

REALI
Essays in the Phil

MA
Univers

PE
Univers

KA
Univers

Realism in Action
by leading repres
theory and phi
of the social sci
of social action,
sophy of scienc
causes in action
and folk-psych
of the topics
participants. A
trust, commitm
the possibility
individual soc
social groups
intentionality,
common know
sciences inclu
Verstehen con
constructivist
reconstruction

KLUWER
DOR

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

Editor-in-Chief:

VINCENT F. HENDRICKS, *Roskilde University, Roskilde, Denmark*
JOHN SYMONS, *University of Texas at El Paso, U.S.A.*

Honorary Editor:

JAAKKO HINTIKKA, *Boston University, U.S.A.*

Editors:

DIRK VAN DALEN, *University of Utrecht, The Netherlands*
THEO A.F. KUIPERS, *University of Groningen, The Netherlands*
TEDDY SEIDENFELD, *Carnegie Mellon University, Pittsburgh, PA, U.S.A.*
PATRICK SUPPES, *Stanford University, California, U.S.A.*
JAN WOLEŃSKI, *Jagiellonian University, Kraków, Poland*

VOLUME 321

REALISM IN ACTION

Essays in the Philosophy
of the Social Sciences

edited by

MATTI SINTONEN
*University of Tampere,
Finland*

PETRI YLIKOSKI
*University of Helsinki,
Finland*

and

KAARLO MILLER
*University of Helsinki,
Finland*



KLUWER ACADEMIC PUBLISHERS
DORDRECHT / BOSTON / LONDON

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

ISBN 1-4020-1667-0

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

All Rights Reserved

© 2003 Kluwer Academic Publishers

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.

Printed in the Netherlands.

TABLE OF CONTENTS

FOREWORD	vii
1. REALISM, TRUTH, AND EXPLANATION	
FREDERICK STOUTLAND / What Philosophers Should Know about Truth and the Slingshot	3
WENCESLAO J. GONZALEZ / From <i>Erklären-Verstehen</i> to <i>Prediction-Understanding</i> : The Methodological Framework in Economics	33
USKALI MÄKI / The Archaeological Construction of the Past: Some Realist Moderations	51
GABRIEL SANDU / The Backward Induction Paradox and Epistemic Logic	69
2. PHILOSOPHY OF MIND AND ACTION THEORY	
ROBERT AUDI / The Scope of Motivation and the Basis of Practical Reason	85
MYLES BRAND / Activity and Passivity	99
REX MARTIN / Reasons and Causes: The Case of Collingwood	113
ALFRED MELE / Intending and Trying: Tuomela vs. Bratman at the Video Arcade	129
JUHANI PIETARINEN / Spinoza on Causal Explanation of Action	137
MARTTI KUOKKANEN / On the Structuralist Constraints in the Explanation Scheme of Folk Psychology	155
3. INTENTIONS, TRUST AND SOCIAL INSTITUTIONS	
KAARLO MILLER / Commitments	169
MAJ TUOMELA / The Components of Rational Trust	179
CRISTIANO CASTELFRANCHI / Grounding We-intentions in Individual Social Attitudes	195
FRANK HINDRIKS / Social Groups, Collective Intentionality, and Anti-Hegelian Skepticism	213
SEUMAS MILLER / Social Institutions	233
GEORG MEGGLE / Common Belief and Common Knowledge	251
PEKKA MÄKELÄ and PETRI YLIKOSKI / Others Will Do It: Social Reality by Opportunists	259
ILKKA NIINILUOTO / Science as Collective Knowledge	269

WHAT PHILOSOPHERS SHOULD KNOW ABOUT
TRUTH AND THE SLINGSHOT¹

A champion came out from the Philistine camp, a man named Goliath; he was over nine feet in height...When the Philistine began moving towards him, David ran quickly to engage him. He put his hand into his bag, took out a stone, slung it, and struck the Philistine on the forehead. The stone sank into his forehead, and he fell flat on his face on the ground. So David proved the victor with his sling and stone; he struck Goliath down and gave him a mortal wound, though he had no sword. (First Samuel, chapter 17)

1. INTRODUCTION

The slingshot argument is so-called² because, like the sling David used to slay Goliath, it uses so little to accomplish so much – or so its defenders claim. It has been used to reject the claim that sentences designate propositions or states of affairs, to undermine concepts like “necessarily” or “because”, to show the futility of talk about mental or linguistic representations of reality, and much else besides. I am not going to discuss these uses of the argument, however, but only its use to show that the correspondence theory of truth – the theory that a sentence is true just in case it corresponds to a particular fact – should be rejected on the ground that, given a couple of unobjectionable premises, it can be proved that if a true sentence corresponds to any fact, it corresponds to every fact. That means that every true sentence corresponds to the same thing – which is to say there really is only one fact – which would be a mortal wound for the correspondence theory.

The inspiration for the slingshot is Frege and hence it is often called “The Frege argument”. Frege held that all three kinds of non-logical, truth-relevant expressions – singular terms, predicates, and sentences – have both *Sinn* and *Bedeutung*: each has a sense and each signifies³ (designates, means) something. In the case of a sentence (with a truth value) its sense is the thought one grasps in understanding it, while what it signifies depends only on whether it is true or false: if it is true, it signifies The True, and if it is false, it signifies The False. While non-synonymous sentences differ in their sense, all true ones signify the same thing, and this parallels precisely the conclusion of the slingshot that all true sentences correspond to the same fact. Needless to say, Frege did not hold a correspondence theory of truth, a major reason being that he put forward considerations analogous to those used in the slingshot. The argument itself, however, is not found in his work, but he nevertheless deserves credit for inspiring it and for creating the logical resources which made its formulation and evaluation possible.

The slingshot is a deductive argument, expressible in first order predicate logic with identity, which can be so formulated as to leave no question that, given the principles of inference used, its conclusion follows from its premises. This suffices

to convince some philosophers that it establishes straightaway that the correspondence theory of truth must be rejected. But it also suffices to convince others that the argument can be ignored because it must involve formal tricks with no bearing on philosophical questions like how to understand the concept of truth. Both attitudes are wrong: although the slingshot can be made rigorously valid, its conclusion can be evaded by challenging one or more of its assumptions, not all of which are by any means self-evident. At the same time, reflection on these assumptions sheds a great deal of light on philosophical questions, and anyone who thinks about issues like truth ought to know what is at stake in accepting or rejecting the argument and the assumptions it requires.

My aim in this paper is to state the argument as clearly as I can, focussing on two versions of it (which I shall call the "Davidson version" and the "Gödel version"), and to explain its significance for the correspondence theory of truth. The technical parts are derivative on the work of others⁴, my intention not being to develop that work, but to make it more accessible and to draw out the philosophical significance of alternative ways of stating the argument and evading its conclusion. This does not mean that I have no view of my own. I take my discussion to show that, while there are good reasons not to accept all the assumptions required by the slingshot (either the Davidson or the Gödel version), the correspondence theory of truth by no means escapes its effects. Some versions accept the assumptions the slingshot requires, and they are mortally wounded by it, while others reject them at the price of a correspondence theory so devoid of explanatory force as to be indistinguishable from a deflationist conception of truth.

2. BACKGROUND PRINCIPLES AND DISTINCTIONS

Before considering the argument, I want to discuss the principles on which its evaluation must turn. These principles are philosophically elementary, but it is easy to lose sight of them when considering the slingshot as a formal argument. The decisive thing is to get clear about the philosophical significance of various substitution principles used in different versions of the argument.

The slingshot is directed against any correspondence theory of truth which claims that a true sentence must correspond to some particular fact, where a fact is an entity whose existence makes true a sentence which corresponds to it. While accounts of facts differ in many ways, two things are essential, which Russell put as follows.⁵ First, facts must be "what they are whatever we may choose to think about them", which implies that we can refer to them in different ways and that how we refer to them makes no difference to what they are. Second, a fact must be "the sort of thing expressed by a whole sentence, not by a single name", where this latter point might be put by saying that a fact must be a sentence-like entity. For example, the sentence 'The author of Waverley lived in Scotland' is true not because it corresponds to the author of Waverley or to Scotland but because it corresponds to the fact that the author of Waverley lived in Scotland, where "the fact that" must be followed by a (true) sentence.

The correspondence theory holds that a true sentence must correspond to some *particular* fact because simply corresponding to the facts is far too general. A false

sentence may very well correspond to *some* fact, but it is false because it does not correspond to the right fact – the fact which would make it true – and it is this notion of each true sentence corresponding to a particular fact which the slingshot purports to undermine. It argues that, given that 'the author of Waverley lived in Scotland' corresponds to the fact that the author of Waverley lived in Scotland, then it also corresponds to the fact that Truman lived longer than Roosevelt, that sugar maples turn red in the fall, that Strindberg was born in Stockholm, etc., where one can put any true sentence after 'the fact that'. This could be taken as the claim that any true sentence corresponds to *every* fact, but it can also be taken as the claim that every true sentence corresponds to the *same* fact, since if a true sentence corresponds to the fact that p, where 'p' is any true sentence, then any other true sentence also corresponds to the fact that p. But if every true sentence corresponds to the same fact, then there is only one fact, which rules out any notion of a true sentence corresponding to a particular fact.

If correspondence is a relation between a true sentence and a particular fact, then we must be able to refer both to sentences and to facts. A sentence is what is true (or false), and we refer to a sentence, say 'Strindberg was born in Stockholm', in various ways: by quoting it (as I just did), by describing it ('the last sentence in the previous paragraph'), by nominalizing it ('that Strindberg was born in Stockholm'), and so on. How we refer to sentences is not crucial to our discussion, nor is it crucial that we speak of sentences, rather than statements, utterances, or propositions, as what are true (or false), and I will speak of sentences because that is customary among logicians when constructing formal arguments.⁶

How we refer to facts, however, is crucial. Because facts are sentence-like, sentences must figure in referring to them, and because facts are entities, we must be able to refer to them by using singular terms. This has often led defenders of the correspondence theory to conclude that sentences are themselves singular terms, but a more plausible version is as follows. Take a paradigm correspondence claim:

A. The true sentence that Stockholm is a large city corresponds to the fact that Stockholm is a large city.

It is natural to construe this as involving the relational predicate 'corresponds to' flanked by two singular terms, '*the true sentence that Stockholm is a large city*', which refers to a sentence, and '*the fact that Stockholm is a large city*', which refers to a particular fact. While there is nothing wrong with this way of construing the claim, it does not clarify the crucial role of sentences since it construes the claim as consisting only of a predicate and two singular terms. But it is not difficult to reconstrue the claim so that it consists of sentences and a sentence connective:

B. The true sentence that [*Stockholm is a large city*] corresponds to the fact that [*Stockholm is a large city*].

If we suppress the two sentences in brackets and replace them with variables, the result is a two-placed sentence connective – the 'correspondence connective':

C. The true sentence that p corresponds to the fact that q.

This enables us to reformulate the slingshot as purporting to show that, given obvious assumptions, whatever *true* sentences the correspondence connective connects (whatever true sentences we substitute for 'p' or 'q' in C), the result will be true. In other words, the correspondence connective, appearances to the contrary notwithstanding, is, defenders of the slingshot maintain, fully extensional in that any true sentence in its scope may be substituted *salva veritate* by any true sentence. It is, indeed, they maintain, a truth-functional connective: the truth value of sentences containing it depends only on the truth value of the sentences it connects.

Defenders of the correspondence theory must obviously deny that we get a true sentence no matter what true sentence we substitute for 'p' or 'q' in C. It is equally obvious that they cannot deny that we get a true sentence whenever we substitute the *same* (or a synonymous) true sentence for both 'p' and 'q' in C: the true sentence that p always corresponds to the fact that p. That sets up the crucial question: are there principles which permit the substitution *salva veritate* of *some* but not all true sentences for 'p' and for 'q' in C? For instance, given that the true sentence "Stockholm is a large city" corresponds to the fact that Stockholm is a large city, are there principles that permit the substitution *salva veritate* of *some* true sentences in "the fact that Stockholm is a large city" which do not also permit the substitution of *all* true sentences? Is there something *between* permitting the substitution of only synonymous sentences – in which case the correspondence connective is fully *intensional*⁷ – and permitting the unlimited substitution of true sentence – in which case the connective is fully *extensional*?

To consider that question, however, we must consider the substitution not only of sentences but also of the singular terms and predicates which are their constituents. A singular term is a term which, in a given context, signifies (or purports to signify) a particular individual. Examples are proper names like 'John' or 'Stockholm', pronouns like 'he', 'she' or 'it', demonstratives like 'this' or 'that', and definite descriptions like 'the capital of Sweden' or 'the fact that Stockholm is a large city'. The standard substitution principle for singular terms 'a' and 'b' is that 'a' may be substituted for 'b' *salva veritate* in a sentence just in case 'a' and 'b' refer to the same individual (so that $a = b$). For example, since Stockholm is (identical to) the Capital of Sweden, we may substitute one for the other *salva veritate*: if it is true that Stockholm is a large city, then it is true that the capital of Sweden is a large city. Or to use a classical example, since 'Cicero' and 'Tully' both refer to the same Roman statesman (Cicero = Tully), we may substitute 'Tully' for 'Cicero' in sentences like 'Cicero often gave speeches in Rome'.

The standard principle of substitution for singular terms is based on the sound idea that if a term is used to refer to a particular individual, then any other term which refers to that same individual will do as well. If it is true that Cicero often gave speeches in Rome, then the sentence 'Rome's greatest orator often gave speeches in Rome' will also be true, whether or not Cicero really was Rome's greatest orator, *provided* the point of using the definite description, 'Rome's greatest orator', was simply to refer to Cicero. We may, therefore, substitute co-referring singular terms *salva veritate* in any sentence in any context in which the terms simply refer to the same individual. That also includes the context of the correspondence connective, in which case, from:

D. The true sentence that Cicero often gave speeches in Rome corresponds to the fact that Cicero often gave speeches in Rome

we may infer:

E. The true sentence that Cicero often gave speeches in Rome corresponds to the fact that Tully often gave speeches in Rome.

A predicate contrasts with a singular term in not referring to a particular individual but in being *true of* an individual or set of individuals. The individuals of which a predicate is true comprise the *extension* of the predicate. 'Is a large city', for example is true of Stockholm, New York, London, and all other large cities, which comprise its extension. A predicate can be thought of as what remains of a sentence if we remove singular terms. Thus if we remove 'Stockholm' from 'Stockholm is a large city', the result is the predicate 'x is a large city', which logicians call an "open sentence" since replacing the variable with a singular term produces an ordinary ("closed") sentence. If we remove both singular terms from 'Stockholm is larger than Helsinki', we get the open sentence, 'x is larger than y', which is a relational predicate requiring (in this case) two singular terms to become a (closed) sentence. A quantified sentence like 'Some cities are dangerous' has predicates but no singular terms; it is taken to express that there is at least one individual which is in the joint extension of the predicates 'x is a city' and 'x is dangerous'. 'All cities are dangerous' is taken to express that if any individual is in the extension of 'x is a city', it is also in the extension of 'x is dangerous'.

According to this account, predicates do not *refer* but are rather *true of* various individuals, and hence it would be incoherent to formulate a substitution principle in terms of *co-referring* predicates. The substitution principle Frege proposed (which is valid in extensional contexts) is that predicates may be substituted *salva veritate* if they are *co-extensive* – that is, true of the very same individuals. We cannot substitute 'is a city' for 'is large' since their extensions are not the same: not every city is large nor is every large individual a city. We can, of course, substitute *synonymous* predicates ('is a pair of glasses' for 'is a pair of spectacles') *salva veritate* since they necessarily have the same extension. But the principle also permits the substitution of predicates which are co-extensive as a matter of contingent fact – predicates which as a matter of fact are true of the same individuals. Quine's nice example is 'is a creature with a heart' and 'is a creature with a kidney', which are co-extensive because there are no organisms which have a heart but do not have a kidney. Another example is 'lives in California' and 'lives in the most populous state in the USA'; still another is 'broke her arm in Nelspruit on January 6, 2000' and 'is a grandchild of mine born in 1996'.

It is clear that coextensive predicates may not be substituted *salva veritate* in the context of the correspondence connective, as can be seen by considering examples. For example, from:

The true sentence that Celina broke her arm in Nelspruit on January 6, 2000 corresponds to the fact that Celina broke her arm in Nelspruit on January 6, 2000

we cannot infer:

The true sentence that Celina broke her arm in Nelspruit on January 6, 2000 corresponds to the fact that Celina is a grandchild of mine born in 1996

even though those predicates are co-extensive in that both are true only of my granddaughter, Celina. This is particularly clear if we consider the converse of 'correspond', namely 'make true', for we surely cannot claim that what makes it true that Celina broke her arm in Nelspruit on January 6, 2000 is the fact that she is a grandchild of mine born in 1996. Nor could we claim that what makes it true that the governor of California lives in California is the fact that the governor of California lives in the most populous state.

A more general way of making this point is to note that, although co-extensive predicates, like 'has a heart' and 'has a kidney', are true of the same individuals, to predicate 'has a heart' of an individual is to predicate something quite different from predicating 'has a kidney', and even though any (living) creature which has the one will have the other, any surgeon better know the difference. Similarly, to say of Celina that she broke her arm in Nelspruit on January 6, 2000 is to say something quite different than to say of her that she is my grandchild both in 1996.

The latter example is of special relevance for our topic. It concerns predicates which are coextensive just because they are true of exactly one individual. They are significant because there are contexts in which, if such predicates can be substituted *salva veritate*, then any sentences with the same truth value may also be substituted *salva veritate*. The intuitive point is that if you can substitute predicates just because they are true of exactly one individual, then, *in a context where all that counts is that one individual*, you can also substitute any sentences with the same truth value. Such contexts (which have to be *devised*) are such that to permit the substitution of co-extensive predicates is *thereby* to permit the substitution of co-extensive sentences (that is, sentences with the same truth value). To put it in other terms: since predicates are open sentences, (closed) sentences are a special case of predicates, and hence there are contexts in which permitting the substitution of co-extensive predicates is *thereby* to permit the substitution of co-extensive sentences. This, as we shall see, is central to constructing the slingshot argument.

Let me sum up this discussion so far. Defenders of the correspondence theory cannot hold that the correspondence connective sets up a fully *extensional* context, for that would mean that *any* true sentences may be substituted *salva veritate* for 'p' or 'q' in this schema:

F. The true sentence that p corresponds to the fact that q.

But they do not have to accept the notion that the correspondence connective sets up a fully *intensional* context – which would require that the same (or a synonymous) true sentence must always be substituted for *both* 'p' and 'q' – the reason being that a context in which co-referring terms are substitutable *salva veritate* need not be a context in which true sentences are substitutable *salva veritate*. Given that 'Stockholm' and 'the capital of Sweden' refer to the same city, it does not follow from:

G. The true sentence that Stockholm is a large city corresponds to the fact that the capital of Sweden is a large city

that *any* true sentence may be substituted *salva veritate* for 'the capital of Sweden is a large city'. The principle of the substitution of co-extensive *predicates*, however, works quite differently: co-extensive predicates may not be substituted *salva veritate* in the context of the correspondence connective.

If this were all there is to say about principles of substitution, however, the slingshot would be utterly unpersuasive. What I take to be its crucial assumption is expressed by Davidson as the assumption that we should permit the "substitution of singular terms for others with the same extension,"⁸ and Davidson's use of the slingshot amounts to the claim that if we accept that principle of substitution in the context of the correspondence connective, we have thereby accepted the substitution *salva veritate* of sentences with the same truth value. But if the "substitution of singular terms for others with the same extension" is the same as the principle of the substitution of co-referring singular terms, then Davidson is claiming that any context which permits the substitution of co-referring singular terms *salva veritate* must be a fully extensional context and there is, as we just saw, no reason to accept that. We have been given no reason to think, for example that we cannot substitute *salva veritate* 'Tully' for 'Cicero', or 'the capital of Sweden' for 'Stockholm', in the context of the correspondence connective.

What this shows is that the notion of singular terms with the same *extension* cannot be the same as the notion of singular terms with the same *reference*, which suggests a complication in the notion of singular terms our discussion so far has ignored. I used 'singular term' to refer to names, pronouns, demonstratives and definite descriptions, but that blurs a distinction which is crucial to evaluating the slingshot. Names, pronouns, and demonstratives are (almost always) used as *referring* terms, that is, used simply to refer to a particular individual⁹: their use is, as it were, exhausted by their successful reference. Definite descriptions, on the other hand, have two distinct uses, which Keith Donellan called 'attributive' and 'referential':

A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion, on the other hand, uses the description to enable his audience to pick out whom or what he is talking about and state something about that person or thing. In the first case the definite description might be said to occur essentially, for the speaker wishes to assert something about whatever or whoever fits that description; but in the referential use the definite description is merely one tool for doing a certain job – calling attention to a person or thing – and in general any other device for doing the same job, another description or a name, would do as well. (Donellan, 1966, p. 285)

Donellan illustrates this distinction with the sentence "Smith's murderer is insane". If we utter it in the presence of the murdered Smith but do not know who Smith's murderer is, then we are saying that whoever murdered Smith is insane.¹⁰ That is an *attributive* use of the description, "Smith's murderer". On the other hand, if we utter it in the presence of the person we believe to be Smith's murderer, simply in order to

say of that particular person that she is insane, then what we are saying is true as long as that particular person is insane, whether or not she was Smith's murderer. That is a *referential* use of the description, "Smith's murderer".

If definite descriptions are used referentially, then they (like names, pronouns, and demonstratives) conform to the principle of substitution for co-referring singular terms. I illustrated this above when I argued that we could substitute 'Rome's greatest orator' for 'Cicero', whether or not Cicero was Rome's greatest orator, *provided* we used the terms simply to refer to the particular man Cicero. However, if definite descriptions are used *attributively*, the principle of substitution of co-referring singular terms is not acceptable because there will be contexts (including that of the correspondence connective) in which definite descriptions cannot be substituted *salva veritate*. The reason is that when used attributively, definite descriptions are used not only (and not primarily) to *refer* but to *describe*. As Donellan put it in the quotation above, "A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so." An utterance of 'Smith's murderer is insane' which involves an *attributive* use of the definite description 'Smith's murderer', is true, therefore, only if whoever is Smith's murderer is insane.

This means that substitution of definite descriptions which function attributively is tantamount to the substitution of *predicates*. But we have seen that the principle for the substitution of co-extensive predicates works very differently from the principle for the substitution of co-referring singular terms. Whereas co-referring singular terms may be substituted *salva veritate* in the context of the correspondence connective, co-extensive predicates may not be. We must, therefore, distinguish between a principle for the substitution of definite descriptions as used attributively – which involves *predicate* substitution – and a principle for the substitution of definite descriptions as used referentially – which involves no predicate substitution. Insofar as names, pronouns, and demonstratives are (as they generally are) used referentially, the latter principle will, of course, also apply to them.

This distinction is obscured by Davidson's reference to the "substitution of singular terms for others with the *same extension*".¹¹ Predicates have extension, though (as Davidson and I agree) not reference, and singular terms can also be spoken as having extension, whether used referentially (the extension is the particular individual actually referred to) or attributively (the extension is whichever individual the description happens to fit). But to leave it at that blurs the distinction between referential and attributive uses of definite descriptions, which obscures the point that, although the principle of substitution for co-referring singular terms is acceptable in any context, it does not follow that the principle applies to definite descriptions in their *attributive* use. As we shall see, definite descriptions in their attributive use are substitutable *salva veritate* only in extensional contexts, which is the decisive point for evaluating the slingshot.

The distinctions just made are, from the point of view of formal logic, *pragmatic* points and hence cannot be incorporated directly into formal arguments like the slingshot. But they must be reflected in the formal arguments if they are to be relevant to philosophical issues like truth, and to help with this, I will make some terminological stipulations. From now on, I will use 'singular term' as the generic term for names, pronouns, demonstratives, and definite descriptions. There will be

two species of the genus: names and definite descriptions. I will use 'names' to cover referring terms in the strict sense, hence not only names proper, but pronouns, demonstratives, and definite descriptions understood *referentially*. This means that the principle I called the substitution principle for co-referring singular terms is re-named the substitution principle for co-referring names. I will use 'definite descriptions' to cover singular terms as used *attributively* rather than referentially, and the principle of substitution for co-referring names will not apply to them for reasons just given.

The canonical expression for a definite description in this restricted sense will be the iota sign: ' ιx '. 'The capital of Sweden' will be symbolized as Russell did: ' $(\iota x)Fx$ ', to be read as 'the unique individual x such that x is the capital of Sweden'. The principle of substitution for definite descriptions, abbreviated ' ι -SUB' ('iota-substitution'; it will be formulated later) will specify the conditions under which, given that $(\iota x)Fx = (\iota x)Gx$, it is acceptable to substitute ' $(\iota x)Fx$ ' for ' $(\iota x)Gx$ ' in a sentential context. The principle will, of course, function differently from the principle of substitution for co-referring names since the latter (unlike the former) never involves the substitution of co-extensive predicates.

3. THE DAVIDSON VERSION OF THE SLINGSHOT

The most discussed version of the slingshot (which I will call "Davidson's version") was first published in Church's review of Carnap's *Introduction to Semantics* in the 1943 *Philosophical Review*. (Church, 1943) Carnap had proposed replacing Frege's notion that sentences signify truth values with the notion that they signify propositions, his aim being to reject the unsettling notion that what a sentence signifies depends only on its truth value. Church argued that, on Carnap's own assumptions, if a sentence signifies any proposition, then every true sentence signifies the same proposition, which means that Carnap's connective, "the sentence S signifies proposition p ", is in fact truth-functional and hence no advance over the Fregean idea that sentences signify truth values.

This form of argument has since been used to make analogous criticisms of other proposed connectives, notably by Quine in his criticism of quantified modal logic and by Davidson in his critique of the correspondence theory of truth, which is our subject. Davidson first used the argument in his "Truth and Meaning" (1967)¹² to undermine the proposal that we might think of the meaning of a sentence as something the sentence signifies, and then think of the meaning of singular terms and predicates as the contribution they make to sentence meaning thus construed. That proposal is a non-starter, he claimed, because, "given two reasonable assumptions", every true sentence signifies the same thing, and so does every false sentence, so that the proposal would amount to the absurd claim that "all sentences alike in truth value must be synonymous". The two "reasonable assumptions" referred to are "that logically equivalent singular terms [sic] have the same reference, and that a singular term does not change its reference if a contained singular term is replaced by another with the same reference."

Davidson next used the argument in his paper on "Causal Relations"¹³, also published in 1967, to argue that the notion of a *causal* connective is incoherent

because we can show that it too must be truth functional in that any true sentences can be substituted *salva veritate* for 'p' and 'q' in the schema, 'The fact that p was the cause of the fact that q'. To show that, we need the same two reasonable assumptions, though now reformulated. The assumption that logically equivalent singular terms have the same reference is reformulated as the principle that logically equivalent sentences may be substituted *salva veritate*. The assumption that "a singular term does not change its reference if a contained singular term is replaced by another with the same reference" is reformulated as the principle (discussed in the previous section) that "substitution of singular terms for others with the same extension" is always permissible.

Davidson appealed to these two assumptions again in "True to the Facts"¹⁴ (1969), where he first used the slingshot in discussing truth, and where he used it twice. Its first use was to undermine a deflationary conception of truth which attempts to show, by using the principle, ' $(\forall p)(\text{the sentence that } p \text{ is true} \leftrightarrow p)$ ', that the truth predicate is redundant. Davidson argued that the principle would enable us to eliminate 'is true' only if its variable ranges over entities "that sentences must be construed as naming," but, he went on, "There are very strong reasons, as Frege pointed out, for supposing that if sentences, when standing alone or in truth functional contexts, name anything, then all true sentences name the same thing", which shows that a redundancy theory based on (objectual) quantification can't work.

Davidson's other use of the argument in "True to the Facts" dealt explicitly with the role of facts in this schema:

The sentence that p corresponds to the fact that q

which leads to a correspondence theory of truth, he noted, if we add that "a statement is true if there is a fact to which it corresponds." (p. 41) The problem with the schema is to determine what we could substitute for 'p' and 'q', the difficulty being that if we substitute anything other than the same sentence for each, then any true sentence will do, so that "if a sentence corresponds to one fact, it corresponds to all." Then he states the argument itself, using the same two assumptions:

...If a statement corresponds to the fact described by an expression of the form 'the fact that p', then it corresponds to the fact described by 'the fact that q' provided either (1) the sentences that replace 'p' and 'q' are logically equivalent, or (2) 'p' differs from 'q' only in that a singular term has been replaced by a coextensive singular term. The confirming argument is this. Let 's' abbreviate some true sentence. Then surely the statement that s corresponds to the fact that s. But we may substitute for the second 's' the logically equivalent '(the x such that x is identical with Diogenes and s) is identical with (the x such that x is identical with Diogenes)'. Applying the principle that we may substitute coextensive singular terms, we can substitute 't' for 's' in the last quoted sentence, provided 't' is true. Finally, reversing the first step we conclude that the statement that s corresponds to the fact that t, where 's' and 't' are any true sentences. (1984, p. 42)

As Searle notes, the argument is given with "breathtaking speed" (Searle, 1985, p. 221) but before spelling it out, let me say a few more words about Davidson's attitude to the correspondence theory of truth.

In "True to the Facts", Davidson defended a correspondence theory in spite of endorsing the slingshot. What he took the slingshot to show was that appeal to *facts* has no place in an account of truth because, given the two reasonable assumptions, "Descriptions like 'the fact that there are stupas in Nepal', if they describe at all, describe the same thing: The Great Fact. No point remains in distinguishing among various names of The Great Fact when written after 'corresponds to'", which amounts to "ontological collapse" as far as the notion of facts is concerned. (1984, 42f.) To respond to this by rejecting the two assumptions is to "leave the frying pan of extensionality for the fires of intension", which means there would be precisely as many facts as there are distinct true sentences, so that the correspondence theory could claim no more than what makes it true that p is the fact that p. While there is every reason to accept that claim, there is no reason to think it could explain anything – either what it is for a sentence to be true or why a sentence is true – and hence no reason to think it supports a correspondence theory of truth.

What Davidson defended in 'True to the Facts', therefore, was a correspondence theory which makes no use of facts. He took this to be possible on the basis of Tarski's truth theory, which explains truth in terms of satisfaction and explains satisfaction as a relation between expressions and sequences of objects. The fact that in Tarski's theory, true sentences are satisfied by *all* sequences is thus a reflection of what the slingshot shows, but the fact that non-sentential expressions can be satisfied by some but not all sequences permits an explanatorily relevant specification of the relation of language and world.

The semantic concept of truth as developed by Tarski deserves to be called a correspondence theory because of the part played by the concept of satisfaction; for clearly what has been done is that the property of being true had been explained, and non-trivially, in terms of a relation between language and something else.... All true sentences end up in the same place (as satisfied by all sequences), but there are different stories about how they got there; a semantic theory of truth tells the story for a particular sentence by running through the steps of the recursive account of satisfaction appropriate to the sentence. (1984, 48f.)¹⁵

Davidson soon ceased to defend the correspondence theory of truth in any form, the break becoming explicit in "A Coherence Theory of Truth and Knowledge" (1983, reprinted in LePore, 1986). In a sense, this was only a terminological change, because the main reason he gave for rejecting the correspondence theory was that it thinks of sentences as representations, which makes sense only if there are facts to be represented, which the slingshot shows to be untenable. Since this is the view he held in "True to the Facts", where he defended the theory ("it deserves not elimination but elaboration"(p. 54)), it looks as if he has only changed the label on the ground that it is misleading to call a conception of truth which makes no use of facts a correspondence theory. But the change also reflected greater appreciation of the significance of the fact that Tarski's theory defines truth only in the context of a particular language and hence yields a different definition of the truth predicate for each language and, moreover, has nothing to say about what each of those differently defined predicates has in common. What conception of truth Davidson presently holds, given his firm rejection not only of the correspondence theory but of epistemic theories of truth is another question, which I will not pursue here.

Let me now give a precise formulation¹⁶ of Davidson's version of the slingshot, which presents it as a formally valid argument proving that if we accept *any* instance of the scheme, 'the sentence that p corresponds to the fact that q' we must, given a couple of "reasonable assumptions", accept *every* instance of the scheme which results from substituting any true sentence for 'p' and any true sentence for 'q'. We can accomplish the same end more smoothly if we take 'p' to stand for a particular true sentence (for example, 'Stockholm is a large city' – abbreviated as 'S'), and hence use 'S' instead of the variable 'p' after 'the sentence that', thus construing the scheme as expressing a *one-place* sentence connective. Hence I define 'C' (the "correspondence connective") as follows:¹⁷

'Cq' =_{df.} 'The true sentence S corresponds to the fact that q'

The task then is to show that, given the reasonable assumptions, we may, in the context of C, substitute *salva veritate* any true sentence for 'q', which amounts to saying that, whatever anyone may have thought, the connective is fully extensional and hence useless in a substantive account of truth. I will follow Davidson's statement of the argument, abbreviating 'the x such that x is identical with Diogenes', as 'the x such that Fx' (so that 'Fx' stands for 'x is identical with Diogenes') and using the iota sign '(ιx)' for 'the x such that'.¹⁸

The inference rules for this version of the slingshot are the standard rules of predicate logic with the addition of two substitution principles, which correspond to Davidson's two assumptions. The first permits the substitution, in the context of C, of logically equivalent sentences – sentences which have the same truth value in every model (or possible situation) – a principle I abbreviate as SUB-LE, whose formal statement is as follows:

SUB-LE: given that 'p ↔ q' holds just in case 'p' and 'q' have the same truth-value in every model, then from 'p ↔ q' and 'Σ(p)' you may infer 'Σ(q)', where 'Σ(q)' is the result of replacing at least one occurrence of 'p' in 'Σ(p)' by 'q'.

The second substitution principle corresponds to Davidson's assumption that we may substitute "co-extensive singular terms" in the context of C, which clearly functions as a principle for the substitution of singular terms in the *generic* sense, which includes definite descriptions. I will (following Neale) call the principle 'iota-substitution' and abbreviate it as ι-SUB. It permits the substitution of a definite description for a definite description, of a definite description for a name and of a name for a definite description, and hence its formal statement requires three rules:

ι-SUB:	$(ιx)p = (ιx)q$	$(ιx)p = a$	$(ιx)p = a$
	$\frac{\Sigma[(ιx)p]}{\Sigma[(ιx)q]}$	$\frac{\Sigma[(ιx)p]}{\Sigma[a]}$	$\frac{\Sigma[a]}{\Sigma[(ιx)p]}$

where 'Σ(ψ)' is the result of replacing at least one occurrence of 'φ' in 'Σ(φ)' by 'ψ':

Here is the Davidson slingshot set out formally:

- | | |
|---|-----------------------|
| (1) $p \leftrightarrow q$ | premise |
| (2) Cp | premise |
| (3) $p \leftrightarrow ((ιx)(Fx) = (ιx)(Fx . p))$ | (Logical equivalence) |
| (4) $C[(ιx)(Fx) = (ιx)(Fx . p)]$ | 2, 3 SUB-LE |
| (5) $(ιx)(Fx . p) = (ιx)(Fx . q)$ | 1 |
| (6) $C[(ιx)(Fx) = (ιx)(Fx . q)]$ | 4, 5 ι-SUB |
| (7) $q \leftrightarrow ((ιx)(Fx) = (ιx)(Fx . q))$ | (Logical equivalence) |
| (8) Cq | 6, 7, SUB-LE |

Let me explain the proof informally. Line 1 uses '↔' for the biconditional and can be read as "'p' and 'q' are materially equivalent", that is, have the same truth value. Line 2 abbreviates "The sentence S corresponds to the fact that p". The aim of the proof is to show that it is permissible to substitute 'q' for 'p' in "The sentence S corresponds to the fact that p" ('Cp') even if the only thing 'q' and 'p' share is that both are true.

The inference from line 1 to line 5 occurs outside the scope of C and depends only on the definition of 'ιx' together with standard rules of predicate logic with identity. Reading '(ιx)Fx' as "the unique object x such that x is F" means that '(ιx)(Fx . p)' is to be read as "the unique object x such that x is F and p". If 'p' is true, this will be true of the unique object x such that x is F; but if 'p' is false, there will be no such thing as the unique object x such that x is F and p of which the formula can be true.¹⁹ This means that the role of 'p' in the reference of '(ιx)(Fx . p)' depends *only* on whether 'p' is true or false, and hence substituting 'q' for 'p' will not change the reference of '(ιx)(Fx . p)', provided 'q' and 'p' have the same truth value. According to line 1, 'q' does have the same truth value as 'p', and hence '(ιx)(Fx . p)' and '(ιx)(Fx . q)' have the same reference, which is to say that '(ιx)(Fx . p) = (ιx)(Fx . q)', which is just line 5.

Lines 3 and 7 express the logical equivalences which are used to get lines 4 and 8 respectively. Since lines 3 and 7 are both outside the scope of C, reference to SUB-LE or ι-SUB is not required for their justification, which depends only on predicate logic with identity and parallels that given for line 5 in the previous paragraph. Line 3 says that 'p' and '(ιx)(Fx) = (ιx)(Fx . p)' are logically equivalent, which is to say they have the same truth value in every model or possible situation. The justification for this (as we saw in the previous paragraph) is that '(ιx)(Fx . p)' has as its reference the unique object x such that x is F and p; if 'p' is true, this will simply be the unique object such that x is F; but if 'p' is false, there will be no such thing as the unique object such that x is F and p. The identity on line 3 holds,

therefore, whenever 'p' is true (because the right hand side designates the same as the left) but does not hold whenever 'p' is false, which means that the expressions which flank the logical equivalence sign have the same truth value in every possible situation. (The justification for line 7 is identical with 'q' substituted for 'p'.)

The remaining inferences occur within the scope of C, and hence their justification requires reference to the additional substitution principles. Line 4 is inferred from lines 2 and 3 by using SUB-LE to replace 'p' in 'Cp' (line 2) with what line 3 says is logically equivalent to it, namely $(\lambda x)(Fx) = (\lambda x)(Fx \cdot p)$. Line 6 is inferred from lines 4 and 5 by substitution of definite descriptions: since (by 5) $(\lambda x)(Fx \cdot p) = (\lambda x)(Fx \cdot q)$, we may, given ι -SUB, substitute $(\lambda x)(Fx \cdot q)$ for $(\lambda x)(Fx \cdot p)$ in line 4, which yields line 6. Finally, line 8 as inferred from lines 6 and 7 by using SUB-LE to replace $(\lambda x)(Fx) = (\lambda x)(Fx \cdot q)$ in 'C[($\lambda x)(Fx) = (\lambda x)(Fx \cdot q)$]' with what line 7 says is logically equivalent to it, namely 'q'.

4. CRITIQUE OF THE DAVIDSON VERSION OF THE SLINGSHOT

The only way to evade the Davidson slingshot is to reject at least one of the substitution principles it uses. The sole argument Davidson gives for the principles is that their rejection would make the correspondence connective fully intensional. But that assumes that the connective must set up either a fully extensional or a fully intensional context and, as we saw in Section 2, there is no reason to accept that.

Barry Taylor sketched out the best kind of defense of the principles when he wrote that "the evident consequences of the traditional conception of the descriptum of a sentence as the complex of entities relevant for its truth" include two claims: first, "sentences so closely connected as to be guaranteed by logic alone to share a truth-value cannot differ in truth-relevant entities, and so must share their descriptum" and, second, "sentences which, like 'Cicero orated' and 'Tully orated', differ merely in the manner they choose to specify the same truth-relevant entity cannot diverge in the complex of such entities they describe." (Taylor, 1985, p. 30) The first of these claims speaks in favor of SUB-LE, the second in favor of ι -SUB.

Taylor's defense of ι -SUB would be cogent if the principle concerned only singular terms like the ones he cites, *names* such as 'Cicero' and 'Tully, whose function is solely to refer and which can, therefore, be used uncontroversially in identity sentences. But ι -SUB also applies to definite descriptions and, as we saw in section 2, we cannot assume a single substitution principle for all singular terms, but must distinguish between a principle for names and a principle for definite descriptions, which neither Davidson nor Taylor do. I will postpone discussion of this crucial point, however, because it applies to the Gödel version of the slingshot in the same way it does to the Davidson version.

The best known criticisms of Davidson's slingshot have rejected SUB-LE rather than ι -SUB, the most prominent example being Barwise and Perry (1983). They have been followed by John Searle, who wrote in *The Construction of Social Reality* that "my method of investigation is to examine the structure of the facts that make our statements true and to which they correspond when they are true" (1995, p. 2), which gives him a weighty reason for defending a fact-based correspondence theory. Searle's objection to SUB-LE is that its use entails that true sentences correspond to

facts which have constituents which are not matched by any constituents of those sentences. He correctly points out that the inference from lines 2 and 3 to line 4, for instance, would license the inference from:

A. The true sentence that Stockholm is a large city corresponds to the fact that Stockholm is a large city

to:

B. The true sentence that Stockholm is a large city corresponds to the fact that the unique individual such that he is President of the U.S. is identical to the unique individual such that he is President of the U.S. and such that Stockholm is a large city.

This is completely silly, he contends, because the true sentence that Stockholm is a large city can't possibly correspond to any facts about the President of the U.S.; the President and his identity are quite irrelevant as far as the fact that Stockholm is a large city are concerned. To think otherwise is to fail to "respect the intuitive notions of 'fact', 'truth', and 'correspondence'" according to which "the truth maker for the statement that snow is white is the fact that snow is white."²⁰

The trouble with this is that Searle cannot consistently hold that the only way to refer to the truth maker for 'snow is white' is to use the description 'the fact that snow is white' because that would be to construe the correspondence connective as fully intensional and hence rule out *any* alternative way of referring to a fact. That would deprive the notion of fact of any explanatory power and would render otiose Searle's project of investigating "the structure of the facts which make our statements true." Searle, like anyone who thinks of a fact as an entity in the world, *has* to admit alternative ways of referring to facts, and hence he has to make use of some principles of substitution. Having rejected SUB-LE, it is unclear what principles he does accept, though he appears to give tacit assent to ι -SUB, which, as we shall see, is extremely hazardous to the correspondence theory.

In my view, Taylor's claim that "sentences so closely connected as to be guaranteed by logic alone to share a truth-value cannot differ in truth-relevant entities, and so must share their descriptum" is a cogent defense of SUB-LE. The fact that logicians invent fancy alternative ways of constructing sentences which are logically equivalent to sentence 'p' doesn't entail that the fancy new sentences are 'about' something distinct from 'p' (have a different 'descriptum'). Sentence B above is not 'about' the unique individual who is President of the U.S. because, to put it roughly, what permits the use of that predicate to construct a sentence [B] which is logically equivalent to sentence A is that the predicate is being used in B to say that presidents of the U.S. are self-identical, and that is no more 'about' presidents of the U.S. than about anything else.

There is in any case no intuitively clear notion of what a *sentence* is 'about', the only precise notion we have being the model-theoretic one, and from that point of view, SUB-LE is not objectionable. If two sentences 'p' and 'q' are such that there is no model in which they differ in truth value, so that in every *possible* situation in

which 'p' is true so is 'q' is, it is difficult to understand what it would mean to say that what makes 'p' true (if anything) is different from what makes 'q' true. I think, therefore, that rejecting SUB-LE is not the way to avoid the slingshot. This deserves more discussion, however, and I will return to it below.

5. THE GÖDEL VERSION OF THE SLINGSHOT

The Gödel version of the slingshot is based on Gödel's contribution to the *Library of Living Philosophers* volume on Russell (Gödel, 1944). Gödel did not state the argument but gave some pointed suggestions about what assumptions were needed and how the argument would go. There have been infrequent discussions of it since, and recently Stephen Neale gave it a rigorous formulation, which I shall use. (Neale, 1995). But first let me quote the relevant passage from Gödel's paper:

An interesting example of Russell's analysis of the fundamental logical concepts is his treatment of the definite article 'the'. The problem is: what do the so-called descriptive phrases (i.e., phrases such as, e.g., 'the author of Waverley' or 'the king of England') denote or signify and what is the meaning of sentences in which they occur? The apparently obvious answer that, e.g., 'the author of Waverley' signifies Walter Scott, leads to unexpected difficulties. For, if we admit the further apparently obvious axiom, that the signification of a complex expression, containing expressions which have themselves a signification, depends only on the signification of these constituents (and not on the manner in which this signification is expressed), then it follows that the sentence 'Scott is the author of Waverley' signifies the same thing as 'Scott is Scott'; and this again leads almost inevitably to the conclusion that all true sentences have the same signification (as well as all the false ones). Frege actually drew this conclusion.... (p. 128)

In a footnote, Gödel gave his suggestion on what the proof should assume (in addition to the "apparently obvious axiom" mentioned in the quotation), namely "that ' $\phi(a)$ ' and the proposition 'a is the object which has the property ϕ and is identical to a' mean the same thing". He also noted that "one would have to use the fact that for any two objects a and b, there exists a true proposition of the form $\phi(a, b)$ as, e.g., $a \neq b$ or $a = a \cdot b = b$ ", but this is a hint on how to construct the argument and not an assumption it must make. (The hint is embodied in lines 2 and 10 of the proof.)

The assumption that ' $\phi(a)$ ' means the same as 'a is the object which has the property ϕ and is identical to a' amounts to the introduction and elimination of the iota sign, for in symbols (with a letter switch) it says that 'Fa' and ' $a = (\iota x)(x = a \cdot Fx)$ ' "mean the same thing". Gödel didn't say what he meant by "mean the same thing" but Neale interprets it as "signify the same thing" and calls it the "Gödelian equivalence", calling the substitution principle it involves "iota-conversion" (ι -CONV) because it involves converting sentences containing names (in my terminology) into sentences containing definite descriptions and vice versa. Put formally, ι -CONV consists of two substitution principles, one for introducing the iota and one for eliminating it:

$$\begin{array}{l} \iota\text{-CONV: } \frac{\Sigma[x/a]}{a = (\iota x)(x = a \cdot \Sigma[x])} \quad \frac{a = (\iota x)(x = a \cdot \Sigma[x])}{\Sigma[x/a]} \end{array}$$

This principle plays essentially the same role in the Gödel version that SUB-LE plays in the Davidson version, although ι -CONV is weaker than SUB-LE because the latter entails the former but not vice versa, but also because (to put it informally) whereas SUB-LE introduces new predicates (since the logical equivalence holds between formulae of the form 'p' and ' $(\iota x)(Fx) = (\iota x)(Fx \cdot p)$ '), the Gödelian equivalence holds between formulae of the form 'Fa' and ' $a = (\iota x)(x = a \cdot Fx)$ ' and hence introduces no new predicates (or names) into sentences. It follows that philosophers like Searle, who object to the Davidson version because it assumes SUB-LE, cannot use that objection against the Gödel version, and Neale takes this to be its great strength.

The compositionality assumption mentioned in the quotation from Gödel is a principle of substitution for singular terms. It is clear, from Gödel's examples and from the use to which the argument is put, that singular terms here include definite descriptions and that the principle is ι -SUB – the same principle used in the Davidson slingshot.

Here is the Gödel slingshot set out formally:

- | | |
|---|---------------------|
| (1) Fa | premise |
| (2) $a \neq b$ | premise |
| (3) Gb | premise |
| (4) $a = (\iota x)(x = a \cdot Fx)$ | 1, ι -CONV |
| (5) $a = (\iota x)(x = a \cdot x \neq b)$ | 2, ι -CONV |
| (6) $b = (\iota x)(x = b \cdot a \neq x)$ | 2, ι -CONV |
| (7) $b = (\iota x)(x = b \cdot Gx)$ | 3, ι -CONV |
| (8) $(\iota x)(x = a \cdot Fx) = (\iota x)(x = a \cdot x \neq b)$ | 4, 5, ι -SUB |
| (9) $(\iota x)(x = b \cdot Gx) = (\iota x)(x = b \cdot a \neq x)$ | 6, 7, ι -SUB |
| (10) C(Fa) | premise |
| (11) C($a = (\iota x)(x = a \cdot Fx)$) | 10, ι -CONV |
| (12) C($a = (\iota x)(x = a \cdot x \neq b)$) | 11, 8, ι -SUB |
| (13) C($a \neq b$) | 12, ι -CONV |
| (14) C($b = (\iota x)(x = b \cdot a \neq x)$) | 13, ι -CONV |
| (15) C($b = (\iota x)(x = b \cdot Gx)$) | 14, 9 ι -SUB |
| (16) C(Gb) | 15, ι -CONV |

This version of the slingshot proceeds in two stages. The first (lines 1 to 9) involves substitutions outside the context of C, the second (lines 10-16) substitutions inside it. The aim of the proof is to show that if S corresponds to the fact that Fa (line 10), it also corresponds to the fact that Gb (line 16), for any 'G' and 'b' for which 'Gb' is true. For example, given that 'the author of Waverley lived in Scotland' corresponds to the fact that the author of Waverley lived in Scotland, it follows that it corresponds to any fact – say, the fact that Stockholm is a large city or that Hesperus is illuminated by the sun. That would mean that the correspondence connective is fully extensional (or truth-functional), so that any true sentence corresponds to every fact.

In the first stage, Gödel's hint about the role of 'a ≠ b' is used in line 2 and then substitutions are made, using ι -CONV AND ι -SUB. The inference from line 1 ('Fa') to line 4 ('a = $(\iota x)(x = a \cdot Fx)$ ') essentially involves, first, substituting for 'a' the definite description ' $(\iota x)(x = a)$ ', so that 'Fa' becomes $F(\iota x)(x = a)$ ('the unique x which is identical to a is F') and then rewriting the latter as 'a = $(\iota x)(x = a \cdot Fx)$ ' ('a is identical to the unique individual x such that x is identical to a and is F). Since no distinction is made between the status of 'a' (a name) and ' $(\iota x)(x = a)$ ' (a definite description), we can use a definite description to interpret line 1 as:

A. The author of Waverley lived in Scotland

and hence interpret line 4 as:

B. The author of Waverley is identical to the unique individual who is (identical to) the author of Waverly and who lived in Scotland.

Reflection on A and B shows that in any possible situation in which A is true, B is also true, and vice versa.

Analogous considerations apply to the inference from line 2 to line 5. The right hand side of line 5 is a definite description which can be read as 'the unique individual x which is such that it is identical to a but not to b'; and that must be a, given (line 2) that it is a which is not identical to b. (The inference from line 2 to line 6 is the same except that 'b' has been switched for 'a'.) Since these inferences assume that names and definite descriptions have the same status, we may interpret line 2 as:

C. The moon is not identical to the sun

and hence interpret line 5 as:

D. The moon is identical to the unique individual which is (identical) to the moon and is not identical to the sun.

And, again, reflection shows that in any possible situation in which C is true, so is D, and vice versa.

Lines 8 and 9 use ι -SUB in an obvious way, given that this stage of the proof requires no distinction between the substitution of definite descriptions and the substitution of names. Line 8 follows from lines 4 and 5 (and line 9 from lines 6 and 7) in virtue of the principle that if a = b and a = c, then b = c.

The second stage of the proof uses the results of the first stage, and while these results are entirely acceptable, it does not follow that the substitutions made in them in the second stage are acceptable unless it is assumed that the same principles of substitution hold outside and inside the correspondence connective, which would beg the question. But before considering this point, let me finish my informal account of the proof.

At line 9 the proof assumes 'C(Fa)' in order to show that 'C(Gb)' can be inferred from it, which would establish that any true sentence corresponds to every fact. The proof infers line 11 from line 10 by the same use of ι -CONV as in the inference of line 4 from line 1. It infers line 12 from lines 8 and 11 by a straightforward use of ι -SUB, with ' $(\iota x)(x = a \cdot x \neq b)$ ' playing the role of 'a' in 'if a = b and a = c, then b = c'.

The proof infers line 13 from line 12 by the use of ι -CONV, now as an elimination rule rather than as an introduction rule. It infers line 14 from line 13 by a use of ι -CONV like its use in the inference from line 2 to line 5. It infers line 15 from lines 9 and 14 by ι -SUB, with ' $(\iota x)(x = b \cdot x \neq a)$ ' as the formula which plays the role of 'a' in 'if a = b and a = c, then b = c.' Finally Line 16 is inferred from line 15 by the use of ι -CONV as an elimination rule, exactly as in the inference from line 12 to line 13.

6. CRITIQUE OF THE GÖDEL VERSION OF THE SLINGSHOT

Since the Gödel version of the slingshot is valid in that, given the two substitution principles it assumes, its premisses entail its conclusion, the only way to evade it is to challenge at least one of those substitution principles. Although the principles are acceptable in extensional contexts, to assume they are, therefore, acceptable inside the scope of C would beg the question. I shall show that there are, indeed, good reasons for denying that ι -SUB is acceptable inside C and that attempts to restrict it in order to make it acceptable are ruled out by ι -CONV. Since the Gödel slingshot requires both principles, its conclusion can be evaded. It does not follow that the correspondence theory of truth is home free, for, as I will argue in the final section, evading the slingshot exacts a heavy price.

If ι -SUB were a principle of substitution only for *co-referring* singular terms ('names' in my terminology), there would be, as we have seen, no objection to its use even in non-extensional contexts, but it also applies to definite descriptions whose function in the proof reflects their *attributive* use in assertions. Definite descriptions, that is, function in the proof not simply to *refer* to a particular individual (not to pick out in some way or other a particular individual we want to describe), but to be *true of* whichever unique individual happens to fulfill the description (so that the description itself is essential, not which particular individual it happens to be true of). They function, that is to say, like predicates, except (since they are *definite* descriptions) for the important condition that they be true of exactly

one individual (though not of some particular individual however described). This means that ι -SUB is a principle which permits not only the substitution of co-referring singular terms but also the substitution of co-extensive predicates.

That this is the case can be seen if we look carefully at the proof. Consider, for example, line 12, which is inferred from lines 8 and 11 by using ι -SUB to replace '(ιx)($x = a \cdot Fx$)' in line 11 with '(ιx)($x = a \cdot x \neq b$)' to get line 12, the justification being the identity expressed in line 8 between those two definite descriptions. The difference between '(ιx)($x = a \cdot Fx$)' and what has been substituted for it, '(ιx)($x = a \cdot x \neq b$)', is that what follows ' ιx ' in the former is ' $x = a \cdot Fx$ ', whereas what follows it in the latter it is ' $x = a \cdot x \neq b$ '. But those are *predicates* (open sentences), which are, indeed, co-extensive, but their substitution means that ι -SUB permits the substitution of co-extensive predicates (and not only the substitution of co-referring terms).

We can reinforce this point by considering line 8, which has the identity sign between the definite descriptions, '(ιx)($x = a \cdot Fx$)' and '(ιx)($x = a \cdot x \neq b$)'. The question is, what sort of *identity* is expressed here? It is misleading to say that each of these definite descriptions *refers* to the same particular individual; a better answer is that the *extensions* of each of the descriptions is the same – that they are true of exactly the same individuals – though in this case (because they are *definite* descriptions) that means exactly one individual. What these definite descriptions are both true of, however, is not some *particular* individual picked out for description; they are both true of whichever individual happens to belong to the extension of both descriptions (whichever individual both descriptions happen to be true of). ι -SUB is better thought of, therefore, not as a principle which permits the substitution of *co-referring* terms but as a principle which permits the substitution of *co-extensive* terms, which include co-extensive *predicates*. The identity expressed in line 8, for example,

$$8. (\iota x)(x = a \cdot Fx) = (\iota x)(x = a \cdot x \neq b)$$

holds only because ' $x = a \cdot Fx$ ' and ' $x = a \cdot x \neq b$ ' are coextensive predicates, and they are coextensive just because both of them are true only of a. ' $x = a \cdot Fx$ ' is true of a and only of a, provided a is F (which it is by line 1 of the proof), while ' $x = a \cdot x \neq b$ ' is true of a and only of a, provided a is not identical to b (which it is by line 2 of the proof). Line 8, therefore, is not an identity sentence in the usual sense: it does not express the identical reference of two names, but rather the sameness of extension of two descriptions.

Another way of showing that ι -SUB permits the substitution of co-extensive predicates is to consider it in the light of Russell's theory of descriptions. Russell held that a definite description must be distinguished from a name because its object may not exist, which he illustrated with the famous example, 'The King of France is bald' (the French monarchy having been abolished when Russell wrote). He further held that this sentence must have a truth value, in which case it must be *false* because there is no King of France. This means that it is misleading to use the iota notation in the usual way (with ' Fx ' for ' x is King of France' and ' Bx ' for ' x is bald'), that is, $B(\iota x)Fx$, to be read as 'the unique individual who is King of France is

bald', because that doesn't make it clear that the sentence is false because there is no King of France. Russell, therefore, treated the sentence as an existential quantification, whose perspicuous paraphrase is "There is a unique individual who is King of France and who is bald", expressed in logical notation, as follows.

$$(\exists x)((Fx \cdot (\forall y)(Fy \rightarrow y = x)) \cdot Bx)$$

Since there is no King of France, the sentence is false (because it is false that $(\exists x)(Fx)$), and as an existential quantification, it includes no singular terms.

Russell argued that definite descriptions should always be treated in this way, so let us apply his theory to the slingshot, this time to the Davidson version; consider the inference from lines 4 and 5 to line 6:

$$4. C[(\iota x)(Fx) = (\iota x)(Fx \cdot p)]$$

$$5. (\iota x)(Fx \cdot p) = (\iota x)(Fx \cdot q)$$

$$6. C[(\iota x)(Fx) = (\iota x)(Fx \cdot q)]$$

Line 6 is inferred from line 4 by substitution in accordance with the identity expressed in line 5: line 6 differs from line 4 in that '(ιx)($Fx \cdot q$)' has replaced '(ιx)($Fx \cdot p$)', which means that the *predicate* ' Fx and q ' has replaced the *predicate* ' Fx and p ' (the two predicates being coextensive because, in virtue of line 1 in the proof (' $p \leftrightarrow q$ '), ' p ' and ' q ' are coextensive, that is, have the same truth value).

These comments are based, however, on formulations which use the iota notation, which Russell found misleading, so let us see what happens if we rewrite a couple of them using his theory of description:

$$4^* C[(\exists x)((Fx \cdot p) \cdot (\forall y)((Fy \cdot p) \rightarrow y = x)) \cdot (\iota x)(Fx = x)]$$

$$6^* C[(\exists x)((Fx \cdot q) \cdot (\forall y)((Fy \cdot q) \rightarrow y = x)) \cdot (\iota x)(Fx = x)]$$

This rewriting only reinforces the point about the substitution of co-extensive *predicates* because 6^* differs from 4^* in that ' p ' has replaced ' q ' twice, with both ' p ' and ' q ' functioning as predicates.²¹ But it also reinforces the point I made above that identity sentences like these, which involve definite descriptions, are not *identity* sentences in the usual sense; for when lines 4 and 6 are re-written in accordance with Russell's theory of descriptions as 4^* and 6^* , neither of them are identity sentences (though they contain identities). This means that on Russell's theory of descriptions, there is a sense in which, strictly speaking, we cannot even *formulate* ι -SUB because the theory rules out the possibility of flanking the identity sign with definite descriptions.

What I just said is misleading, however, in that Russell did permit the formulation of sentences like 4, '(ιx)($Fx) = (\iota x)(Fx \cdot p)$ ', but he did so *only* if they were understood as *abbreviations* for sentences like 4^* . That entails that substitutions which appeal directly to identity sentences like '(ιx)($Fx) = (\iota x)(Gx)$ '

are acceptable *only* in extensional contexts, which means ι -SUB is an acceptable principle of inference only in extensional contexts. (It is, indeed, a principle which can be derived from the standard rules of inference in predicate logic with identity.)²²

Russell's theory of descriptions makes evident what I take to be independently defensible, namely that ι -SUB permits the substitution of co-extensive predicates and hence that its unrestricted use is acceptable only in extensional contexts.²³ Either version of the slingshot, therefore, can be evaded by refusing to accept ι -SUB on the ground that permitting its use in the context of the correspondence connective is to assume that the context is fully extensional, thus assuming what was supposed to be proved.

The reason critics and defenders alike tend to miss this point is because 'singular term' is so widely used to apply to both names and definite descriptions, and hence there is a strong inclination to think that since co-referring names can surely be substituted *salva veritate* in the context of C, so can definite descriptions, even when their formal treatment reflects their attributive (and not their referential) use. This is reinforced by speaking of the *extension* not only of definite descriptions but of singular terms generally, or by speaking of what singular terms in general *stand for* or *represent*, terminology which obscures the distinction between using a term referentially and using it attributively, which is a distinction the formal treatment of singular terms must reflect to be relevant to philosophical issues like truth.

The inclination to treat names and definite descriptions alike is made a matter of principle in the Gödel version of the slingshot because of its use of ι -CONV, which sanctions the move back and forth between names and definite descriptions, though its real effect is to treat all singular terms as (attributively used) definite descriptions. It might be possible to formulate a restricted substitution principle for singular terms in the context of C if ι -CONV were rejected, the restriction being that ι -SUB is acceptable for names and for definite descriptions, provided their treatment reflects their referential rather than their attributive role. What ι -CONV does, however, is insure that the treatment of all singular terms reflects their attributive role, and hence if ι -CONV is accepted, ι -SUB has to be rejected to evade the slingshot.

Although Davidson's version does not use ι -CONV, it is also unable to use a restricted version of ι -SUB which would rule out the substitution of co-extensive predicates, the reason being that it must use SUB-LE. While critics like Searle are incorrect in rejecting the use of SUB-LE outside extensional contexts (its use does not entail that a context is extensional), it is like ι -CONV in insuring that the treatment of all singular terms will reflect their attributive rather than their referential role. We can see this if we examine the use made of SUB-LE in the Davidson slingshot, for example in the inference from line 2, 'Cp', to line 4, 'C[(ιx)(Fx) = (ιx)(Fx . p)]'. That makes sense only if the definite descriptions which flank the identity sign in line 4 are not taken to *refer* to the same particular individual, but to set conditions such that whatever individuals meet the one will meet the other (in this case such that exactly one individual will meet them). What is claimed, therefore, is not identity of reference of names but sameness of extension of definite descriptions, which is all that can be claimed in proofs which use SUB-LE as this one does. This is why Davidson speaks of the *extension* of singular terms,

which reflects their attributive use as being true of objects, rather than their referential use as referring to them.

Neale argues that the Gödel version of the slingshot is to be preferred to the Davidson version because it uses the weaker ι -CONV rather than the stronger SUB-LE. The latter is, of course, stronger than the former because it entails but is not entailed by it, but as far as evading the slingshot is concerned, there is no difference. Although SUB-LE in itself is not objectionable, it has the same role in the slingshot as ι -CONV, namely to insure that all singular terms are treated as having extensions rather than referents. But to treat singular terms as having extensions is to permit the substitution of coextensive predicates, which shows that ι -SUB is unacceptable in context C, which means that both versions of the slingshot can be evaded by rejecting ι -SUB on the grounds that there is no reason to accept it except in contexts already known to be extensional.

7. CONCLUSION: CONSEQUENCES FOR THE CORRESPONDENCE THEORY OF TRUTH

A heavy price has to be paid for evading the slingshot by rejecting ι -Sub, namely that definite descriptions (though not names) may not be substituted in the context of the correspondence connective. Thus given that 'Stockholm is a large city' corresponds to the fact that Stockholm is a large city, it does not follow that it corresponds to the fact that the capital of Sweden is a large city or to the fact that the birthplace of Strindberg is a large city. To use another example: although 'Hesperus is illuminated by the sun' corresponds to the fact that Phosphorus is illuminated by the sun (since 'Hesperus' and 'Phosphorus' are names of the same planet), it does not follow that it corresponds to the fact that the evening star is illuminated by the sun (since 'the evening star' is a definite description) even though 'Hesperus', 'Phosphorus' and 'the evening star' all signify the planet Venus.

Defenders of the correspondence theory may well find this constraint on substitution overly restrictive and arbitrary and hence try to evade the slingshot by rejecting SUB-LE and ι -CONV instead of ι -SUB. I argued earlier that rejecting SUB-LE is not the way to evade the slingshot, and there is even stronger reason not to reject it simply to save ι -SUB. The reason is that to reject SUB-LE is to deny that logically equivalent sentences can be substituted *salva veritate* in the context of C, while to accept ι -SUB is to assert that (merely) coextensive predicates can be substituted *salva veritate* in that context, which is to reject a weaker substitution principle in order to save a stronger one. It is true that the Davidson slingshot cannot exploit that fact to prove that ι -SUB sets up an extensional context since the proof also needs SUB-LE, but the Gödel version can, since it needs only ι -CONV, and hence this strategy for evading the slingshot won't work unless ι -CONV is also rejected. But ι -CONV is even weaker than SUB-LE, and while rejecting it evades the Gödel version, it is hard to believe that another formal proof could not be constructed, which uses a principle of inference which functions like ι -CONV and exploits the substitution of coextensive predicates permitted by ι -SUB in order to show that C must be extensional if ι -SUB is used.

The function of ι -CONV (and a function of SUB-LE) is to convert sentences with names into logically equivalent sentences with definite descriptions, and vice versa, and its effect is to treat all singular terms as having extensions rather than references, which is the formal reflection of their being used attributively rather than referentially in assertions. Rejecting ι -CONV and SUB-LE would make it possible to treat singular terms so as to reflect their referential use, which suggests permitting their substitution in the scope of C if and only if they are being used referentially, thus avoiding the substitution of co-extensive predicates, which makes the slingshot a deadly weapon.

The question is whether principles could be devised to do this – to permit the substitution of definite descriptions if and only if that does not also permit the substitution of co-extensive predicates. Such principles are implicit in a Russellian theory of facts, which construes facts as composite entities consisting of a particular, referred to by a name, and a universal or property, referred to by a predicate (or consisting of several particulars and properties). The key move here is to construe a predicate not as having an extension consisting of the objects of which it is true, but as referring to a property, so that predicates can be substituted if and only if they refer to the same property. Thus 'creature with a heart' and 'creature with a kidney' have the same extension but they do not refer to the same property (since what it is to have a heart is different from what it is to have a kidney, as transplant surgeons well know), and hence they are not inter-substitutable in the context of C.

What this key move does is to assimilate predicates to singular terms and hence to assimilate the substitution rules for predicates to those for singular terms. Just as one can substitute 'a' for 'b' only if 'a' and 'b' are co-referring (so that $a = b$), so one can substitute the predicate 'is an F' for the predicate 'is a G' if and only if the predicates are co-referring. Against this background, accepting ι -SUB will not load the slingshot because ι -SUB will be applicable only when the substitution of definite descriptions involves the substitution of predicates which refer to the same property; it will not be applicable when it involves the substitution of (merely) co-extensive predicates.

This Russellian way of evading the slingshot, however, exacts at least as heavy a price as simply rejecting ι -SUB. For one thing, it will require a version of the correspondence theory for which correspondence to the facts is a matter of "correspondence-as-congruity" rather than "correspondence-as-correlation" (to use the terms from Pitcher, 1964, p. 10) because it requires structural isomorphism between names and particulars and between predicates and properties. This imposes on defenders of the correspondence theory the burden of defending a structural isomorphism between language and reality, which undermines the most common defense of the theory, namely that it is nothing more than common sense whose denial involves extravagant metaphysics.

Davidson always assumes that fact-based versions of the correspondence theory adopt the Russellian point of view, which he views not as a strategy to avoid the slingshot but as the primary reason for rejecting correspondence theories. Thus he wrote in "True to the Facts" (Davidson, 1984, p. 49) that "The failure of correspondence theories based on the notion of fact traces back to a common source: the desire to include in the entity to which a true sentence corresponds not only the

objects the sentence is 'about' (another idea full of trouble) but also whatever it is that the sentence says about them," where the latter expresses the notion of a fact as including objects plus properties. He reiterated this 30 years later in his response to Neale in the *Library of Living Philosophers* volume, agreeing that the Gödel slingshot can be evaded "as Russell's semantics did it, by making properties part of facts and so the entities that correspond to predicates" (Hahn, 1999, p. 667) but going on to say that "This is a course against which I have argued on the grounds that it cannot be incorporated into a satisfactory theory or definition of truth, and entities that are made up of abstract entities can hardly be thought of as empirical truth-makers."

The first objection stated here, which seems to me uncontested, is that the Russellian idea of treating predicates as referring to properties is not consistent with Tarski's treatment of predicates as satisfied by sequences of objects, which entails that there can be no appeal to Tarski's truth theory in defense of a correspondence theory which uses Russellian facts.²⁴ The second objection, that Russellian facts are peculiar entities in that they consist not only of particulars referred to by singular terms but also of abstract objects referred to by predicates, can be dealt with only by defenders of the correspondence theory who are prepared to develop it as an elaborate metaphysical theory.

Davidson made another criticism of the Russellian point of view in "True to the Facts" when he mentioned "the well-explored consequence that it becomes difficult to describe the fact that verifies a sentence except by using the sentence itself" ("verifies" here is used in the sense of "make true"). (Davidson, 1984, p. 49) This suggests that the aim of permitting alternative ways of referring to facts without loading the slingshot cannot after all be realized by introducing Russellian facts because taking predicates to refer to properties puts an equally severe restriction on the ways we can refer to properties.

The reason is that, although the notion of different predicates having the same extension is straightforward, the notion of their referring to the same property is not, and this raises the question of the conditions under which different predicates refer to the same property. That turns out to be essentially the same question as when different instances of 'the fact that a is F' refer to the same fact. In both cases it can be assumed that *synonymous* expressions refer to the same property or the same fact respectively; but beyond that, the question of whether 'the predicate F' refers to the same property as 'the predicate G' turns on whether those expressions are being used referentially or attributively. But that is the very distinction which motivated the Russellian notion that predicates refer to properties, the idea being that taking facts to include properties insures that singular terms are never substituted in the context of C unless they are functioning so as to reflect a referential use. But now it turns out that the introduction of properties as the referents of predicates cannot insure that singular terms are substituted in the context of C only when used referentially unless the predicates themselves are used referentially and not attributively. That raises the question of what insures that predicates are being used referentially, and to insure that, further entities to be referred to will have to be introduced, the result being an infinite regress, which shows that nothing has been accomplished by the introduction of properties as the referents of predicates.

I conclude, therefore, that the strategy of evading the slingshot by keeping t-SUB but developing principles which rule out its use to substitute (merely) co-extensive predicates in the scope of C does not yield a better version of the correspondence theory than simply giving up t-SUB. In either case, the only substitution which can be made for 'q' in 'p corresponds to the fact that q' are substitutions for genuine referring terms which are constituents of 'q'. In practice this will include at most names because there is no way of establishing that a definite description is used referentially rather than attributively in the context of C. Even if it is coherent to take a predicate as referring to a property, it is hopeless to try to distinguish between when a predicate is being used to refer to a particular property and when it is being used attributively to set conditions some particular property might meet. The same is true of the use of definite descriptions in the context of the correspondence connective. There is no way to distinguish between the referential and the attributive use of 'the capital of Sweden' in:

The true sentence 'Stockholm is a large city' corresponds to the fact that the capital of Sweden is a large city'.

But if 'the capital of Sweden' is not being used referentially, then this claim sets up the slingshot. Hence my conclusion that the slingshot can be evaded but at the price of forbidding any substitutions in the scope of the correspondence connective other than names for names.

This has two consequences which I will state but not develop. One is that the constraints the correspondence theory must meet to evade the slingshot are not met by most fact-based versions of the theory. This is true for any version of 'scientific realism' which takes it to be an empirical matter what the structure of the truth-making facts are, one example being the influential metaphysical realism once defended by Putnam.²⁵ It is also true for any version of the theory which tries to establish substantive conclusions about such matters as the nature of mental states, the structure of social reality, or the constituents of action, by inquiring into the structure of the facts which make claims about them true. The reason is that, given the constraints, all that can be said about the structure of the fact which makes p true, beyond that it is that fact that p, would involve the substitution of co-referring names, and that will yield no conclusions which are either substantive or interesting.

The other consequence is that any versions of the correspondence theory which survive these constraints will vary in no significant way from deflationist conceptions of truth, which maintain that all there is to be said about the distinction between what it is to be true and what it is to be false is summed up in the Equivalence Principle' – 'p' is true if and only if p – together with various accounts about how to apply the principle to contexts where what is said to be true cannot be expressed but only referred to in complex ways.²⁶ This does not mean we must stop speaking in terms of true sentences corresponding to the world. The truth of a sentence surely depends on what is said in asserting it (which varies with meaning and context) and on what the world is like, which is to say that its truth depends on correspondence to what the world is like. So 'Stockholm is a large city' corresponds to what the world is like, but if we want to say more precisely to what it

corresponds, we can say only that it corresponds to the fact that Stockholm is a large city. Those who try to say anything more substantive than that will find themselves in the company of Goliath and other victims of the fabled slingshot.

University of Uppsala and St. Olaf College

NOTES

¹ A version of this paper was discussed at a seminar at Uppsala University where it benefited from the comments of John Bacon, Krister Segerberg, and, especially, Tor Sandqvist, whose remarks both saved me from some bad mistakes and suggested the use of the referential-attributive distinction to evaluate the slingshot. I also received very helpful comments from Gary Iseminger, Owen Jones, and Peter Pagin. I am particularly grateful to Paul Needham, who inspired the paper, who saved me from some blunders, who made me see that that the crucial move in the slingshot is the substitution of co-extensive predicates, and whose conversation and writings were tutorials on extensionality.

² The argument evidently acquired this marvelous name from Barwise and Perry (1983).

³ I follow Gödel, who suggested in Gödel, 1944 that Frege's 'bedeuten' be translated as "signifies". It's a fairly ordinary term and it relates Frege to medieval discussions of similar notions which are typically translated as "signify" ("signification").

⁴ Particularly Needham, forthcoming; Neale, 1995; and Neale and Dever, 1997. Neale, 1995 has an excellent and comprehensive bibliography of discussions of the slingshot.

⁵ "The Philosophy of Logical Atomism" reprinted in Russell (1956), p.182.

⁶ Though not crucial to evaluating the slingshot, the issue of what are taken to be true or false lies just beneath the surface of my central points, and the issue is surely very important for wider questions about an adequate conception of truth.

⁷ In this respect it is just like a (*de dicto*) belief context: John's belief that Stockholm is a large city can be characterized only as the belief that Stockholm is a large city – not even as the belief that the capital of Sweden is a large city (for John may not believe that Stockholm is the capital of Sweden).

⁸ The formulation is found both in "Causal Relations" (Davidson, 1980) and in "True to the Facts" (Davidson, 1984). Specific references are given in the next section.

⁹ Or occasionally, particular individuals considered as a unit, as with the demonstrative 'those'.

¹⁰ It is not necessary for the attributive use that we not know who Smith's murderer is, but the idea is more easily grasped when that assumption is made.

¹¹ Just as it is obscured by speaking of the substitution of singular terms for others which 'stand for', or 'designate' the same things. The common use of the term "represent" in this context is even more obscuring since it can indicate not only both *extension* and *reference* but also *sense*, as when it is said that a false sentence may represent what is not the case.

¹² Davidson, 1984, pp. 17-36. The quotations are from p. 19. Davidson wrote there that "the argument derives from Frege", citing Church's *Mathematical Logic*.

¹³ Davidson, 1980, pp. 149-162. The quotations are from pp. 152f.

¹⁴ Davidson, 1984, pp. 37-54. The quotations are from pp. 39-42.

¹⁵ It is worth noting that although Davidson defended this notion of a correspondence theory without facts by reference to Tarski's semantics for quantificational logic, it is in fact closer to older correspondence theories than to the fact-based ones which became current only after the latter part of the 19th century (central figures being Meinong and Russell). Earlier philosophers did not think of a sentence as signifying a fact, state of affairs, or other sentence-

like entity. They thought of a sentence as signifying the object (or objects) which its subject term signified and predicating something of that object (or objects). This is not Tarski, but closer to the Tarskian point of view Davidson used than to the fact-based correspondence he rejected.

¹⁶ My formulation is indebted to Neale, 1995 and to Needham (forthcoming).

¹⁷ Note that Davidson formulates the correspondence connective as "The *statement* that p corresponds to the fact that q", not as "The *true sentence* that p corresponds to the fact that q." I use 'sentence' rather than 'statement' as explained above, and I use "true sentence" to simplify the exposition but it does not affect the proof. My version requires that true sentences go in for both 'S' and 'q' on the ground that fact-based correspondence theories generally deny that false sentences correspond to anything. Davidson's version of the slingshot requires that what goes in for 'q' have *the same truth value* as 'S', which yields the claim that a false sentence corresponds to the fact that q, provided 'q' is a false sentence: that would make the fact that q a *false* fact, which is something correspondence theories want to avoid. The proof goes through either way.

¹⁸ Standard statements of the argument (like Davidson's) use sentences of the form '(tx)(x = a)' because in standard predicate logic, singular terms ('names' in my terminology) like 'a', which can flank the identity sign, are taken to have a reference, so that '(tx)(x = a)' is understood to refer to one and only individual. I use '(tx)Fx' to simplify the formulation of the argument, but for the argument to go through, 'Fx' must be understood to have the form 'x = a'.

¹⁹ If we take 'Fx' to signify 'is President of the U.S.' and 'p' to signify 'Today is Friday', then '(tx)(Fx & p)' can be read as 'the unique individual such that he is President of the U.S. and today is Friday'. If today *is* Friday, that description (at this writing) signifies Bill Clinton; if today is not Friday, it doesn't signify anybody, since in order for the description to signify something it has to be Friday today.

²⁰ 1995, p. 224. This is Searle's exact argument but with a different example.

²¹ This rewriting presumes that the crucial definite descriptions – '(tx)(Fx . p)' and '(tx)(Fx . q)' – have what has come to be called "narrow scope", for only if they are interpreted in that way are lines 4 and 6 entailed by their premises. If those definite descriptions are interpreted as having "wide scope", then lines 4 and 6 would be written thus:

4** $(\exists x)((Fx . p) . (\forall y)((Fy . p) \rightarrow y = x) . C(tx)(Fx = x))$

6** $(\exists x)((Fx . q) . (\forall y)((Fy . q) \rightarrow y = x) . C(tx)(Fx = x))$.

The inference from line 4** to line 6** involves no predicate substitution in the context of C, which means that it does not beg the question (by permitting predication substitution in the context of C), but it also means that the argument will not show that C is extensional since all the inferences will be outside the scope of C. This does not constitute a way of *evading* the slingshot, however, because it cannot be claimed that lines 4 and 6 are rightly interpreted as 4** and 6**, the reason being that it is obvious that 4** and 6** do not follow from the premises cited in the proof.

It might be argued that the way to deal with the slingshot is simply to argue that when it comes to C, definite descriptions should, in general, be taken to have wide scope. I am tempted by the idea that the referential use of definite descriptions could be reflected formally by giving them wide scope, and their attributive use by giving them narrow scope. But I have not been able to make sense out of the correspondence connective when definite descriptions are interpreted as having wide scope. (For discussion of the bearing of scope issues on slingshot-like arguments cf. Føllesdal, 1969.)

²² Neale (1995, p. 786) made this point when he wrote that Whitehead and Russell demonstrated that, "although descriptions are not genuine singular terms (in their system), if a predicate F applies to exactly one object (i.e. if it has exactly one object in its extension), in

truth-functional (i.e. extensional) contexts the description '(tx)Fx' can be treated *as if* it were a singular term for derivational purposes.... [This] is a derived rule of inference that can be used in truth-functional contexts...."

²³ Neale recognized this when he wrote that the Gödel slingshot is able to show that the supposedly non-extensional correspondence connective is extensional by employing the fact that "descriptions (as standardly understood) contain *formulae* as proper parts; by permitting the interchange of such devices when their contained formulae are satisfied by the same object, one is essentially permitting the interchange of formulae; and once some weak additional inference principle is assumed (t-CONV), the formulae in question can be drawn out of their iota-governed contexts to make the purportedly non-truth-functional S-connectives provably truth-functional." (1997, p. 151) His "formulae" refers to open sentences, and his point is that t-SUB permits the substitution of open sentences provided they "are satisfied by the same object", that is, have the same extension, which is to set up a fully extensional (truth-functional) context. Neale did not make this central to his critique, however, which is both puzzling and unfortunate from the point of view of clarity about the slingshot.

²⁴ This applies, for example, to Ilkka Niiniluoto's claim that Tarski's truth theory is "an explicate of the classical (fact-based) correspondence theory of truth." (Peregrin, 1999, p. 91) He argues that Tarski's account needs to be completed (improved?) by adding Carnap's claim (in *Introduction to Semantics*) that sentences designate propositions (rather than truth values), but he says nothing about Church's use of the slingshot to undermine this Carnapian idea. Nor does he consider how Tarski's notion that predicates are true of sequences can be consistent with Carnap's notion that they designate properties, or how Tarski's view that true sentences are satisfied by all sequences can be reconciled with the correspondence claim that true sentences correspond to particular facts. Niiniluoto's view strikes me as a hopeless case of wanting to have your cake and eat it too – of using a Carnapian ontology of properties, propositions and facts but failing to defend it against slingshot-type attacks on the ground that a *Tarskian* truth theory is immune to such attacks.

²⁵ For other examples of what the constraints exclude see Neale and Dever, 1997, pp. 156f.

²⁶ In "Do We Need Correspondence Truth?" (Peregrin, 1999, pp. 81-90) I have defended such a deflationist conception against those who claim that we cannot get along without correspondence truth.

REFERENCES

- Barwise, J. and J. Perry: 1983, *Situations and Attitudes*, MIT Press, Cambridge.
- Church, A.: 1943, 'Review of R. Carnap, *Introduction to Semantics*', *The Philosophical Review* 52, 298-304.
- Davidson, D.: 1980, *Essays on Actions and Events*, Clarendon Press, Oxford.
- Davidson, D.: 1984, *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford.
- Davidson, D.: 1986, 'A Coherence Theory of Truth and Knowledge', E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford.
- Donellan, K.: 1966, 'Reference and Definite Descriptions', *The Philosophical Review* 75, 281-304.
- Føllesdal, D.: 1969, 'Quine on Modality' in D. Davidson and J. Hintikka (eds.), *Words and Objections*, Reidel, Dordrecht.

- Gödel, K.: 1944, 'Russell's Mathematical Logic' in P. A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, Open Court, Chicago.
- Hahn, L. E., (ed): 1999, *The Philosophy of Donald Davidson*, Open Court, Chicago.
- Neale, S.: 1995, 'The Philosophical Significance of Gödel's Slingshot', *Mind* **104**, 761-825.
- Neale, S and J. Dever: 1997, 'Slingshots and Boomerangs', *Mind* **106**, 143-163.
- Needham, P.: 1994, 'The Causal Connective', J. Faye, U. Scheffler, and M. Urchs. (eds.), *Logic and Causal Reasoning*, Akademie Verlag, Berlin, 67-89.
- Needham, P.: (forthcoming), *Law and Order: Problems of Empiricism in the Philosophy of Science*.
- Peregrin, J.: (ed.) 1999, *Truth and Its Nature (If Any)*, Kluwer Academic Publishers.
- Pitcher, G.: (ed.) 1964, *Truth*, Prentice-Hall, Inc.
- Russell, B.: 1905, 'On Denoting', *Mind* **14**, 479-93.
- Russell, B.: 1956, *Logic and Knowledge*, R.C. Marsh (ed.), George Allen and Unwin, London.
- Searle, J.: 1995, *The Social Construction of Reality*, Allen Lane – The Penguin Press, London.
- Taylor, B.: 1985, *Modes of Occurrence: Verbs, Adverbs, and Events*, Blackwell, Oxford.

FROM *ERKLÄREN-VERSTEHEN* TO *PREDICTION-UNDERSTANDING*:
THE METHODOLOGICAL FRAMEWORK IN ECONOMICS¹

Within the methodology of social sciences, a classic controversy opposes *Erklären* (explanation) to *Verstehen* (understanding), and within the methodology of economics, a more recent debate confronts *prediction* to *understanding*. However, both disputes have a common ground insofar as the prominence of *Erklären* or *prediction* ordinarily goes in favor of the unity of science and the accent in *Verstehen* or *understanding* normally conducts towards the diversity of science (i.e., a distinct characterization of science in the social sciences, in general, and economics, in particular).

On the one side, those who stress the importance of *Erklären* and those who defend *prediction* as the main goal and principal test of economics tend habitually to see the economic method in tune with the method of natural sciences. And, on the other side, those authors who highlight the role of *Verstehen* or emphasize the function of *understanding* generally tend to distinguish social sciences (among them, economics) from natural sciences. Meanwhile, those who prioritize understanding the economic phenomena over predicting them tend to insist on the methodological difference between the social and the natural sciences.

Raimo Tuomela studied the first controversy in his analysis of G. H. von Wright's book *Explanation and Understanding*², which deals with the methodology of social sciences and constitutes a sophisticated version of the dichotomy between *Erklären* and *Verstehen*³. Tuomela's analysis is useful for clarifying that approach, one of the latest stages in that methodological dispute – the sixth one – , which has clear repercussions to the methodological framework of economics and, indirectly, affects the role of prediction in social sciences, in general, and in economics, in particular. His position will be seen here from the point of view of causality and teleology, looking at its possible impact upon the methodological structure of the social sciences and the problem of prediction in economics. In this regard, it should be taken into account that prediction is frequently used as a test of economics as a science and, habitually, that character affects the issue of the scope of economics⁴.

After clarifying the methodological framework of *Erklären* and *Verstehen*, the attention will be on the question of the methodological debate which confronts specifically *prediction* to *understanding*. The problem will be addressed following the recent dispute between Alexander Rosenberg and D. N. McCloskey. Keep in mind the methodological controversies within economics, which show clear differences between the mainstream tendency – the neoclassical economics – and the critics of this position (especially Herbert A. Simon, winner of the Nobel Prize in Economics in 1978 and the new institutional school). This methodological debate in economics involves a tension between two orientations: i) emphasis on the importance of the success of economic predictions⁵, and ii) a more realistic attitude,

SOCIAL INSTITUTIONS

My aim in this paper is threefold. First, I will make some suggestions concerning the nature of social institutions. This will consist chiefly in distinguishing between social institutions and other related social phenomena, and describing some of the defining features of institutions. Second, I will present a critique of John Searle's account of social institutions.¹ Third, I will sketch an alternative teleological account of institutions.

In the opening section I distinguish between, what I will term, joint action, spheres of activity, conventions, social norms, rules, social groups, organisations and (finally) institutions. I should also point out that I have offered detailed treatments of these notions – excluding institutions – elsewhere (Miller, 1992a, 1992b, 1997a, 1997b, 2001).

I. A TAXONOMY OF SOCIAL ACTION AND SOCIAL ENTITY TYPES

Joint actions consist of the individual actions of a number of agents directed to the realisation of a collective end. A collective end – notwithstanding its name – is a species of individual end; it is an end possessed by each of the individuals involved in the joint action. However it is an end, which is not realised by the action of any one of the individuals; the actions of all or most realise the end. Examples of joint action are two people lifting a table together, and two men jointly pushing a car. So joint actions can be analysed in terms of individualist notions.²

Moreover, I hold that these individualist notions do not necessarily bring with them a normative dimension, other than the minimal normativity of instrumental rationality. By instrumental rationality I mean the rationality of undertaking an action which is the means to one's end; I ought to do x for the reason that x is the means to my end, e. Accordingly, I conclude that joint actions do not necessarily have a normative dimension (other than in this minimal sense).

We can further elaborate the notion of joint action, by introducing the notion of *joint activity*. Roughly speaking, a joint activity is a complex of differential, interlocking, joint actions and individual actions directed to some overarching collective end. So building cathedrals and sailing ocean liners are joint activities.

We can further distinguish between different, albeit connected and overlapping, *generic* kinds of joint activity, including communicative, economic, educative, sexual and religious activity. The repetition over time, and duplication in space, of any one of these different kinds of generic joint activity can give rise to a more or less connected, and more or less continuous, stretch of joint activity which I will term a *sphere of activity*. So the ongoing series of economic transactions across Finland constitutes a sphere of activity.

Spheres of activity (by stipulative definition) are regulated by *conventions*. They thereby take on different specific forms according to the specific conventions which structure them. Often, though not necessarily, they are also regulated by explicit rules, including laws. And they are also often – but again, not necessarily – regulated by social norms. Here I distinguish between rules, conventions and social norms.³

A *rule* is a (typically explicitly formulated) directive issued by some authority or other to undertake a certain course of action in certain circumstances. So laws are a species of rule. A convention is a regularity in action performed by a set of individual agents and directed to a collective end. More precisely, conventions are *joint procedures*, where a procedure is (roughly speaking) a habit; the automatic repetition of an action is a recurring situation. So a joint procedure is a procedure which is jointly followed because it realises a collective end.

Moreover, in my view conventions *qua conventions* are normatively neutral – other than in the minimal sense involved in instrumental rationality. I maintain that conventions are normatively neutral notwithstanding the fact that agents can be held to have failed to conform to any given convention. On my account a conventional regularity in action does not constitute a *standard of behaviour* to which agents *ought* to conform. Naturally, an agent by not conforming might fail to realise the collective end that he has; the end which is realised by conformity to the convention. However, that would be a mere failure of instrumental rationality.

So conventional actions are not necessarily the content of directives issued by authorities. Nor are they necessarily explicitly formulated anywhere. Thus driving on the left hand side is a convention in Australia, though not in Finland; it is a regularity in action among Australians, which secures the collective end of avoiding car collisions. As it happens, this convention is also a law, though it might not have been.

Social norms are regularities which are also norms; agents believe that they *ought* to conform. Here the 'ought' is not that of mere instrumental rationality; it is not simply a matter of believing that I ought to conform because it serves my purpose. Some conventions and all rules are also norms in this strong sense. For example, the convention and the law to drive on the left is a norm; people feel that they ought to conform. This strong (and wide) sense of 'ought' *includes* – but is not exhausted by – the so called *moral* 'ought'.

We need now to make a threefold distinction between social groups, organisations and institutions. The defining elements of social groups are particular *individuals* standing in certain relations – including normative relations (in the strong sense) – to one another. The defining elements of organisations are (embodied) particular *roles* standing in relations to one another. By embodied it meant that these roles are filled by individual persons – though not by any individuals in particular.⁴

So organisations consist of an (embodied) formal structure of interlocking roles.⁵ These roles can be defined in terms of tasks, procedures (in my above sense) and conventions. Moreover, unlike social groups, organisations are individuated by the kind of activity that they undertake, and also by their characteristic functions or ends. So we have governments, universities, business corporations, armies, and so on. Perhaps governments have as an end or goal the ordering and leading of

societies, universities the end of discovering and disseminating knowledge, and so on.

It is important to note that on my (stipulative) definition of organisations they are, *qua organisations*, non-normative entities. In this respect they are analogous to conventions, as I have defined conventions. So being an organisation is not of itself something that is either morally good or bad, any more than being a convention is in itself morally good or bad. I can consistently hold this while maintaining that organisations, as well as conventions, are a pervasive and necessary feature of human life, being indispensable instruments for realising collective ends. Collective ends are a species of individual end; but merely being an end is in itself neither morally good nor morally bad, any more than being an intention or a belief are in themselves morally good or morally bad.

While my definition of an organisation does not include any reference to a normative dimension, most organisations do as a matter of contingent fact possess a normative dimension. As was the case with conventions, this normative dimension will be possessed (especially, though not exclusively) by virtue of the particular moral (or immoral) ends that an organisation serves, as well as by virtue of the particular moral (or immoral) activities that it undertakes.⁶

Organisations with the above mentioned normative dimension are *social institutions*.⁷ So institutions are often organisations, and many systems of organisations are also institutions. For example, capitalism is a particular kind of economic institution, and in modern times capitalism consists in large part in specific organisational forms, including multi-national corporations. Nevertheless, some institutions are not organisations, and do not require organisations. For example, the English language is an institution, but not an organisation. Moreover, it would be possible for a language to exist independently of any organisations specifically concerned with language.

Institutions are (by stipulative definition) normative entities, defined in part in terms of normatively described ends, and in part in terms of activities governed by social norms. So institutions are unlike organisations. Nevertheless, there are important relations between many institutions and organisations. Firstly, as we saw above, many institutions are also organisations, or have organisational elements. Secondly, organisations exist, or at least ought to exist, to facilitate institutional purposes, and ideally, *morally desirable* institutional purposes.

We saw above that social groups necessarily possess institutions. Moreover, there can be no institutions outside social groups.⁸ Further, institutions logically presuppose spheres of activity. An institution is (at least) a structure of conventions and social norms which regulates some form of generic joint activity in accordance with more or less discernible ends. Communicative activity is not necessarily institutional activity, but typically linguistic activity is. Spiritual experiences are evidently not necessarily institutional phenomena, but religious activity – including non-organisational religious activity – is typically institutional in character. Extramarital sexual activity is not necessarily institutional in character, but marital sex is. Periodic bartering between explorers and members of local tribes is non-institutional economic activity, whereas Finland's export of paper and import of cars within the framework of EU guidelines is institutional activity.

A final feature of institutions is what I will term (following Giddens, 1984⁹) their enabling capacity.¹⁰ Institutions not only constrain joint activity, they also facilitate it in ways that allow new and different actions and relationships to come into existence. For example, and famously, the institution of language enables speakers to construct entirely new sentences. In this regard a key feature of institutions is their use of, what I term, *joint institutional mechanisms*. I will discuss joint institutional mechanisms in the final section of this paper.

In the next section I will turn to a consideration of Searle's account of institutions. Before doing so, let me stress one final point. The taxonomy I have offered divides the social action types and entities thus taxonomised into two basic kinds: those with a normative dimension and those that are (in themselves) normatively neutral. So actions (including joint actions), spheres of activity, conventions and organisations do not per se have a normative dimension (in my strong, wide sense), while social norms, rules, social groups and institutions do.

2. SEARLE ON INSTITUTIONS

According to Searle, institutional facts are (ultimately) physical objects, or states of affairs or events, upon which, what he terms, status-functions have been collectively imposed (Searle, 1995). I say "ultimately" because although status-functions can be collectively imposed on pre-existing institutional facts, any such iterated structure of status-functions must terminate at some point in physical objects or events (or more precisely, must terminate in what he terms, brute facts) (*ibid.*, p. 27, p. 34).

Something has a status-function if it has, or those who use it have, deontic powers. Thus a police officer has a status-function, and therefore a set of deontic powers, including rights to stop, search and arrest people under certain conditions. A five dollar bill is a piece of paper (a physical object) the bearers of which have various deontic powers, including the right to exchange the bill for goods to the value of five dollars. These status-functions, and therefore deontic powers, have been collectively imposed in the sense that the relevant members of a community accept or agree to or otherwise treat the objects or persons that possess these status-powers as if they do in fact possess them. But in accepting or so treating, for example, the police officer as if he has the right to arrest people, the police officer comes to have that right. By Searle's lights, if no-one ever paid any attention to police officers they would cease to have any deontic powers and therefore any status-function; indeed they would cease to be police officers. Similarly, if no-one was prepared to exchange five dollar bills for goods then these bits of paper would cease to have any status-function, and the bearers of them would cease to have any deontic powers.

Searle's account of institutions makes use of four basic notions: (1) imposition of functions; (2) the deontic powers of institutional persons and objects; (3) the distinction between constitutive and regulative rules; (4) collective intentionality.¹¹ Let us look more closely at these notions, beginning with functions and deontic powers.

2.1. Functions and Deontic Powers

Searle's notion of function concerns what it is that a group collectively imposes on a physical phenomenon. For example, if members of a community began to sit on a log then the log would in effect have become a bench. So a function – that of being used to sit on – would have been collectively imposed on a physical object.

On Searle's conception all functions ultimately depend on the existence of physical objects on which functions are imposed. However, some functions – such as the function of being a chair – depend on the *specific* physical properties of the object on which they are imposed. Thus a log can become a bench only if it has a certain size and shape. By contrast, the key feature of *institutional* facts is that they involve functions, which cannot be imposed simply by virtue of the specific physical properties of the phenomena on which they are imposed. Rather possession of the function exists by virtue of the *collective* character of the imposition. (Searle, 1995, p. 39.) So being a chair is not an institutional fact; rather its functionality exists by virtue of its specific physical properties. On the other hand, being a medium of exchange is an institutional fact; its functionality supposedly exists by virtue of collective imposition rather than specific physical properties.

It should be noted that according to Searle functions can be individually imposed on physical objects by individuals. This would be the case if, for example, Robinson Crusoe alone used the log as a bench. Further, if a log had the function of a bench *collectively* imposed on it, then its being a bench would be a social fact. (*Ibid.*, p. 88.) However, the fact that the log was a bench would not be an *institutional* fact; for its being a bench does not depend on the *collective* character of the imposition of this function.

According to Searle, in the case of institutional facts the imposition of a function consists not only in the assignment of a function, but also of a status with deontic properties (*ibid.*, pp. 100–101). These deontic properties – rights and duties and the like – are regarded by Searle as powers. If a police officer has a right to arrest you, then he has a power to do so; if you have a right to goods worth five dollars in exchange for your five dollar bill, then you have a power to receive those goods. And, to reiterate, these status functions or powers are possessed by virtue of the agreement or acceptance of the community; the powers are collectively imposed and maintained.¹²

The tight connection that Searle makes between function imposition and deontic properties is problematic. Searle asserts that the notion of a function necessarily brings with it values and deontology. Thus if the purpose of the heart is to pump blood then it ought to pump blood. (Searle, 1995, pp.15–16, p.19) No doubt purposes and ends can be assessed in the sense that they can be said to have been realised or to have failed to be realised. Moreover, if one has an end then, other things being equal, one ought to enact the means to that end. So functions, goals and the like bring with them normativity in the minimal sense pertaining to instrumental rationality. But the claim that functions and purposes bring with them normativity in any stronger sense has not been shown. Specifically, Searle has not shown that his deontic properties, such as rights and duties can be derived from the notion of a function. Nor has he shown that deontic properties can be derived from his notion of function taken in conjunction with collective intentionality. Indeed, it would seem

that his own examples of tool using social animals is sufficient to demonstrate that deontology is not generated by functionality, whether collectively imposed or not.

A more plausible source of *some* deontic properties is rules – in my above defined sense of that term viz. explicitly formulated directives issued by authorities to perform certain kind of actions under certain circumstances. However, Searle claims that ‘rules’ in the very general sense of *general policies* are capable of abuse, and hence have a normative status (*ibid.*, p.48). I do not accept the proposition that general policies or habits necessarily generate *standards* of conduct – as opposed to mere regularities in behaviour – and therefore deontic properties. Certainly Searle offers no demonstration of his claim.

At any rate I suggest that this above-described non sequitur – moving from individual or collective imposition of functional properties to deontic properties – infects Searle’s account of status-functions. For it is perfectly possible for a person or an object to have a set of functions – and be treated as having those functions – without any associated (non-instrumental) deontology. For example, a set of individuals might use a certain sort of relatively rare shell as a medium of exchange, and do so notwithstanding the fact that no-one had any desire to possess these shells independent of the fact that they could be used as a medium of exchange. In this scenario all that is required is that each exchanges shells for goods, and goods for shells, intends to continue to do so, and believes that all the others do and intend likewise. Of course it would add greatly to the stability of this arrangement if these shells were somehow authorised as an official medium of exchange, and if a (rule constituted) system of rights and duties in relation to the exchange of these shells was introduced and enforced. However, such a deontological structure is not a necessary feature of the system of exchange.¹³

Moreover, even where a deontological framework of the kind described by Searle has been adopted, the relevant deontic powers do not subsume, take the place of, provide the basis for, or go hand in hand with, the substantive functions of the objects or persons in question. Here by *substantive* functions I mean the functions which are *central* and necessary to the object or person being the kind of thing that it is. Consider an incompetent surgeon who is incapable of performing a successful operation on anybody, and who largely avoids doing so, or when he absolutely has to, always ensures that he is part of a team comprised of other competent surgeons and nurses who actually do the work. By virtue of being a fully accredited surgeon this person has a set of deontic powers, including the right to perform surgery, and others have deontic powers in relation to him, including the right that he performs operations competently and with due diligence. Moreover, these deontic powers are maintained in part by (say) the Royal College of Surgeons, his colleagues and the community. However, the surgeon simply does not possess the substantive functions of a surgeon. The deontology is there, but the substantive functionality is not. Accordingly, it is simply false to claim, as Searle must do, that he is a surgeon. (Searle, 1995, pp.49–50)¹⁴ If someone cannot perform, and knows nothing about, surgery he cannot be a surgeon, irrespective of whether he has the highest professional qualification there is, is treated as if he were a surgeon, and indeed is widely believed to be the finest surgeon in the land.

It might be thought that this kind of example suits my argument by virtue of the physical character of the role of surgeon. But my point can be made in relation to

non-physical roles and objects. Consider a philosophical article published in a leading journal. This is an institutional fact. Now recall the recent notorious Sokal hoax. Sokal, a physicist, decided to write a piece of gibberish and submit it to a learned journal specialising in postmodernist writings. The journal accepted his piece, and published it. So the piece of writing was institutionally certified as being a very high quality philosophical article. Moreover, many who read it believed this to be the case. In reality, however, it was nothing of the kind; it was gibberish and intentionally created by Sokal as such. So the ‘philosophical article’ possessed the relevant collectively imposed deontic properties, and therefore on Searle’s account constituted the institutional object, philosophical article. However, this piece was not a philosophical article; for it lacked the intellectual properties necessary for it to be so. Accordingly, the collective imposition of deontic properties was not sufficient to generate the relevant institutional fact. Institutional authorisation and collective intentionality cannot turn entirely incompetent surgeons into real surgeons; nor can they turn gibberish into real philosophical articles.

I conclude that in relation to institutional facts Searle’s notion of deontic power – whether collectively imposed or not – does not necessarily subsume, take the place of, provide the basis for, or go hand in hand with, substantive functionality. Substantive functionality and deontic powers can come apart. But let us turn now to Searle’s notion of collective intentionality, and the related notions of social acceptance and social agreement.

2.2. *Collective Intentionality, Collective Acceptance and Constitutive Rules*

Searle’s apparatus of status-functions is created and sustained by, what he terms, collective intentionality. As we have just seen, his view that the substantive functions of objects and persons is subsumed by, or necessarily goes hand in hand with, deontic powers is incorrect. Moreover, it is now clear that collective intentionality cannot generate the substantive functions of many institutional persons and objects. What of the relationship between collective intentionality and deontic powers? Let us get a little clearer on the notion of collective intentionality.

Collective intentionality is for Searle a primitive notion expressed by locutions such as “we-intend”. A we-intention is not reducible to an individual intention, nor to an individual intention in conjunction with other individual attitudes such as individual beliefs. (*Ibid.*, pp.24–26) I do not accept Searle’s conception of primitive we-intentions. As argued elsewhere, I take the view that an analysis is possible using only individualistic notions, including especially the above-mentioned notion of a collective end. In any case Searle’s notion of primitive we-intentionality is under-theorised. He has not provided an account of it, analytical or otherwise.

But that aside, what of the role of collective intentionality – however it is to be understood – in relation to the creation and maintenance of deontic powers? Consider the deontic powers associated with the utterance of sentences, such as the right of hearers that speakers would aim at the truth. Searle must hold that such a right can only exist if a group adopts some sort of we-attitude eg we-accept or we-agree or we-belief or we-intention, to these deontic powers. And presumably the

same goes for other moral rights, such as the right to life, the right to freedom, and so on.

Now *prima facie* hearers have a moral *right* that speakers aim at the truth, and a moral *right* to life, and so on, quite independently of whether the community we-intends that this be the case, or accepts that it is, or whatever. So the *source* of at least some central deontic properties, namely, some moral properties, is evidently not collective intentionality.

Naturally, whether or not deontic properties, including rights, are *respected* might ultimately turn on the attitude of the community. So general community 'acceptance' of deontic properties, including rights – in a weak sense of acceptance – is necessary if a specific deontological framework is to take effect in that community. So much is trivially true.¹⁵

I conclude that Searle has not established that all, or even most, deontic properties associated with institutional objects and persons are (so to speak) *ontologically* dependent on collective intentionality.¹⁶

However, are there other cases of deontic properties – especially those involved in *authority* relations – which might be ontologically dependent on collective intentionality.

Consider Peter Sellars in the movie, *Being There*. Sellars plays the role of a gardener who for various reasons begins to be treated by the staff of the President, and ultimately by everyone, as if he were the President of the USA. Eventually, he can even have run for office and be elected. Unfortunately, he has no understanding of the political system, or of relevant policies, and has no leadership qualities whatsoever. Nevertheless, it seems to be the case that the gardener can become, the President by virtue of collective acceptance.

Institutional *authorities* are vulnerable to a degree that other institutions, such as language, are not. As Searle points out, the communist government of Russia turned out to have clay feet (*ibid.*, p.91). Once people chose not to obey its directives, it was finished; it simply ceased to function or exist as a government. However, it is difficult to see how the English language could go out of existence in such spectacular fashion; for it depends on millions of often-disconnected communicative interactions between millions of different people.

There is a reason for the vulnerability of institutional authorities. In the special case of institutional authorities deontic properties are, as Searle holds, *ontologically dependent* on collective acceptance; no collective acceptance, no deontic properties. The point here is not simply that (say) rulers cannot *exercise* their right to rule if their right to rule is not collectively accepted. Rather a ruler does not even possess a *right* to rule unless she is able to exercise authority over her subjects. This seems to be a general feature of the deontic properties of those in authority. So Searle is right about institutional authorities, but wrong about other institutional forms.

Moreover, Searle is also correct to speak of the rights the rights possessed by institutional authorities as being *powers*. Indeed, the actions of those in authority, including those whose authority is not morally legitimate, constitute in large part the *exercise of power*, and not merely the exercise of authority. But from this it does not follow that all social actions somehow necessarily involve exercises of (social) power. In particular, it does not follow that actions in conformity to conventions and social norms are necessarily exercises of power, either on the part of those who

conform, or on the part of those upon whose conformity one's own conformity is dependent. Perhaps conventional actions are for the most part rational and voluntary actions performed in the service of mutual interest. (See Lewis, 1969, and Miller, 1992b)

I now want to focus attention on the connection between collective acceptance and substantive functionality. I will begin with the special case of institutional authorities.

What is it to collectively accept or collectively treat someone as if he or she is a leader? Presumably it is for people to obey her directives because they believe that they ought to do so. But now what is the actual or realised substantive functionality of a leader? Surely it is in large part getting people to obey one's directives because they believe that they ought to do so. So in the case of leaders, and other institutional authorities, I suggest that Searle's notion of collective acceptance is redundant; it collapses into actual or realised substantive functionality. To possess the functional properties of a leader is (in large part) to have one's directives obeyed by one's followers. But to have one's directives obeyed by one's followers is to be collectively accepted.

What of the relation of collective acceptance and substantive functionality in the case of other kinds of institution, such as money or language? Once again I suggest that Searle's notion of collective acceptance collapses into the actual or realised substantive functionality of the institution.¹⁷ If people exchange dollar notes for the purpose or end of receiving goods (and exchange goods for the purpose of receiving dollars) then they have 'accepted' dollars in the only important sense of that term; so there is no need for an additional notion of collective intentionality or collective acceptance or treating dollars as if they were dollars. A similar point can be made for language. If speakers and hearers utter a set of specific noises for the purpose or end of referring to specific things, communicating thoughts to others, and so on, then they have 'collectively accepted' the language in the only important relevant sense of that term. That is, the substantive functionality of language has been realised; the speakers and hearers are using language.

And there is this further point. Speakers and hearers – and persons using a medium of exchange etc. – typically act in conformity with conventions, and indeed, social norms. So the substantive functionality of these institutions is realised in accordance with conventions and social norms. Accordingly, there is collective acceptance in the sense of conformity to conventions and social norms. When we use language or a medium of exchange we typically do so in part by conforming to the relevant conventions and social norms. So much is obvious.

A final point about Searle's notion of collective intentionality and its relation to rules. In his later formulations it turns out that the we-intentional state in question is conventional in character; he speaks of conventional powers (Searle, 1995, 104). Searle no-where offers an analysis of conventions. However, he does distinguish them from rules. For Searle, conventions, by contrast with rules, are arbitrary (*ibid.*, p.28). Moreover, as we saw above, he takes the view that rules are necessarily normative phenomena.

There are a number of unanswered questions here. What exactly is the nature of rules and of conventions, and precisely how are these phenomena to be differentiated?

While Searle does not explain the nature of conventions or the nature of rules in general, he does offer an account of the distinction between constitutive rules and regulative rules. Moreover, constitutive rules have a key role in his account of institutions. Unlike regulative rules, constitutive rules "create the possibility of certain activities" (*ibid.*, p.18). Constitutive rules have the form 'X counts as Y in C' where Y is the function imposed on X in conditions C. (*Ibid.*) So constitutive rules are, according to Searle, very important in the construction of social institutions.

Now it is by no means clear that some *conventions* could not have the form "X counts as Y in C". On Searle's view of conventions as arbitrary, a convention-governed imposition of Y on X (if it were possible) would necessarily be a matter of arbitrary choice. But this seems to be so in many instances. Consider a convention among military strategists to use certain shaped pebbles to stand for troops in their strategising concerning troop movements.

More important, it is by no means clear that the notion of a *rule* – as opposed to a certain realised rule *content* viz. that persons count something as something else – is actually doing any work here. I argued above that function imposition does not necessarily bring with it values and deontology. Now I suggest that function 'imposition', whether collective or not, does not require rules, or even conventions. Consider the above-mentioned example of military strategists collectively imposing functions on physical objects. If they used toy soldiers that looked like real soldiers then the arbitrary character of their decision would diminish. So by Searle's lights there would not be a convention. But equally there might be no rule. It might be a one-off episode, never to be repeated. The point is that it is the notion of counting one thing as another thing that is doing the work; whether they count x as y in accordance with a convention or rule or by virtue of some other collective device is not important.

This latter point is really a special case of the more general point made earlier, namely, that collective acceptance collapses into actual or realised substantive functionality. For what is critical is that the relevant individuals perform the action viz. treat the gardener as if he were President, count the pebble as a soldier, use the paper as a medium of exchange, and so on. What is critical is not constitutive rules, but rather substantive functionality.

In the opening section of this paper I located the notion of a social institution in relation to that of an organisation, and in the introduction to this paper I offered a preliminary description of social institutions. In this section I have provided a critique of Searle's account of institutions. In the next section I will return to my own view.

3. SOCIAL INSTITUTIONS

As earlier indicated, I agree with Searle's general view that social entities, including social institutions, logically depend on social actions. However, I do not subscribe to Searle's account according to which the social actions in question are essentially the collective imposition of status functions on pre-existing social objects, and, ultimately, natural objects. Rather I view institutions as structures of conventions and norms that regulate various species of generic joint activity, and I view institutional actions as actions performed in accordance with the conventions, norms, ends, roles etc that constitute institutions.

Moreover, I hold the view that institutions should be in part be marked off from other related phenomena, such as organisations, by virtue of their normative dimension. And I accept Searle's view that this normative dimension includes the ascription to persons and objects of status bearing deontic properties. However, as argued above, I do not accept that the source of those deontic properties is collective acceptance, other than in the special case of institutional *authorities*.

In relation to institutional authorities, I endorse Searle's view that the relevant deontic properties are powers, and that those in authority are engaged in the exercise of these powers. But, contra Searle, I do not accept that the exercise of power is constitutive of other social actions, such as actions in conformity to conventions or social norms.

Moreover, in my view, the normative dimension of institutions, including hierarchical institutions comprised of authority relations, involves more than Searle's deontic powers. For one thing, it involves the felt ethical or moral worth of the ends or purposes of institutions, and specifically the kinds of subjectively felt moral goods that institutions produce or fail to produce. Accordingly, the definition of an institution will typically include a description of the (alleged) moral good that it purports to produce. For example, governments should provide leadership, universities purport to produce knowledge and understanding, language enables the communication of truths, marriages facilitate the caring and moral development of children, economic systems ought to produce material well-being, and so on.

Moreover, these moral goods, or at least believed moral goods, are, normatively speaking, the *collective* ends of institutions, and as such they conceptually condition the social norms that govern, or ought to govern, the constitutive roles and activities of members of institutions, and therefore the deontic properties that attach to these roles. To use one of Searle's examples, a police officer has certain deontic powers of search, seizure and arrest, but these powers are justified in terms of, and generated by virtue of, the moral good, law and order (say) that it is, or ought to be, the role of the police officer to maintain.

Further, and as I have argued above, a defining property of an institution is its substantive functionality, and so contra Searle, a putative institutional entity with deontic properties, but stripped of its substantive functionality, typically ceases to be an institutional entity, at least of the relevant kind; would-be surgeons who cannot perform surgery are not surgeons. Here, by substantive functionality, I have in mind the specific defining ends of the institution or profession. In the case of institutions,

including professions, the defining ends will be collective ends; they will not in general be ends that an individual could realise by his or her own action alone. In short, the theory of institutions, and of any given institution, is a *teleological* theory.¹⁸

Further, institutions in general, and any given institution in particular, require both a teleological *descriptive* theory, and a teleological *normative* theory. Naturally, whether or not my commitment to teleological descriptive theories of institutions is warranted depends on empirical facts. If it turned out, for example, that most or all institutions did not have collective ends that were regarded either as intrinsic moral goods, or the means to intrinsic moral goods (derived moral goods) – that is, the participating agents did not in fact seek to realise the relevant putative defining collective ends – then my teleological descriptive theory would be false; I would have to abandon it. Moreover, the falsity of the teleological *descriptive* theory would put pressure on the acceptability of any teleological *normative* theory of institutions. If it turned out that no institution at any time or place *in fact* involved to any extent the pursuit of the relevant kind of collective end that was an *objective* (intrinsic or derived) moral good, then this would make it implausible to claim that institutions nevertheless in general *ought* to aim at collective ends that are objective (intrinsic or derived) moral goods.

But it seems quite clear that participants in many institutions do as a matter of fact seek to realise collective ends that are both believed to be moral goods, and are in fact moral goods, and that these collective ends are sufficiently fundamental to be regarded as definitive. Consider the institution of the English language. If the relevant speaker/hearers ceased to have as a collective end the successful communication of their beliefs, intentions and the like, then the English language would go out of use. Again, consider the institution of the nuclear family. A mother and a father typically (jointly) act having as an important end the material and moral wellbeing of their child. For example, one parent might work and the other perform a child minding role. It is obvious that the well-being of the child is not simply an unintended consequence of their actions; it is something that they explicitly have *in mind*, and indeed, in all probability, they will explicitly explain or justify their actions in terms of their contribution to this collective end. Now consider educational institutions. Teachers and students typically have as a collective end the attainment by the students of intellectual skills and knowledge. That the students come to possess these skills and knowledge is an end which teachers and students not only jointly pursue, but explicitly claim that they are pursuing in school mottos, discussions with students, and so on. Indeed, often the most powerful criticism that a student or a teacher believe that they can make of a particular class is that the students do not learn anything.

In relation to the above it might be argued, firstly, that nevertheless there are some institutions, especially economic institutions, where there are no such collective ends, and, secondly, that I have failed to identify the most important social consequences of the kinds of institutional activity that I have discussed.

This second point is hard to rebut in the absence of an account of what these social consequences are supposed to be. Candidate accounts include those propounded by marxists and functionalists. According to a functionalist, the most important social consequence of child rearing might be to ensure an orderly and

harmonious society. According to a marxist, the most important social consequence of schools might be to inculcate an ideology that enables class exploitation.

I accept that sometimes the most important consequences of social institutions are consequences that are not also collective ends of the kind that I have described. But this does not show that these collective ends are not definitive of social institutions, and that the social consequences in question are definitive of institutions. Air pollution and associated ill health might turn out to be the most important social consequence of the use of taxis in New Delhi, but nevertheless, taxis are to be defined as modes of transport – people use taxis in order to get from one place to another – and not in terms of their contribution to air pollution. Nevertheless, air pollution is a huge problem in New Delhi, and, if possible, taxis should be built so as to limit their polluting effects. Similarly, if some schools are primarily sites of ideological indoctrination, then this needs to be looked into and changed; but the claim that *necessarily* schools are predominantly sites of ideological indoctrination is simply false.

Further, quite often macro-level social *consequences* are in fact a collective end of institutions. A single teacher and her students have as a collective end the education of these particular students. The realisation of this collective end contributes to the realisation of a larger collective end, namely, the education of all the students in the school; and the realisation of this larger collective end contributes to the contribution of a still larger collective end, namely, the education of all the students in the whole education system. What we have here is what I have described elsewhere as a layered structure of joint action. (*Ibid.*)

What of the other complaint; that some institutions, notably the economic system, do not centrally involve collective ends. I suggest that the best way to understand organisational action, including in the case of business organisations, is in terms of *quasi*-joint action. There is a collective end eg. the production of cars, but there is also an individual end eg. securing a wage, and production of cars exists as a collective end, but it depends on the existence of the individual end to secure a wage. For no one would be prepared to work if they were not paid. So while the collective end is not merely a means to the individual end, nevertheless the collective end is not sustainable without the individual end. Receiving a wage is a necessary, but not a sufficient condition, for pursuing the organisations collective ends.

Perhaps the most influential model of economic behaviour is that put forward by rational choice theorists. However, rational choice theory is a theory that purports to supply *explanations* of social action. By contrast, my account is an account of what social action *consists in*. It is consistent with my account of business organisations as layered structures of quasi-joint action, that the ultimate motivation of the actions of economic actors is some benefit to themselves; but my account does not require that this be so.

On the other hand, the rational choice model faces problems that my account does not face. Consider the problem posed by free-riding to rational choice accounts that seek to explain the production of public goods in terms of the individual actions of rational self-interested actors. As Mancur Olsen (1965) famously demonstrated, such rational self-interested actors would free ride, rather than make their contribution to the production of the public good. By contrast, my account of

organisations is not committed to self-interested actors. Indeed, all institutions, including most organisations, involve conformity to social norms, and conformity to social norms is not able to be understood simply in terms of self-interested motives.

As far as the economic system itself is concerned eg. the modern capitalist system, it is inevitably a mix of competition and cooperation. However, the conventions, laws and institutional regulations that constitute the framework within which competitive economic behaviour takes place ought to be in large part contrived in terms of collective ends that are objective moral goods, and especially the provision of material well-being to the participating economic actors. The fact that the current capitalist system is inadequate in this respect merely serves to illustrate the need for institutional reform. Arguably, its inadequacies pertain more to economic justice than economic efficiency. For there are profound questions of economic justice, both in relation to the differential economic benefits that go to those explicitly engaged in economic activity within the existing system eg. salaries of many chief executive officers of multinational versus the salaries of 'sweat shop workers' in some of the factories in Asia owned by those multinationals, and in relation to those who are largely excluded from the existing economic system. Consider in the latter connection, many women in contemporary western societies, and many citizens of currently existing nation states of Africa.

A different kind of objection to my teleological account of institutions is that it fails to accommodate manifestly morally unacceptable institutions, such as the institution of slavery. Assume that slavery in (say) the USA was set up as an institution for the purpose of providing cheap labour. There are two salient questions. First, what is the defining collective end(s) of this institution of slavery? Second, in what sense, if any, was this end a believed and/or actual moral good? Presumably, the ultimate collective end of this institution was the material well-being of the relevant *non-slave* population. This was *believed* to be a moral good. Obviously, in so far as it was an end that excluded the slave population it was, at the most, a *limited* moral good. Moreover, this collective end involved the systematic violation of the human rights of the slaves. So whatever the slave owners might have believed, the institution was morally unacceptable by virtue of the immoral means used to realise its collective end. But was there in fact a layered structure of *collective* ends to which the slaves contributed by engaging in joint activity?

There might be a tendency to deny this because the slaves were *coerced* into performing the tasks that they performed. That the slaves were coerced is obviously true, but it is not necessarily irrelevant. An agent who is coerced into performing an action, might nevertheless, still intentionally perform the action, and perform it in order to realise the relevant end. For example, if I threaten to shoot you if you do not shoot Fred, and you intentionally fire your gun having as an end to kill Fred, then you intentionally killed Fred, albeit under mitigating circumstances. Matters would be different if you intentionally fired the gun, but did not have the death of Fred as an end; for example, if you intentionally fired but in doing so nevertheless tried to miss him. Similarly, if a group of slaves individually worked in the cotton field having as an end the harvesting of the cotton in that field, then they were engaged in joint action directed to a collective end. On the other hand, if each worked individually without (say) coordinating their actions in the service of the collective

end, then there is no joint action. It is an empirical question whether or not slaves engaged in joint activity directed to collective ends, and if they did, to what extent they did. The answer to this empirical question is probably in the affirmative, at least in many cases; for in all probability slaves need to engage in a considerable extent of joint activity directed to collective ends, if the institution of slavery is to work.

Naturally, the fact that a group of persons is forced to have collective ends that only benefit their masters – as well as being forced to intentionally perform actions which are deeply inimical to their well-being – is bound to be profoundly damaging to their moral personalities; more damaging than if they performed the actions, but did not pursue the ends. This is in part because the ends that one spends most of one's life pursuing are bound to be in part constitutive of one's moral personality. To the extent that the teleological account of institutions pays due heed to the importance of collective ends – including the collective ends of slaves – it is, if anything, better equipped than other accounts to make sense of the damaging and immoral character of the institution of slavery. At any rate, it now appears that the institution of slavery is not a counter-example to the teleological account of institutions.

Thus far I have spoken in terms of the theory of institutional action where institutions have been taken to be different and separate 'entities'. However, there is also a need for a theoretical account of the *interrelationships* between different institutions. It is clear that on my teleological account of institutions any given institution is to be understood in terms of the collective end or ends to which its activities are and/or ought to be directed. However, there still remains the question of the relationships between institutions. One issue concerns the extent or degree of any required relationship. Another concerns the nature of the required relationship. As far as the extent of the relationships is concerned, in the post-Enlightenment West this interaction between institutional organisations belonging to the same society has typically taken place in the context of a commitment to a basic separation between them. Governments must stand apart from corporations lest public and private interests are confused, and corporations must stand apart from one another in the interests of competition.

As far as the *nature* of the relationship between institutions is concerned, this is presumably to be determined primarily on the basis of the extent to which the differential defining collective ends of institutions are complementary rather than competitive, and/or the extent to which they mesh in the service of higher order ends.

Finally, the substantive functionality – now *collective end(s)* – of an institution has a distinctive feature, which is in itself a (further) defining property of institutions. This feature helps to explain the sense in which institutional action – as opposed to merely conventional or norm governed action – is, or can be creative. The institutional creativity I want to emphasise is not that made so much of by Searle, namely, social construction by means of social acceptance. Rather my concern is with the capacity of institutions to not only constrain, but to create the possibility of, novel, or new or unique individual and social actions. The most obvious example of this is the oft-cited capacity of languages to enable the generation of entirely new sentences.

Such institutional action is often the result of what I will call, a *joint mechanism*.¹⁹ Joint mechanisms are also a defining feature of organisations. Note also that joint mechanisms are in themselves normatively neutral. Joint mechanisms, like organisations, only take on a normative dimension in institutional contexts. Joint mechanisms consist of: (a) a complex of differentiated but interlocking actions, and; (b) the result of the performance of those actions. 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 voting mechanism. Voters vote for candidates. So there is interlocking and differentiated action. 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 is voted in. That there is a result is (in part) constitutive of the mechanism. Moreover, joint mechanisms are a pervasive feature of organisations and institutions. Examples include, voting mechanisms, tossing a coin, courts of law, governments and so on.

Charles Sturt University

NOTES

¹ Searle, 1995.

² See Miller, 1995, for a discussion of the main analyses of joint action. See also Tuomela, 1995.

³ Tuomela operates with a different taxonomy, namely, what he calls s-norms (expectation based) and r-norms (agreement based). See Tuomela, 1995, Chapter 10 op.cit.

⁴ See Tuomela, 1995, op.cit. Chapter 8 for one account of roles.

⁵ See Harre, 1979, pp. 37-43 for an account of structure.

⁶ Here and elsewhere in this paper I do not bother to explicitly distinguish objective from subjectively felt moral considerations. Suffice it to say that social norms by definition involve subjectively felt moral considerations. See Miller, 1997, op.cit.

⁷ For another account of institutions see Harre op.cit. pp. 97-100.

⁸ Though an institution can exist across different social groups.

⁹ I am not suggesting that Giddens notion of an enabling capacity is precisely the same as mine.

¹⁰ Institutions have further properties, including the capacity to reproduce themselves, and to exist in trans-societal forms.

¹¹ Searle ascribes a certain priority to the institution of language – language is somehow constitutive of institutional reality. Searle, 1995, chapter 3. Searle account of institutions also involves, what he calls, background abilities. *ibid.* chapter 6.

¹² For a related 'collective acceptance' view see Raimo Tuomela's "Collective Acceptance, Social Institutions and Social Reality"

¹³ Searle in a later paper (1997) claims that the point of the deontic power is to enable the performance of the function. But the function of money viz. exchange can be performed by shells (without deontic status), and the function of surgeons by non-accredited persons with surgery skills. That is, it is false that deontic powers are necessary for the performance of these functions.

¹⁴ Searle oscillates between claiming that there is a collective imposition of substantive functionality and claiming that there is a collective imposition of *additional, ancillary* deontic powers. I do not dispute the latter.

¹⁵ As Raimo Tuomela points out ("Searle, Collective Intentionality and Social Institutions" p.2) the notion of collective acceptance as used by Searle is multiply ambiguous and in need of conceptual analysis. Tuomela (with W Balzer) offers his own account in Tuomela and Balzer, 1999.

¹⁶ In Searle's terms, "the Y content is imposed on the X element by collective acceptance" (Searle, 1995, p.104).

¹⁷ Raimo Tuomela makes a similar point in his 1997, p.2.

¹⁸ I develop this account in detail in *Social Action: A Teleological Account* (Cambridge University Press, forthcoming)

¹⁹ The following discussion is a version of what appeared in the last section of Miller, 1992a.

REFERENCES

- Giddens, A.: 1984, *The Constitution of Society*, Polity Press, Cambridge.
- Harre, R.: 1979, *Social Being*, Blackwell, Oxford.
- Lewis, D.: 1969, *Convention*, Harvard University Press, Cambridge, Ma.
- Miller, S.: 1992a, "Joint Action", *Philosophical Papers* **xxi** no.3.
- Miller, S.: 1992b, "On Conventions", *Australasian Journal of Philosophy* **70** no.4.
- Miller, S.: 1995, "Intentions, Ends and Joint Action", *Philosophical Papers* **xxiv** no.1.
- Miller, S.: 1997a, "Social Norms", in R. Tuomela and J. Hintikka (eds.), *Social Action* (vol.2), Kluwer, Dordrecht.
- Miller, S.: 1997b, "Individualism, Collective Responsibility and Corporate Crime", *Business and Professional Ethics Journal* **16** no.4.
- Miller, S.: 2001, *Social Action: A Teleological Account*, Cambridge University Press, Cambridge.
- Olson, M.: 1965, *Logic of Collective Action*, Harvard University Press, Harvard.
- Searle, J. R.: 1995, *The Construction of Social Reality*, Penguin, London.
- Searle, J. R.: 1997, "Reply to Raimo Tuomela", *Philosophy and Phenomenological Research* **lvii**, 451-454.
- Tuomela, R.: 1995, *The Importance of US*, Stanford University Press, Stanford.
- Tuomela, R.: 1997, "Searle on Social Institutions", *Philosophy and Phenomenological Research* **lvii**.
- Tuomela, R. and W. Balzer: 1999, "Collective Acceptance and Collective Social Notions", *Synthese* **117**, 175-205.
- Tuomela, R. (forthcoming): "Collective Acceptance, Social Institutions and Social Reality", *Journal of Economics and Sociology*.
- Tuomela, R. (unpublished manuscript): "Searle, Collective Intentionality and Social Institutions."