a far wider, indeed in principle unlimited, range of counterexamples.[4] Suppose that Tom, confusing me with another person, incorrectly believes that I lost my wallet, even though the description "my lost wallet" is vacuous, since, in fact, I never *had* a wallet to lose. Now, if Tom says to me, "I hereby promise to find your lost wallet", I can without contradiction say,

(7) Tom promised to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else.

What I *cannot* say without contradiction, however, is,

(8) Tom ought to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else.

If I never had, and so I never lost, a wallet, then it cannot be the case that Tom *ought* to find my lost wallet, but it *can* be the case that Tom, confused as he was, has *promised* to me to find my lost wallet. And since, in that circumstance, "Tom

[4] In what follows, I am assuming the 'attributive' reading of definite descriptions, whose existence no one, as far as I know, doubts. I note in passing that Searle belongs to those who believe, in addition, that the 'attributive' reading of descriptions is the *only* semantically (as opposed to pragmatically) available one. See Searle (1979).

promised to find my lost wallet" will be true even though "Tom ought to find my lost wallet" is not true, the former statement cannot entail the latter, contrary to what Searle's deducibility claim requires. Similarly, suppose that Tom, mistakenly believing that someone killed Mary, says, "I hereby promise to arrest the person who killed Mary" even though the description "the person who killed Mary" is vacuous, since, in fact, Mary is alive and well. I could then without contradiction say,

(9) Tom promised to arrest the person who killed Mary, but no one killed Mary—in fact, Mary is alive and well.

What I could *not* say without contradiction, however, is,

(10) Tom ought to arrest the person who killed Mary, but no one killed Mary—in fact, Mary is alive and well.

If Mary is alive, then it cannot be the case that Tom *ought* to arrest the person who killed Mary, but that does not prevent Tom from mistakenly believing that Mary is dead, and so from *promising* to arrest the person responsible for her death. And since, in these circumstances, "Tom promised to arrest the person who killed Mary" may be true even though "Tom ought to arrest the person who killed Mary" cannot be true, the former of these statements cannot entail the latter, contrary to what Searle's deducibility claim requires. In short, the existence of vacuous descriptions, no less than the existence of vacuous names, makes the inference from "X promised to p" to "X ought to p" invalid, and thus shows Searle's fundamental claim that promises necessarily create obligations to be untenable. Indeed, since the vast majority of linguistically possible definite descriptions are such that merely knowing what they mean does *not* amount to knowing whether or not they are vacuous, I suspect that the number of promises that have been exchanged in the course of human history without *any* obligations having thereby been created must be staggering.

The pertinent generalization could be formulated as follows. Call 'content-clause' the clause p that, in statements of the form "X promised to p" and "X ought to p", specifies the content of the promise or of the obligation that such statements ascribe. Then, if the content-clause p contains a proper name or a definite description and is such that, when it occurs *un*-embedded, it is false whenever that name or that description are vacuous, the effect of *embedding* p in matrix sentences of the forms "X promised to p" and "X ought to p" is necessarily the falsity of the latter matrix sentence but not necessarily the falsity of the former matrix sentence. Assuming *one* of the senses of the distinction between extensionality and non-extensionality that Searle himself recognizes[5] (the sense in which a sentence is extensional with respect to an item it contains if and only if it entails that the purported referent of that item exists), this can be re-expressed as follows: if

[5] See Searle 1983: 181ff; 2004: 174ff.

the content-clause of a promise-ascription or of an obligation-ascription is independently certifiable as extensional with respect to the names or descriptions it contains, then the promise-ascription is *non*-extensional with respect to those names or descriptions, even though the obligation-ascription *is* extensional with respect to those names or descriptions. In short, the difference between "X promised to $_{p}$[. . .r. . .]" and "X ought to $_{p}$[. . .r. . .]", where p is a content-clause and r a name or description inside that content-clause, is that, when p is extensional with respect to r, "X ought to $_{p}$[. . .r. . .]" is *also* extensional with respect to r, even though "X promised to $_{p}$[. . .r. . .]" is *not* extensional with respect to r.[6] The claim that promises necessarily create obligations results, then, from the failure to recognize that statements about obligations are extensional in a way in which statements about promises are non-extensional. For it is precisely that difference that makes it possible for statements of the latter kind to be true even though statements of the former kind are false, and it is, in turn, that possibility that prevents statements of the latter kind from entailing statements of the former kind.

It is strictly speaking unnecessary, but it may nevertheless be useful, to point out that the kind of problem exposed above could not be avoided by deploying the sort of strategy that Searle had used in rebutting certain early objections to his derivation that were appealing to the notion of '*prima facie* obligation'. Those objections were aiming to dispute Searle's claim that promises entail obligations by relying on the observation that the obligation a person assumes in making a promise may sometimes be in conflict with *other* obligations that that person independently has, and may consequently have to be discharged in order for such a conflict between obligations to be resolved. As Searle has pointed out, that observation, though correct in itself, cannot be construed as an *objection* to his derivation: two obligations cannot be in conflict unless they both *exist*, and an obligation cannot, logically, be discharged unless it is already in existence; so, the fact that an obligation incurred by an act of promising may have to be subsequently discharged in order for a conflict between obligations to be resolved has no tendency to show that it was non-existent *at the time* of promising; and the bringing into existence of an obligation by virtue an act of promising *at the time* of that act of promising is all that Searle's derivation, properly understood, requires.[7] Notice, however, that this sort of move would be entirely inapplicable to the family of counterexamples considered above. When a vacuous name or a vacuous description figures in the content-clause of a promise, *no obligation whatsoever* is created at the time of promising regarding anything mentioned in the content-clause of the promise (if, for example, I never lost, since I never had, a wallet, and Tom says to me, "I hereby promise to find your lost wallet," no obligation of Tom's to find my non-existent lost wallet is thereby created,

[6] Please note that I take no stand here on the question whether promise-ascriptions or obligation-ascriptions are extensional or non-extensional in senses of that distinction *other* than the one specifically explained above.

[7] Searle's more general misgivings about the standard interpretation of the notion of '*prima facie* obligation' are presented in detail in Searle (1978).

no matter what Tom might imagine); but an obligation that has never existed cannot, logically, be discharged, and the question whether the person making such a promise has or does not have any *other* obligations with which his *non-existent* promissory obligation might be in conflict makes no sense. Consequently, the cases considered above cannot be assimilated to cases where a promissory obligation does exist but is *subsequently* discharged because it conflicts with independently assumed obligations; and so, Searle's way of deflating objections to his derivation that rely on situations where conflicts between actually existing obligations are possible is of no use in solving the problem posed by the cases presented above.

Granting the obvious point that, in the cases presented above, no obligations are generated, someone (though, I presume, not Searle) might make a desperate final attempt to save Searle's derivation by claiming that, in the same cases, no promises are made either—in other words, that if, for example, Tom says to me, "I hereby promise to find your lost wallet," and it turns out that I never lost, since I never had, a wallet, it not only follows that Tom has not thereby acquired any obligation, but also that he has not even made any *promise*. That this is false is evident not only from the already noted fact that, in the described circumstances, Tom's utterance could without contradiction be reported by a statement such as "Tom promised to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else", but also from the fact that, in the same circumstances, Tom's utterance would be subject to a variety of forms of appraisal that would be unintelligible *except* on the assumption that it *was* the utterance of a promise. For example, knowing better than Tom whether or not I ever owned a wallet, I might respond to his utterance of "I hereby promise to find your lost wallet" by truly and intelligibly saying to him, "Well, Tom, the promise you just made cannot be fulfilled, since I never had, and so I never lost, a wallet." But that perfectly true and intelligible remark would have to be counted either as just unintelligible or as blatantly false if Tom had *not* made the promise to which I would be referring, and if, in particular, the proposition that I never had, and never lost, a wallet had the power to entail the proposition that Tom did not *promise* to find my lost wallet. What the proposition that I never had, and never lost, a wallet does entail is, precisely, that Tom cannot *fulfil* the promise he has made, not that he cannot *make* the promise he has made. But fulfilling a promise and making it are two quite different things, and the fact that a promise cannot be fulfilled does not imply that it is not a promise, any more than the fact that an assertion cannot be true implies that it is not an assertion, or the fact that a request cannot be complied with implies that it is not a request. Rather, a promise can be fulfilled *or* unfulfilled only if it is already in existence, just as an assertion can be true or false only if it is already in existence, and a request can be complied or not complied with only if it is already in existence. Indeed, Searle's familiar and fundamental distinction between the *success* of an illocutionary act and the *satisfaction* of its propositional content (see, e.g., Searle 1991) was designed precisely in order to acknowledge these elementary facts; but none of these elementary facts could be acknowledged if the desperate strategy now under consideration was to be systematically applied, since its

application would have the effect of systematically misdescribing what are, in fact, conditions of propositional content satisfaction *as if* they were conditions of illocutionary act success. It would require claiming, for example, that the existence of a King of Germany is not only a condition on the *fulfillment* of the promise to arrest the King of Germany, but also a condition on the *existence* of the promise to arrest the King of Germany; that it is not only a condition on the *truth* of the assertion that the King of Germany has been arrested, but also a condition on the *existence* of the assertion that the King of Germany has been arrested; that it not only a condition on *complying with the request* that the King of Germany be arrested, but also a condition on *making the request* that the King of Germany be arrested; and so on. Since no one (and least of all Searle) would presumably be prepared to accept any of these claims, it should readily be granted that the presence of vacuous names or vacuous descriptions in the content-clause of a statement ascribing a promise does *not* necessitate the falsity of that statement. And since, on the other hand, the presence of vacuous names or vacuous descriptions in the content-clause of a corresponding statement ascribing an obligation does necessitate the falsity of that statement, there is, as far as I can see, no escape from the conclusion that ascriptions of promises do not *entail* ascriptions of obligations.

4. CONCLUSION

I have first argued that, even if it were true, as Searle claims, that statements about promises entail statements about obligations, it would not thereby follow, as he also claims, that non-evaluative statements entail evaluative ones (and that, therefore, the "naturalistic fallacy" is not a fallacy), since (as the instability of his own categorizations indicates) the conditions of application of the labels 'evaluative' and 'non-evaluative' to statements about promises or to statements about obligations are far more theory-sensitive than they would have to be if the latter claim was to follow from the former one. Concentrating, then, on the former claim alone, I have argued that it is not in fact the case that statements about promises entail statements about obligations, since statements about promises can be true whereas corresponding statements about obligations are false when vacuous names or vacuous descriptions occur in the clauses that specify the content of the promises or of the obligations that such statements ascribe—since, in other words, statements about promises can be non-extensional whereas statements about obligations are extensional with respect to names or descriptions appearing in their respective content-clauses.

As indicated in the introductory section, the claim that statements of the form "X promised to p" entail corresponding statements of the form "X ought to p", apart from its early involvement in Searle's discussion of the "naturalistic fallacy," occupies a central place in the accounts of sociality and rationality recently developed in Searle's *The Construction of Social Reality* and *Rationality in Action*, where it is offered as a particularly clear instance of certain general

theses that are fundamental to those accounts (in particular, of the thesis that linguistic acts are the foundations of many extralinguistic rights and obligations, and of the thesis that the rights and obligations thus generated are prime examples of desire-independent as opposed to desire-dependent reasons for action). In arguing that Searle's specific claim about promises and obligations cannot, at least as formulated, be sustained, I do not, of course, mean to imply that the above-mentioned theses, or some related theses advanced in these profound and wide-ranging works, should be discarded. On the contrary, I am convinced of their significance, and regard the present discussion as a preliminary step in seeking better ways of articulating their content and implications. If one abandons the dogmatic policy of accepting as worthy of one's theoretical attention only those inferential relations that can be *deductively* validated, one will have no difficulty in acknowledging the obvious fact that people do have the tendency to defeasibly infer "X ought to p" from "X promised to p", in a way in which, for example, they do not have the tendency to infer, even defeasibly, "X ought to p" from "X expressed the wish to p". The problem, then, is to explain, in a naturalistic spirit, what it is that warrants inferences of the former sort without warranting inferences of the latter sort, *even though* neither the former nor the latter conform to deductive canons. And I see no reason to suppose that, once that naturalistic explanation is had, it will not cohere with Searle's fundamental ideas about sociality and rationality. For, as Searle would probably agree, sociality and rationality are not entirely, or even largely, deductive affairs.[8]

REFERENCES

Searle, J.R. (1964). How to derive 'ought' from 'is'. *The Philosophical Review*, 73: 43–58.

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J.R. (1978). *Prima facie* obligations. In J. Raz (ed.), *Practical Reasoning*. Oxford: Oxford University Press.

Searle, J.R. (1979). Referential and attributive. *The Monist*, 62: 190–208.

Searle, J.R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, J.R. (1991). Meaning, intentionality, and speech acts. In E. Lepore and R. Van Gulick (eds.), *John Searle and his Critics*, Oxford: Blackwell.

Searle, J.R. (1995). *The Construction of Social Reality*. New York: The Free Press.

Searle, J.R. (1997). Responses to critics of *The Construction of Social Reality*. *Philosophy and Phenomenological Research*, 57: 449–458.

Searle, J.R. (1998). *Mind, Language, and Society*. New York: Basic Books.

Searle, J.R. (1999). Fact and value, 'is' and 'ought', and reasons for action. In G.O. Mazur (ed.), *Twenty-Five Year Commemoration to the Life of Hans Kelsen (1898–1973)*, New York: Semenenko Foundation.

Searle, J.R. (2001). *Rationality in Action*. Cambridge, MA: MIT Press.

Searle, J.R. (2004). *Mind: A Brief Introduction*. Oxford: Oxford University Press.

---

[8] I am grateful to Margaret Gilbert and Leo Zaibert for their helpful comments on an earlier draft.

INDEX

THEORY AND DECISION LIBRARY

SERIES A: PHILOSOPHY AND METHODOLOGY OF THE SOCIAL SCIENCES

Editors: W. Leinfellner (*Vienna*) and G. Eberlein (*Munich*)

1. G. Zecha and P. Weingartner (eds.): *Conscience: An Interdisciplinary View*, Salzburg Colloquium on Ethics in the Sciences and Humanities (1984). 1987 ISBN 90-277-2452-0
2. R.W. Scholz: *Cognitive Strategies in Stochastic Thinking*. 1987 ISBN 90-277-2454-7
3. H. Nurmi: *Comparing Voting Systems*. 1987 ISBN 90-277-2600-0
4. M. Schmid and F.M. Wuketits (eds.): *Evolutionary Theory in Social Science*. 1987 ISBN 90-277-2612-4
5. C.L. Sheng: *A New Approach to Utilitarianism*. A Unified Utilitarian Theory and Its Application to Distributive Justice. 1991 ISBN 0-7923-1301-1
6. F. Forman: *The Metaphysics of Liberty*. 1989 ISBN 0-7923-0080-7
7. G. Bernard: *Principia Economica*. 1989 ISBN 0-7923-0186-2
8. J.W. Sutherland: *Towards a Strategic Management and Decision Technology*. Modern Approaches to Organizational Planning and Positioning. 1989 ISBN 0-7923-0245-1
9. C. Vlek and G. Cvetkovich (eds.): *Social Decision Methodology for Technological Projects*. 1989 ISBN 0-7923-0371-7
10. P. Hoyningen-Huene and F.M. Wuketits (eds.): *Reductionism and Systems Theory in the Life Sciences*. Some Problems and Perspectives. 1989 ISBN 0-7923-0375-X
11. K.G. Grunert and F. Ölander (eds.): *Understanding Economic Behaviour.* 1989 ISBN 0-7923-0482-9
12. G. Antonides: *The Lifetime of a Durable Good*. An Economic Psychological Approach. 1990 ISBN 0-7923-0574-4
13. G.M. von Furstenberg (ed.): *Acting under Uncertainty.* Multidisciplinary Conceptions. 1990 ISBN 0-7923-9063-6
14. R.L. Dukes: *Worlds Apart*. Collective Action in Simulated Agrarian and Industrial Societies. 1990 ISBN 0-7923-0620-1
15. S.O. Funtowicz and J.R. Ravetz: *Uncertainty and Quality in Science for Policy*. 1990 ISBN 0-7923-0799-2
16. J. Götschl (ed.): *Erwin Schrödinger's World View*. The Dynamics of Knowledge and Reality. 1992 ISBN 0-7923-1694-0
17. R. von Schomberg (ed.): *Science, Politics and Morality*. Scientific Uncertainty and Decision Making. 1993 ISBN 0-7923-1997-4
18. J. Baron: *Morality and Rational Choice.* 1993 ISBN 0-7923-2276-2
19. M. Allais and O. Hagen (eds.): *Cardinalism*. A Fundamental Approach. 1994 ISBN 0-7923-2398-X
20. E. Carlson: *Consequentialism Reconsidered*. 1995 ISBN 0-7923-3547-3
21. J. Götschl (ed.): *Revolutionary Changes in Understanding Man and Society*. 1995 ISBN 0-7923-3627-5
22. A. Camacho: *Division of Labor, Variability, Coordination, and the Theory of Firms and Markets.* 1996 ISBN 0-7923-3832-4
23. R. Hegselmann, U. Mueller and K.G. Troitzsch (eds.): *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. 1996 ISBN 0-7923-4125-2

24. J. Nida-Rümelin: *Economic Rationality and Practical Reason*. 1997 ISBN 0-7923-4493-6
25. G. Barbiroli: *The Dynamics of Technology*. A Methodological Framework for Techno-Economic Analyses. 1997 ISBN 0-7923-4756-0
26. O. Weinberger: *Alternative Action Theory*. Simultaneously a Critique of Georg Henrik von Wright 's Practical Philosophy. 1998 ISBN 0-7923-5184-3
27. A. Carsetti (ed.): *Functional Models of Cognition*. Self-Organizing Dynamics and Semantic Structures in Cognitive Systems. 2000 ISBN 0-7923-6072-9
28. J. Nida-Rümelin and W. Spohn (eds.): *Rationality, Rules, and Structure*. 2000 ISBN 0-7923-6326-4
29. J. Klüver: *The Dynamics and Evolution of Social Systems*. New Foundations of a Mathematical Sociology. 2000 ISBN 0-7923-6443-0
30. V.A. Lefebvre: *Algebra of Conscience*. Revised Edition with a Second Part with a new Foreword by Anatol Rapoport. 2001 ISBN 0-7923-7121-6
31. J. Götschl (ed.): *Evolution and Progress in Democracies.* Towards New Foundations of a Knowledge Society. 2001 ISBN 1-4020-0063-4
32. N.J. Saam and B.Schmidt (eds.): *Cooperative Agents.* Applications in the Social Sciences. 2001 ISBN 1-4020-0190-8
33. B. Verbeek: I*nstrumental Rationality and Moral Philosophy*. An Essay on the Virtues of Cooperation. 2002 ISBN 1-4020-0639-X
34. J. Klüver: *An Essay Concerning Sociacultural Evolution*. Theoretical Principles and Mathematical Models. 2002 ISBN 1-4020-0750-7
35. T.J. Fararo and K.Kosaka: *Generating Images of Stratification*. A Formal Theory. 2003 ISBN 1-4020-1500-3
36. O. Dahlbäck: *Analyzing Rational Crime –Models and Methods*. 2003 ISBN 1-4020-1657-3
37. E. Oeser: *Evolution and Constitution*. The Evolutionary Selfconstruction of Law. 2003 ISBN 1-4020-1784-7
38. A. Carsetti (ed.): *Seeing, Thinking and Knowing. Meaning and Self-Organisation in Visual Cognition a Thought.* 2004 ISBN 1-4020-2080-5
39. N. Gontier, J.P.v. Bendegem, D. Aerts (eds.): *Evolutionary Epistemology, Language and Culture. A Non-Adaptationist, Systems Theoretical Approach*. 2006 ISBN 1-4020-3394-X
40. J. Klüver, C. Klüver: *On Communication. An Interdisciplinary and Mathematical Approach.* 2007 ISBN 1-4020-5463-7
41. S.L. Tsohatzidis (ed.): *Intentional Acts and Institutional Facts*. Essays on John Searle's Social Ontology. 2007 ISBN 978-1-4020-6103-5