Volume 7, Number 1
2013

ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

http://abstracta.oa.hhu.de

d|u|p

abstracta
Linguagem, Mente & Ação

ISSN 1807-9792

Volume 7, Number 1
2013

Contents

abstracta
Linguagem, Mente & Ação

Editorial

In the last few years, the philosophical community has witnessed an interest in highlighting the role of social factors in cognition and its development. This gave rise to new interdisciplinary research projects not only in philosophy, but also in social and psychological sciences. The introduction of theoretical notions such as "shared worldviews" or "shared intentionality" blurred the classical demarcation lines between social and psychological phenomena, and urged a more permissive interpretation of the relation between their respective ontologies. This special issue of Abstracta aims to contribute to this emerging field of research with contributions that approach the relation between psychological and social phenomena from a variety of perspectives.

Lo Presti focuses on the complex relation between social ontology and situated cognition. In particular, Lo Presti argues that there is a dependency relationship between the two, both at a methodological and a phenomenological level. Research on social ontology, he argues, depends on research on social cognition. At the same time, social phenomena influence social cognitive processes and interaction, which in turn influence social phenomena.

Lauer discusses the role of shared worldviews and social identities in the explanation of intentional collective actions. On the basis of actual examples of sectarian conflicts and ethnic violence, she argues that the fuzziness of shared worldviews and social identities prevents ascription of such representations to single individuals. As a result, the behavior of individual agents should be explained on the basis of their individual mental representations, rather than by reference to shared worldviews and social identities.

Finally, Townsend, who examines the dynamic nature of both inter-and intra-group relations, argues that the personhood of groups is not only dependent on the efforts of group members, but also on the attitudes of members of the wider discursive community, within which a given group is situated and operates.

Contributions in the second part of this issue are regular original research papers and cover a large array of topics ranging from ethics to metaphysics. Harbecke offers a compatibilist account of mental causation based on Yablo's seminal distinction between determinate and determinable properties. Harbecke argues that explaining the behavior of single agents on the basis of their specific mental properties rather than by reference to collective representations requires the mental properties in question to be causally efficient. The core of his "new compatibilism" is that mental properties are patterns, which stand in determinate-to-determinable relation to physical properties. Qua being both non-distinct and non-identical, mental properties do not compete in the production of behavioral effects. A way out of the well-known causal exclusion problem is thus suggested.

Spielthenner focuses on normative practical reasoning and examines whether this type of reasoning can be logically conclusive. More specifically, Spielthenner argues that practical arguments are non-trivially ambiguous since they can, at a given time, express different pieces of practical reasoning, each of which has a different logical status.

Edward criticizes Raz's argument that it is impossible for there to be a genuine amoralist and that there is consequently no philosophical puzzle of the amoralist. He offers three possible interpretations of Raz's argument and argues that none of them is acceptable, casting, in this way, doubts on Raz's initial argument.

Finally, Horn focuses on the allegiance between disjunctivism and naïve realism and argues that linguistic arguments against private or internal meanings do not imply perceptual directness. On these grounds, it is argued that the espousal of direct realism – naïve or not – does not require adherence to disjunctivism.

The Editors.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 5–17, 2013

Social Ontology and Social Cognition*

Patrizio Lo Presti

Department of Philosophy and Cognitive Science, Lund University, Sweden
Patrizio.Lo_Presti@fil.lu.se

Abstract
The aim of this paper is to show that there is a reciprocal dependency relationship between social cognition and social ontology. It is argued that, on the one hand, the existence conditions of socially meaningful objects and of social groups are about sucjets' social cognitive processes and interactive patterns and, on the other hand, social cognitive processes and interactive patterns are modulated by socially meaningful objects and social groups. I proceed from a historically informed distinction between social ontologies – between what might be called constructivist and emergentist theories of social reality. I then distinguish three theories of social cognition, theory-theory, simulation theory, and interaction theory, and argue that the first distinction and the latter map onto each other. Finally I argue that the reciprocal dependency between social ontology and social cognition can be justifiably though of as causal in Di Paolo et. al.'s (2010) sense of "downward" or "circular" causation. It is concluded that the dependency between social ontology and social cognition pertain to both a methodological and a phenomenal level. First, research on social ontology depends on research on social cognition; and, secondly, social phenomena, involving socially meaningful objects and groups, influence social cognitive processes and interaction, which in turn influence social phenomena.

1 Introduction: The construction and emergence of the social

How can the contingent empirical fact that we live in a world of nations, cultures, religions, families, and other forms of social relationships that seemingly have causal efficacy on each individual's life, be accommodated with the reductionist realist paradigm prevalent today? As John Searle (2006, p. 13) put it, how, in a world constituted by particles in fields of force, can it be that some carbon based organism after 5 billion years of evolution have created a world of money, property, and government?

These questions form the core of the subject matter of social ontology, a discipline that since Searle's *The Construction of Social Reality* (1995) has surged analytical philosophy and given rise to lively debates. Social ontologists are concerned with the existence conditions of *social phenomena*. Social phenomena are phenomena involving subjects and social relations which 'give rise to' families, groups, organizations, nations, and so on, or units of agency with concomitant roles, rules, norms, and functions.

This paper focuses on the 'give rise to'-relation between subjects and social phenomena. To that end, as an introduction, it is informative to put the 'give rise to'-relation in historical perspective.

*This research was carried out as part of the NormCon project, funded by the European Science Foundation's EUROCORES program EuroUnderstanding, *Understanding and Misunderstanding: Cognition, Communication and Culture*.

Social ontology peaks a longstanding research program that is by and large neglected in todays analytical theorizing. Most analytical philosophers conceive of social phenomena as products of a certain kind. They usually proceed by analyzing the 'give rise to'-relation in terms of the mental states or speech acts of individuals. Social phenomena, in this tradition, are products of the agency of a multiplicity of individuals because, roughly, individuals either knowingly instantiate type-identical mental states or because individuals together declare that social relations or objects have a certain meaning. Thus social phenomena are *constructed*, either out of mental components – beliefs, intentions, desires – the contents of which are shared by individuals who believe that they so share (Tuomela 2003, 2007; Bratman 2009; Gilbert 2006) – or out of speech act components – declaratives, performatives – the utterance of which create social phenomena – families, organizations, nations, money, and so on – if people accept the declarations (Searle 1995, 2006, 2010). Social phenomena, involving socially meaningful objects and relationships, are in the contemporary eye, then, socially constructed. However, this focus on social constructivism in social ontology is itself only the tip of a historical iceberg.

By the turn of the last century there was another conception of the social reality. Emile Durkheim (1895/1972, p. 69) wrote, "Whenever any elements combine and, by the fact of their combination produce new phenomena, it is evident that these phenomena are not given in the elements" According to Durkheim, social and collective phenomena, society at large, are emergent phenomena. Society is a *product of activities* of people but it is no more reducible to individual mental states than life is reducible to mitosis. Georg Simmel had similar ideas about social phenomena when (1910/1971, p. 134) he wrote that people *play society*. That is, according to Simmel as I understand him, social phenomena emerge at the junction of social interactions, and society at large emerges from the social interactions of people who 'play' different roles. I will call this conception of social reality 'emergentist' on the basis that its proponents conceive of social reality as emergent from activity, rather than necessarily constructed intentionally out of speech acts or believed sharing of mental states in the constructivist sense of the previous paragraph.

The above short exercise in the history of social ontology serves to distinguish two alternative accounts of social reality: the *constructivist* and the *emergentist* accounts. Constructivism is the view that people together, through believed sharing of mental states – intentions, goals, commitments, and so on – or through declarative or performative speech acts, *create* social phenomena; social reality is relative to the mental states of individuals aimed at the construction of social relations and objects. This is the view found in much of today's analytically informed social ontology. We find this view in Tuomela's (2007) approach according to which, the existence conditions of social groups, for instance, involve that subjects believe about each other that they believe they are forming a social group. Thus people whose mental states are appropriately related are in position to *create* a social group, or a unit of activity, intentionality, and in general a 'we' of cognition and action. We find constructivism in Gilbert's (1989) semantics of the first person plural pronoun 'we', according to which the referent of 'we' is subjects who have expressed willingness to form a 'we' under conditions of common knowledge. According to Gilbert it is through expressions of appropriate kind under appropriate conditions that collectives, what she calls 'plural subjects', are created. Consider also Searle's (1995, 2010) view that something is a social object only if it is declared to have a function beyond its physical features. On Searle's account, the social functions of objects and persons are relative to the intentionality and recognition of individuals that objects have the functions in question; people must recognize, for instance, that "we accept that these pieces of paper count as money and give their owner the right to buy stuff", and the pieces of paper must be declared to have

that status, in order for the pieces of paper to count as money.[1] In contrast, on the emergentist view, social phenomena are acted out and emerge from social interactions, and are irreducible to mental state or speech act components. From this perspective there are no necessary conditions about acceptance of speech act contents, collective mental states, or believed sharing of mental states, for the emergence of social objects or phenomena. Rather, social reality is the product of patterns of interaction in the sense of being irreducible to individuals' activities or cognitive states and processes; social reality arises from interaction.

So far I have spoken loosely about 'social objects', 'social phenomena', and 'social units'. To clarify, I take social *objects* to be the set of particulars whose existence depends in the constructivist or emergentist sense on social interaction and cognition. Thus rabbit pelt, to use an example of Tuomela's, is not a social object, but when used under the right conditions, e. g., as medium of exchange in social interaction, then it is a social object. If patterns of social interaction recur in which a rabbit pelt is treated as an exchange medium without there being any point at which a rabbit pelt is declared to qualify as a medium of exchange or if there is no believed sharing of acceptance that a rabbit pelt is a medium of exchange, then we can say that the social object, the rabbit pelt as medium of exchange, is an *emergent* social object. In contrast, if declarations or believed sharing of mental states about rabbit pelts is what causes rabbit pelts to count as money, then we can say that the social object, the rabbit pelt as an medium of exchange, is *created* in the constructivist sense. Now, a social *unit* I take to involve similar genealogical processes as social objects, but 'units' or 'unities' relate to social relations rather than to objects. Thus, a family is a social unit, as is a subculture, and in general every instance of a social relation where subjects involved act or cognize as a 'we'. That the genealogy of social units is similar to that of social objects means that just like a rabbit pelt can become a medium of exchange through repeated patterns of social interaction or through declaration or sharing of mental states of acceptance or belief, so a family or a gang can be created through matrimony or vows of allegiance, or emergent in recurring interactions where people, as Simmel would have it, play their respective roles. Lastly, social *phenomena* I take to be phenomena involving social units and objects. Thus it is a social phenomenon that the euro is the medium of exchange in some European countries and that women usually do more household work than men, where money and households are social objects and Europe, and women and men as social groups, are social units. That women's salary is generally lower than men's is a social phenomenon, which if true, is a social fact.

In summary, there are two conceptions of the 'give rise to'-relation between subjects and social phenomena. On the one hand, one can conceive of social phenomena as grounded in declarations or agreement among individuals that certain objects, persons, and relationships are to have certain social statuses, or one can conceive of them as emergent from recurrent interactive patterns not necessarily involving such declarations and agreement.

In the next section, we will see that it is central for any attempt to understand the 'give rise to'-relation between subjects and social phenomena, that is, to understand theories in social ontology, to also understand what the theories presuppose with regard to underlying cognitive states and processes of subjects involved in social phenomena. To provide that understanding I now turn my focus on social cognition.

[1]Constructivism is a widely held approach in social ontology research, and recounting most or even many of its proponents requires too much space. But see, for instance, Gilbert (1990, 2000, 2009), Bratman (1992, 1993), Searle (2006, 2007), Schweikard and Pettit (2006), List and Pettit (2011), Tuomela (2002, 2003).

2 Cognitivism and non-cognitivism

Research on social cognition is research on how people understand others and their social surroundings. More specific, research in social cognition is concerned with cognitive processes that enable subjects to make sense of interaction in social arenas – such as churches, banks – and to act in accordance with how one is meant to act in social arenas. In this section, I will examine three theories of social cognition that can be divided into cognitivist and non-cognitivist theories. My aim here is not adjudication between, but clarification of, these theories with the aim to show how they map onto aforementioned distinctions between constructivism and emergentism in social ontology.

Since the effectuation of false-belief tests (Perner and Wimmer 1983), which are taken to show that children understand that mental states of others can deviate from their own once they have acquired folk-psychological concepts such as 'belief' and 'desire', what has been called theory-theory (TT) has drawn many adherents (Baron-Cohen 1995). According to TT, intersubjective understanding is backed up by mental state-attribution justified by recognition that in certain arenas or in certain interactions people usually have beliefs, intentions, desires, and so on, with a certain content. For instance, when someone reaches for the cookie jar he or she usually has a desire for cookies, a belief that there are cookies in the jar, and an intention to take a cookie from the jar. According to TT, understanding the other person, arriving at the meaning of his or her movement, essentially involves knowing something about cookies and cookie jars and from these premises inferring the other person's intention in terms of his or her beliefs and desires. Inferring intentions in terms of beliefs and desires presupposes having folk-psychological concepts signifying mental states. Since the process is described in terms of inferences, this theory of social cognition is that subjects form a theory about the other person's mental states. Baron Cohen (1995, pp. 3–4) writes, "it is hard for us to make sense of behavior in any other way than via the mentalistic … framework … [A]ttribution of mental states … is our natural way of understanding the social environment" (cf. Toby and Cosmides 1995).

In contrast to TT, and in the wake of neuroscientific research on brain areas functioning as so called 'mirroring' or 'resonance' systems (Gallese and Goldman 1998; Gallese 2005, 2007), a theory according to which social cognitive processes are simulative processes has been suggested. According to simulation theory (ST), people understand each other and their social world by means of running a simulation 'as if' oneself were the other or were in a similar social situation in which an observed other is situated. According to ST, social cognition is not underlain by subjects' mounting of interpretative or inferential processes with the other's behavior or environmental cues as premises, yielding as conclusion what the other means with his or her action or what socially significant environmental cues signify. Rather, "the state ascribed to the target is ascribed as a result of the attributor's instantiating, undergoing, or experiencing, that very state." (Goldman and Sripada 2005, p. 208). Thus, and although in ST there is the notion of agents ascribing mental states and meanings to others, the form of this ascription processes is subjunctive, 'as if', rather than in an inferential theory-like form (Goldman 2005b; Gallese 2005).

Criticism has been mounted against both TT and ST for their commitment to what has been called the *mentalistic assumption*. The mentalistic assumption is that, "mentalizing, or mindreading, underlies basically all social understanding and interaction" (Michael 2011, p. 561). The term 'mindreading' refers to the ascription of mental states involved in social understanding according to both TT and ST. The worries are that, first, if TT or ST were correct, then one should find in phenomenology a corresponding sensation of the theorizing

or simulative processes. But we seem not to be undergoing such phenomenological states when we understand others or our social world (Gallagher and Zahavi 2008, p. 176). And, secondly, even if one assumes that folk-psychological theories or simulative processes are implicit, it still seems that our conceptualizations of the processes underlying social cognition are misleading. Strictly speaking, "there is no neuronal subjunctive" (Gallagher 2007, p. 361). That is, if sub-personal processes of simulation or theorizing involving pretense, 'as if' states, instrumental for mindreading are to function as the explananda of social cognition, then we must conceive of those sub-personal processes as *pretending* and *using* information about the other or the social environment to form a model. But 'pretense' and 'use' are personal-level concepts. Therefore, TT and ST cannot be understood as true descriptions of social cognitive processes, neither at a personal explicit or sub-personal implicit level of description.

The alternative account of social cognition that critics propose is called interaction theory (IT). According to IT, mentalizing or mindreading, i.e., ascriptions or even simulations of mental states do not necessarily underlie social cognition. Rather than the third-person observational, theorizing, or simulative stances TT and ST ascribe to social cognizers, social cognition emerges according to IT in second-person interactions (Gallagher 2008b, pp. 164–5). It is interactive processes themselves, with others and social surroundings, that constitute social understanding and, furthermore, give rise to social meaning (p. 167; cf. De Jaegher et. al. 2010). It is not necessary that interactive processes be supplemented by inferences of simulations.

It should be clear why research on social cognition is important for research in social ontology. Since social ontology produces analyses of social facts and properties – analyses of the existence conditions of such things as money, nations, religion, and families – and since the analyses that are on the table analyze such facts and properties in terms of speech acts, sharing of mental states, or interaction, it is obvious that to understand how social facts and properties can exist it is necessary to understand how subjects can understand each others' mental states, speech act, and actions. Thus social cognitive processing characterizes at least one aspect of the 'give rise to'-relation between subjects and social phenomena.

Since my aim is to show that there is a reciprocal dependency relationship between social ontology and social cognition, both methodologically and phenomenally, I will now try to elucidate, in light of preceding two sections, how questions asked in the two domains map onto each other. In the following sections, I will argue that not only are questions in the two domains linked, but also social phenomena and social cognitive processes themselves exert influence on each other.

Without trying to settle the issue between theory-theory, simulation-theory, and interaction-theory, I suggest that one can distinguish two main approaches to social cognition: *cognitivism* and *non-cognitivism*. Cognitivist approaches to social cognition are characterized by the mentalistic assumption, that is, by their commitment to the claim that social understanding necessarily involves ascription of mental states to others. Non-cognitivist approaches to social cognition are characterized, negatively, by rejection of the mentalistic assumption and, positively, by the claim that perceptual or interactive processes are sufficient for social understanding. Here perceptual and interactive processes are to be understood as *inherently* sense-making. By inherently sense making I mean that it is not necessary that the processes be supplemented by cognitive processes such as inferences or simulations in order for social understanding to be enabled.

From this distinction we can draw two clarifying conclusions regarding commitments of theories in social ontology. First, if communication necessarily involves understanding the meaning, intentions, and communicative intentions of speakers, and if sharing of mental states

necessarily involves mental state-ascriptions as a result of simulation or inference, then constructivist accounts of social reality presupposes a cognitivist account of social cognition. Constructivist accounts of social reality presuppose a cognitivist account of social cognition since Searle (1995, 2006, 2010), whose social ontology is presently one of the most influential, bases his theory on declarative speech acts. And Searle clearly states (1983, p. 166) that, "what one communicates is the content of one's representations", implying that to understand the speech acts with which social objects and units are constructed one must understand and ascribe mental states to speakers. Furthermore, Tuomela, whose social ontology is one of the more prominent amongst those based on believed sharing of mental states (2003, 2007), explicitly states (2007, p. 188) that only collective acceptance that an object has a social meaning and is meant to be used in a certain way can account for the object having that meaning. Second, if it is sufficient for subjects to engage in social interaction that they have intentions the contents of which refer to others, but does not necessarily entail ascription, from theorizing or simulation, of mental states to others, then emergentist accounts of social reality presuppose a non-cognitivist account of social cognition. Emergentist accounts of social reality presuppose a non-cognitivist account of social cognition since Durkheim and Simmel, who I take to be the pioneers of emergentist social ontology (cf. Tollefsen 2002; Gilbert 1989; Greenwood 2003), denied what is now called the mentalistic assumption and claimed that it is people's interactions that constitute social entities. For instance, Simmel (1908/1971, p. 8) wrote, "consciousness of the abstract principle that he is forming society is not present in the individual", suggesting that there need be no (ascription of) beliefs or communicative intentions involved in the emergence of social phenomena.

It is fair to say that the unearthing in this section of the presuppositions of theories in social ontology of theories in social cognition suggests a straightforward mapping of questions asked in the two fields. That is, a constructivist social ontology presupposes that a cognitivist approach to social cognition is supported, whereas an emergentist approach to social ontology does not. The emergentist approach to social ontology is supported by non-cognitivist theories of social cognition. Therefore, constructivist and emergentist social ontology hinge on the plausibility of cognitivist and non-cognitivist theories of social cognition (although a precise forecast for respective social ontologies' ability to handle falsification of theories of social cognition on which they depend cannot at this point be given).

3 Downward causation

The mapping of the socio-cognitive onto the socio-ontological does not simply entail that research on social ontology is aided by research on social cognition. It also entails the reverse relation, that research on social cognition is facilitated by research on social ontology. Furthermore, I will argue in this section that social cognition substantively, not as a research object alone, is facilitated by cognizers being 'situated' in a social reality in a sense to be clarified. It is desirable to first of all investigate the nature of the relation holding between social objects and units and social cognition.

Ezequiel A. Di Paolo and colleagues (2010) use the notion of 'circular' or 'downward' causation: a causal relation holding between emergent entities and low-level processes that give rise to those entities. An emergent entity is described as one "whose characteristics are enabled but not fully determined by the properties of the component processes" (p. 40). This emergent entity in turn "introduces ... modulations to the boundary conditions of the lower-level processes that give rise to it" (p. 41). Remember that social phenomena emerge from social cognitive or interactive processes (speech acts, shared mental states, or socially directed

action, depending on what ontology of social reality is preferred). Now, if there is a downward or circular causal relationship between social cognitive processes and social phenomena, that would mean that social cognitive and interactive processes give rise to social phenomena which in turn influence the social cognitive and interactive processes. That is, if social objects and units causally influence people's understanding of each other and their social environments, then that might be understood as social phenomena having downward causal efficacy with regard to the processes of social understanding (or misunderstanding) – the processes from which social phenomena emerge. I will soon illustrate the possibility of the downward, circular relation with two examples. Empirical findings will also be adhered to. But first, let me emphasize that of primary interest for present purposes is vindicating that a downward, circular influence between, on the one hand, social cognitive states and processes and social interactions, and, on the other, social objects and units involved in social phenomena, does obtain. The nature of this relationship is of secondary interest. I will henceforth leave undecided whether this relation is causal in nature and focus instead on the plausibility of the obtaining of the relation.

To illustrate the possibility of downward or circular influence between emergent social phenomena and social cognitive and interactive processes, consider the following example. You're in church with lots of other people filling the rows. The organ is playing and along the aisle two persons are walking solemnly. They stop when they reach the altar and repeat sentences pronounced by the priest. This situation makes sense if you recognize that you are at a wedding. But also, you recognize, or understand, that you are at a wedding by focusing on the social objects and other agents' behavior in this situation. For instance, the altar has a certain meaning, e. g., it is treated as a place where wedding ceremonies, baptisms, and so on, take place. Other objects in the situation have other socially relevant meanings, e. g., the arrangement of benches, peoples' clothing, and so on. Importantly, the set of social objects, units and people in the situation seems to play a central explanatory role in accounting for your, as cognizer, grasping of the social meaning of the situation. But the social meaning of the objects, units, and people is, reflexively, emergent from social cognitive and interactive processes. That is, on the one hand, social phenomena, weddings for instance, emerge from social sense-making and interaction, while, on the other hand, the social phenomen also determine social sense-making and interaction. Social cognition and interaction partly determines the constitution of social reality and the constitution of social reality partly determine social cognitive state and process and social interaction.

Consider an altered version of the wedding example. Suppose that you're giving a lecture at a conference. As you're explaining one of your slides, two persons solemnly dressed as if on a wedding stride towards you between the conference attenders. Something in this situation is terribly wrong, and the obvious reason is that there has been a *misunderstanding*. Why does the appearance of bride and groom *not* make sense? The social environment does match the social interaction; conference halls are standardly not *meant* to house matrimonies, conferences have emerged as gatherings for exchanges of ideas, not for weddings.

The import of these examples is this: objects and persons are *meant* to function, to be used, and to act in certain ways. The existence condition for these functions and roles – these social meanings, the very structure of social environments – is, we have seen, the occurrence of appropriate social cognitive states, processes and interactions of the people involved – their communication, beliefs, or repeated interactions depending on preferred social ontology. What the examples show is that the emergent social objects, roles, functions, and units enable social understanding (or misunderstanding). Understanding and misunderstanding of others and socially relevant objects and events are social cognitive states. Therefore, we can conclude that

social reality, emerging from social cognitive processes and interactions, influence social cognitive processes and interactions and produce states of social understanding (or misunderstanding). Thus social cognitive states and processes and social phenomena circularly influence and perpetuate each other; they form a social circuit, a circular system of reflexive determination.

In the final section before concluding, I will reconnect conclusions drawn about the circular relationship between social cognitive and interactive processes and social phenomena to orthodox contemporary social ontology. The aim will be to find support for my argumentation in some prevalent theories of social ontology. But first, lets summarize our central findings. Two conclusions can be drawn at this point.

First, social cognitive and interactive processes involving objects and persons with social meanings, emergent from social cognitive and interactive processes, can produce social understanding and misunderstanding; i.e., influence social cognition and interaction. Without going into too much detail, this conclusion presupposes that social cognitive processes have access to perceptual, proprioceptive, affective, and other subsystems. This is because identification of social objects and social statuses of others requires access to cues indicating such social meaning – e. g., wedding rings, police badges, uniforms. Also, interaction in accordance with how one is socially meant to interact requires proprioceptive afferent and efferent signals in execution and evaluation of appropriate action. Similarly, affective states, e. g., disapprobation and approval, likely play a role in and are indicative of social understanding and misunderstanding. This does not mean that whenever social understanding or misunderstanding occurs there is some perception of social objects, or proprioceptive or affective state to which explanations of the former necessarily refer. It means that the occurrence or non-occurrence of the latter influence the production of the former. In the next section, I will exemplify how perception can be recruited in social cognition to achieve social understanding or misunderstanding.

Second, the examples considered can be multiplied, and what they show be generalized. Thus take any situation involving social objects or persons with social meaning and shift between introducing and removing them. The prediction is that subjects in the situations imagined will be further removed from or closer to being able to make sense of others and their social surroundings. Since virtually all agency and cognition is agency and cognition situated in a socially meaningful world, questions posed and answered with regard to social agency and cognition and social ontology seem inexorably linked. So there is not only a *methodological* advantage for research in social ontology to be sensitive to research in social cognition, and vice versa, it is also predicted that, *phenomenally*, social cognition and action is sensitive to the ontology of the social environments in which they occur.

I hope to have clarified a sense in which social reality, on the one hand, and social cognition and agency, on the other, partly codetermine each other in a circular manner.

4 Social cognition and social interaction

I said above that the conclusion that social reality influences social cognitive and interactive processes presupposes perceptual access to socially meaningful objects, events, and persons. In this final section, I want to pursue the implications of this presupposition. It will appear that there is support in contemporary orthodox social ontology of my claim that perception of social objects influences social understanding and social interaction.

Searle (1995, p. 85) writes, “we have status indicators in the form of marriage certificates, wedding rings, and title deeds” which serve ‘epistemic functions’ (p. 120). For Searle, social reality is an epistemically objective, even if ontologically subjective, reality (pp. 8–12). This means that whereas some entities, for instance stones and trees, are ontologically objective in

the sense that their existence is independent of what anyone thinks about such entities, other entities, for instance dollar bills and marriage, depend for their existence on people's assigning and recognizing a social meaning of pieces of paper and social relationships. The 'epistemic functions' of some objects, for instance wedding rings and title deeds, can be thought of as *indicating* what role or function an object or person carries. This is an interesting line of thought connecting to my argument that socially meaningful objects and persons have downward influence on cognition and agency. Because, if, as Searle claims, what meanings persons and objects are bestowed with are indicated, then perceptual access to such indicators certainly implies access to information about social meaning. To conclude that social entities have an influence on social understanding is close to home.

Consider the example of being Secretary General of the U.N. For Searle, having this status *means* that the person has a range of actions open for him or her – a set of 'powers' (p. 106). Interestingly, if social meaning – or social statuses and functions, in Searle's terminology – is indicated by objects so that subjects are epistemically justified in identifying social meaning when perceiving the indicating objects, and if social meanings entail an appropriate way of interacting, then subjects with perceptual access to indicators are in position to make sense of others and their social surroundings. Searle seems never to have seen or have been interested in this implication.

But is it justified to claim that social objects that exert influence, through perception, on social cognitive and interactive processes? Outside of research in social ontology, empirical support for that claim can be found. Shaun Gallagher (2008a) argues that perception of socially meaningful objects is 'smart'. By smart Gallagher means that perception need not be supplemented by other cognitive processes for an observer to make sense of perceptual input (pp. 539–40). Perception of socially relevant objects is 'direct', according to Gallagher, in the sense that there need not be inferential steps or simulative processes premised on perceptual input; perceptual input is in itself sufficiently informative for recognition of something as a car, rather than recognition of something as a car being an inference from perceptual input of metallic mass in a certain shape.

Gallagher have developed the direct perception account, introduced by J. J. Gibson, in the last decade (Gallagher 2001, 2004, 2007, 2008a, 2008b). Although answering how precisely perception can be direct is not a *sine qua non* for my argument to go through, I will give a review of experiments carried out by Marcel (1992), reported by Gallagher and Marcel (1999), about how focus on, and agency in, socially significant situations enhance cognition and agency. Reviewing the experiments is only a way of showing that socially significant objects and events can influence social cognition through perception, and thus that the downward or circular relation between social reality and social cognition and interaction is empirically supported.

Marcel (1992) distinguished three levels of intention formation: intentions in *abstract decontextualized*, in *pragmatically contextualized*, and in *socially contextualized* agency. Abstract decontextualized agency is "detached from what would ordinarily be considered a significant context" (Gallagher and Marcel 1999, p. 9), for instance handling a cylinder shaped object in an experimental setting. Pragmatically contextualized agency is "performed in the course of a natural activity whose purpose arises from personal projects and concerns" (*ibid.*), for instance dishing a teacup. Socially contextualized agency "has a meaning defined by cultural categorizations … and represent states of the self in regard to others" (*ibid.*), for instance serving friends cups of tea at a tea party. What Marcel found was that patients suffering from ideomotor apraxia, that is, persons with difficulty in executing intentions in body movement, had near normal abilities in socially contextualized agency, whereas they had great difficulties

in abstractly decontextualized settings. This led Gallagher and Marcel to conclude that when subjects' intentions are guided by focus on socially significant objects and events involving their social relations to other people their cognitive and agentive performance is enhanced (p. 12; cf. Leontiev and Zaporozhet 1960). Hence there is experimental data in support of the claim that socially meaningful objects and persons influence social cognition and interaction. Since most, if not all, theories in social ontology agree that social objects and meaning emerge from or are constructed in social cognitive and interactive processes, we can conclude that there is empirical support for the claim that there is downward or circular influence between socially meaningful objects and persons, on the one hand, and the social cognitive and interactive processes from which social meaning emerge, on the other.

Reconsidering Searle's notion of social statuses being indicated, it seems we have found in experiments on social cognition a basis for the conclusion of my argument. That is, the social ontology of situations in which people act and cognize influence social cognition and interaction, while social cognition and interaction influence the social ontology of situations in which people act and cognize.

5 Conclusions

Social ontology is about the existence conditions of social phenomena; phenomena involving two or more subjects, their relations and interactions, and often socially meaningful objects involved in interaction. Social cognition is about the sense-making processes, interactive, perceptual, simulative, or theory-like, that enable subjects to understand social phenomena.

In the first sections of this paper, we have seen that social reality is given rise to by social cognitive and interactive processes. The 'give rise to'-relation from such processes to social phenomena can be characterized in several ways. Social objects or units involved in social phenomena can be *created* through speech acts or believed sharing of mental states accepted by a group as the group's goals, beliefs, and so on. Social phenomena can also *emerge* from repeated social interactions in absence of any declarative or performative speech-acts, or believed shared acceptances of goals, beliefs, and so on. The former, constructivist, sense implies some ascription of mental states among people involved in the creation, what in social cognition is called 'mindreading', whereas the latter implies a history of recurrent social interactions not necessarily involving mindreading. The upshot of these implications is methodological: they suggest that research in the domains of social ontology and social cognition is inexorably linked – theoretical presuppositions in one domain are depends on results in the other. A prognosis and desideratum of the state of debate in social ontology and social cognition respectively is, therefore, that only an account that consistently and coherently integrates creation and understanding of social meaning of objects and persons will and should lead the way for future research.

In the latter sections, it has become clear that beyond the desideratum that researchers on social ontology and cognition coordinate efforts, and beyond the prediction that such coordination is fruitful for future research, there is a real reciprocal dependency between social cognitive and interactive processes, on the one hand, and socially meaningful objects and persons, on the other. Thus, a second prognosis and desideratum provided by this paper is that only accounts of social ontology and cognition providing explanation and prediction of social phenomena's and social cognitive processes' reflexive influence on each other will and should lead the way for future research.

In conclusion, whether we analyze the 'give rise to'-relation from subjects to social phenomena in constructivist or emergentist terms, the proposition that I have argued in favor of

suggests itself: social reality is partly determined by underlying social cognitive and interactive processes, and social cognitive and interactive processes are in turn partly determined by the structure of social reality. The result is that the constitution of social reality and the progression of social understanding and interaction can be understood as co-dependent and co-determining.

Acknowledgements: I would like to thank two anonymous referees at Abstracta for crucial insights and criticisms.

References

Baron-Cohen, S. (1995) *Mindblindness: An essay on autism and theory of mind*, Cambridge: MIT Press.

Bratman, M. (1992) 'Shared Cooperative Activity', *The Philosophical Review* 101, 327–341.

Bratman, M. (1993) 'Shared Intentions', *Ethics* 104, 97–113.

Bratman, M. (1999) *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge: Cambridge University Press.

Bratman, M. (2009) 'Modest sociality and the distinctiveness of intention', *Philosophical Studies* 144: 149–165.

Butterfill, S. (in press) 'Interacting mindreaders', *Philosophical Studies*.

Davidson, D. (2001) 'Agency', In D. Davidson (ed.) *Essays on Actions and Events*, Oxford: Oxford University Press.

De Jaegher, H. (2009) 'Social understanding through direct perception? Yes, by interacting', *Consciousness and Cognition* 18, 535–542.

De Jaegher, H., Di Paolo, E.A. & Gallagher, S. (2010) 'Can social interaction constitute social cognition?', *Trends in Cognitive Science* 14, 441–447.

Di Paolo, E.A., Rohde, M. & De Jaegher, H. (2010) 'Horizons for the Enactive Mind: Values, Social Interaction, and Play', In J. Stewart, O. Gapenne and E.A. Di Paolo (eds.) *Enaction: Toward a New Paradigm for Cognitive Science*, Cambridge: MIT Press.

Durkheim, E. (1950) 'Les règles de la methode sociologique', In A. Giddens (ed.) *Emile Durkheim: Selected Writings* (1972), New York: Cambridge University Press.

Gallagher, S. (2001) 'The practice of mind: theory, simulation or primary interaction?', *Journal of Consciousness Studies* 8, 83–108.

Gallagher, S. (2004) 'Situational Understanding: A Gurwitschian Critique of Theory of Mind', In L. Embree (ed.) *Gurwitsch's Relevancy for Cognitive Science*, Dodrecht: Springer.

Gallagher, S. (2007) 'Simulation trouble', *Social Neuroscience* 2, 353–365.

Gallagher, S. (2008a) 'Direct perception in the intersubjective context', *Consciousness and Cognition* 17, 535–543.

Gallagher, S. (2008b) 'Inference or interaction: social cognition without precursors', *Philosophical Explorations* 11, 163–174.

Gallagher, S. & Marcel, A.J. (1999) 'The Self in Contextualized Action', *Journal of Consciousness Studies* 6, 4–30.

Gallagher, S. & Zahavi, D. (2008) *The phenomenological mind*, London: Routledge.

Gallese, V. (2005) "Being like me': Self-other identity, mirror neurons and empathy', In S. Hurley and N. Chater (eds.), *Perspectives on imitation* vol. 1, Cambridge: MIT Press.

Gallese, V. (2007) 'Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition', *Philosophical Transactions of the Royal Society B* 362, 659–669.

Gallese, V. & Goldman, A.I. (1998) 'Mirror neurons and the simulation theory of mind-reading', *Trends in Cognitive Science* 2, 493–501.

Greenwood, J.D. (2003) 'Social Facts, Social Groups and Social Explanation', *Nôus* 37, 93–112.

Goldman, A.I. (2005) 'Imitation, Mind Reading, and Simulation', In S. Hurley and N. Chater (eds.), *Perspectives on imitation* vol. 2, Cambridge: MIT Press.

Goldman, A.I. and Sripada, C.S. (2005) 'Simulationist models of face-based emotion recognition', *Cognition* 94, 193–213.

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

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

Gilbert, M. (1990) 'Walking Together: A Paradigmatic Social Phenomenon', *Midwest Studies in Philosophy* 15: 1–14.

Gilbert, M. (2006) 'Rationality in Collective Action', *Philosophy of the Social Sciences* 36: 3–17.

Gilbert, M. (2009) 'Shared intention and personal intentions', *Philosophical Studies* 144: 167–187.

Lavelle, J.S. (2012) 'Theory-Theory and the Direct Perception of Mental States', *Review of Philosophy and Psychology* 3, 213–230.

Leontiev, A.N. & Zaporozhet, A.V. (1960) *Recovery of Hand Function*, London: Pergamon.

List, C. & Petitt, P. (2011) *Group Agency: the possibility, design, and status of corporate agents*, Oxford: Oxford University Press.

Marcel, A.J. (1992) 'The personal level in cognitive rehabilitation', In N. Von Steinbüchel, E. Pöppel and D. Cramon (eds.) *Neurophysiological Rehabilitation*, Berlin: Springer.

Michael, J. (2011) 'Interactionism and Mindreading', *Review of Philosophy and Psychology* 2, 559–578.

Miller, K. & Tuomela, R. (1988) 'We-Intentions', *Philosophical Studies* 53, 367–389.

Scweikard, D. & Pettit, P. (2006) 'Joint Actions and Group Agents', *Philosophy of the Social Sciences* 36, 18–39.

Searle, J. (1969) *A Theory of Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.

Searle, J. (1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

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

Searle, J. (2003) 'Social Ontology and Political Power', In F.F. Schmitt (ed.), *Socializing Metaphysics: The Nature of Social Reality*, Oxford: Rowman & Littlefield.

Searle, J. (2006) 'Social Ontology: Some Basic Principles', *Anthropological Theory* 6, 12–29.

Searle, J. (2010) *Making The Social World: The Structure of Human Civilization*, Oxford: Oxford University Press.

Simmel, G. (1908) 'Exkurs über das Problem: Wie ist Gesellschaft möglich?', In D.N. Levine (ed.) *George Simmel: On individuality and social forms* (1971), Chicago: Chicago University Press.

Toby, J. & Cosmides, L. (1995) 'Foreword', In S. Baron-Cohen (author) *Mindblindness: An essay on autism and theory of mind*, Cambridge: MIT Press.

Tollefsen, D.P. (2002) 'Collective Intentionality and the Social Sciences', *Philosophy of the Social Sciences* 32, 25–50.

Tuomela, R. (2002) *The Philosophy of Social Practives: A Collective Acceptance View*, Cambridge: Cambridge University Press.

Tuomela, R. (2003) 'Collective Acceptance, Social Institutions, and Social Reality', *American Journal of Economics and Sociology* 62, 123–165.

Tuomela, R. (2005) 'We-Intentions Revisited', *Philosophical Studies* 125, 327–369.

Tuomela, R. (2006) 'Joint Intention, We-Mode and I-Mode', *Midwest Studies in Philosophy* 30, 35–58.

Tuomela, R. (2007) *The Philosophy of Sociality: The Shared Point of View*, Oxford: Oxford University Press.

Tuomela, R. & M. Tuomela (2005) 'Cooperation and trust in group context', *Mind & Society* 4, 49–84.

Wimmer, H. & Perner, J. (1983) 'Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception', *Cognition* 13, 103–128.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 19–37, 2013

'Social identity' and 'shared worldview': Free riders in explanations of collective action

Helen Lauer

University of Ghana, Legon, Ghana
helenlauer@yahoo.com

Abstract

The notions 'worldview' and 'social identity' are examined to consider whether they contribute substantively to causal sequences or networks or thought clusters that result in intentional group actions. Routine reference to such purportedly key components of agents' intentions are presumed to help explain their collective actions. But problems emerge when we consider the theoretical details of attributing one worldview and identity to each individual, or a shared worldview to a whole community. Where does one worldview, or type of identity, leave off and another begin? Comparable fuzziness surfaces when we inspect the notion of distinct worldviews as inherently incommensurable, or distinct social identities as inherently antagonistic. Three proposed explanations of sectarian conflict or ethnic violence are analysed as examples of theories that link intentional group behaviour to the worldviews and social identities of the individuals directly involved. But as will be shown, it is not facts about worldviews and identities as such, but historically specific facts and contingent circumstances that impinge upon those individual agents' welfare (as well as their beliefs and values) which need to be examined in order to explain their group-motivated behaviour—be it violent, conciliatory, or otherwise.

1 Introduction[1]

If you want to explain how a human being is transformed into a suicide bomber, or how one neighbourhood group turns into a killing machine against another, it is generally conceded that the process under scrutiny is considerably less transparent than the sequences of thought involved when two people paint a house, or take a walk together (Bratman 1993, Gilbert 1990, Velleman 1997), or when a group goes blueberry picking every summer (Tuomela 2003) or when two rival teams try to defeat each other for the championship (Turner 2003). Indeed the process whereby people throw down everything and resort to violence in the name of a group allegiance remains obscure, and initially appears distinct from these other examples of coordination.

Since the end of the twentieth century international relations scholars, social psychologists, political scientists, and anthropologists have been preoccupied with situations where people are engaged in ethnic conflict, tribal war, sectarian violence, without being formally conscripted and institutionally authorised to do so by any sovereign state openly declaring its responsibility and control over the killing operations (Christie 1998, Eller 2005, Horowitz 1985). Theories that address such phenomena have been proliferating in the social sciences and in foreign policy circles without scrupulous philosophical attention.

[1] An ancestor of this paper appeared as "Worldviews and Identities: How not to explain intentionally collective actions," in *Legon Journal of International Affairs* (LEJIA) May 2007 Vol. 4 no.1 pp. 43-65.

The recent analytic literature devoted to scrutinizing theories of collective intentionality has developed a specialisation within action theory, wherein contributing decision theorists, philosophers of language and of mind have tried to characterise the collective sort of thought which "creates and maintains institutions," that is, the "we-attitudes" and "we-modes" of intention constitutive of norm-following (Tuomela 2003: 153, 162). Theorists in this vein have proposed models of interactive knowledge and coordination (Bratman 1993, 1999; Chant and Ernst 2008, Searle 2008, Tuomela 2005) to reveal the cognitive working behind global teleconferences and the stock market. Analytic precursors to these recent projects attempted to uncover the linguistic foundations for the kind of rule-following that makes all language participation possible including the meaning of moral imperatives (Wittgenstein 1958, Sellars 1963).[2] Other analysts have attempted to secure a metaphysical rather than linguistic or epistemic foundation for the assertability of statements referring to groups. Ruben (1982) for instance deduced the existence of a species of natural kinds as the necessary ontological commitment for correct statements of fact about stable social groups. However, such models were devised to explain the possibility and persistence of institutional arrangements given individual allegiances that are assumed to be fixed. For example Raimo Tuomela (2003: 136), among the most prodigious formalists in this area, brackets any variables concerning group dynamics by dubbing the group referent in his functions as "egalitarian," and treating it as a dummy constant, in order to study the logic of collective intentionality among individuals whose group commitments constitute a fixed status quo.

In this essay the focus is on collective behaviour that includes "struggles for social justice" (Hardin 1995, Honneth 1998). Violent group actions are intended to challenge, confront, and disrupt prevailing norms and institutions of the status quo rather than to sustain them. It is not obvious that the models of "collective acceptance" and "we-mode" of thinking (Tuomela 2003: 136, 144) ascribed to agents to explain how social institutions are composed and maintained are appropriate to represent episodes of unprecedented violence and social protest.

The data I am addressing here register as "sudden … episodic outbursts shattering periods of apparent tranquillity" (Horowitz 1985: 13). Such behaviour constitutes for some observers "a return to the primitive" (Freeman 1998: 15). To study the complex anatomies of social rupture constituted by groups' non-compliance with prevailing norms of dominance and subordination, one must look to the work of social and political philosophers contributing to a literature which has mushroomed in the last fifty years alongside the geographic dispersion of people from post-colonial societies in Asia, Africa and the Caribbean, migrating into former colonizing and neo-colonizing nations of the global North. These discussions are about cross cultural interchanges within a single socio-political network (Kymlicka 1997). Here theories about the ontogeny and moral status of groups have been used as vehicles to re-examine principles of distributive justice and civic rights, fairness, equity, and political obligations under the impact of changing demographics called 'multiculturalism' (Fraser & Honneth 2003). But again, no ready guidelines or precedents are on offer in these theories for analysing the causal mechanics or the logic that motivates individuals *in extremis* to engage in unprecedented violence due to their allegiance to subjugated or dominant groups.

The following reflections offer no positive account of why individuals engage in group violence. Rather, I scrutinize a general approach to explaining group conflict which rests on the assumption that violent behaviour is an inherent symptom of the way the combatants perceive

[2]I am grateful to Stephen P. Turner (2003) for reviving my attention to Sellars' classic analysis of 'ought' statements in his convincing analysis.

the world and themselves. I will demonstrate three snapshots of theories about ethnic and sectarian violence that display this assumption, but only as examples of my narrow concerns about treating social conflict as an *a priori* consequence of the way humans identify themselves as social beings. My snapshots are intended neither to do any significant damage nor full justice to the well reputed and widely received theoretical projects from which they are drawn. After defining terms in section 2, section 3 demonstrates the question begging effect of appealing to the group identity of agents in conflict in order to explain why they are in group conflict. I will introduce an argument from a World Bank consultancy report produced by the political scientist Donald L. Horowitz (1998) whose seminal book on ethnic conflict has lasting influence. Horowitz spells out the psychological roots of ethnic affinity as a means of accounting for its flipside, ethnic conflict. He proposes an analysis of ethnic affinity whereby "ethnic conflict is one phenomenon and not several" (1985: 53); but in doing so he recognises that "ethnic affiliations are located along a continuum of ways in which people organise and categorize themselves" (1985: 55). Hence Horowitz is neither simplistic nor essentialist in his analysis of the great swathes of data he selectively collects to document ethnic conflicts throughout Africa, Asia and the Caribbean, searching for root patterns and categories of behaviour to yield an "explanation that will hold cross culturally" (1985: xi). In section 4 I continue to examine this unilateral view of how to explain ethnic conflict, but the theorist relies on deductive method rather than enumerative induction, to reveal the nature of social identity as inherently adversarial. There my example derives from another seminal paper on ethnic conflict by Walker Connor (1972), the social scientist celebrated among nationalist movement activists (Conversi 2004), who coined the term 'ethnonationalism' (Zuelow 2002). I will show why an axiom of group identity which Connor regards as fundamental to personhood appears paradoxical because it defines group conflict among persons as logically inevitable. Section 5 addresses another problem with drawing up a universal template fitting all motivations for social conflict. I sketch the central argument of Axel Honneth (1992, 2001, with Fraser 2003), the widely acclaimed political philosopher at the centre of the debates about social justice and political recognition. Honneth's work is championed as uniting and empowering activists engaged in all sorts of struggle for social justice everywhere—from gay rights in Delhi to pastoralists' entitlement to land in Darfur (Babiker 2006, Deng 1995)—but ironically, Honneth's account of the fundamental need for social recognition can be interpreted as elitist and provincial. In closing I will show why the vocabulary of shared identity appears to force Connor to commit a fallacy of misplaced concreteness (Ryle 1949) when he proposes social engineering to reform combatants' worldviews, their perceptions, and their core sense of identity, as a means of ending ethnic conflict (Connor 1972: 353).

In all three of these examples it is the very nature of personhood, defined in different ways, which is presented as the key factor, irreducible to contingent circumstances, which is held chiefly responsible for all types of group conflict. These depictions of the *sui generis* nature of personhood as an essential contributor to group conflict counter views that regard group-identified violence as the result of rational choice (Hardin 1995) or as a response to extreme economic deprivation (Agnew 1992). The trouble with talking about the nature and logical structure of personhood *as such* is that it shifts our analytic focus away from the specific and very practical conditions, historical episodes, and material circumstances which undermine people's mutual trust, self confidence, and which feature in the way people perceive their options. But the deficit inherent in this shift of focus cannot just be declared; it has to be shown. And of course it is a patent truism—too trivial to bear remarking upon—that discovering facts about a region's political history, its relation to foreign powers, its changing economic

dynamics, will be key to understanding the causes and solutions to conflicts among people who occupy that region.[3] My objection to referring to the structure of every person's subjective 'worldview' as the key causal factor in all group conflict is that it diffuses the possibility of making such discoveries and mitigates their potential impact. Talk of worldviews and identities *per se* lends a misleading sense of locality and concreteness to the sources of individuals' decision making, as if actions could be wholly accountable in light of what agents themselves perceive and believe to be true and important.

On each of these accounts, the beliefs picked out as the primary source of protracted conflict are purported to prevail independently of the particular historical, cultural and contingent facts that distinguish groups and their purportedly characteristic behaviour patterns. But an explanatory model may fail to account for a certain kind of behaviour if it deflects attention away from the contingent events and relations unfolding at the specific times and places which comprise the reasons *why* people view the world as they do and therefore make the particular choices that they do. This is because the reasons that underlie intentional behaviour may involve factors of which the agents themselves are not aware. This point suggests there is something wrong with methodological individualism, which insists that the warrantability of a social explanation presupposes every substantive statement is translatable into one that attributes properties, relations and dispositions to individual agents and their situations (Brodbeck 1954, James 1985, Watkins 1973: 179). On the other hand, problems of coherence arise when worldviews are treated as emergent properties belonging to a group of individuals, or to any specifiable locus other than individual agents, as will be discussed in section 3.

Perhaps the objection raised here to positing worldviews in explanatory discourse would be better directed at the notorious difficulty of the subject matter itself; the dilemma at issue is succinctly captured by the mixed metaphor familiar to social psychologists: "people are both shaped by and the architects of their worlds" (Shepherd and Stephens 2010: 353). Making the phenomena under explanation in the social sciences even more intractable, the subject matter talks and sometimes explains itself. People's own accounts of what they get up to may not provide the whole story, but surely the agent's version of the facts must figure large in explaining what he aims to be doing, especially when his behaviour appears desperately counterproductive. To clarify situations quite so murky as this, surely it is innocuous to posit an abstract inner repository of beliefs and attitudes housed in a conceptual framework or system, and to associate that framework with an individual agent's most profoundly influential circle of social influences, allocating their centrality by positing the individual's 'social identity' as intrinsic to the framework's core. Positing worldviews might have instrumental value even if it lacks ontological status and epistemic respectability. But if we have no way of learning characteristics of such a framework (the agent's 'worldview') or its core (the agent's 'social identity') except by analogy from properties already and routinely ascribed to the individual agent's particular thoughts, then no clear explanatory purpose is served by making more than nominal references to a conceptual framework or to its core.[4]

[3] I am very grateful to the anonymous referees of *Abstracta* for stressing this point, and for other substantive comments on an earlier draft of this essay.

[4] A comparable though distantly related point (as well as others parallel to observations here concerning worldview individuation and its attribution to a community) is made by Frederick Suppe (1979: 218) in his exhaustive critical review of the (now passé) reference to *weltanschauungen* in twentieth century debates about the nature and dynamics of conflicting scientific theories. Of course the connection is oblique: my concern here is with worldviews as they get attributed as part of the data; not in the way that worldviews have been purported to characterize or qualify the content and tempo of scientific practice and theory change. Another point of departure for this analysis is Donald Davidson's seminal objections to the contrast between 'scheme' and 'content' (2001 (1974)).

This last point seems to recommend methodological individualism, contrary to the remark concluding the paragraph above. Overall, I do not expect the considerations gathered here will decide a preference for either methodological individualism or collectivism, for instrumentalist or interpretivist or an inductive realist reading of social explanation (Salmon 1992). I only mean to show the disappointments encountered by treating purportedly essential components of an agent's personhood or psyche or outlook as either necessary or sufficient for explaining why that agent has gone to radical extremes of group-affiliated behaviour on a given occasion. I propose that these disappointments are salient regardless of what view one holds about the underlying logic or overall significance of a social scientific explanation, or the levels of category by which it must be constructed, or the methods by which it should be pursued, or the means by which it should be assessed. I do not mean that the notions under interrogation are philosophically noxious in themselves. As will be sketched in section 2, both 'worldview' and 'social identity' carry a weighty spectrum of connotations accumulated from a range of respected literatures.

2 Defining terms

The term 'worldview' has had such an imposing career that its history has been tracked by several surveyors of different disciplines (Naugle 2002, Olthius 1989, Tuche 2008, Vidal 2008, Wolters 1989). General consensus and the *Oxford English Dictionary* indicate it first appeared in 1790, early in the tradition of German idealism when Kant coined 'Weltanschauung' in *Kritik der Urteilskraft.* (Naugle 2002, Tuche 2008: 1, Wolters 1989: 15).[5] The English term appeared in 1858 again (Tuche 2008: 1) in a popular treatment of Christian theology, and ever since it has played a key role in the ecumenical vocabulary that relates Christian doctrine with other phenomenological and scholarly traditions. For instance Wolters (1989: 15) recounts the extensive use of 'worldview' by the Dutch neo-Calvinists in the late nineteenth century, to contrast everyday beliefs (practices and rituals, folkloric, doctrinal, faith-based and commonsense convictions) of ordinary people, with the claims sourced in formal science and metaphysical speculation (arranged in highly abstract, systematic theories) labelled 'philosophy'. The interface of worldviews with philosophies as a way of signalling different levels and styles of abstraction continues to this day in Belgium under the aegis of the Leo Apostel Centre in Brussels (Aerts *et al* 2011, Note *et al* 2009). Other German words that bear a family resemblance to Weltanschauung to refer to workaday thought and practice is 'Lebensform' (form of life, or way of life) and its cognate 'Lebenswelt' (life-world). The former appears famously in Wittgenstein's *Philosophical Investigations* (1953: p. 11 article 23, p. 88 article 241) to characterise language participation in its broadest conceptual and practical respects. Neto (2011: 76) observes that at the time Wittgenstein was lecturing, 'Lebensform' was a term commonly used in biology. Husserl (1936) is generally credited with first using 'Lebenswelt' to highlight and centralise the subjective experience of belonging in a collective (Carr 1970). He thereby embellished the received approaches in the sociology of his day. Anthropologists ever since have generally recognised the phrase 'differences in worldview' as synonymous with 'cultural differences' (Geertz 1973: 93).

In order to register my narrow compass of concerns as clearly as possible I will maintain a distinction, somewhat pedantically, between two obvious and often conflated connotations of worldview as used in psychology and sociology, popularized philosophical and theological discussions (Schutz 1973, Note et al (eds.) 2009). I will make use of these subscripts just for

[5]The first translator into English of *Kritik der Urteilskraft* purportedly in 1858 rendered this as 'worldview'. I have not been able to locate with confidence the name of that first translator.

awhile, as it is intended only to keep the analysis more precise than it might be otherwise. I do not mean to imply that these senses cannot be co-extensive;[6] nor need the contrast be signalled indefinitely. It is only a rough and ready contrast after all, and cannot be well-defined in any case.

Hereafter 'worldview$_{ef}$' connotes the sense in which certain mental states or belief contents are presumed to provide an uneliminable 'empirical foundation' (= ef) for interpreting everyday experiences and forming intentions.[7] This is just a way of hypostasizing the range of know-how that gets called socialisation or social conditioning, combined with abilities regarded as innate. Ostensibly, some basic thoughts and categories are required in a transcendental[8] sense in order to form judgments and negotiate survival. It remains a lively area of research in developmental psychology to determine when and how basic functions and their later derivatives are generated throughout a person's cognitive maturation. Worldview$_{ef}$ means to be exhaustive. Let us assume these common life skills, motor capacities, doctrines, theories and imperatives are responsible for all the different governance structures, architectures, cuisines, kinship arrangements, canons of education, legal systems, mythologies, stories of divinity, rituals of worship and courtship, that comprise the objective social realities sustained by different cultural traditions (Sayre-McCord, 1991).[9] Part of this basic layer or foundation of thought or sub-propositional awareness might be the roots of one's social identity. I will introduce what I mean by 'social identity' in this picture momentarily. Worldviews$_{ef}$ feature in conflict episodes in the following way. The influential theorists whose views are only partially sampled later in this essay—Connor (1972), Horowitz (1985, 1998) and Honneth (2001, 2007)—regard the co-existence of multiple worldviews$_{ef}$ as the chief cause of belligerent hostility among people living under one state apparatus. They regard the propensity to engage in conflict as *sui generic* to personhood, positing the propensity for social struggle as intrinsic to having any social identity whatsoever (Connor 1972; Honneth 2001; Huntington 1993). These diverse views of social conflict sweep aside contingent circumstances as incidental to episodes of violence, attributing social conflict instead to fixed universal features of human cognition. But it seems that Connor's appeal to every individual's worldview$_{ef}$ as the source of social strife deteriorates into paradox; and both Connor's and Honneth's theses fail to accommodate all the data—as will be pursued in later sections 3 and 5.

In contrast 'worldview$_{ic}$' connotes a purposefully produced and revisable 'ideal construction'(= ic) *about* one's way of life which undergoes conscious revision in the light of new evidence and changing normative principles that one selectively gains throughout one's life. The ideal construction is a development and elaboration out of the empirical foundation, but some aspects diverge. This contrast between 'worldview$_{ef}$' and 'worldview$_{ic}$' is consistent with a measurable hypothesis widely received by social psychologists, viz. what people actually

[6]Paul Snowdon's (2003) analysis of Ryle's (now tacitly) received distinction between 'knowing how' and 'knowing that' inspired this caveat. Perhaps we will discover that descriptions of 'pure' and 'practical' reasoning are two ways of denoting the same cognitive processes in certain incidents of reasoning, as it was discovered that the 'the Morning Star' is identical to 'the Evening Star'.

[7]Varieties of meaning or cognitive holism, and internal assumptions of such theses, have received considerable opposition. See Davidson's criticism of distinguishing an overall structure for all one's thoughts from their particular contents, in his seminal 'On the Very Idea of a Conceptual Scheme' ([1974] 1984). And for a more vehement, sustained objection to various semantic and conceptual holism theses including Davidson's, see Fodor and LePore (1992, 1993).

[8]I allude here to the form of transcendental argument characterized by Jerry Fodor and Ernest LePore (1992: 261) as follows: "Any argument of the form 'F-ing is impossible unless P; F-ing actually occurs; therefore P'."

[9]Speculative controversy surrounds the source of this variety: are these differences genetically encoded or learned with language? (Du Plessis, 2001) Are there logical differences between worldviews$_{ef}$ or do they follow from laws of physics? If so, which logical laws; which laws of physics? (Aerts 2000; Penrose 1990; Stapp 1993)

think and do diverges from what they say when asked about what they think and do.[10] In this second sense, a worldview$_{ic}$ "is a framework that ties everything together, that allows us to understand society, the world, and our place in it … help[ing] us to understand, and therefore to cope with, complexity and change … and help[ing] us to make critical decisions which will shape our future" (Heylighen 2000: 1). It is often supposed not only that worldviews$_{ic}$ *can* be revised, but that they *should* be revised—to enhance human welfare, or to fulfil other identifiable goals. For instance Walker Connor (1972) advises homogenizing every individuals' worldviews$_{ic}$ to improve national (and thereby global) stability. The African moral philosopher Kwame Gyekye (1997) advocates Walker Connor's proposal (1972) to foster in a population's worldview$_{ic}$ a sense of "meta-nationality" to reinforce a nation-state's prospects for stability by ensuring that through language unification and other measures, the entire population shares a "single psychological focus" (Connor, 1972: 353). Not all theorists share this optimism about the feasibility of harmonizing national policies designed specifically to dissipate destabilizing influences in society. Some political theorists hold a contrary *a priori* assumption that volatile schisms between group identities are endemic to the very process of socialisation. Perhaps the most influential popularisation of this position is Samuel P. Huntington's (1993) 'clash of civilisations' thesis.

At first glance, the suggestion that we can and should be encouraged to reconstruct our worldview$_{ic}$ begs the question of why we are having so much trouble getting along together in the first place, if indeed both the room for improvement and the direction for achieving it are ostensibly ready to hand. This may be a version of the ancient Greek problem of *akrasia*: as socialised agents we are presumed to harbour all the capacities required to move towards perfecting the human quality of our collective lives overall; and yet we do not. Does the rich imagery of 'contrasting worldviews$_{ef}$' contribute to our understanding of why there appears to be a propensity within some cultures to foster provincialism and xenophobia, while other societies encourage equanimity and appreciative confidence in diversity?[11] If worldviews$_{ef}$ do play some causal or mediating role in the formation of intentions, then where do they figure in an accurate map of the social realm? Are worldviews$_{ef\ \&\ ic}$ properties of individual agents; or rather do they exist as features of a social field emerging as and when agents interface? Or do worldviews$_{ef\ \&\ ic}$ belong to enduring loosely cohesive social units transforming over generations?

There are good reasons for regarding the core or central beliefs of a worldview$_{ef}$ and its brighter normative '-ic' version as the social 'identity$_{ef}$' and 'identity$_{ic}$', respectively, of those individuals sharing those worldview$_{ef\ \&\ ic}$. For it is uncontroversial that having any identity at all includes thinking sometimes about one's interactions with other people. Such thinking about human interactions, real or potential, presumably involves sharing or refusing to share certain attitudes, beliefs, preferences and repulsions of other people with whom one most frequently interacts. This is why having one or more specifiable identities seems integrally linked to sharing one or more worldviews$_{ef}$ with other people. And of course if we are to learn from one another—for instance to be introduced to wholly new ideological perspectives which will inspire us to change our outlook on the world, maybe even to yield an outright conversion

[10]Social psychologist David Matsumoto (2006: 35) demonstrated empirically that a person's "culture" (way of life, practices and ideologies as well as knowledge traditions, political organisation and everyday habits) is distinct from his "cultural worldview" (belief system *about* his culture) which is comprised of "social construction of reality expressed in consensual ideologies … learned through the media and cultural elites, authority figures and opinion leaders."

[11]Kwasi Wiredu (1998) argues convincingly that in West Africa, indigenous institutions of democratic governance foster and reinforce the political will to cooperate and accommodate, which is structurally pre-empted by the multi-party electoral systems in the traditions of western republics.

of faith—we must be able to share more than one worldview$_{ic}$ together. Correlatively, we can create together an image of our ideal self—so there must be an identity$_{ic}$ which in some respects can be shared; presumably this is what makes it possible for life coaches to do business.

More pertinent to the matter of evaluating explanatory models of group behaviour is the question of how we might speak of individuals' sharing a worldview or social identity (either the –ef or the –ic sense of these notions) in a way which doesn't boil down to saying something which is either vacuously true or plainly false. On inspection, this image of a shared worldview$_{ef}$ containing the social identity of group members will not do at all for real life. To begin with, it gravely distorts social reality by failing to concede the variety of obligations, expectations, contrary and contradictory beliefs and priorities attributable to a whole community, no matter how small or how closely knit it is (Crehan 2002). Even when people are elected deliberately to function as representatives and endeavour to speak and act consciously on behalf of their group, individuals' knowledge sets will have to diverge. (For instance you may know, as your neighbours may not, how many of your siblings abroad have children.) Some individuals, located in a certain kind of cross-cultural "frontier or boundary situation," may be privy to two or more worldviews$_{ef}$ simultaneously (MacIntyre 1987: 388). Such might be the case for formally educated people leading complicated, cosmopolitan lives in postcolonial cities. So here a different theoretical problem arises: there seems no principled way to set any upper limit of worldviews$_{ef}$ or worldviews$_{ic}$ and identities that might belong to one such individual, nor to distinguish among the variety that might be shared between two such individuals.

Reflecting on these situations, it is obvious that overall replication of individuals' thought sequences is not a prevailing feature of shared worldviews$_{ef}$ anywhere. So the mere absence of mirror imaging between our priorities and projects cannot by itself account for the kind of reasoning that leads us to resort to violent conflict. Talking about agreement between our beliefs obviously cannot mean a literally perfect match; so *similarity* of our beliefs must be all that is intended by saying that we share a worldview$_{ef}$. But pressing upon the 'similarity' of our beliefs cannot warrant presuming that people risk misunderstanding each other to the point of violent animosity, if their beliefs are not similar enough to say that they share a worldview$_{ef}$. If belonging to the same worldview$_{ef}$ is required for us to understand each other, and if belonging to the same worldview$_{ef}$ means by definition that we must share similar beliefs, then I could never learn anything totally new from you. And in fact people routinely do learn very new things from one another, whether they have lived at close quarters or at very different times or places from each other. This becomes more obvious as the phenomenon called globalisation accelerates. Given today's telecommunication technology, the same worldview may be exposed in some sense to an indeterminable number of individuals everywhere at once. To be publicly accessible to the extent that a single coherent belief system has the degree of global influence apparent nowadays, a worldview$_{ic}$ must be timeless and borderless, and without any fixed location. But then it is not clear how to attribute distinct and wholly incompatible worldviews to individuals or to groups in violent conflict, except as just another way of saying that they are in obvious (if only because violent) conflict.

Defining a worldview$_{ef}$ in the loose and familiar way proposed above, prompts the question of how to tell when our beliefs have become *dissimilar* enough to say that there obtains not one worldview between us but two. There appears to be no fixed worldview-neutral standpoint from where a referee could distinguish which beliefs are shared by different groups of people, in order to select which statements of fact and descriptions of events belong to one particular worldview$_{ef}$ and not another. Even with respect to the one worldview$_{ef}$ that we share, it is

not obvious how to sort those beliefs and principles that are essential to our maintaining the same perspective from those convictions that are negotiable without our risking a breach in our worldview.[12] In sum, it is appropriate to wonder whether the notions of 'shared' and 'contrasting' worldviews$_{ef}$ or worldviews$_{ic}$ really amount to anything more than observing that between certain people there is broad agreement, and between others broad disagreement, on many topics. Because, contrary to what is assumed by the theorists studied in sections 2–4, it is by no means obvious how to trace the sole cause of unwieldy group behaviour directly to the fact that members share one *type* of worldview$_{ef}$ or social identity and not another, as if it were the type of worldview or identity *itself* which is causing them to be cantankerous or belligerent.

Before moving away from the definition of terms, it is important to justify analysing the terms 'worldview' and 'identity' as intimately related—either as empirical givens or normative constructions. I can leave off the subscripts now. Because human lives are so interdependent, there are many contributors to the literature on social and political identity who see some sense of group or national identity as interdependent with having a personal identity (see Miščević 2000).

Still, one might take exception to treating the terms 'worldview' and 'identity' as interchangeable. Somehow one expects more of the notion of identity. Yet with respect to the sense in which an identity and a worldview are properties we can share as members of the same social group, the best that it seems one can do by way of distinguishing them is to suggest that one's identity is somehow the most basic and fundamental or the core part of one's worldview. But this distinction is very brittle. Because if we start making rigorous demands about how to tell in a principled way when to count one thought or subliminal experience as *more basic* than another, then the contrast between the core of a belief system, vs. its midfield or periphery, threatens to break down. Thus the insistence by social psychologists upon distinguishing 'identity' from 'social identity' (Cantwell and Martiny 2010: 319-320) is suspect. The same problem of essential contestability arises when trying to demarcate the difference between which choices are personal vs. which to count as public, or for that matter what counts as a physical fact in contrast with a social fact, as Hacking (1997) and Turner (1999) have noted in reviewing Searle's doctrine of social reality. The ambiguity of the first dichotomy is pertinent to whether the notion of a worldview (which is public) must be kept distinct from the notion of one's identity (which is personal); so an example is called for to illustrate the essential contestability (MacIntyre 1973) of such a contrast: Consider one's choice of a life partner, and how one manages one's sexual gratification as an adult. Nothing could be more personal and profoundly intimate as choices go in American society; but for most people throughout contemporary West Africa such concerns are at the heart of social life and remain of central importance to the entire community. For many female individuals worldwide, relinquishing sexual gratification altogether is not even remotely a matter of personal choice; it is a rigid standard of proper breeding and of becoming socialised as a marriageable member of a reputable family within the community (Shweder 2004). But I will not pursue these provocative issues here, least of all the more trenchant epistemological problem of systematically ordering beliefs according to their apparent relative *basic*-ness.[13]

[12] If we start making rigorous demands about how to tell in a principled way when to count one thought as *more basic* than another, then the contrast between the core of one's belief system *versus* its midfield or its periphery threatens to break down. In this paper there is no space to deal adequately with the dense difficulty of ordering thoughts according to their apparent relative *basic*-ness.

[13] The problem is taken up in Nicole Note *et al* (eds) 2009. In these considerations it is sufficient to regard a thought A as more basic than B in the sense that psychologists indicate that having the thought A is somehow causally responsible

3 Citing worldviews as primary causes and the risk of circularity

Vagueness is not the cause of the vacuity that emerges when we refer to an inscrutable worldview in order to explain why the people who share it are acting for inscrutable reasons. Generally, vagueness in itself is not a fault of key terms used in social studies. Many concepts related to the description of human affairs are irredeemably vague ('family', 'gift', 'virtuous', 'self-evident'). They must be so; otherwise their intended versatility will fall short of applying to the full range of divergent value systems and knowledge traditions (MacIntyre, 1973). Even if we could specify precisely a set of parameters to determine when a worldview should count as shared, and therefore accessible to whom, the problem of vacuous question begging claims will remain if we try explaining the reason for an action by citing the agent's worldview as its primary cause. To show this, briefly: Suppose you are identified as being a member of community X because you clearly share the X worldview. As a social scientist I have systematically observed that you do Y and not Z on Mondays, and thus I can document your belonging to the X community, as I can cite background studies which establish that to do Y and not Z on Mondays is a defining trait of the X-centric worldview. But then to explain *why* it is that X-centric people go for Y and always refrain from doing Z on Mondays, nothing is illuminated by my referring to the fact that this is a key feature of the X-centric worldview.[14]

The question begging illustrated in this abstract narrative might seem to be just an artifice, until one considers the following political scientist's depiction of how psychological affiliation causes ethnic hostility of all kinds. Donald Horowitz (1989) explains the structure and strategy of ethnic conflict by illuminating a particular property as unique to ethnic group identity. Horowitz claims that belonging to an ethnic group means that one is "thinking of oneself in a special way." That is, ethnic identity entails thinking of oneself as "possessing properties that are characteristic or representative of the social category that [one's] group embraces" (Horowitz 1998: 16). This identification with others in the group is what "leads" individuals to "submerge" themselves into the "collective identity" of other people in whom "they see themselves." Horowitz cites psychological studies confirming this tendency to engage in self sacrificing behaviour for those whom I see as extensions or images of myself (Horowitz 1998: 16). He cites further studies showing that it is members of my own group that I am most likely to assume share my tastes, aptitudes, values and beliefs, even in the absence of relevant facts about them. And he stresses that the converse is also true. Horowitz cites even more studies displaying that my being attracted to a person tends to make me exaggerate what I have in common with that person. So Horowitz concludes that a tight interdependence has been empirically established as existing between my being attracted to you and my treating you as belonging to my ethnic or social group. Now since my being attracted to you causes me to be "biased in your favour," then all the fundamental elements of a psychological account are at hand to explain how my ethnic identity can cause me to indulge in self destructive behaviour.

To illuminate the essential content of this explanation let us review it in outline: I am disposed to feel an affinity with you because I am attracted to you; and my thinking of you as belonging to my group causes me to be attracted to you all the more. This is because my being attracted to you causes me to exaggerate our similarities; so I see you as belonging to my

for an individual's thinking B, to paraphrase a connotation of basic-ness suggested by David M. Rosenthal, Coordinator, Interdisciplinary program in cognitive science of the Graduate and Research Center of the City University of New York. In conversation, 2006.

[14] I owe the substance of this point to Bernard Williams (1972: 35) where he complains about a standard functionalist defence of ethical relativism.

group just in case I find you attractive. And that is what the empirical data confirms: that I am inclined to find you attractive because I think of you as part of my group, and it is because I see you as part of my group that I am inclined to find you attractive. Which presumably means I am prepared to fall for you literally as well as figuratively. So we have the structure of Horowitz's portrayal of the psychological structure of ethnic conflict. A tighter and more ineffectually question begging circle would be hard to imagine.

Apart from the emptiness of this account, it is difficult to avoid evidence that contradicts Horowitz's cited studies. Many people are habitually attracted to others all the more because of their obvious and blatant differences—culturally, temperamentally, and physically. More germane is the conceptual problem of how to differentiate people into groups based on their *perceived* similarities and differences; for then the criteria determining whether someone belongs to ethnic group A and not B will necessarily shift and vary depending upon who you ask.

Nevertheless, for a good number of social and political commentators following Walker Connor (Conversi, ed. 2003), it is the sheer perceived variety of identities or worldviews that remains the prime cause of destabilizing conflict in the world today. Walker Connor (1972) has grown influential through declaring that statistically, when it comes to the instability of nation states, it is "multi-ethnicity" in the *intangible* sense, that trumps all other "tangible" factors (including religion, culture, GDP, geography, climate, population size, historical impact of colonisation, type of economy, style of governance). This leads him to declare that "the prime cause of political disunity is the absence of a single psychological focus shared by all segments of the population" (Connor 1972: 353).

4 Citing worldviews as incommensurable and the risk of paradox

We must assume that people ordinarily do share worldviews to some extent, even if they are diametrically opposed to each other; because to expose the divergence between two contemporary opinions or perspectives or agenda, there must be some common landscape or backdrop against which the opposing views and projects can be set in order to expose their glaring contrasts. So for instance consider the antagonism of worldviews in South Africa before 1994. It seems it would have been obvious to assign mutually exclusive mental universes to the presiding president at that time F. W. de Klerk, and the succeeding president Nelson Mandela (1994) who was then being shunted in and out of solitary confinement cells—on the grounds that their beliefs and values were so totally polarized for their lifetimes. But attributing to these two South Africans strictly isolated, mutually exclusive worldviews would be mistaken. It's not as if during his 27 years of maximum security imprisonment Mandela never shared with his oppressors any factual beliefs at all about the apartheid system. Mandela knows far more intimately about the life-threatening effects of apartheid law than those who enforced it. The worldviews of the ANC militant and of the last apartheid chief executive did intersect—indeed they collided—in the most penetrating and horrifically graphic ways. A crucial moment in world history was Mandela's insisting that he would die before he would acquiesce to the injustice of the very same system over which P. W. Botha and later de Klerk held executive authority.[15] Mandela was not condemning some other parallel government system in an incommensurable worldview, in

[15]In the internationally renowned Rivonia trial, Nelson Mandela declared on Monday 20 April 1964, in Pretoria's Palace of Justice: "During my lifetime I have dedicated myself to this struggle of the African people. I have fought against white domination, and I have fought against black domination. I have cherished the ideal of a democratic and free society in which all persons live together in harmony and with equal opportunities. It is an ideal which I hope

which even the names 'President Botha' and 'President de Klerk' could have no conceivable significance or meaning remotely like the one understood in the world of his oppressors. In stark contrast, when it came to the crunch, Pieter Willem Botha finally resigned from office; and after him Fredrick Willem de Klerk did nothing to defend the system which since 1989 had vested in him so much prestige and power. On the contrary de Klerk has even been credited with dismantling those same institutions without a blink when it appeared most prudent to him and his kind to do so. Unless we recognise that it was *the same set of material conditions and legal statutes* about which the convictions, feelings, intentions, character traits and actions of these three personalities crossed each other so diametrically, we will not capture the extent of contrast between them as moral agents and as historical figures. Permit a simplistic analogy: you and your neighbour may share faith in the same Lord Jesus even though you are prepared to die by the Holy Word, while he carries it around and quotes from sacred scripture strictly to win friends and influence people. But you both share the same faith in some crucial—if only nominal—sense; were it not the case then there would be no way that your sincerity could contrast so severely with his hypocrisy, as it surely and objectively does.

So far we have been considering that worldviews at their outer reaches appear to be very flexible; one worldview may have to be distributed among contemporaries interacting with each other no matter how insuperably different their political, social and personal value orientations may be. Yet some of the same theorists we are discussing (Conversi (ed.) 2003; Connor 1972; Fraser and Honneth 2003) who recognise worldviews as malleable and receptive to reform, are also impressed by the ineffable tenacity of the central beliefs of a person's worldview called his or her (ethnic or social) identity.

One's ethnic identity cannot be defined by legal statute nor political party allegiance nor religious ordinance. Walker Connor stresses that ethnic identity is not an essentially "cultural assimilation;" it is rather "profoundly psychological" (Connor 1972: 341-342). He says an individual can "shed" all the customs and mannerisms that count as "tangible" or "overt cultural manifestations," but that does not rid the person of who he or she really is. Connor chides us that it is:

> … superficial … to predicate ethnic strife upon language, religion, customs, economic inequity, or some other tangible element … what is fundamentally involved in such a conflict is that divergence of basic identity which manifests itself in the 'us-them' syndrome. (Connor, 1972: 341, 344)

Connor stresses that "the idea of 'us' requires 'them'." If having a sense of belonging would be impossible without a sense of not-belonging, then the 'us *versus* them' dichotomy is unelim-inable and polarization is fundamental to a person's "basic identity" (Connor, 1972: 341). And if knowing who I am requires knowing who I am *not*, polarization will be seen as intractably, unavoidably and incorrigibly personal. Quoting Connor again: "the ultimate answer to the question of whether a person is one of us or one of them seldom hinges upon adherence to overt aspects of culture" (Connor, 1972: 341). He urges that we realise the "primary cause" of ethnic conflict is not a matter of culture or economics or politics but of this polarity which is *sui generis* and "fundamental" to personhood. "The prime cause of political disunity is the

to live for and to achieve. But if needs be, it is an ideal for which I am prepared to die." That day he was sentenced to life imprisonment. Quoted by Helen Joseph, ed. (1981).

absence of a single psychological focus shared by all segments of the population" (Connor, 1972: 353).[16]

But this kind of analysis leads to paradox. The sort of groups that depend for their allegiance upon this polarizing feature of their members' basic identity might be called *exclusionary groups*. To belong to an exclusionary group means to regard every member of any other group as being one of 'them' and not 'us'. Now consider building a coalition or confederation of all such exclusionary groups; let us call this Exclusionation, or E for short. You can belong to confederation E provided that you do not feel that you belong to it. That is because recognising yourself as belonging to any other group besides the one which is essential to your basic identity must mean that you do not really belong to an exclusionary group in the first place, and so you would not properly qualify for membership in confederation E either. But according to Connor, subscribing to the *us versus them* polarity is essential to having any basic identity at all. Thus, everybody must belong to some exclusionary group. And so, hypothetically, each and every one of us would have to belong to such a confederation E, on the sole condition that we refuse to recognise that we do. Such a paradox appears to follow from Connor's positing the "us versus them syndrome" as an intrinsic feature of our basic identity.[17]

5. Primordialist claims[18] about ethnic identity and the risk of provincialism

As mentioned in the introduction, another influential social commentator besides Walker Connor has proposed a quite different psychological feature as requisite for social identity, which also appears mistaken although it does not degenerate into blatant paradox. Axel Honneth (1992, 1995, 2001, 2007) is widely appreciated in political philosophy for positing a primordial human need for recognition as the primary root cause of all struggles against social, economic and political injustice. Honneth claims that our sense of self respect and integrity generate initially from our receiving social approval from significant others, and that to maintain self respect throughout life we must constantly seek facsimiles of that primal acceptance upon which our survival and security depended in our infancy. He cites as primary evidence the direct injury that we all feel from degrading treatment or from public humiliation, from belittling verbal abuse or from outright physical assault. Indeed we all know first hand that self esteem can be damaged by receiving poor treatment. So Honneth argues the contra-positive equivalent must also be self evident. That is, since the absence of respect from others can trigger the experience of injury to our self esteem and personal integrity, then it must also be the case that sustained self-esteem and personal integrity depend upon our sensing that other people recognise us as worthy of their respect. Thus he regards the primal need for recognition as essential in all fights for social justice.

But Honneth's psychoanalytic template for the foundation of social morality seems to apply to only a narrow range of society where individuals learn to view their welfare as entailing the obliging deference of others. He seems to overlook the possibility that individuals can harbour self-esteem and personal integrity even though from birth their subordinate status does not

[16]Connor brings to mind the poignancy observed by a BBC correspondent Matthew Price who remarked ruefully, "Neither side understands the other," when reporting an Israeli tank raid killing 12 Palestinian civilians in their residential neighbourhood in the Gaza Strip. The quote was broadcast on the *BBC Worldservice* from Beit Hanoun in Gaza, November 12, 2006.

[17]I owe this initiative of applying the Russellian paradox to explore the consequences of positing a polarizing component as essential to our 'basic identity' to the example set by Romane Clark's (1980) elegant criticism of the view that belief contents must have a propositional structure.

[18]I owe to Anton du Plessis (2001) the summary classification of causal theories explaining ethnic conflict into "primordialist" and "constructivist," which I understand as jargon in the literature of international relations (IR) reflecting the more general social scientific dichotomies, 'nature *versus* nurture' or 'innatist *versus* behaviourist'.

entitle them to expect that significant others will recognise or treat them as worthy of respect. Women generally occupy such a position in most societies. And both men and women in South Africa, who forged a new definition of citizenship through the second half of the 20th century, did so despite the annihilating, profoundly abusive conditions that prevailed throughout their lives and which had been legalised long before their birth.

Significant in this regard is the title chosen for the chronicle of an articulate ANC activist, Naboth Mokgatle (1971): *The Autobiography of an Unknown South African*. Mokgatle vividly describes his regimen for cultivating Black Consciousness which inspired his powerful contribution to the success of the anti-apartheid struggle. He writes that he intentionally provoked abuse, contriving to routinely expose himself to the worst excesses of physical brutality and psychological humiliation by breaking pass laws conspicuously so that he would be caught by authorities and dealt with in the degrading way apportioned to his identity under apartheid law. He says this discipline taught him to lose his fear of police and prison (Mokgatle, 1971: 216-224). Contrary to Honneth's theory, Mokgatle's self respect and esteem that fuelled his fight for justice was strengthened precisely as he focussed his awareness on the abuse, insults and injuries perpetrated by significant authorities who actively deprived him of respectful recognition and intentionally disqualified his moral worth. Nor was this a one-off peculiarity unique to the psychology of one South African. The ANC ideology deliberately schooled activists in the anti apartheid struggle to divorce themselves from interaction with liberal whites eager to offer them recognition and political reverence, precisely because such recognition and approval was regarded as a weakening influence that threatened to undermine the uncompromising autonomy of the Black Consciousness Movement, an autonomy which was necessary to genuinely overturn (rather than to subtly perpetuate by colluding with) the white supremacist status quo.[19] It can be argued that the political unity that was achieved for one brief but glorious historical moment in 1994 for all South Africans was made possible because the prevailing mood was to embrace with unrestrained admiration the variety of worldviews and identities that reflected the whole population, celebrated as a multi ethnic rainbow. The transcendent political unity did not emerge by eliminating or sublimating the population's multiple psychological perspectives, nor by erasing from people's thoughts their awareness of their diverse ethno-nationalist identities and loyalties, as Walker Connor and others following him have prescribed.

5 Conclusion

Among those who regard social engineering as a feasible way of improving community relations, it is popular to suppose that the best way to influence group behaviour (e.g. to dismantle sectarian or ethnic hostility) is to reconstruct and expand the worldviews and identities of the population directly engaged in and victimized by the conflict. Walker Connor is widely influential in propounding the view that worldview reform will directly yield a safeguard against destabilizing elements for a nation state, even if adjustments of all the other superficial or "tangible" elements have failed—be they efforts to revise economic conditions, cultural prejudices, political affinities, or religious exclusions (Connor 1972: 353).

[19]This was spelled out to me in conversation by one of the founders of the Black Consciousness Movement, the ANC activist and colleague of the late Steven Biko, Dr. Mamphela Ramphele, New York City, August 1984. This approach of renouncement has served both Dr. Ramphele's nation and her global recognition very well; among many other accomplishments including honorary doctorates and various influential World Bank posts, in 2004 she was voted 55th of the Top 100 Great South Africans.

In closing, I propose that the foregoing considerations of sections 1–4 cumulatively suggest that this approach reveals a version of Ryle's fallacy of misplaced concreteness (1949), encouraged by the worldview and intrinsic identity model of group conflict. I don't mean to go too far with the allegory, lest it detract from the point here: Purportedly distinct, geographically remote worldviews_{ef} and worldviews_{ic} seem to intersect with each other in countless unpredictable ways.

Indeed worldviews understood as conditioned belief systems and as normative schedules of voluntary commitment are frighteningly telekinetic and porous. For a graphic illustration, recall in 2006 when the life of a young Lebanese woman was jeopardized in a quiet residential neighbourhood of Baalbeck in the Bekaa Valley of her country (Spector 2009). Her home and school and her future were destroyed, not because of the way she perceives herself or her world, but because of the way her identity and her world were perceived by an evangelical preacher in Texas who successfully lobbied the US House of Representatives to escalate the weapons of mass destruction being sent to support America's proxy war using Israel's defence forces in the Middle East (Kurtzer 2010). The example highlights graphically the drawback in focussing on the content of individuals' worldviews *as such* in order to explain their behaviour: talk of worldviews suggests that people's primary reasons for acting depend ultimately upon how they *perceive* their world and not how their world is in fact. It is true that people in a conflict zone (or after a trauma) often act out of exaggerated anxiety over imagined threats. Nonetheless it is typically the case that a girl who runs away hysterically from a moving army tank that is firing shells does so in order to avoid getting hit by explosive shells, not because she fears the sensation of getting hit by explosive shells, nor because she is upset by her *perception* of a moving army tank firing shells.

In turn, the Texan preacher's success in mobilizing three and a half thousand Christian Zionists to rally in Washington DC to export more explosive shells had such a devastating impact in Lebanon not because of the accuracy of his beliefs about the Lebanese; but because of the war mongering climate dominating America in June 2006, and the inevitable vulnerability of US Congressional representatives to pressures of public opinion (Chafets 2007). The point here is that the priorities and beliefs—that is, the identities and worldviews—of politicians half a world away might have to be considered to make sense of why Lebanon was set on fire in August 2006. Quizzing those caught up in defending against the relentless bombardment, or speculating about the structure of their thinking, might illuminate first hand experiences and perspectives of the latest war in Lebanon, but there is no reason to suppose that this line of inquiry will necessarily reveal the cause of that conflict.

References

Aerts, Diederik et al (2011) (eds.) *Worldviews, Science and Us*, London: World Scientific.

Agnew, Robert (1992) 'Foundation for a general strain theory of crime and delinquency', *Criminology* 30.1, 47–87.

Arnold, Millard (ed.) (1978) *Steve Biko: Black Consciousness in South Africa*, NY: Random House.

Babiker, Mustafa (2006) "African Pastoralism through Anthropological Eyes: Whose Crisis?" IN *African Anthropologies: History and critique and practice*, Ntatangwi, M. et al (eds.) Dakar: CODESRIA and London: Zed, 170–187.

Bekker, Simon (2001) 'Identity and Ethnicity', IN *Shifting African Identities*, Bekker, S.B. et al (eds). Pretoria: Human Sciences Research Council, 1–6.

Boaten, Abayie (1999) "Ethnicity and Ethnic Conflicts in Africa: Ghana's Example," *Pan African Anthropological Association Conference:* Anthropology of Africa and Challenges of the Third Millennium: Ethnicity and Ethnic Conflicts.

Bratman, Michael (1999) *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge: Cambridge University Press.

Bratman, Michael E. (1993) 'Shared Intention', *Ethics*, 104.1 October, 97–113.

Brodbeck, May (1954) 'On the Philosophy of the Social Sciences', *Philosophy of Science*, 21.2, April, 140–156.

Cantwell, Allison M. & Sarah E. Martiny (2010) 'Bridging Identities through Identity Change', *Social Psychology Quarterly* 73, 319–320.

Carr, David (1970) 'Husserl's Problematic Concept of the Life-World', *American Philosophical Quarterly*, 7.4, 331–339.

Chafets, Zev (2007) *A Match Made in Heaven: American Jews, Christian Zionists, and one man's exploration of the weird and wonderful Judeo-Evangelical Alliance*, New York: Harper Collins.

Chant, Sara Rachel & Zachary Ernst (2008) 'Epistemic Conditions for Collective Action', *MIND* 117.467, 549–573.

Christie, Kenneth (1998) (ed.) *Ethnic Conflict, Tribal Politics: A Global Perspective*, London: Routledge.

Clark, Romane (1980) "Not Every Act of Thought Has a Matching Proposition," IN *Midwest Studies in Philosophy Volume V: Studies in Epistemology.* Peter A. French, et al (eds.) Minneapolis: University of Minnesota, pp. 509–524.

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

Connor, Walker (1972) "Nation-Building or Nation-Destroying?" *World Politics*, 24.3, 319–355.

Connor, Walker (1978) 'A nation is a nation, is a state, is an ethnic group is a . . .' *Ethnic and Racial Studies*, 1.4 October, 377–400.

Conversi, Daniele (2004) (ed.) *Ethnonationalism in the Contemporary World: Walker Connor and the study of nationalism*, London: Routledge.

Crehan, Kate (2002) *Gramsci, Culture and Anthropology*, Berkeley: University of California Press.

Crehan, Kate (2006) "Hunting the Unicorn: Art and Community in East London," IN *The Seductions of Community: Emancipations, Oppressions, Quandaries*, Gerald Creed (ed.) Sante Fe: SAR Press, pp. 163–180.

Cutler, A. Claire (1999) "Locating 'Authority' in the Global Political Economy," *International Studies Quarterly*, 43.1, 59–81.

Davidson, Donald (2001 (1974)) "On the Very Idea of a Conceptual Scheme," IN *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.

Deng, Francis M. (1995) *War of Visions: Conflict of Identities in the Sudan*, Washington DC: Brookings Institute Press.

Du Plessis, Anton (2001) "Exploring the Concept of Identity and World Politics," *Politics of Identity and Exclusion in Africa: From Violent Confrontation to Peaceful Cooperation*, Seminar Report no. 11, 13–25, Johannesburg: Konrad Adenhauer Foundation.

Ekeh, Peter (1975) "Colonialism and the Two Publics: A Theoretical Statement," *Comparative Studies in Society and History*, 17, 91–112.

Eller, Jack David (2005) (ed.) *Violence and culture: a cross-cultural and interdisciplinary approach*, Belmont, California: Thomson/Wadsworth.

Fodor, Jerry & Ernest Lepore (1992) *Holism: A Shopper's Guide*, Oxford: Blackwell.

Fraser, Nancy & Alex Honneth (2003) *Redistribution or Recognition? A political-philosophical exchange*, London: Verso.

Freeman, Michael (1998) 'Theories of Ethnicity, Tribalism, and Nationalism', IN *Ethnic Conflict, Tribal Politics: A Global Perspective*, Kenneth Christie (ed.), London: Routledge, 15–34.

Geertz, Clifford (1973) *The Interpretation of Cultures*, New York: Basic Books.

Gilbert, Margaret (1990) 'Walking Together: a paradigmatic social phenomenon', *Midwest Studies in Philosophy* VI, 1–14.

Guibernau, Montserrat & John Hutchinson (2004) "History and National Destiny," *Nations and Nationalism*, 10.1, 1–8.

Gyekye, Kwame (1997) *Tradition and Modernity*, New York: Oxford University Press.

Hacking, Ian (1997) 'Searle, Reality and the Social: a review symposium on John R. Searle, *The Construction of Social Reality.* London; Allen Lane, 1995', *History of the Human Sciences* 10.4, 83–110.

Halliday, Fred (1995) "International relations and its discontents," *International Affairs*, 71.4, 733–746.

Hardin, Russell (1995) *One for All: the logic of group conflict*, Princeton: Princeton University Press.

Honneth, Alex (1992) "Integrity and Disrespect: Principles of a Conception of Morality Based on the Theory of Recognition," *Political Theory*, 20.2, 187–202.

Honneth, Alex (2001) "Recognition: The Epistemology of Recognition," *Proceedings of the Aristotelian Society*, 75.1, 111–126.

Honneth, Axel (1992) 'Integrity and Disrespect: Principles of a Conception of Morality Based on the Theory of Recognition', *Political Theory*, 20.2, 187–202.

Honneth, Axel (1995) *The struggle for recognition: the grammar of social conflicts*, Cambridge: Polity Press.

Honneth, Axel (2001) 'Recognition: The Epistemology of Recognition', *Proceedings of the Aristotelian Society*, 75.1, 111–126.

Honneth, Axel (2007) *Disrespect: the normative foundations of critical theory*, Cambridge: Polity Press.

Horowitz, Donald L. (1985) *Ethnic Groups in Conflict*, Berkeley & Los Angeles: University of California Press.

Horowitz, Donald L. (1998) "Structure and Strategy in Ethnic Conflict," *Annual World Bank Conference on Development Economics* Washington DC April 20–21, 1998.

Huntington, Samuel P. (1993) "The Clash of Civilizations," *Foreign Affairs*, 72.2, 22–28.

Husserl, Edmund (1936 (1970)) *The Crisis of European Sciences and Transcendental Phenomenology*, transl. David Carr, Evanston: Northwestern University Press.

James, Susan 91985) *The Content of Social Explanation*, Cambridge: Cambridge University Press.

Kaya, Ahyan (2005) Book review: *Ethnonationalism in the Contemporary World: Walker Connor and the study of Nationalism*, (ed.) Daniele Conversi. IN *Nations and Nationalism* 11.3, 463–485.

Kurtzer, Daniel (2010) *A Third Lebanon War: Contingency Planning Memorandum no. 8.* July. New York: Council on Foreign Relations Centre for Preventive Action.

Kutz, Christopher (2000) 'Acting Together', *Philosophy and Phenomenological Research* 61.1, 1–31.

Kymlicka, Will (1997) *Multicultural Citizenship*, London: Oxford University Press.

MacIntyre, Alisdair (1973) 'The Essential Contestability of Some Social Concepts', *Ethics*, 84.1, 1–9.

MacIntyre, Alisdair (1987) 'Relativism, Power, and Philosophy', IN *Philosophy: End or Transformation?* Baynes, K. et al. (eds.), Massachusetts: MIT Press, 385–411.

Mamdani, Mahmood (2003) 'Making sense of political violence in Postcolonial Africa', IN *Ghana in Africa and the World Today*, Toyin Falola (ed.) New Jersey: Africa World Press, 689–711.

Matsumoto, David (2006) 'Culture and Cultural Worldviews: Do verbal descriptions about culture reflect anything other than verbal descriptions of culture?' *Culture & Psychology* 12.1, 33–62.

Miščević, Nenad (2000) *Is National Identity Essential for Personal Identity?* Chicago: Open Court.

Mokgatle, Naboth (1971) *The Autobiography of an Unknown South African.* Berkeley: University of California Press.

Monagan, Sharmon Lynnette (2010) 'Patriarchy: Perpetuating the practice of female genital mutilation', *Journal of Alternative Perspectives in the Social Sciences* 2.1, 160–181.

Naugle, D.K. (2002) *Worldview: The History of a Concept*, Grand Rapids, Michigan: Wm B. Eerdmans.

Neto, Norberto Abreue Silva (2011) 'The uses of "forms of life" and the meanings of life', IN *Forms of Life and Language Games*, Gálvez, Jesús Padilla & Margit Gaffal (eds.) Piscataway, New Jersey: Transaction Books, 75–106.

Note, Nicole et al. (2009) (eds.) *Worldviews and Cultures: Philosophical reflections from an intercultural perspective*, Rueil-Malmaison: Springer Scientific.

Olthius, James H. (1989) 'On Worldviews', IN *Stained Glass: Worldviews and the Social Sciences*, Marshall, P. et al (eds.) New York: University Press of America, 26–40.

Root, Michael (1986) "Davidson and Social Science," IN *Truth and Interpretation: Perspectives on the philosophy of Donald Davidson.* Ernest LePore (ed.) Oxford: Blackwell, pp. 272–304.

Ruben, David-Hillel (1982) 'The Existence of Social Entities', *Philosophical Quarterly* 32.129, October, 295–310.

Ryle, Gilbert (1949) *The Concept of Mind*, Chicago: University of Chicago Press.

Said, Edward W. (2001) 'The Clash of Ignorance', *The Nation*, October 22 http://www.thenation.com/doc/20011022/said. Accessed June 23, 2006.

Salmon, Merrilee H. (2002) 'Philosophy of the Social Sciences', *Introduction to the Philosophy of Science*, Salmon, M.H., et al. (eds.) New Jersey: Prentice-Hall, 404–425.

Schutz, Alfred & Thomas Luckmann (1973) *The Structures of the Life-World*, trans. R.M. Zaner & T. Engelhardt, Evanston: Northwestern University Press.

Searle, John R. (2008) 'Language and Social Ontology', *Theory and Society* 37.5, October, 443–459.

Sen, Amartya (2006) "The Interview," with Carrie Gracie. *BBC Worldservice.* August 5.

Shepherd, Hana R. & Nicole M. Stephens (2010) 'Using Culture to Explain Behavior: An integrative cultural approach', *Social Psychology Quarterly* 73, 353–354.

Shweder, Richard Allan (2004) 'What About Female Genital Mutiliation? And Why Understanding Culture Matters in the First Place', IN *Engaging Cultural Differences: The Multicultural Challenge in Liberal Democracies*, Shweder, R.A. et al. (eds.) New York: Russell Sage Foundation, 216–222.

Smith, Anthony D. (1999) *Myths and Memories of the Nation*, London: Oxford University Press.

Spector, Stephen (2009) *Evangelicals and Israel: The Story of American Christian Zionism*, Oxford: Oxford University Press.

Suppe, Frederick (1979) *The Structure of Scientific Theories*, 2nd edition, Urbana: University of Illinois.

Tollefson, Deborah (2002) 'Collective Intentionality and the Social Sciences', *Philosophy of the Social Sciences* 32.1, 25–50.

Tuche, D. (2008) 'Worldview, Challenge of Contextualization and Church Planting in West Africa – Part 1: Definition of Worldview and the Historical Development of the Concept', *Global Missiology* Special issue: 'Contextualization', July. http://www.globalmissiology.net/ Accessed March 5, 2013.

Tuomela, Raimo (2003) 'Collective Acceptance, Social Institutions, and Social Reality', *American Journal of Economics and Sociology*, 62.1, January, 123–165.

Tuomela, Raimo (2005) 'We-intentions Revisited', *Philosophical Studies*, 125.3, September, 327–369.

Turner, Stephen P. (1999) Review essay: 'Searle's Social Reality: the Construction of Social Reality', *History and Theory*, 38.2, 211–231.

Turner, Stephen P. (2003) 'What do We Mean by "We"?' *ProtoSociology*, 18–19, 139–162.

Velleman, J. David (1997) 'How to Share an Intention', *Philosophy and Phenomenological Research*, 57, 29–50.

Vidal, Clément (2008) 'Wat is een wereldbeeld? (What is a worldview)?' IN *Nieuwheid denken. De wetenschappen en het creatieve aspect van de werkelijkheid.* Van Belle, H. & J. Van der Veken (eds.) Leuven: Acco.

Watkins, J.W.N. (1953) 'Ideal Types and Historical Explanation', IN *Readings in the Philosophy of Science*, (eds.) Feigl, H. & M. Brodbeck, New York: Appleton-Century-Crofts.

Watkins, J.W.N. (1973) 'Methodological Individualism: A Reply', IN *Modes of Individualism and Collectivism*, (ed.) O'Neill, J., London: Heineman, 179–184.

Wendt, Alexander (1992) "Anarchy Is What States Make of It," *International Organisation* 46.2, 321–347.

Williams, Bernard (1972) *Morality: An Introduction to Ethics*, Cambridge University Press.

Wittgenstein, Ludwig (1958) *Philosophical Investigations*, transl. G.E.M. Anscombe, New York: MacMillan.

Wolters, Albert M. (1989) 'On the Idea of Worldview and Its Relation to Philosophy', IN *Stained Glass: Worldviews and the Social Sciences*, Marshall, P. et al (eds.) New York: University Press of America, 14–25.

Zimmerman, Ulf. (2001) Book review of Lawrence E. Harrison & Samuel P. Huntington (eds.) *Culture Matters: How Values Shape Human Progress.* NY: Basic Books (2000). http://www.h-net.org/reviews/showrev.cgi?path=26284981759538 Accessed September 2006.

Zuelow, Eric (2002) Book review of *Ethnonationalism in the Contemporary World: Walker Connor and the Study of Nationalism.* Daniele Conversi (ed.), New York: Routledge, (2002). The Nationalism Project. http://www.nationalismproject.org/books/ Accessed October 2006.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 39–53, 2013

Being and Becoming in the Theory of Group Agency

Leo Townsend

University of Cape Town
South Africa
leo@tenderscan.co.za

Abstract

This paper explores a bootstrapping puzzle which appears to afflict Philip Pettit's theory of group agency. Pettit claims that the corporate persons recognised by his theory come about when a set of individuals 'gets its act together' by undertaking to reason at the collective level. But this is puzzling, because it is hard to see how the step such a collective must take to *become* a group agent – the collectivisation of reason – can be taken without them already *being* an agent. I explore this puzzle by recounting Pettit's account of the emergence of group agents. According to Pettit this process has two stages: a first stage in which a collective incurs the distinctive pressure exemplified by the Doctrinal Paradox, and a second in which the collective responds to that pressure by instituting decision-making mechanisms designed to secure collective rationality. After arguing that this second, response stage in Pettit's account is not coherent, I conclude with the tentative suggestion that the personhood of groups should be seen as depending not only on the efforts of group members but also on the recognitive attitudes of other persons in a wider discursive community.

Introduction

Over the past ten years Philip Pettit has developed a novel form of realism about group agents.[1] Drawing on formal results obtained in the field of judgement aggregation, Pettit argues that suitably designed social groups should be recognised as 'corporate persons'[2] with 'minds of their own'.[3] Such conclusions have been advanced before, but what is especially striking about Pettit's theory is that it resists capture in either of the two dominant traditions through which they have previously been promulgated. The first of these traditions is the 'authorisation theory' associated with Hobbes, according to which a group agent is formed when a collective authorises a privileged sub-group to speak and act for the whole. Pettit contends that, unlike his own view, this conception of group agents is only a 'redundant realism', since 'everything the recognition of such a group agent entails is already expressible in individual-level language'.[4] But Pettit also wishes to distance his theory from the other venerable tradition of theorising

[1] Pettit (2002; 2003; 2007; 2009), List & Pettit (2006; 2011). Throughout this paper I refer to the theory of group agency as Pettit's theory, even though it has been largely worked out in collaborative publications (most regularly with Christian List). This is because Pettit is the common thread in all the publications I am focused on, and, judging by his individually-authored work, it is most likely he who is responsible for the sections of the co-authored work which are of primary interest to me here.

[2] Pettit (2002), List & Pettit (2011)

[3] Pettit (2003)

[4] List & Pettit (2011: 8)

about group agents, which he calls 'animation theory'. Animation theory, which Pettit traces to the influence of Hegel, holds that 'group agents are emergent entities over and above the individuals who compose them'[5], formed when a transcendent force 'animates them and gives rise to a single centre of agency'.[6] On account of his commitment to 'methodological individualism', Pettit insists that this sort of view is metaphysically objectionable and intolerably mysterious. The view he defends thus seeks to be both non-redundant and unmysterious, and my hope for this paper is that it casts some doubt on whether Pettit successfully achieves this.

Pettit claims that the corporate persons recognised by his theory of group agency are 'made, not born'[7], and this raises the question of just who it is who makes them. His original treatment of the Doctrinal Paradox and subsequent forays into organisational design made in collaborative work with Christian List[8] suggest that it is the group itself that somehow 'gets its act together', transforming itself from a collection of individuals into an autonomous rational agent.[9] But this creates something of a bootstrapping puzzle for his account, for it is hard to see just how the step that groups must take in order to *become* agents – the 'collectivisation of reason' – can be taken without the group in question already *being* an agent.

My aim is thus to see whether Pettit succeeds in providing a non-redundant, unmysterious and *non-circular* account of the process whereby a collective incorporates into an agent or person.[10] In Section 1, I explore what Pettit considers to be the first stage of this process, in which a collective incurs a certain kind of *pressure* – the pressure to incorporate, to become a group agent. For Pettit, this stage of the process is exemplified by the much-discussed Doctrinal Paradox of judgement aggregation, and so the question I pursue in this first section is whether the paradox can have any traction for a collective that is not already incorporated. In Section 2, I turn to the second stage in Pettit's account, in which the collective responds to the pressure to incorporate by undertaking to exercise reason at the collective level. Pettit thinks there are a variety of decision-making procedures which might accomplish this 'collectivisation of reason', and the question I ask about these is whether they too might actually presuppose an already-incorporated group agent. In the end I think that Pettit fails to dissipate the bootstrapping problem bruited above, and, in Section 3, I diagnose this failure as the product of a tendency to overlook the role that inter-subjective attitudes in a broader discursive community might play in instituting a group's status as an intentional subject or corporate person.

1 The Pressure to Incorporate

The lesson of the Doctrinal Paradox, and Pettit and List's[11] feted impossibility theorem, is that two requirements we may wish to place on the formation of collective attitudes are in fact in tension with each other.[12] The first is that the collectively-held attitudes ought to be sensitive to the attitudes of the collective's members; the second is that the attitudes collectively

[5]Ibid., p. 9

[6]Ibid., p. 9

[7]Pettit (2007: 495)

[8]Pettit (2002); List & Pettit (2002; 2011)

[9]For the sake of clarity I use the term 'collective' here to denote an un-incorporated but potential group agent, and 'group' for an already incorporated group agent.

[10]Worries over the possible circularity of accounts of collective intentionality have been explored before, notably by Tollefsen (2002b) who targets the views of Tuomela (1995) and Gilbert (1996). As far as I know Pettit's theory of group agency has not been subjected to similar critique.

[11]List & Pettit (2001)

[12]Though Pettit's recent work (esp. List & Pettit 2011) focuses predominantly on the impossibility theorem, I concentrate here on the Paradox. I take it that Pettit's reason for this recent shift in focus is that the impossibility theorem somewhat more *precise* than the Paradox: it sharply delimits the extent of the difficulty for attitude aggregation (as not simply a problem for majoritarian aggregation functions, but all such functions), and it allows Pettit to identify exactly

endorsed should comprise a set that would be rational to hold. Now Pettit's key claim is that certain kinds of collectives will face an overwhelming pressure to give priority to this second requirement, even though doing this may lead them to violate the first requirement – they may end up espousing views which many (even all) of their members individually reject. To make sense of Pettit's account of how group agents emerge we must first make sense of this 'pressure' to exercise reason at the collective level.

1.1 Paradox and Pseudo-paradox

The Doctrinal Paradox is usually illustrated with the following toy jurisprudential example. A defendant is being sued for damages arising from an alleged breach of a contract. Legal doctrine dictates that the verdict of the court – as to whether or not the defendant is liable – must depend on just two issues: (1) whether a contract existed prohibiting the defendant from performing a certain action, and (2) whether or not the defendant performed that action. Such a case will be turn out to be paradoxical if the views of the judges have a certain kind of structure, such as the following:[13]

Judge	**Issue 1: Contract?**	**Issue 2: Breach?**	**Outcome: Liability?**
A	Y	N	N
B	N	Y	N
C	Y	Y	Y
Majority	Y	Y	N

The reason this profile is paradoxical is that the outcome of the case depends on the decision-making protocol adopted. If the court allows a majority of outcome votes to determine the overall outcome – sometimes called the 'conclusion-based approach' – the defendant is not liable; if they apply legal doctrine to the majority-favoured issue votes – the 'premise-based approach' – then the defendant is liable.

Pettit has devoted considerable effort to generalising the scope and significance of the Doctrinal Paradox in this form – showing how it can arise for a great variety of groups in a great variety of circumstances, and that it can arise even without the mediating influence of 'doctrine'.[14] But for all his generalising efforts, Pettit has, to my mind, neglected the task of *circumscribing* the difficulty confronted by groups in paradoxical scenarios. For it is possible – and, I think, instructive – to generate profiles which *resemble* paradoxical cases but in which, intuitively, there is no real problem for the group. As we shall see, reflection on such 'pseudo-paradoxes' can make it tempting to think that the pressure to collectivise reason will only be felt by groups which are already incorporated.

Imagine that a television station wishes to know whether viewers would like the Olympic diving event to replace one of the two programmes – Game of Thrones (GT) or Mad Men (MM) – ordinarily screened between 7pm and 9pm on a certain night. Upon interviewing a (very) small sample audience, the following profile is generated:

---

what the various options for overcoming that difficulty are (i.e, the different ways of relaxing the conditions which generate the impossibility theorem). From the perspective of my discussion here, however, these considerations are both moot: I am concerned only with the general tension between collective rationality and individual responsiveness (which the Doctrinal Paradox reveals vividly), and I am happy to admit that collective rationality can be secured by the various decision-making procedures Pettit proposes. Moreover, since the Paradox gets us to reflect on the way in which a particular kind of collective, in the course of its discursive activity, might encountering and respond to this 'discursive dilemma', it is arguably more amenable to the question I am concerned with here – the question of the *process* through which group agents emerge.

[13] Kornhauser & Sager (1993: 11)

[14] See esp. Pettit (2003)

Person	Replace (GT or MM) with diving?	Drop GT?	Drop MM?
A	Preferred	Dispreferred	Preferred
B	Preferred	Preferred	Dispreferred
C	Dispreferred	Dispreferred	Dispreferred
Majority	Preferred	Dispreferred	Dispreferred

This profile may well present a problem for the TV station, but it does not obviously amount to a paradox for the TV-watchers. A majority of the individuals interviewed preferred that regular programming be interrupted in order for the diving to be screened, but most did not want Game of Thrones to be dropped, and most did not want Mad Men to be dropped. Is 'the majority' confused here, even though each individual is quite clear and consistent? No, because the apparent inconsistency involved in 'what the majority wants' vanishes when we observe that on each issue the majority preference is constituted by a unique pair of individual preferences. Although it is inconsistent to wish for either Game of Thrones or Mad Men to be dropped, but neither Game of Thrones nor Mad Men to be dropped, the rational requirement that such inconsistency be avoided only applies to unitary entities. Since the majority featured in each column is, strictly speaking, a *different* majority each time, the bottom line of this profile reflects disagreement amongst the people interviewed rather than inconsistency.

Why can we not then say, to return to the standard example of the Doctrinal Paradox, that there is no failure of rationality in the court's verdicts derived from majoritarian voting on all issues? This would mean that the court judges that (a) a contract existed, that (b) a breach occurred, that (c) if a contract existed and a breach occurred then the defendant is liable[15], and that (d) the defendant is not liable. This sounds irrational only because we have surreptitiously slipped from talk of what 'a majority of judges' thinks to talk of what 'the court' thinks; if we remember that the view of the court is only an arithmetic function of the individual judges' views then this appearance of irrationality disappears. In other words, to say that there is a failure of rationality appears to beg the question of the court's status as an entity in its own right – one which is subject to certain constraints of rationality.[16] But if paradox only arises for groups that are already rational entities in their own rights, then it seems it could not provide the impetus for the emergence of such entities.

1.2 Purpose and Paradox

Clearly, Pettit needs a principled way of distinguishing paradox from pseudo-paradox, and a way of deeming the bottom line inconsistency in the judicial profile pernicious without presupposing that the panel of judges is already incorporated. How might this be done? Perhaps by drawing on the idea that certain social collectives are *purposive*.

The sample group of TV-watchers is what Pettit calls a 'mere collection'[17]: they have no common purpose or even any knowledge of one another's contribution to the interview process in which they each participate. By contrast, the court is a 'purposive collectivity'[18] required to *act* in designated ways – imposing penalties, awarding damages, sentencing, etc. – which uphold legal doctrine. Because the court's judgements are 'near the coal-face'[19] of

[15]This is based on the assumption that the judges agree on the relevant legal doctrine.

[16]Such question-begging is, arguably, exhibited in Pettit's tendency to describe *the set of views* which issue-by-issue majoritarian voting would deliver as 'irrational' instead of simply inconsistent (see esp. Pettit 2002, 2003; List & Pettit 2002).

[17]List & Pettit (2011: 31-2)

[18]Pettit (2002, 2003)

[19]Pettit (2009: 77)

their action they cannot, on pain of jeopardising the court's agential capacities, be inconsistent: 'Let an agent try to act on inconsistent representations or motivations… and there will be a straightforward breakdown: actions will be supported that cannot be realised'.[20]

This distinction between purposive collectivities and mere collections seemingly helps Pettit to make non-circular sense of the pressure to incorporate. A collective might be recognised as purposive – it might have certain goals which it needs to pursue – and because of this, *as a response*, it ensures that its collective views are rational. According to Pettit, standards of rationality are 'nothing more or less than desiderata of agency: standards such that agents will generally do better as agents by robustly satisfying them'.[21] In order to be able to effectively *act* in fulfilment of its purpose, a purposive collectivity must take steps to ensure that its collectively-held attitudes are, at a minimum, mutually consistent.[22] Thus it is the panel of judges' purposiveness which distinguishes them from the TV-watchers, and which ensures that they find themselves susceptible to the distinctive pressure exemplified by the Doctrinal Paradox. In short, the fulfilment of purpose requires agency, and agency requires rationality – where, crucially, what is 'required' is not pre-required but can be provided responsively.

One might still wish to press the original bootstrapping objection here, by claiming that identifying something as a purposive collectivity must presuppose that it is a rational agent, since a purpose could not be fulfilled by a collective that was not already a rational agent. But this line of objection is mistaken: it elides the distinction between potentiality and actuality, overlooking the fact that treating something as a purposive collectivity can itself serve to instil or awaken the very rational-agential capacities required for the fulfilment of its purposes. This kind of bootstrapping, according to which certain capacities are *brought about* by being taken to be *already there*, is often thought to underpin the developmental processes in individuals. For example, some theorists[23] have claimed individuals come to be trustworthy only by being trusted – that is, by being treated as though they were already trustworthy. Similarly, it has been claimed that actual fitness to be held responsible for one's actions results from a process of 'responsibilization'[24] – of being treated as though one were fit to be held to responsible. Such bootstrapping seems to serve a powerful purpose in the individual case and it should not be ruled out in the case of social groupings.

A better line of objection targets the purported link between purposiveness and the requirement of agency. An agent, according to the conventional picture Pettit adopts, is something with beliefs and desires, and the capacity to intervene in the world for the fulfilment of its desires in line with its beliefs[25]. But collectives are composed of individuals who are themselves agents, so why can't the individuals comprising the collective co-ordinate their actions so as to serve the purposes of the collective? Purposiveness may require agency but it is not clear at this point why it requires the collective's own agency.

To answer these questions we need to attend more carefully to the notion of purpose involved in the idea of a purposive collectivity. Once again we may find it instructive to reflect on the differences between the TV-watchers and the panel of judges. Why are the TV-watchers not counted as purposive? Certainly they are interviewed for a purpose. The

[20]List & Pettit (2011: 25).

[21]Pettit (2007: 497)

[22]According Pettit, to be rational one must form attitudes which the evidence favours ('attitude-to-evidence' rationality), one must not have inconsistencies in one's body of attitudes ('attitude-to-attitude rationality'), and one must tend to act for the realisation of one's preferences in line with one's beliefs ('attitude-to-action rationality').

[23]E.g., McGeer (2002); Pettit (1995)

[24]Garland (2001)

[25]Pettit (2002, 2007, 2009); List & Pettit (2011)

crucial difference between them and the panel of judges seems to be that it is not *their* purpose but the TV-station's which their stated preferences are rung in to serve. (To see this, imagine that the interviewees pooled their resources and bought the TV-station; then the pattern of their judgements on the scheduling issue would be *their* problem, and the profile would be paradoxical.) By contrast, the panel of judges seems to have a purpose of its own – something which the individuals on the panel collectively pursue – so perhaps it is this proprietary purpose which prompts certain collectives to form their own centres of agency.

What then does it mean for a collective to have a purpose of its own? Presumably it means that there exists some goal that the collective strives to achieve or some role that it seeks to perform. But one thing it cannot mean, if Pettit's account of the emergence of group agents is to be coherent, is that the collective has a kind of *attitude* of its own – something like a group intention. For if the purpose of a purposive collectivity were tied to a group-level intention then it would seem that we would already have an incorporated group agent – something which is already an intentional subject in its own right – and so the having of a collective purpose could not be integral to the process of becoming such an agent.

Pettit sidesteps this concern by appropriating a key concept from the literature on collective intentionality – the concept of 'joint intention'.[26] According to Pettit, a joint intention can be distinguished from a group intention since it is a configuration only of individual attitudes, such that:[27]

(1) each of the individuals in the collective intends that they together promote some purpose,
(2) each of the individuals intend to play their part in fulfilling that purpose,
(3) each of the individuals forms these intentions (1&2) because they believe others have formed similar intentions, and,
(4) this is all a matter of common awareness.

The notion of joint intention appears to provide Pettit with a route out of the bootstrapping worry, at least insofar as the worry concerns the pressure to collectivise reason. The first clause (1) tells us that, where there is joint intention, each of the members of a collective intend that they – the collective – jointly serve some purpose. *Intending-that* (as opposed to *intending-to*) is conceived by Pettit as a matter of the individuals each favouring a state of affairs over which they have at least partial control.[28] This ensures that each of the members has an individual interest in taking steps to preclude the sorts of inconsistencies in collective-level judgements which can (as the Doctrinal Paradox reveals) result from perfect individual responsiveness. This is because, as Pettit tells us, inconsistencies cannot provide a sound basis for the fulfilment of the collective's purpose. The pressure felt by certain collectives to collectivise reason is precisely the realisation that their purposes cannot be served if inconsistencies in group-level judgements are allowed to arise. But when commitment to these purposes is understood in terms of joint intention it is an entirely individual affair – there is no group intentional subject presupposed.

2 The Response to the Pressure

Perhaps then a conceptual wedge can be driven between the notion of members of a collective severally sharing a purpose and the notion of a group with its own purpose. If so, then the fact that members of a collective share a purpose does not imply that the collective is already

[26] Bratman (1999); Tuomela (1995); Gilbert (2001)

[27] List & Pettit (2011: 33)

[28] As Pettit, drawing on Bratman's (1999: Ch 8) seminal statement of the concept, puts it, 'I intend that something happen only if… I want it to happen and I am in a position to do something about it' (Pettit & Schweikard 2006: 23).

incorporated – and hence it is possible that its purposiveness, understood in terms of joint intention, might call for incorporation without presupposing it. This would mean that we can make sense of the pressure to incorporate in a way that does not make it unnecessary (i.e., such that the only possible subjects of that pressure are already incorporated). But this does not yet mean that the coherence of Pettit's account of the emergence of group agents is secured, because it is not clear at this point how a collective could possibly *respond* to that pressure without already being a group agent.

We might worry that what is required as a response to the pressure exemplified by the Doctrinal Paradox is the collective's own response, where this is understood as some sort of group action. But such a response could not of course be part of the account of the emergence of group agency, since it would quite clearly require the collective to already be an agent. So what is needed is a member-level story – one which doesn't pre-require but might still be able to institute a group agent. Pettit seeks to provide such a story by endorsing a second notion prevalent within the collective intentionality literature – the notion of *joint action*, which he distinguishes from group agency.

Pettit claims that several individuals can co-participate in a joint action on the basis of their joint intention without constituting a group agent.[29] But although it need not, joint action can, according to Pettit, give rise to group agents: a collection of individuals can jointly act in certain specific ways so as to transform themselves into a group agent.[30] This is the second, response stage in Pettit's account of how group agents come into existence, and its tenability depends on whether the steps that need to be taken to transform a collection of individuals into a group agent are indeed candidates for joint action. In this section I tackle this issue by considering first whether collectivising rationality is *itself* a possible joint action, and second (because I think it is not) whether collective rationality is something which could be secured within a group indirectly, *via* the members' joint action.

2.1 Collectivising Rationality as Joint Action

What steps must be taken for a collective to transform itself into a group agent? We know from Pettit's discussion of the Doctrinal Paradox that he thinks they need to *collectivise rationality*, but is this a potential joint action? More particularly, can each individual member *intend that* they together collectivise rationality? This might seem a peculiar question, especially in light of the fact that we are not here questioning whether collectivising rationality is something that groups can do (our question is about *who* does it, and, specifically, whether it can only be done by a group agent). But if groups do collectivise rationality, and we agree with Pettit and List's[31] supervenience thesis – that the attitudes and actions of the group supervene only on the attitudes and actions of individual members – then it seems the collectivisation of rationality must be something which individual members can intend. Intending-that meant favouring a certain state of affairs – in this case, that the collective collectivises rationality – and committing to playing one's part in that state being realised. But since, by the supervenience thesis, individuals cannot help but be involved in anything the collective does (including becoming rational at the collective level), it seems that individuals can certainly so intend, and hence that the collectivisation of rationality can be a joint action.

We should, however, be careful here, for it is not obvious that simply being a part of something amounts to 'playing one's part'. Consider the following example. Imagine that,

[29] Pettit & Schweikard (2006: 19)
[30] Ibid., p. 33
[31] List & Pettit (2006, 2011)

for whatever reason, I wish to discover *completely by accident* that my partner is having an affair (perhaps I wish to separate, but think it is only legitimate for victims of adultery to instigate these things, and I also don't want to snoop or pry). Since it could *happen* that I discover completely by accident that my partner is having an affair, it seems plausible that it is a happening I could wish for. But is it something I could intend? Certainly I could not *intend to* make this discovery since what is discovered is not wholly within my personal sphere of influence (i.e., some part of it is in *another's* sphere of influence). Could I *intend that* I accidentally discover my partner's infidelity? I do not think so, even though it is plausibly a state of affairs I could favour, and one, were it to be realised, which would necessarily involve me. The reason I cannot intend it (in the sense of intending-that) is that the way in which I would have to be involved is not as an agent.[32] Since what has to happen for the wish to be realised has to happen *accidentally*, I cannot intend to actively play any role in bringing it about; playing a role would, precisely, defeat the fulfilment of that wish.

I think that something similar goes on when we try to conceive of the members of a collective intending that they together collectivise rationality. Collectivising rationality, as the Doctrinal Paradox teaches us, is in direct conflict with individual responsiveness. What this means is that, although this collectivisation must necessarily involve individuals, it does not cast them in the sort of role they can intend beforehand to perform. To commit to the collectivisation of rationality is to commit to being less than an individual agent in the exercise of one's group role; in acting and thinking *for* the group one is not thereby performing as an autonomous agent, but as part of the group's subpersonal machinery. When one intends, one takes up an attitude towards a potential future action of one's own (this can be part of a joint action), but what is required in the collectivisation of rationality is not the performance of one's own action, but the performance of someone else's action – the group's.[33] This means that the collectivisation of rationality is not a potential joint action; it is a singular process whereby a new centre of rationality and agency is formed, and this process is not one in which group members can intelligibly intend to participate.

2.2 Collectivising Rationality via Joint Action

Perhaps Pettit is able to sidestep this concern too, since he does not quite say that collectivising rationality is itself the joint action through which groups emerge. Nonetheless he does believe that joint action can effect the collectivisation of rationality, and so bring about group agents:

> 'First, the members act jointly to set up certain common goals and to set up a procedure for identifying further goals on later occasions. Second, the members act jointly to set up a body of judgements for rationally guiding action in support of those goals, and procedures for rationally developing those judgements further as occasion demands. And third, they act jointly to identify those who shall act on any occasion in pursuit of the goals'[34]

What is the difference between this and jointly performing the action of collectivising rationality? The key difference seems to lie in the introduction of 'procedures', by which is meant more or less explicit rules to guide the internal machinations of the group. If by joint action these procedures can be inculcated, and if action conforming to these procedures will count as collectively rational, then it appears that joint action can, in the end, bring about group agents (though that action is not itself collectivising rationality).

[32] In Pettit's terms, I am not 'in a position to do something about it'.

[33] Cf. Rovane (2004)

[34] Pettit & Schweikard (2006: 33)

But what will these all-important procedures be? Pettit considers the virtues of a number of functionally explicit organisational procedures, including the 'sequential priority rule' in which logically interconnected issues are placed in a sequence and issues only go to a vote so long as they are not already determined by their logical relation to prior, already-decided issues. In the end Pettit thinks that though they guarantee consistency in group-level judgement, the rigidity of functionally explicit procedures such as the sequential priority rule may cause groups endorsing them to violate some other demands of rationality: [35]

> 'When I realize that some propositions that I believe entail a further proposition, the rational response may well be to reject one of the previously accepted propositions rather than to endorse the proposition entailed. Those are the undisputed lessons of any coherence-based methodology and the group that operates under a sequential priority rule, or under any variant, will be unable to abide by them'[36]

So if functionally explicit procedures are unable to secure collective rationality, what sort of organisational mechanisms does Pettit think groups should adopt in order to transform themselves into group intentional subjects? To answer this we need to attend to a distinction Pettit makes between 'reason' and 'rationality'. As we saw above, rational standards, for Pettit, are nothing other than the desiderata of agency; rational creatures are those who tend to believe what the evidence favours, whose store of beliefs and preferences are largely consistent, and who tend to act in line with their intentions. They are veracious, consistent and enkratic because this is the best way to be an agent. But none of this requires a rational creature to attend to the propositional contents of her attitudes and form meta-propositional attitudes about them – asking, for example, whether this belief is *true*, whether that preference *implies* another, whether these judgements are mutually *consistent*, and so on. According to Pettit, simple rational agents can get along adequately without being able to attend in this higher-order way to their thought; their rationality – the extent to which they meet the standards of rationality laid out above – is 'secured by their make-up or design'.[37] But sophisticated rational agents, who have the capacity to form attitudes about their attitudes, can go beyond this, reinforcing their rationality by means of *reason*:

> 'To be able to reason… is to be able to conduct an intentional activity that is designed – and perhaps explicitly intended – to raise the chance of satisfying [rational] desiderata'[38]

Pettit thinks that the robust rationality of groups will only be secured if the group can succeed in *reasoning*, and he thinks that there is at least one sort of organisational procedure which can achieve this: the straw vote. A straw vote is simply a non-binding vote, through which a group's judgement can be registered without being concretised; the views thus reached can always be revised later if the group so decides. A straw vote can be used to temper the inflexibility of the likes of the sequential priority rule, thus allowing groups to be more sensitive

[35]Functionally explicit decision-making procedures like the sequential priority rule and the distributed premise-based approach are also, arguably, susceptible to a being-becoming circularity worry similar to the kind I explore here. Determining an issue sequence or assigning extra weight to the views of designated experts within the group are themselves important decisions which the group faces and which may themselves depend on prior issues. If so, then in order to avoid an endless regress, the group must seemingly be already able to decide the issue of sequencing / expertise assignment before these procedures can be deployed. But then it looks like the purported mechanisms of collectivising rationality actually presuppose a group that is already collectively rational (and able to make decisions). This concern has been developed by O'Madagain (2012), but since the problem does not appear to afflict the procedure which Pettit truly favours – the straw vote – I am not pursuing it here.

[36]Pettit (2007: 511)

[37]Ibid., p. 495

[38]Ibid., p. 499

to what rationality (not just consistency) requires. For example, a group might vote on each proposition in a sequenced set, and, if their voting generates an inconsistency, they 'consider all the different possible ways in which previously formed attitudes or the new attitude could be revised so as to restore consistency… [and] take a vote… on which of the possible revisions to make'.[39]

When groups deploy a straw vote procedure they are clearly doing what Pettit would call reasoning. This is because they are concerned with the consistency of the judgements they have defeasibly adjudicated – standing ready, should inconsistency be discovered, to make the appropriate revisions *at any point within the overall body of judgment* (depending on what the evidence favours). In this way the straw vote allows the collective to be more robustly rational, since the collective is required to reflect on the overall rationality of their collective judgements.

Once again, this much can be readily accepted, since the degree to which rationality is or is not attained by a collective is not our primary concern. Our concern is with the process by which collective rationality is attained and, more specifically, with whom it is that institutes and handles that process. The suspicion we have been pursuing is that the process by which rationality is collectivised is a process which pre-requires a group agent, and so cannot be the process through which group agents emerge. This means that the question we need to ask about the sort of collective reasoning implicit in straw vote-type procedures is whether it is best seen as an achievement of the individuals in the collective or of an already-formed group agent.

Pettit's discussion of the reason/rationality distinction as it pertains to individuals suggests that when that distinction is transposed to the social level, it must be the group itself that reasons:

> 'The rationality of the simple creature is realised sub-personally so far as there is nothing *the creature* can do in order to improve its rational performance… We reasoning creatures transcend this limitation. We do not have to rely on the processing for which our nature programs in order to be rational. *We* can do something about it… The transcendence that reasoning achieves gives [individuals] a degree of personal control over [their] own rational performance'[40]

If the exercise of reason in individuals is a personal-level phenomenon it would seem that the exercise of reason in collectives must be a group-level phenomenon – and the control over its rational performance which such exercise strives for is something the group itself achieves. Though individuals participate in the straw vote procedure through which group reasoning is enacted, their participation is not agential in the sense required for this reasoning to be a plausible joint action; they are not the primary units of agency which direct this activity and which the activity itself serves. It is the rationality of the group itself which the activity seeks to reinforce and it is the agency of group itself which does the reinforcing. The contributions of individuals to this activity are not just subvenient but *subservient* to the rational agency of the group when it is instituted and sustained by means of reasoning. This means that when Pettit claims that group agents are not born but made through a process of group reasoning he introduces a puzzle about the genesis of group agents that he does not appear to have the resources to resolve.

[39]List & Pettit (2011: 63)

[40]Pettit (2007: 502, italics added)

3 Conclusion: Recognition and Group Agency

Pettit's account of the coming into being of group agents depicts this process as having two distinct stages. In the first stage we find that certain collectives, whose members share a common purpose, will eventually encounter situations of the sort exemplified by the Doctrinal Paradox – situations which pressurise them into taking steps to satisfy the demands of rationality at the collective level. This pressure can be seen as the realisation on the part of members of the collective that unless they make their collectively-held attitudes rational, they will find themselves unable to effectively pursue the purpose they all share. In the second stage, the members respond to this pressure by jointly acting to commit themselves to organisational mechanisms guaranteeing the rationality of their collective views; in this way, they make themselves into an autonomous group agent. I have sought to cast doubt especially upon the coherence of this second stage in the process: it does not seem plausible to me that individuals could jointly act so as to become, effectively, subpersonal machinery in a larger person.

If I am right then Pettit's account of the emergence of group agents is not compelling, or at least is not complete. In this final, concluding section I look beyond Pettit's own theory to see whether some assistance can be brought in from elsewhere; my tentative suggestion is that this can be found in what Pettit would likely consider the most unwelcome of places.

3.1 A Hegelian Dilemma

The question we have been pursuing – 'Who makes a group agent?', or 'Who is to collectivise reason?' – is strikingly reminiscent of a question Hegel asks in the *Philosophy of Right*. Reflecting on the potential unity of 'a people', Hegel asks 'Who is to frame a constitution?'[41] The question seems to pose the following dilemma:

> 'either the would-be framers are an atomistic collection of individuals, who thus lack the moral unity that a political constitution presupposes, or they already identify themselves as a people, a *Volk*, in which case they already exist as a constituted "moral unity," whose unwritten constitution exists in its *Volksgeist*'.[42]

The first option makes the framing of a constitution impossible, while the second makes it unnecessary. Our question, 'Who is to collectivise reason?' can be seen to create a similar dilemma: either the would-be collectivisers are an atomistic heap of individuals who lack the unity of purpose that the very process of collectivising reason requires (which makes the process impossible), or they are already a unified group that can think and act as an integrated whole (which makes the process unnecessary).[43] The various efforts made by Pettit and his collaborators to finesse the being-becoming circularity – especially the suggestion that joint action can account for the emergence of group agents – can thus be read as attempts to show up this dilemma as false. But, as we have seen, the jury (or the panel of judges, as the case may be) is still out on the success of these attempts.

How does Hegel resolve the dilemma posed by his question? The dilemma is false, he claims, since it presupposes that a constitution is something which can be *made*:

> 'it is absolutely essential that the constitution should not be regarded as something made, even though it has come into being in time. It must be treated rather as

[41] Hegel (1967: S273)

[42] Rehg (2011)

[43] It is worth emphasizing that though they bear striking *structural* resemblances to one another, Hegel's dilemma and Pettit's should not be thought of as relating to the very same issue. Hegel is considering a very different sort of unity (moral unity, as opposed to unity of purpose and activity) and a very different sort of social creature (a 'people', as opposed to an organised, purposive collective) from that with which Pettit is concerned.

> something simply existent in and by itself, as divine therefore, and constant, and so as exalted above the sphere of things that are made'.[44]

In short, the dilemma vanishes the moment we accept that a people is not something which comes to be 'constituted' by their own or anyone else's efforts; constitutions cannot be made but can only be made explicit or, perhaps, amended. It is senseless to try to institute a constitution since the only possible subjects of a constitution are those who are already constituted as a people ('sharing a Volksgeist'), and hence already the holders of a constitution (though it may be implicit). But if constitutions cannot be instituted by human effort then where do they come from? Hegel's suggestion is that they are 'divine' – that they are the product of what he elsewhere calls 'objective spirit'. What this means is that a people cannot bootstrap themselves into existence by instituting their own moral unity:

> 'we cannot be exercising an *instituting power* without taking ourselves to have been *instituted* as a "we." We cannot just institute ourselves into our instituting powers. Therefore, one has to acknowledge a primacy of… objective spirit over the subjective spirit'.[45]

3.2 A Hegelian Solution for Pettit

Is a structurally similar move available to Pettit? If he were to give up on the claim that groups are 'made, not born' and acquiesce to some analogue of the Hegelian account of the constitution of a people, would a route out *his* dilemma be forged? It certainly would, since it would allow him to say that the collectivisation of reason is something which simply happens – perhaps organically, perhaps through some or other 'divine' means – but not which something anybody (the group, or the members of a collective) has to *achieve*. And once the group is up and running ('born'), the sorts of mechanisms we have been discussing can then intelligibly feature as means by which the group sustains itself.

But what is equally certain is that Pettit would emphatically reject this sort of solution. Indeed (as noted in the introduction), one of what he considers to be the chief virtues of his account of group agency is that it can be sharply distinguished from the mysterious 'animation theory' about groups:

> 'the broadly Hegelian thought… [that] individual members give rise to a single agent only when a counterpart force comes on stream with its transforming effect'[46]

I do not wish to consider the prospects for foisting 'animation theory' upon Pettit. Instead I want to consider the extent to which a different 'broadly Hegelian' approach might be of some use to Pettit in bypassing his being-becoming trouble. The approach is inspired by Robert Brandom's[47] social pragmatist reading of Hegel's master concept of 'mutual recognition', and what I think it helps to show is how the emergence of group agents might be a more *social* process than Pettit allows. Specifically, it seems plausible to think that the recognitive attitudes of others in a wider discursive might have some role to play in the constitution of a group intentional subject.

We can begin with something which Pettit clearly accepts about the notion of a rational agent (or indeed a person): that it is a normative status. What this means is that being a rational agent (or person) implies that one is normatively constrained in certain ways and also that one

[44] Hegel (1967, S273)
[45] Descombes (2011: 388)
[46] List & Pettit (2011: 73-74)
[47] Brandom (2002, 2009)

is normatively entitled to adopt certain attitudes and take certain courses of action. But whom or what institutes these normative statuses? What makes these obligations and entitlements *genuine* or binding? According to Brandom, it is part of Kant's notion of the enlightenment ideal of 'autonomy' that we realise that no such normative status would be genuine were it to be imposed from without.[48]

The point can be illustrated by analogy with the chief development of enlightenment political philosophy – the idea of the social contract. Just as social contract theorists claimed that the legitimacy of political authority over citizens derives from those citizens' autonomous choice to *subject themselves* to that authority, so too do the norms which govern our thought and talk more generally only get a grip on us if they are somehow self-imposed. But this creates a puzzle, since it is hard to see how the norms of reason to which rational creatures are subject could be both self-imposed and genuinely binding: if we are the masters of our own boundedness then are we really *bound*? It is this puzzle – generated by Kant's dual insistence on normativity and autonomy – which Brandom sees Hegel as seeking to address via his notion of mutual recognition.

The solution wrought by Hegel involves the notion of a community:

> 'Hegel's principle innovation is his idea that in order to follow through on Kant's fundamental insight into the essentially *normative* character of mind, meaning, and rationality, we need to recognize that normative statuses such as authority and responsibility are at base *social* statuses'.[49]

The enlightenment ideal of autonomy entails that normative statuses are instituted by human attitudes, and for Brandom's Hegel, the key attitude is that of 'recognition' – the *treatment* of others as bearers of those statuses. What we need in order to actually be authoritative, or responsible, is to be recognised as such by others. But in seeking to be so recognised we are also recognising – we are according a certain kind of authority to those others we seek to be recognised by, namely, the authority to recognise us. According to Brandom's Hegel, normative statuses are only instituted when recognition is in this way symmetrical, or 'reciprocal'.

Reciprocal recognition thus provides not only a distinctively social account of what it is to *be* the bearer of certain normative statuses, it also provides a distinctively social account of what *becoming* the bearer of such statuses involves. Coming to be a good chess player, to borrow Brandom's own example, is a matter of coming to be recognised as such by those one recognises as such. Reciprocal recognition refers not just to an inter-subjective attitudinal *pattern* but to a certain sort of inter-subjective *process*:

> 'Achieving the status of being a good chess player is not something I can simply do by coming subjectively to adopt a certain attitude towards myself… It is up to me whom I recognise as good chess players, in the sense in which I aspire to be one. But it is not then in the same way up to me whether I qualify as one of them… To earn their recognition in turn I must be able to play up to their standards… My recognitive attitudes can define a virtual community, but only the reciprocal recognition by those I recognise can make me actually a member of it, accord me the status for which I have implicitly petitioned by recognising them'[50]

Could something like this model of reciprocal recognition in principle be exploited by Pettit to overcome the problem of how group agents emerge? I think it could. For one thing it

[48]Brandom (2009: Ch 2)
[49]Ibid., p. 66
[50]Ibid., pp. 70-71

need not involve any of the metaphysical mysteriousness Pettit associates with Hegelianism: for Brandom's Hegel, *Geist* or 'spirit' is nothing other than the 'realm of discursive activity' which is 'synthesized by reciprocal recognition'.[51] Perhaps then we can see reciprocal recognition as achieved in part by the joint action of members in a collective: group members can jointly act to petition a 'virtual community' for recognition as an intentional subject. If they succeed – if the community defined by their joint act of petitioning reciprocates by recognising them as an intentional subject – then their status as such is secured. Pettit's individualistic, 'performative' approach to personhood[52] obscures this possibility, making it seem as if intentional subjects could only be self-made or born, and this is what lands him with the bootstrapping difficulty I think he struggles to overcome. Group agents cannot bootstrap themselves into existence but they can take themselves to be a group agent and strive for this status to be recognised by a community which their own attitudes have delineated. If there is something to be said for this proposal then perhaps we can conclude that group agents are neither born nor made, though it may take a village to raise one.

References

Brandom, R. (2002). *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*. Cambridge: Harvard University Press.

Brandom, R. (2009). *Reason in Philosophy: Animating Ideas*. Cambridge: Harvard University Press.

Bratman, M. (1999). *Faces of Intention*. Cambridge: Cambridge University Press.

Descombes, V. (2011). 'The Problem of Collective Identity: The Instituting We and the Instituted We' in Laitinen, A & Ikaheimo, H. (eds). *Recognition and Social Ontology*. Leiden: Brill.

Garland, D. (2001). *The Culture of Control*. Chicago: University of Chicago Press.

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

Hegel, G.W.F. (1967). *Philosophy of Right*, trans. T. M. Knox, New York, Oxford University Press

Kornhauser, L. A. & L. G. Sager. (1993). 'The one and the many: adjudication in collegial courts'. California Law Review, 81:1–59

List, C & Pettit, P. (2002). 'Aggregating Sets of Judgements: An Impossibility Result,' *Economics and Philosophy* 18: 89–110

List, C & Pettit, P. (2006). 'Group Agency and Supervenience', in *Southern Journal of Philosophy* (Spindel Conference 2005), 44: 85–105

List, C & Pettit, P. (2011). *Group Agency*. Oxford: Oxford University Press.

McGeer, V. (2002). 'Developing Trust', *Philosophical Explorations* 5:1 (2002) 21-38

O'Madagain, C. (2012). 'Group Agents: Persons, Mobs or Zombies', *International Journal of Philosophical Studies* 20: 2. pp 271-287.

Pettit, P. (1995). 'The Cunning of Trust,' *Philosophy and Public Affairs*, 24: 202–225

Pettit, P. (2002). 'Collective Persons and Powers', *Legal Theory,* 8 (2002), 443–470

Pettit, P. (2003). 'Groups with Minds of Their Own', *Socializing Metaphysics: the Nature of Social Reality*, F. Schmitt, ed., Lanham, MD: Rowman & Littlefield: 167–93.

Pettit, P. (2007). 'Rationality, Reasoning and Group Agency,' *Dialectica*, Vol 61, 2007, 495-519

[51] Brandom (2009: 72)

[52] List & Pettit (2011: 171-2)

Pettit, P. (2009). 'The Reality of Group Agents', in *Philosophy of the social sciences : philosophical theory and scientific practice*, C. Mantzavinos, ed., Cambridge, UK; New York : Cambridge University Press: 67–91.

Pettit, P and Schweikard, D. (2006). 'Joint Actions and Group Agents', *Philosophy of Social Sciences*, 36: 18–39

Quinton, Anthony. (1976). 'Social Objects', *Proceedings of the Aristotelian Society* 75: 1–27

Rehg, W. (2011). Review of 'Recognition and Social Ontology'. *Notre Dame Philosophical Reviews*. Online: http://ndpr.nd.edu/news/27905-recognition-and-social-ontology/

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 55–71, 2013

Mental Causation and the New Compatibilism

Jens Harbecke

Philosophy of Science
Witten/Herdecke University, Germany
jens.harbecke@uni-wh.de
www.jensharbecke.com

Abstract

Twenty years ago Stephen Yablo developed his original theory of mental causation, which has drawn much attention ever since. By providing a detailed reconstruction of Yablo's approach, this paper first demonstrates that a certain line of critique that has repeatedly been brought forward against Yablo over the last two decades misconstrues the core idea of the model. At the same time, the reconstruction reveals that Yablo's approach is probably the first explicit version of the "new compatibilism" within the philosophy of mind. This fairly young family of theories essentially claims a non-identity as well as a non-distinctness of mental and physical phenomena. The second part of the paper then demonstrates that the new compatibilist approaches in general, and Yablo's theory in particular, even though they can resist much of the popular criticism, fall prey to a new theoretical trilemma once the nature of the respective *analysantia* is made explicit. Finally, a model of the psychophysical relation is developed that may allow the new compatibilists to escape the mentioned trilemma.

1 Introduction

Twenty years ago Stephen Yablo (1992b; 1992a; 1997; 2000; 2001) developed his original theory of mental causation, which has drawn much attention ever since.[1] Yablo's most cited article on the topic (1992b) first develops a distinction between determinable and determinate properties as it is known from traditional classifications of colors and geometrical forms. For instance, "crimson" and "bordeaux" are sometimes described as determinates of the determinable "red"; "being rectangular" and "being square" are considered determinates of the determinable "being quadrangular".

In a second step, Yablo suggests that "(...) mental properties stand to their physical realizations in the relation that quadrangularity bears to squareness, or that colors bear to their shades (...)." (1992b, 256) In Yablo's view, this insight holds the key for a convincing solution to the problem of mental causation. Determinable and determinate properties are in no causal competition with each other. This is true even in situations, in which instances of both properties are already causally sufficient for a corresponding effect. Consequently, the widely discussed "causal exclusion principle" (cf. premises (4), (4*) and (4**) in sec. 2) cannot be applied to cases where the two candidate causes are related by determination.

According to a popular interpretation of Yablo, the following argument nicely captures the line of thought just sketched and is therefore at the heart of the author's approach:

[1] A check on GoogleScholar in March 2013 mentions 440 quotations of Yablo (1992b).

	(I)	All properties that are related by determination do not compete for causal influence.
	(II)	All mental properties are related by determination to their subvenient physical properties.
∴	(III)	No mental property competes with any of its subvenient physical properties for causal influence.

Whilst the validity of this argument is not in dispute[2], the two premises have been debated in various ways over the last twenty years. Some authors have challenged the claim that determinates and determinables do not compete for causal influence (cf. McGrath 1998, 171; Gillett and Rives 2005, 496; Walter 2007, 238-240), which in turn has motivated others to defend this claim in more detail (cf. Shoemaker 2001, 432; Wilson 2009, 166-168). Again other authors have denied that the relation between mental and physical properties is that of determination (cf. Ehring 1996, 471-476; Worley 1997, 283/284, McGrath 1998, 170; Funkhouser 2006, 562/563; Walter 2007, 221; Haug 2009, 7-8), to which again others have responded with a defense of Yablo (Wilson 2009, 162-165).

One of the central aims of the present paper is to show that a good portion of the debate represented by these authors actually misconstrues the core idea of Yablo's model. In contrast to the popular opinion, it is not the case that Yablo's theory is based on an argument of the form (I)-(III). Rather, the notion of classical determination plays a merely heuristic role for Yablo. The real work in the theory is done by the essentialist analysis of events developed in Yablo 1992a (407-425, 430-436), 1992b (261-265), and most detailed in his 1987. Consequently, the mentioned criticisms of premises (I) and (II) are of little relevance for Yablo's approach. After two decades of debate on the theory, this point requires emphasis.

The second central aim of this paper is to support the claim that Yablo's theory is a variation of what can be described as a "new compatibilism" within the philosophy of mind, i. e. an approach that assumes both a *non-identity* and a *non-distinctness* of mental and physical phenomena. Depending on various possible *analysanda* presupposed, this approach has received several different formulations by different authors.

The third central aim is to show that the new compatibilism in general, and Yablo's theory in particular, faces a new fundamental trilemma. All existing versions of the new compatibilism either collapse into a strong dualism, a strong monism, or an unsatisfactory neutralism with respect to their *analysantia* if they want to maintain the contention about the non-identity and non-distinctness of the *analysanda*. Since, in light of this trilemma, the new compatibilism is an unsatisfactory approach at this point, a forth hypothesis to be defended investigates a potential way out for the new compatibilists.

The structure of the paper is the following. Sec. 2 briefly sketches the problem of mental causation that Yablo and the new compatibilists aim to solve. Sec. 3 isolates the reasons why Yablo's argument has often been interpreted as equivalent to (I)-(III). Moreover, some popular objections against premises (I) and (II) are reconstructed. Secs. 4 and 5 focus on the actual dialectics of Yablo and show that the generally valid objections against (I) and (II) do not affect the core of Yablo's approach.

Section 6 relates Yablo's version to other variations of the new compatibilism. The mentioned trilemma is highlighted in 7, before a potential solution is explicated in section 8.

[2]A rigorous formalization would be: $\forall\phi\forall\psi(\mathcal{D}\psi\phi \rightarrow \neg\mathcal{C}\phi\psi)$, $\forall\phi\forall\psi(\mathcal{M}\phi\wedge\mathcal{P}\psi\wedge\mathcal{S}\phi\psi \rightarrow \mathcal{D}\psi\phi) \therefore \forall\phi\forall\psi(\mathcal{M}\phi\wedge\mathcal{P}\psi\wedge\mathcal{S}\phi\psi \rightarrow \neg\mathcal{C}\phi\psi)$, where $\mathcal{D}(\_,\_)$ = . . .determines. . ., $\mathcal{C}(\_,\_)$ = . . .causally competes with. . ., $\mathcal{S}(\_,\_)$ = . . .supervenes on. . ., $\mathcal{M}(\_)$ = . . .is a mental property, $\mathcal{P}(\_)$ = . . .is a physical property.

2 The problem of mental causation

This section briefly reviews the so-called "problem of mental causation". The problem is usually considered as consisting of four jointly incompatible assumptions. Two general formulations of these assumptions are the following, both of which appear frequently in the literature. The first formulation can be termed the "token formulation".

(1*)	Some mental events have physical effects.
(2*)	No mental event is identical to any physical event.
(3*)	All physical events have a sufficient physical cause.
(4*)	Physical events are not systematically causally overdetermined.

The second version can be described as the "type-" or "property formulation".

(1**)	Some mental properties are causally sufficient for some physical properties.
(2**)	No mental property is identical to any physical property.
(3**)	For every instance x of a physical property P_1, there is an instance y of a further physical property P_2, such that y's being an instance of P_2 is causally sufficient for x's being an instance of P_1.
(4**)	Instances of physical properties are not systematically causally overdetermined.

Each assumption of the formulations seems to be defensible in isolation. In particular, mental events and properties play an indispensable role when we interpret actions of agents. Assuming that mental events never cause anything would amount to the contention that virtually all our interpretive practices are pointless. The inacceptability of this consequence suggests (1*)/(1**). On the other hand, as empirical science seems to tell us, any physical event has a complete sufficient physical cause, suggesting the truth of (3*)/(3**).[3]

The conjunction of assumptions (1*)/(1**) and (3*)/(3**) implies that mental events must (i) either be redundant causes of physical events or (ii) they must themselves be physical events. The latter option has turned out difficult to defend at least with respect to the type formulation. To be sure, a quasi consensus exists in philosophy of mind that the mind depends on the physical. One way to express such a hypothesis is by the notion of supervenience:

(S)	Necessarily, for all x and all mental properties M, if x instantiates M, then there is also a physical property P such that x instantiates P and, necessarily, all individuals instantiating P also instantiate M.

The problem is that a supervenience relation between sets of properties is not sufficient for an identity of mental to physical properties. In fact, it has turned out difficult to identify a single paradigmatic mental property that is coextensional with some (natural and non-disjunctive) physical property (cf. Putnam 1967). This insight can be expressed by the following hypothesis:

(M)	For all mental properties M and all physical properties P, if P is sufficient for M, there is always at least one x that is an M but not a P.

Since coextensionality is necessary for identity, (M) implies (2**) (and the falsity of what we described as "option (ii)"; i. e. the claim that mental events are identical to physical events). Moreover, if a fine-grained model of events is presupposed (cf. Kim 1973), (M) also implies (2*). The reason is that, according to this model, for two events to be identical it is necessary for the properties of the events be identical.

[3] It should be noted, however, that the doctrine of the causal completeness of the physical is less obvious than its aquiescent acceptance in the philosophical literature suggests (cf. Montero 2003; Primas 2007). Bishop and Atmanspacher (2010) have argued that the notion of causation is inconsistent with the fundamental laws of physics insofar as these laws have no direction of time, hence no past and future, hence no cause and effect.

This leaves only the first of the two options (i) and (ii): Mental events, i. e. instances of mental properties, can only be redundant causes of physical events, i. e. instances of physical properties. This claim, however, contradicts assumption (4*)/(4**), so that the full set of assumptions (1*)/(1**)–(4*)/(4**) is inconsistent. The reason for holding on to (4*)/(4**) is that redundant causes *per definitionem* do not make any difference to what goes on in the world, and it seems extravagant to assume the existence of things that make no causal difference whatsoever.

From the inconsistency it follows that any feasible theory of mental causation has to reject at least one of the assumptions, where this rejection will have to be defended as demanding the lowest overall theoretical costs. In the 1970s, psychophysical functionalism was widely considered a dominant strategy in this respect (cf. Putnam 1967, 1975; Fodor 1974). Functionalism accepted that mental events overdetermine effects with physical causes. After a while, however, and mainly through the work of Jaegwon Kim (1979, 1985) it became clear that functionalism brings much higher costs than initially expected. Kim's work led many authors to embrace an identity theory and, thereby, to reject premise (2*)/(2**). In the late 1990's however, several authors have shown that Kim's version of an identity theory is not conservative and ultimately eliminates the mental (cf. for instance Bontly 2002 and Harbecke 2008). This consequence has struck many philosophers as unacceptable as well.

Yablo and authors who below will be described as the "new compatibilists" (sec. 6) have sought to evade this general dilemma between functionalism and the identity theory. The next section focuses on Yablo's proposal for an analysis that resists both option i) and option ii).

3 Classical Determination

As mentioned in sec. 1, a popular line of thought interprets Yablo's approach as essentially based on argument (I)-(III). The reason is that the notion of "determination" plays a central role in Yablo's presentations of his theory.

In particular, Yablo develops a thought experiment in his 1992b (257) and 1992a (423) that helps to formulate a *reductio* proof against premise (4**) and, thereby, for premise (I). In this thought experiment, a pigeon "Sophie" is conditioned to peck on red triangles. However, the triangles presented to Sophie are not only red, they are scarlet. If the causal exclusion principle applies generally, it may seem that the redness of the triangle is a different and competing cause of Sophie's pecking than the triangle's scarletness. But, of course, that's absurd. Yablo concludes that "(...) the [exclusion] principle does not apply to determinates and their determinables – for we know that they are not causal rivals." (1992b, 259)[4] Hence, Yablo accepts premise (I).

Premise (II) Yablo considers established on the basis of the widely accepted thesis of psychophysical supervenience (S) and the thesis of a multiple realization of mental properties (M) (cf. sec. 2 and Yablo 1992b, 254/255). Yablo defines determination as follows:

(D)	A property P determines a property Q just in case "for a thing to be P is for it to be Q, not *simpliciter*, but in a specific way." (1992b, 252)

In Yablo's understanding, (D) is basically implied by the conjunction of (S) and (M). To the rhetorical question whether mental properties are in fact determinables of certain physical properties just as colors are determinables of their shades, Yablo answers: "Yes, at least that is my conjecture." (1992b, 256) Hence, it seems that Yablo also subscribes to premise (II).

[4] Already MacDonald and MacDonald 1986 had made this observation. However, in contrast to Yablo, the MacDonalds presuppose the identity of instances of determinables and determinates.

As mentioned above, various authors have challenged premises (I) and (II). For instance, Gillett and Rives (2005, 491) and Walter (2007, 239) have argued that Yablo's example of the pigeon Sophie shows something important about potentially competing explanations, but little to nothing about ontology. Metaphysically speaking, the exclusion principle applies without exception: Determinable properties are screened-off from their effects by their determinate properties. Hence, if read metaphysically, premise (I) is false (cf. also McGrath 1998, 171).

Premise (II) has been attacked by arguments of the following form (cf. Ehring 1996, 471-476; Worley 1997, 283/284, McGrath 1998, 170; Funkhouser 2006, 562/563; Walter 2007, 221; Haug 2009, 7-8):

	(IV)	All determination relations satisfy the higher-order properties in set $\mathfrak{D}$.
	(V)	The relation that mental properties bear to their subvenient physical properties does not satisfy the higher-order properties in set $\mathfrak{D}$.
$\therefore$	(VI)	*Not*: The relation that mental properties bear to their subvenient physical properties is that of determination.

Ever since William Ernest Johnson (1921, 174) introduced the distinction of determinates and determinables into philosophy, several authors have aimed to delineate precisely set $\mathfrak{D}$.[5] From the results of these investigations, it must indeed be concluded that the relation between mental properties and their physical supervenience bases is not that of determination.

As an example, consider the fact that all determinates falling under the same determinable are incompatible with another: no object can be scarlet and bordeaux at the same place at the same time. This incompatibility principle does not seem to apply to the supervenience bases of mental properties. It seems possible in principle that a person can instantiate two neural supervenience bases of the same mental state at the same time. With arguments of this kind, premises (IV) and (V) seem to be confirmed, and therefore the truth of (VI). However, (VI) is just the negation of (II), suggesting that the argument (I)-(III) is not sound.

4 Metaphysical Determination

Although the arguments against premises (I) and (II) presented in sec. 3 have probative force, they turn out to be of little importance for Yablo's overall theory. The reason is that in his version of argument (I)-(III) Yablo refers to a different notion of determination than the one presupposed by the argument (IV)-(VI). Morever, Yablo's theory is not primarily concerned with the causal connection between mental and physical properties, but with mental and physical events.

Yablo's definition of determination (D) seems to match the criterion of "asymmetric sufficiency" that had been identified by Johnson (cf. also Ehring 1996, 470). However, whilst for classical determination this sufficiency describes a relation between concepts, Yablo intends it as a metaphysical relationship as the following example shows (1992b, 253). Assume that for the property of having a temperature of 95° Celsius, there exists a maximally specific microphysical property K such that everything that has K also has a temperature of 95° Celsius, but not vice versa. According to Yablo, the relation between these two properties is a prototypical determination relation. However, no conceptual analysis will reveal that K is sufficient for having a temperature of 95° Celsius.

> "[And] since it is only the metaphysics that matters to causation, we should discount the traditional doctrine's conceptual component and reconceive determination in wholly metaphysical terms. " (1992b, 253)

[5] For reviews of the results of these investigations, cf. Harbecke 2008, 183-191; Funkhouser 2006, and Sanford 2008.

This shows that Yablo considers the metaphysical determination relation with which he is primarily concerned as distinct from Johnson's classical determination relation. Consequently, Yablo can accept premises (IV) and (V), which refer to classical determination (see also Yablo 1997, fn. 22). At the same time, he can accept the truth of premise (II), since conclusion (VI) does in fact contradict (II). In short, argument (IV)-(VI) poses no threat to Yablo's position.

Another argument promoted by Gillett and Rives (2005, 491) and Walter (2007, 239) attacks premise (I) (cf. 3). The authors claim that the exclusion principle (4**) applies in the metaphysical sense to properties connected by determination, even if two explanations referring to a determinate and its determinable property may well be accepted out of explanatory purposes at the same time.

One reason[6] why Yablo's analysis is not affected by this argument either is that the author is clearly not concerned with a second-order relation between properties but a first-order relation between events (cf. Yablo 1992a, 407-425, 430-436, Yablo 1992b, 260-280, Yablo 2001, 66/67). That is, Yablo is concerned not with problem (1**)-(4**), but with problem (1*)-(4*). As a consequence, whether the exclusion principle (4**) applies to properties related by determination is not a question that Yablo's theory has to concern itself with. At least more argument would be required to show that Yablo has to confront this question.

This suggests that the two kinds of criticism often brought forward against Yablo misunderstand the core aim and content of the author's theory. As a consequence, the original theory is left widely unharmed by the critical arguments, even if these are of great importance in themselves. At the same time, this insight of the immunity of Yablo's theory against the popular objections does not yet answer questions about the positive content of this theory. In particular, it is not clear at this point why Yablo can accept the truth of premise (4**) and still reject premise (4*). It has also not been clarified what role argument (I)-(III) ultimately plays for Yablo's theory. To provide answers to these questions is the aim of the following section.

5 Cumulative Subsumption

According to Yablo, the key for solving problem (1*)-(4*) is the insight that mental and physical events stand in a first-order determination relation. The author's definition of this relation is analogous to (D):

(d)	"[An event] p determines [an event] q iff: for p to occur (in a possible world) is for q to occur (there), not simpliciter, but in a certain way." (1992b, 260)

The central challenge for Yablo now consists in showing that (i) this definition is interpretable in a metaphysical way, (ii) that events related by (d)-determination do not compete for causal influence, and (iii) that mental and (the corresponding) physical events are in fact related in this way. Secs. 5.1, 5.2, and 5.3 discuss these challenges consecutively. For what follows, Yablo's notion of an "individual essence" is central, which he analyzes in terms a "cumulative subsumption of categorical properties".[7] Both notions are presented by Yablo in his 1992b (sec. 5), and most extensively in his 1987 and 1992a (sec. 2).

[6] A second reason for rejecting arguments similar to those of Gillet & Rives and Walter has been developed by Bennett (2003). Bennett shows that there is a disanalogy between prototypical cases of overdetermination and cases in which one of the causing properties is sufficient for the other.

[7] In 1997 Yablo uses the notion of an "inclusion of influence" and in his 2000 and 2001 that of "intensive parthood"; however, it is clear that the author intends these synonymously as the originally developed notion of "subsumption/refinement/strengthening" explicated in 1992b, 1987, and 1992a.

5.1 Definition

According to the ordinary understanding, the essence of a thing is the set of properties without which the thing cannot exist. Not all essential properties are contained in a thing's essence, however. Identity properties such as "being identical to x" or type-properties such as "being of the same type as x" are determined by the essence of a thing, but they do not themselves contribute to the way a thing is.

Essences also provide a set of conditions that must be satisfied to be a certain thing. For many things x and y, more conditions must be satisfied to be x than to be y. For instance, the *Rosetta Stone* should be distinguished from the *Rosetta Granite*, which has just the same properties as the *Rosetta Stone* but lacks the properties of having been discovered by Pierre-Francois Bouchard and of having provided the key for the deciphering of hieroglyphic writing. Similarly for events, the blocking of the valve should be distinguished from the sudden blocking of the valve, not the least because the two may have different causal effects.

In Yablo's view, the inclusion relation between essences of events helps to make sense of (d). If q's essence is included in p's essence, then p "subsumes" q ($p \geq q$). Event p "determines" event q ($p > q$) if the inclusion is proper (1992b, 262).[8] Since only categorical properties can be contained in essences, something must be said about what distinguishes categorical from hypothetical properties. Yablo makes the following first attempt to launch the distinction:

> "[A] property is *categorical* if its possession by a thing x at a possible world is strictly a matter of x's condition in that world, without regard to how it would or could have been; other properties, for example counterfactual and modal properties, are *hypothetical*." (Yablo 1992b, 261/262)

Despite its intuitive appeal, the definition turns out to be circular. The problem is that it ultimately describes a property as categorical if it is possessed by thing x, without regard to how x would or could have been in other possible words *in categorical respects*. This makes the circularity obvious. To solve this problem, Yablo uses the inclusion relation itself to distinguish categorical from hypothetical properties.

(K)	"[A property] C is categorical only if: necessarily, for all p and q such that $p \geq q$, p has C iff q does." (Yablo 1992b, 263)

This definition excludes all identity properties from the set of categorical properties. Moreover, with the set of cumulative properties it is now possible to explain the content of (d). If for p to occur (in a possible world) is for q to occur (there), not simpliciter, but in a certain way, then the essence of p subsumes that of q.

Applied to events, this idea implies that, if p determines q, then q occurs in more worlds than p. Suppose, for example, that p and q are coincident events that instantiate property set $\{F, G, H, I\}$ in the actual world. Suppose further that $\mathcal{P} := \{F, G, H\}$ is the essence of p and that $\mathcal{Q} := \{F, G\}$ is the essence of q. Event q now can occur in some possible worlds in which p does not occur, because its essence is poorer, whilst the opposite is false.

5.2 Causal competition

With the example of two events p and q developed in the last section, it becomes clear why Yablo rejects the claim of a causal competition between events related by determination. If p and q have essences $\mathcal{P}$ and $\mathcal{Q}$, then it is impossible for p to screen off q and vice versa, since this would amount to p (q) screening off itself.

[8] What Yablo 1992b calls "subsumption" what is described as "refinement" in his 1987 and as "strengthening" in his 1992a. The intended meaning is the same, however. Cf. fn. 7 also.

Despite the fact that the coincident events p and q do not screen each other off, they are not identical, since q occurs in some possible worlds in which p does not occur. With these points in place, Yablo can show how two non-identical events that both have influence on the same physical effect, need not be in causal competition with each other in the sense of a "causal overdetermination" (cf. premise (4*)).

5.3 Mental Determination

The solution for problem (1*)-(4*) now seems straightforward. The author assumes that mental events are typically determined by physical events (note that the two claims (s) and (m) are analogous to (S) and (M)):

(s) "Whenever a mental event m occurs, there occurs also a subsuming physical event p, that is, a physical event whose essence includes m's essence." (Yablo 1992b, 268)

(m) "For every mental event m, and every physical event p which subsumes m, p subsumes m properly and so determines it." (1992b, 270)

If mental and physical events are in fact related in this way, then they are not identical. But as sec. 5.2 showed, they can also not compete for causal influence. With this conclusion, problem (1*)-(4*) is solved.[9]

The general picture that emerges from this hypothesis is that mental events necessarily coincide with physical events, i. e. each mental event always instantiates the same categorical properties as some physical event. In this sense, mental events are *not distinct* from their subvenient physical events. However, since mental events are to be distinguished from physical events in hypothetical respects, they are also *not identical* to them. Mental events have "impoverished" essences relative to physical events.

5.4 Evaluation

The main claims of Yablo's theory are now clearer and it is more obvious why it is not threatened by premise (4*). This result complements the claims made in sec. 4, according to which Yablo can accept the conclusiveness of argument (IV)-(VI) and he is not affected by the truth of premise (4**).

However, if determination in terms of cumulative subsumption is the relation that does the main work for Yablo's theory, it may seem confusing that Yablo spends much time discussing examples of classical determination and an argument similar to (I)-(III) especially in 1992b (252-260). The reasons for this dialectical step become clearer once the whole of Yablos theory as represented by 1987; 1992a; 1992b; 1997; 2000, and 2001 is taken into account. Yablo's reply to the question whether "(...) mental properties stand to their physical realizations in the relation that rectangularity bears to squareness, or that colors bear to their shades?" (1992b (256) is intended *heuristically* and not in the sense of a factual equivalence. The goal of introducing the "Sophie" thought experiment is merely to show that there is at least one second-order relation whose *relata* do not compete causally with respect to further events. This is then taken as a reason to search for a first-order relation with the same features.[10]

[9] A question that cannot be answered here is whether Yablo actually rejects a premise of the argument or whether he assumes an ambiguity in the premises. It may seem at first that the author rejects premise (4*) thereby claiming an overdetermination of physical events. The definition of causal proportionality (cf. 1992b, 273-279) suggests, however, that Yayblo accepts an "overdetermination" of causal influence but rejects an overdetermination in terms of genuine causality. For a discussion of this distinction, cf. Harbecke 2008, 300-309.

[10] Already in the introduction to his 1992b, Yablo is very cautious in asking: "What if mental phenomena are determinables of physikal phenomena in *something like* the traditional sense (...)?" (250, emphasis added) This makes it

6 A new compatibilism

By describing mental and physical events as *non-identical* and *non-distinct*, Yablo initiatied a theoretical tradition that can be described as a "new compatibilism" in the philosophy of mind.[11] It is a "compatiblism", because this theory claims a compatibility of mental causation with a psycho-physical non-identity. It is "new" because it should be carefully distinguished from older compatibilist approaches such as functionalism à la Fodor and Putnam, a pre-established harmony à la Leibniz, or a Cartesian dualism.

What unites several theoretical approaches under this term is the fundamental notion of an inclusion, or containment, relation between certain entities. For the notion of an inclusion, recall some simple definitions of set theory. Two sets M and N are identical iff: $M \subseteq N$ and $N \subseteq M$. Two sets M and N are distinct iff: $M \cap N = \emptyset$. If $M \subset N$, then M and N are neither identical nor distinct. The new compatibilists use this simple idea in one way or other to describe the relation between mental and physical phenomena.

In Yablo's theory, this idea occurs in the definition of a proper subsumption of *event* essences. Interestingly, Kim in his 1993 claimed that events such as Sebastian's stroll and his leisurely stroll are "different, if not entirely distinct, events. Not entirely distinct since the latter *includes* the former." (45)

Several further authors have argued analogously for the claim that mental *properties* are neither identical nor distinct from physical properties. According to this idea, the set of causal powers constitutive of a mental property is always contained in the set of causal powers constitutive of some physical property (cf. Worley 1997; Wilson 1999, 2009; Shoemaker 2001; Clapp 2001; Ehring 2003).[12] According to the causal theory of properties, every natural property is associated with a set of conditional powers (cf. Shoemaker 1980, 212/13). A property P equips objects, that instantiate P with a conditional power $K(U, E)$ if instances of P in circumstances U bring about instances of E, whilst realizations of U in isolation do not bring about instances of E. Powers whose circumstantial condition U is empty are called "powers simpliciter".

In some cases, the conditional powers that are bestowed by a property are a subset of a further property that are bestowed by a second property. According to authors such as Shoemaker (2001), Clapp (2001), Ehring (2003) and Wilson (1999, 2009), the relation between mental and subvenient physical properties should be understood in this way: the powers associated with any mental property are a subset of powers associated with some physical property. Since the two properties are not identical, but also not distinct, this theory should be viewed as versions of the new compatibilism.

Another approach claims that mental phenomena are "wholly constituted" by physical phenomena, without the constitution relation implying identity (Pereboom and Kornblith 1991; Pereboom 2002; Gillett 2003). More concretely, it proposes that "(...) mental causal powers are wholly constituted of physical causal powers; they are neither identical to (nor are they necessary and sufficient for) them, nor wholly independent of them. That's why

clear that the example of classical determinables and determinates plays merely a heuristic role. In his 1992a, the author even develops the theory of cumulative subsumption first, and then discusses the notion of determination.

[11] To my knowledge, the term "new compatiblism" is not in use so far in the philosophy of mind. I adapt the term "compatibilsm" from Terence Horgan, who uses the notion of "causal compatibilism" to describe the hypothesis that "(...) mental causation via nonphysical properties can co-exist with physical causation even if the physical realm is causally closed (...)." (Horgan 1997, 166) See also Karen Bennett, who uses the same term to describe the hypothesis that "(...) no effect can have more than one sufficient cause unless it is overdetermined." (Bennett 2003, 473) The way "compatibilism" is used in these contexts is to be distinguished from its use in the debate on free will.

[12] The original idea of an inclusion of sets of causal powers definitional of properties is already found in Fales 1982 and 1990, 239-243. However, Fales did not intend this idea as a theory of mental causation.

they don't compete." (1991, 143) This "constitutionalist" version of the new compatibilism presupposes mental causal *powers* as the *analysandum*.[13] Mental powers are taken to be wholly constituted by physical powers, which reflects the inclusion idea and which seems to rule out a distinctness of the two. Nevertheless, the former are not identical to the latter, as they can be multiply constituted (cf. *op. cit.* 138). It follows that mental powers do not compete with causal powers and both have a place in the causal nexus (cf. *op. cit.* 143).

A more recent approach developed by Dardis (2008) is based on the claim that instantiations of mental properties as "form properties" constitute mental events always in conjunction with physico-material properties. According to Dardis, every mental event necessarily instantiates a complex material property P_M. This is the property of being of the kind of matter that underlies the mental event. This material property, i. e. typically the property of being made of neurons, is not sufficient for the corresponding mental state, however. Additionally, a form property is required that determines a certain structure of the matter (cf. *op. cit.* 158-161). The form property is a mental property M_F. "Mental properties dovetail with matter properties custom-made to supplement their work." (173). Only the synthesis of P_M and M_F makes a corresponding mental event.

Since Dardis accepts a psychophysical supervenience hypothesis (cf. *op. cit.*, 135/136), any mental event m has to be coinstantiated with a physical event p. Such a physical event presumably instantiates also a matter property P_M, but additionally a physical form property P_F. Since m and p are different with respect to M_F, resp. P_F, they cannot be identical. However, since they share the instantiation of P_M they are not distinct and they cannot compete causally in a strong sense. This again reflects the inclusion idea and it shows that Dardis' approach is a version of the new compatibilism.

7 A new trilemma

The new compatibilism can escape the general dialectical dilemma that was diagnosed in sec. 2 for the debate on mental causation. It can embrace premise (2*) or (2**) respectively. At the same time, the new compatibilists typically disarm exclusion principles such as (4*) and (4**) by showing that the relevant mental and physical phenomena are not distinct. If mental and physical phenomena are not distinct, then they cannot causally overdetermine physical effects in the strong sense of (4*) and (4**). Rather, they overdetermine physical effects in an unproblematic way. Consequently, (1*)-(4*) and (1**)-(4**) can all be true and the new compatilists are able to avoid the classical criticisms against both reductive and non-reductive theories of mental causation (cf. sec. 2).[14] This makes the new compatibilism a serious candidate for a theory of mental causation.

Nevertheless, the following arguments show that the new compatibilism faces a new kind of trilemma which applies to the different versions of the theory in different ways. For instance, recall that in Yablo's proposal, the essence $\mathcal{M}$ of any mental event m is a subset of an essence $\mathcal{P}$ of some physical event p. Both essences $\mathcal{M}$ and $\mathcal{P}$ contain only categorical properties. The question is whether these categorical properties can be characterized more specifically.

[13] In the original text of Pereboom and Kornblith, it is not always clear what the *analysandum* is supposed to be. In another paragraph, the authors discuss mental states (1991, 134): "[W]e hold that token mental states are physically constituted, but not identical to, the token physical states which constitute them." However, this does not seem to change anything about the fundamental idea of the constitutionalist theory.

[14] Just as for Yablo, it is not immediately clear whether the new compatibilists in general reject premise (4*) or (4**), or whether they claim that the two formulations of the problem of mental causation contain an ambiguitiy. Both interpretations are possible.

If all the categorical properties contained in $\mathcal{M}$ are mental properties (if there be such), then the essence $\mathcal{P}$ of the physical event p contains as a subset at least some genuinely mental properties. This, however, questions the physical character of p, for how can p be physical if it is partly essentially mental? Secondly, this reading of Yablo's theory would seem to postulate a strong property dualism that claims an independent existence of mental and physical properties. With such a strong property dualism it may no longer be clear that Yablo can actually accept the completeness premise (3*). This would only be possible if he also accepted that the instantiation of property set $\mathcal{P} \setminus \mathcal{M}$ is already sufficient to bring about the physical effect of p. In this case, the subset of $\mathcal{P}$ that is equivalent to $\mathcal{M}$ would be redundant for p's causal role with respect to its physical effects. This redundancy once again calls into question whether those aspects of p that are mental have anything to do with the causal effects of p.

On the other hand, if the categorical properties in $\mathcal{M}$ are physical properties exclusively, then it is unclear why m should be considered a mental event at all and not just another, and less comprehensive, physical event p^*. If all of the properties contained in the essence of m are physical, there is nothing particularly mental about m any more. It is clear that, in this way, a causal exclusion of mental events by physical events is avoided. However, at the same time, mental causation is explained away, rather than explained. The nature of the cause is completely physical, and mental causation has given way to physical causation.

Finally, if the properties contained in both p and m should be considered neither mental nor physical, but perhaps "neutral", a similar problem occurs. Now the mental event m is completely made up of neutral properties, and so is the physical event p. As a consequence, the only thing that remains "mental" about m is its denomination. In the metaphysical sense, mental causation, and indeed physical causation, has disappeared. Some events cause other events, but characterizing these events as "mental" or "physical" has no ground in how things are in themselves. The distinction becomes merely pragmatic. Again, it seems as though in Yablo's theory mental causation is ultimately explained away.

Note that this conclusion does not depend on the distinction of categorical vs. hypothetical properties. Rather, it is induced by the notion of the mental or physical *nature* of an event. The nature of an event is directly dependent on the (categorical) properties it instantiates and, hence, on its essence. It is due to this that mental causation disappears once only physical and/or neutral properties are allowed into its essence.[15]

The trilemma can be developed analogously for the inclusion theory of mental and physical properties. This theory explained a supervenience relation between mental and physical properties, and the multiple realization of mental properties by physical properties, through the fact that the conditional powers associated with a mental property M constitute a subset of several sets of conditional powers associated with physical properties P_1, P_2,... .

However, if all conditional powers associated with the mental property M are mental powers, then it is unclear why these can be contained in sets of powers P_1, P_2,... . Either the powers associated with M are causally redundant in P_1, P_2,... or premise (3**) is false. If,

[15]Yablo seems to anticipate this trilemma already when he says: "Someone might of course ask, why any physical [event] p should have the mentally consequential *kind* of physical property, but this is easily explained. Consider the bearing of supervenience on *mental* events: for each of m's mental properties, supervenience assigns it a necessitating physical property. But it is hard to think what m's physical properties could be if not those of some physical event p which subserved it." (1992b, 267/267) In this quote, Yablo suddenly introduces a supervenience relation between *properties* to explain the subsumption/inclusion between essences of mental and physical events (note the difference to the supervenience formulation (S) in sec. 2). The problem is that now a critic may demand an explanation as to why this supervenience relation between properties holds. Does it hold in virtue of another inclusion relation, perhaps? Yablo does not answer this question.

however, all conditional powers associated with M are physical powers, then there seems to be no reason to describe M as a mental property. Finally, if the powers associated with M and P_1, P_2,... are neutral, i. e. neither mental or physical powers, then it also becomes unintelligible why they should be contrasted with another in the first place.

For the constitutionalist approach, which had declared mental powers to be wholly constituted by physical powers without identifying the former with the latter, the trilemma takes the following form. Either a mental power M is actually constituted *completely* by physical powers $\mathsf{P}_1, \mathsf{P}_2, \ldots, \mathsf{P}_n$. Then M just is the conjunction of powers $\mathsf{P}_1 \wedge \mathsf{P}_2 \wedge \ldots \wedge \mathsf{P}_n$. However, this conjunction just seems to pick out a complex physical power, so that it remains quite unclear why it should also be described as mental. Or a mental power M is *not* completely constituted by physical powers $\mathsf{P}_1, \mathsf{P}_2, \ldots, \mathsf{P}_n$ because there is something more to it than these powers can provide. The result is a dualist ontology once again, since that which is "more to M" is clearly non-physical. Moreover, unless a corresponding premise analogous to (3) is to be false, that which is "more to M" is likely to turn out redundant with respect to the effects of physical powers $\mathsf{P}_1, \mathsf{P}_2, \ldots, \mathsf{P}_n$. Finally, if both mental and physical powers are neutral, the question is why they should be distinguished in the first place.

Dardis' (2008) approach is affected by the trilemma in the following way. He describes events as essentially constituted by a matter property and a form property. The form property of mental events is always a mental property, so that a mental event m is always constituted by an instantiation of a matter property P_M and a mental form property M_F. Supervenience implies that m is always accompanied by a physical event that in addition to P_M instantiates a physical form property P_F as well. Since m and p differ with respect to M_F, resp. P_F, they are not identical. But since they share the instantiation of P_M they are also not distinct and they cannot compete causally, at least in the strong sense. Nevertheless, Dardis can accept premise (3*) only if he claims that the instantiation of P_M and P_F is already sufficient for physical effects of p. But this again leaves the impression that M_F does not do much causal work, and hence, that m is not a cause. Once again, mental causation would be explained away. This problem would, of course, vanish if M_F could be considered a physical form property or if perhaps both M_F and P_F would be somehow considered neutral. However, the consequences would be just as desastrous for the project of explaining how mental causation works.

With these implications of the general trilemma, the project of the new compatibilism is doubtlessly under threat. It may be granted, that the presented theories all make a promising move to avoid the problem of mental causation on one level by describing mental and physical phenomena as non-identical and non-distinct. However, any deeper analysis of the postulate reveals a new trilemma. The new compatibilism will have to offer a way out of this trilemma if it wants to circumvent defeat. The next section develops a potential solution to this particular challenge of the new compatibilism.

8 A potential way out

The central challenge that Yablo and the new compatibilists face is to provide a credible analysis of the relation between mental and physical events (which they deem to be somewhere in between identity and distinctness) that can avoid the trilemma explicated in sec. 7. This section aims to characterize an idea that may help the new compatibilists in this respect. The idea makes reference to the heuristic notion of a "pattern" that has a certain correspondence to a concept presented by Daniel Dennett (1991).

From an information-theoretic point of view, a pattern can be defined as follows. A series (of dots, numbers etc.) is random if, and only if, the information required to describe the series

accurately (by picking out each element along with its position in the series) is incompressible. In this case, only the verbatim bit map will preserve the series. A series contains a pattern if, and only if, there is a more efficient, compressible description of the series (cf. Chaitin 1975, 48).

Dennett offers a first step to a metaphysical interpretation of a pattern. In his view, a pattern does not only exist relative to an actual description. The only condition required for the existence of a pattern within a series, structure, or configuration of objects is that this series, structure, or configuration is amenable to an information-compressing description (cf. 1991, 34). As long as this condition is fulfilled, the relevant pattern is "real".

Dennett's proposal must be credited general intuitive force. For instance, a description of figure 1 that mentions question marks of a certain shape and size is much more efficient than a detailed list mentioning each heart and ring along with their size and relative positions. Moreover, it seems true that the two question marks are "there" and "real" in some sense.

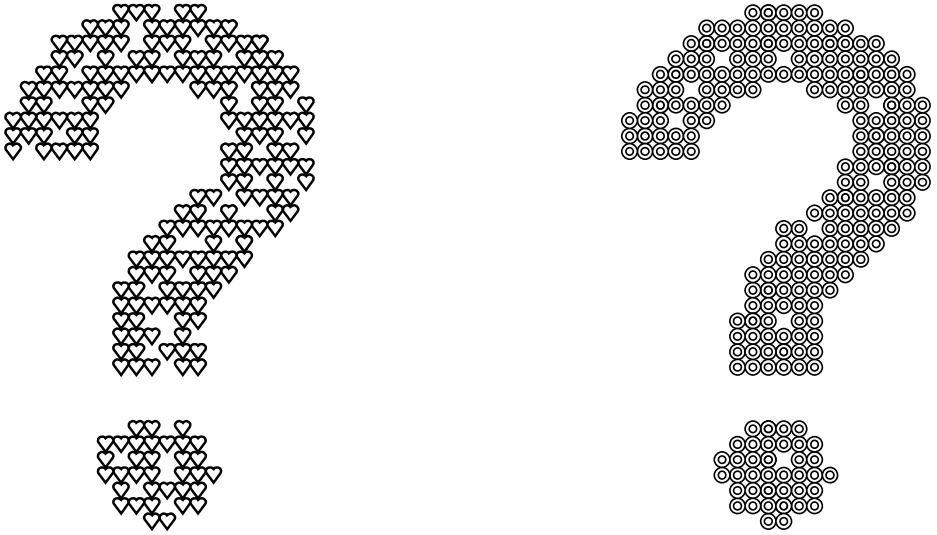

Figure 1: Question marks as real patterns.

One problematic aspect of Dennett's model, however, is the fact that he remains vague as to what "real" and "being there" means in this context. It remains unclear whether patterns are metaphysically real in a substantial sense or only epistemically real, i. e. in the sense that they are inevitably considered as real by a suitable observer. As a consequence, the precise nature of the relation between patterns and the elements constituting them stays opaque. Dennett may consider this actually a virtue of his theory, as it is only intended to provide a framework for more detailed accounts of mental phenomena. However, since being robustly real is a prerequisite for having causal effects in the world, the vagueness is problematic for a satisfactory theory of mental causation. Or in other words, patterns can play a role for a theory of mental causation only if they can be shown to be real in a metaphysically robust sense.

How could such a demonstration of the robustness of patterns be achieved? One strategy is to bring out more explicitly the unfiltered intutions directly provoked by colloquial examples such as the one of figure 1. From an unprejudiced perspective, it seems quite odd to say that the question marks of figure 1 exist only in situations in which someone obeserves or describes them. They are clearly not illusionary as are ghosts walking down the staircase or mirages in the desert. Ghosts and mirages disappear when their observer disappears, but black question

marks printed on a white sheet of paper do not. They continue to be there even if in an instance all minds should cease to exist.

Moreover, already two-dimensional patterns such as question marks have various kinds of typical effects. For instance, they emit light waves of a specific wave length and with a particular structure, they cause the production of characteristic shapes if laid on a xerox machine (where the produced shapes do not necessarily involve hearts and rings), they may also sometimes leave certain kinds of imprints on fingertips and other highly sensitive objects.

Observations of this kind are also induced by three-dimensional patterns. The shape of a cube and the shape of a pop can are both patterns in a sense reminiscient of the question marks of figure 1. Widely independently from the material they are made of, they have various typical causal effects such as characteristic air currents if situated in a wind tunnel, certain imprints they leave in soft sand etc. In short, many two- and three-dimensional patterns are associated with typical causal effects that are widely indepent of the stuff constituting them. Intuitions such as these are always disputable, of course. However, from a coherentist perspective, they may offer at least prima facie reasons to believe that patterns are real in a metaphysically robust sense.

A further evident feature of patterns is the fact that they necessarily require something of which they are made, and that many different, but not all, kinds of matter can serve to constitute, or realise, a given pattern with all its typical causal effects. For instance, the question mark pattern determines that it can be made up of small printed hearts and rings in many different configurations, but not of oxygen molecules (at least not at room temperature). Only if constituted by particular kinds of stuff, the question marks can perform the causal effects associated with them. Note that the hearts and rings, i. e. the stuff that makes up the question marks in the above illustration, are again patterns themselves. In short, patterns seem to have a dual character by combining a causal ("what kinds of effect does the pattern have?") with a constitutive aspect ("what kinds of material can make up the pattern?"). Both aspects are obviously interdependent, but they cannot be reduced to one another.

If these conclusions, and the intuitions on which they are based, are accepted, a particular picture of the nature of patterns emerges. Actual patterns are not really distinct from their realizing elements as they could never be realized without the elements. At the same time, the multiple realizability of the causal profile of patterns suggests that a pattern is not identical to the various configurations realizing it. In particular, the causal behaviour of the question mark on the right of figure 1 would have been the same in virtually all relevant circumstances even if one or two rings would have been replaced by hearts.

The relationship between patterns and the matter constituting them is probably characterized best as a "sufficient approximation" of the constituting elements and their configuration to the boundary conditions definitional to the patterns in question. For two-dimensional patterns, this sufficient approximation is determined by spatio-temporal and material aspects. For instance, it is fairly easy to draw lines defining the outer boundaries of the question marks of figure 1 and, with a suitable procedure, to calculate the degree of the approximation of various configurations of hearts to these lines. A similar procedure for defining the outer boundaries of cubes and pop cans is conceivable.

The upshot of these observations and of the intuitions underpinning them is that, if mental and neural events are like real patterns, the trilemma described in sec. 7 for the relationship between mental and neural events loses its bite. But are mental and neural events pattern-like? Mental events surely involve a causal dimension in the sense that a set of typical inputs and outputs is associated with the mental properties they instantiate. Moreover, they involve a constitutional aspect: They require physical material to be instantiated. And according to a

widespread assumption, many different kinds of matter, but not all, are in principle able to realise a mental event (cf. principle (M) in sec. 2). In these respects, there is a certain analogy between two-dimensional and three-dimensional patterns and mental events/patterns.

Consequently, just as the question marks are not identical to, but also not entirely distinct, from the hearts and rings constituting them, mental events would seem to be not identical with, but also not distinct from, the neural events that realize them. However, this non-identity and non-distinctness is not grounded in the kinds of inclusion relations that lead to the trilemma described in sec. 7. Rather, it imitates what with respect to two- and three-dimensional patterns was described as a "sufficient approximation". Notwithstanding, it excludes a competition between mental patterns and the neural stuff constituting them. In this sense, interpreting mental and neural events as patterns may help to develop a more satisfactory version of the new compatibilist's approaches to mental causation, including Yablo's theory.

The central challenge of this analysis of mental events as pattern instantiations concerns the notion of a "sufficient approximation" that has been described as the criterion for the instantiation of a pattern by the underlying matter. If applied to mental events, the notion would have to involve a complex set of spatial, temporal and material dimensions. Spelling out these dimensions in detail is beyond the scope of this paper. The notions of robust mental patterns and the relation of sufficient approximation at this point only constitute a framework for a more detailed account.

9 Conclusion

This paper first developed a reconstruction of Stephen Yablo's widely discussed theory of mental causation. On the basis of this reconstruction, it was possible to show that many of the popular criticisms brought forward against Yablo misconstrue the core ideas of the approach. Moreover, it was argued that Yablo's theory should be considered a variation of the "new compatibilism" in the philosophy of mind – a theory which describes mental and their underlying physical phenomena as non-identical but non-distinct. However, it was also shown that Yablo's theory falls prey to a new kind of trilemma that huffs virtually all of the approaches within the new compatibilism. In a final step it was suggested that analyzing mental properties as patterns, and mental events as instantiations of mental patterns, would allow the new compatibilists to avoid the trilemma and remain faithful to their most central claims. To provide a plausible specification of the notion of a "sufficient approximation" and the indication of its dimensions turned out to be the central challenge to this strategy of the new compatibilism. Its precise analysis was relegated to future research within the new compatibilist framework.

References

Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs 37*(3), 471–497.

Bishop, R. and H. Atmanspacher (2010). (manuscript).

Bontly, T. (2002). The supervenience argument generalizes. *Philosophical Studies 109*(1), 75–96.

Chaitin, G. (1975). Randomness and mathematical proof. *Scientific American 232*(5), 47–52.

Clapp, L. (2001). Disjunctive properties: Multiple realizations. *The Journal of Philosophy 98*(3), 111–136.

Dardis, A. (2008). *Mental causation: the mind-body problem*. Columbia University Press.

Dennett, D. (1991). Real patterns. *The Journal of Philosophy 88*(1), 27–51.

Ehring, D. (1996). Mental causation, determinables and property instances. *Noûs 30*(4), 461–480.

Ehring, D. (2003). Part-whole physicalism and mental causation. *Synthese 136*(3), 359–388.

Fales, E. (1982). Generic universals. *Australasian Journal of Philosophy Kensington 60*(1), 29–39.

Fales, E. (1990). *Causation and universals*. London: Routledge.

Fodor, J. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese 28*(2), 97–115.

Funkhouser, E. (2006). The determinable-determinate relation. *Noûs 40*(3), 548.

Gillett, C. (2003). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy 100*(11), 591–603.

Gillett, C. and B. Rives (2005). The non-existence of determinables: or, a world of absolute determinates as default hypothesis. *Noûs 39*(3), 483–504.

Harbecke, J. (2008). *Mental Causation. Investigating the Mind's Powers in a Natural World*. Frankfurt a. M.: Ontos.

Haug, M. (2009). Realization, determination, and mechanisms. *Philosophical Studies*, 1–18.

Horgan, T. (1997). Kim on mental causation and causal exclusion. *Noûs, 31, Supplement: Philosophical Perspectives 11, Mind, Causation, and World*, 165–184.

Johnson, W. (1940/1921). *Logic-Part I*. Cambridge: Cambridge University Press.

Kim, J. (1973). Causation, nomic subsumption, and the concept of event. *The Journal of Philosophy 70*(8), 217–236.

Kim, J. (1979). Causality, identity, and supervenience in the mind-body problem. *Midwest Studies in Philosophy 4*(1), 31–49.

Kim, J. (1985). Supervenience, determination, and reduction. *The Journal of Philosophy 82*(11), 616–618.

Kim, J. (1993). Events as property exemplifications. In J. Kim (Ed.), *Supervenience and Mind*, pp. 33–53. Cambridge: Cambridge University Press.

MacDonald, C. and G. MacDonald (1986). Mental causes and explanation of action. *The Philosophical Quarterly 36*(143), 145–158.

McGrath, M. (1998). Proportionality and Mental Causation: A Fit? *Noûs 32*, 167–176.

Montero, B. (2003). Varieties of causal closure. In S. Walter and H. Heckmann (Eds.), *Physicalism and mental causation: the metaphysics of mind and action*, pp. 173–187. Exeter: Imprint Academic.

Pereboom, D. (2002). Robust nonreductive materialism. *Journal of Philosophy 99*(10), 499–531.

Pereboom, D. and H. Kornblith (1991). The metaphysics of irreducibility. *Philosophical Studies 63*(2), 125–145.

Primas, H. (2007). Non-boolean descriptions for mind-matter problems. *Mind and Matter 5*(1), 7–44.

Putnam, H. (1975). Philosophy and our mental life. In H. Putnam (Ed.), *Mind, Language and Reality: Philosophical Papers, vol. 2*, pp. 291–303. Cambridge: Cambridge University Press.

Putnam, H. (1980/1967). The nature of mental states. In N. Block (Ed.), *Readings in philosophy of psychology. Vol. 1*, pp. 223–231. Cambridge: Harvard University Press.

Sanford, D. H. (2008). Determinates vs. determinables. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2008 ed.).

Shoemaker, S. (2001). Realization and mental causation. In C. Gillett and B. Loewer (Eds.), *Physicalism and its Discontents*, pp. 74–98. Cambridge/New York: Cambridge University Press.

Shoemaker, S. (2003/1980). Causality and properties. In *Identity, Cause, and Mind. Philosophical essays,*, pp. 206–233. Cambridge: Cambridge University Press.

Walter, S. (2007). Determinables, determinates, and causal relevance. *Canadian Journal of Philosophy 37*(2), 217.

Wilson, J. (1999). How Superduper does a Physicalist Supervenience need to be? *The Philosophical Quarterly 49*(194), 33–52.

Wilson, J. (2009). Determination, realization and mental causation. *Philosophical Studies 145*(1), 149–169.

Worley, S. (1997). Determination and mental causation. *Erkenntnis 46*(3), 281–304.

Yablo, S. (1987). Identity, essence, and indiscernibility. *The Journal of Philosophy 84*(6), 293–314.

Yablo, S. (1992a). Cause and essence. *Synthese 93*(3), 403–449.

Yablo, S. (1992b). Mental causation. *The Philosophical Review 101*(2), 245–280.

Yablo, S. (1997). Wide causation. *Noûs 31, Supplement: Philosophical Perspectives 11, Mind, Causation, and World*, 251–281.

Yablo, S. (2000). The seven habits of highly effective thinkers. In *Proceedings of the 20th World Congress of Philosophy*, Volume 9, pp. 35–46.

Yablo, S. (2001). Superproportionality and mind-body relations. *Theoria 16*(40), 65–75.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 73–84, 2013

On Normative Practical Reasoning

Georg Spielthenner

The University of Zambia
georg_spielthenner@yahoo.de

Abstract

This article offers an analysis of normative practical reasoning. Reasoning of this type includes at least one normative belief and it has a practical conclusion (roughly, a conclusion about what to do). The principal question I am interested in is whether this type of practical reasoning can be logically conclusive. This issue has received remarkably little philosophical discussion despite the central role this reasoning plays in our everyday discourse about action and in the resolution of ethical problems. I distinguish three kinds of normative reasons. Quasi-normative reasoning, I maintain in Section 1, can only be valid if it is in fact theoretical reasoning, despite appearance to the contrary. In Section 2, I argue that there are two kinds of genuine normative reasoning (purely normative reasoning and hybrid reasoning) that can be logically conclusive. This article also shows that practical arguments are non-trivially ambiguous (which has been largely ignored in the literature) because they can at the same time express different pieces of practical reasoning that have a different logical status.

Practical reasoning is, in short, reasoning about what to do. It is, however, a many-faceted activity that involves at least three different types of arguments, which it is important to keep apart. By 'practical reasoning' one can refer to a pattern of reasoning in which the premises and the conclusion are norms. This type of practical reasoning is studied in deontic logic. A second type of practical reasoning is *instrumental* reasoning. That is, roughly, reasoning to the realization of the agent's ends. In this article, I have nothing to say about these two types of reasoning. But by 'practical reasoning' one often understands a pattern of reasoning that has a normative belief (e. g., a belief that something is obligatory, permitted or forbidden) amongst its premise-states and that has a practical conclusion. John Broome (2001) has called such reasoning *normative* practical reasoning (hereafter, normative reasoning).

Philosophers have written a great deal on instrumental reasoning, and there is also a large literature on the kind of reasoning that is studied in deontic logic. But there is a much more limited literature on normative reasoning. This relative neglect is regrettable because this type of reasoning is not only prevalent in our everyday practical discourse, it is also of considerable ethical relevance. On a widely held view, ethical reasoning consists (at least partly) in a derivation of normative conclusions from ethical principles and statements of facts, which yields exactly the type of reasoning that is under consideration in this article.

The discussion of normative reasoning goes back to Aristotle's sketchy treatment of so-called practical syllogisms. He infers, for instance, from the premises "I should make something good" and "A house is something good" the conclusion that a house is being made (*De Motu*

Animalium 701a16-17).[1] In the 20th century, Georg Henrik von Wright explored the varieties of practical reasoning and he also examined some examples of normative reasoning. He holds, for instance, that the argument "*A* must do *x*. Unless *A* does *y*, he cannot do *x*. Therefore *A* must do *y*" represents a logically valid piece of reasoning (1983, p. 13). Considering the obvious philosophical importance and interest of normative reasoning, it is surprising that it has received relatively little attention in the contemporary philosophical literature. The most interesting recent treatment, of which I am aware, is by John Broome (2001). In this paper, he "tentatively proposes" that normative reasoning is a logically correct (valid) type of practical reasoning, even though he admits that his arguments for this view are slender, and he considers his research as preliminary only (p. 181).

Against this background, the aim of this article is now simple to state. I try to determine whether normative practical reasoning can be logically valid. To achieve this aim, I distinguish between three different types of normative reasoning, which I shall consider one by one. I start (in Section 1) with *quasi*-normative reasoning, Sections 2 and 3 are then concerned with two versions of genuine normative reasoning. *Quasi*-normative reasoning, I shall maintain, is only valid if it is a form of theoretical reasoning. But I shall argue that the two versions of genuine normative reasoning can be logically valid. By showing that practical arguments are non-trivially ambiguous (because they can at the same time express different pieces of practical reasoning that can have a different logical status), this article also demonstrates the importance of distinguishing between *arguments* and pieces of *reasoning* – a distinction that has been widely ignored in the literature on practical reasoning. For simplicity of presentation, I have sought to set out my position in a non-technical way. But for the sake of brevity and clarity, some formalization has been unavoidable.

1 *Quasi*-normative Reasoning

Normative reasoning is, like all reasoning, an activity that takes place in the reasoner's mind and it takes us from existing states of mind, the "premise-states", to a new state, the "conclusion-state", to borrow Broome's (2001, 176) terms. As I have already mentioned, normative reasoning implies that at least one premise-state is a normative belief. Let me illustrate the role of normative beliefs in this type of reasoning by adapting one of von Wright's (1983) examples. Suppose you believe that it is *A*'s duty to make the hut habitable. You also believe that unless *A* heats the hut, he cannot make it habitable; and therefore you conclude that it is *A*'s duty to heat the hut (p. 14). This is a description of a simple piece of reasoning. One of your premise-states is the normative belief that it is *A*'s duty to make the hut habitable. The second premise is a descriptive belief; and the reasoning takes you from the two "premise-states" to a "conclusion-state".

Normative beliefs are, as beliefs in general, psychological attitudes and as such they have *contents*, which I take to be propositions.[2] The content of your first premise (i. e., the proposition that it is *A*'s duty to make the hut habitable) can be *expressed* by the sentence 'It is *A*'s duty

[1] Other examples of Aristotle's treatment of normative reasoning can be found in the *Ethica Nicomachea* 1147a25-30, *De Anima* 434a15-20, the *Metaphysics* 1032b8-10, and *De Motu Animalium* 701a13-14. See also Thornton (1982).

[2] This view is not uncontroversial. Some authors have doubted that the contents of intentional states are always propositions. For instance, Rabinowicz and Rønnow-Rasmussen (2004) hold that also *things* and *persons* can be the contents of such attitudes (p. 393), and von Wright (1963, 160) holds that their object is a *state of affairs*. But others have convincingly argued that beliefs and preferences are *propositional* attitudes. We do, for instance, not prefer coffee to tea as such but rather drinking coffee to drinking tea (Hansson, 2006); and Searle (2001) holds that "all desires have whole propositions as intentional contents (thus 'I want your car' means something like 'I want that I have your car') . . . " (pp. 248-9).

to make the hut habitable'. The point to emphasize now is that expressions of normative beliefs can be given a purely *descriptive* interpretation, despite the fact that they often contain deontic terms. On such an interpretation, the sentence 'It is *A*'s duty to make the hut habitable' can be understood as saying that there exists a norm to the effect that *A* has a duty to make the hut habitable (see von Wright 1983, 131-2). To put it differently, if you believe that there exists the norm that *A* has a duty to make the hut habitable you can *express* this belief by saying 'It is *A*'s duty to make the hut habitable'. Although this sentence contains the term 'duty', it does not express a valuation but is what von Wright (1983) calls a *normative statement*. A normative statement is a statement in the strict sense, i. e., a sentence that is either true or false.[3] In our example, the normative statement is true if there exists a norm that requires *A* to make the hut habitable and it is false if no such norm exists.[4]

In short, when we believe that *p* ought to (can or must not) be done then the content of our belief is sometimes the proposition that there is a norm to the effect that *p* is obligatory (permitted or forbidden), and the *expression* of this belief is then a statement in the strict sense, even if it contains deontic terms such as 'ought' or 'should'. In such cases, I shall argue in the following, the reasoning only *appears* to be normative. I therefore dub it *quasi*-normative reasoning.[5]

But before we proceed, two further introductory remarks should be made. First, it seems that we need to clarify what exactly the conclusion of a piece of practical reasoning is. There is fundamental disagreement in the philosophical literature over this issue. Some hold that it is an action (Anscombe 1963), while others reject this as inadequate and contend that it must be an intention (Broome 2001), a decision (Searle 2001), an imperative (Gensler 1996), a desire (Audi 2001), or a normative judgement (Clarke 1985).[6] However, since nothing substantial in this essay hinges on this issue, we need not try to settle this dispute here, and I shall assume that the conclusion is a desire to do something.[7] Second, since I shall claim that some modes of normative reasoning are valid while others are not, I need to briefly outline when, on my view, practical reasoning (and therefore also normative reasoning) is valid. The details of this are more than we need, but here is the basic idea.[8]

[3]According to von Wright (1983), a normative statement is "a statement to the effect that something ought to or may or must not be done" (p. 67). For example, by saying 'You must not park on this side of the street', we may make the statement that there is a by-law or regulation prohibiting it, which is true or false.

[4]It should be noted that this is not the only possible descriptive interpretation of a deontic sentence. In the example "*A* ought to make the hut habitable. Unless *A* heats the hut, he cannot make it habitable. Therefore *A* ought to heat the hut", the deontic sentence '*A* ought to heat the hut' can plausibly be interpreted as meaning that unless *A* heats the hut it will not become habitable, which is again a statement in the strict sense. See also Raz (1999, 171-177) who discusses the use of normative statements in legal reasoning.

[5]That normative sentences can be given a descriptive interpretation is a widely held view among writers on ethics and there is a vast literature on this topic. The *locus classicus* is probably Hare (1952). He holds that value judgements are often used for conveying purely factual information (p. 112) and he distinguishes different ways in which sentences can be used descriptively – e. g., by using them in an "inverted-comma sense" (p. 165). Among the many other authors who accept this view are Edwards (1955) who claims that 'X.Y. is a good person' can be used for stating that he is loving and free from envy (p. 147), and Stevenson (1944) who argues that saying that *X* is a good college president can be tantamount to saying that he is an industrious and honest executive (p. 208). The distinction between prescriptive and descriptive interpretations of normative sentences has also played a major role in the discussion whether normative conclusions can be derived from purely factual premises – see, e. g., Black (1964) and Genova (1973).

[6]This issue has been extensively discussed by others. See, for instance, Barnes (1983), Clarke (1985) and von Wright (1963; 1983).

[7]As usual in philosophy, that I take 'desire' here as an umbrella term that can stand for any type of evaluative attitude, not only for some strong wishes or cravings.

[8]I have provided a more detailed discussion of practical validity in (Spielthenner 2007). In the present article, I take it for granted that a piece of practical reasoning can be logically valid. This view is by no means beyond dispute. Many logicians and philosophers endorse it (e. g., Kenny 1978 or Broome 2001), but several writers have argued against it

Like reasoning in general, practical reasoning is valid if (and only if) the set consisting of the premises and the negation of the conclusion is *inconsistent*. For definiteness, let me state this basic fact in the following principle of valid practical reasoning.

> (P) A piece of practical reasoning that consists of the premises $P_1, \ldots, P_n$ and the conclusion C is valid *iff* the set $\{P_1, \ldots, P_n, \neg C\}$ is inconsistent.

This basic idea has been expressed in different ways. Searle (2001) holds that the acceptance of the premises of a valid practical argument "*commits* one to the acceptance of the conclusion" (p. 241); according to Gensler (1996, 16), inconsistency of the premises and the conclusion means that we *ought not to* combine accepting the premises with accepting the conclusion; and Richard Hare contends that "he who assents to the premises is compelled not to dissent from the conclusion, on pain of logical inconsistency."[9]

Of course, for such a characterization to be illuminating, an explanation of "inconsistency" should be given. Since practical inconsistency implies that *different* propositional attitudes (beliefs and valuations) are incoherent, it is not plain when the premises of a piece of practical reasoning and the negation of its conclusion are inconsistent. However, for lack of space I shall attempt to make this notion intuitively clear by going through a simple example, rather than explicate it.

Let it be assumed you argue that Catholics should go to church today because they are obligated to go to church on Sundays and today is Sunday. The first premise of your reasoning is the true normative belief that Catholics are obligated to go to church on Sundays. That is to say, the content of your normative belief is the true proposition that Catholics are obligated to go to church on Sundays.[10] No matter what we take the conclusion of your reasoning to be, it is easy to see that you are *not inconsistent* if you believe the premises but deny the conclusion. More concretely, if you believe that Catholics are obligated to go to church on Sundays (i. e., you believe that there is a norm that requires them to do this), and you believe that today is Sunday, but you do not desire that Catholics go to church today, you are *not* in a state of mental incoherence of the same nature as logical inconsistency.[11] According to P, the reasoning is therefore invalid.

In order to see things in a clearer light, let us consider what conclusion *is* entailed by the premises. Schematically, your premises can be stated as follows (putting their contents in brackets and writing 'B' for 'I believe that' because we assume that you verbalize your reasoning to yourself in the first person).

(1) B (Catholics are obligated to go to church on Sundays).

(2) B (Today is Sunday).

If you infer from these premises that Catholics are obligated to go to church today – that is, if your conclusion is 'B (Catholics are obligated to go to church today)', then your reasoning is

(e. g., Mitchell 1990). Since a consideration of this issue would take us beyond the confines of the present work and I have argued for the validity of practical arguments elsewhere (see Spielthenner 2008), I will not pursue this issue further here.

[9] Quoted from Kenny (1978, 75).

[10] Canon 1247 of the *Codex Iuris Canonici* states that "on Sundays and other holydays of obligation, the faithful are obliged to assist at Mass."

[11] That *quasi*-normative reasoning does not allow drawing a practical conclusion has also been held by Mackie (1977). In a somewhat different context he writes that "the statement that a certain decision is thus just or unjust will not be objectively prescriptive: in so far as it can be simply true it leaves open the question whether there is any objective requirement to do what is just or refrain from what is unjust, and equally leaves open the practical decision to act in either way" (pp. 26-27).

valid (because the contents of your premise-states entail the content of your conclusion-state). But your conclusion is *not* practical. It is rather a belief that something is obligated. You can, of course, *express* this belief by saying that Catholics *should* go to church today, but the use of so-called deontic terms does not *eo ipso* render your conclusion *practical*. It is a common mistake to think that expressions that contain deontic terms are thereby normative expressions.

The upshot of what I have said so far in this section is that *quasi*-normative reasoning can be valid, but if it is valid it is not practical reasoning but *theoretical* reasoning. Practical reasoning should provide a reason for doing something, not for believing it. But the reasoning under consideration can provide only reasons for believing that Catholics have an obligation to go to church, not for desiring that they go to church.

I envisage the objection that this result is counter-intuitive. After all, it seems that deontic logic can show that, in our example, the contents of the premises entail a normative conclusion and the reasoning seems therefore genuinely normative. To see that holding this is a mistake, let us formalize the reasoning. Let 's' mean 'Today is Sunday' and let us use 'c' for 'Catholics go to church today'. The deontic operator 'O' is to be read 'it is obligatory that'. If we symbolize the reasoning as suggested we obtain $B(s \rightarrow \mathrm{O}c), Bs \models B(\mathrm{O}c)$. On the left side of '$\models$' (the "double turnstile") are the premises of the reasoning and on its right side is the conclusion. Now, according to the standard system of deontic logic, the content of the conclusion (i. e., Oc) follows from the contents of the premises. That is to say, the argument $s \rightarrow \mathrm{O}c, s \models \mathrm{O}c$ is logically correct and therefore the reasoning $B(s \rightarrow \mathrm{O}c), Bs \models B(\mathrm{O}c)$ is valid.[12] The content of the conclusion appears to be a *norm*, and therefore the reasoning seems to be practical.

My response to this objection is that deontic operators can be (and sometimes *must* be) descriptively interpreted, and that the appropriate logic of *quasi*-normative practical reasoning is deontic logic *descriptively* understood.[13] In the present context, the expression '*B*(Oc)' stands for 'I believe that Catholics are obligated to go to church today'. But on the assumed descriptive interpretation, the content of your belief is the proposition that there exists a norm to the effect that Catholics are obligated to go to church today. The expression of this belief is therefore a statement of fact. Hence, deontic logic cannot be used to show that we can derive a *practical* conclusion from the contents of your premises.

Let us take stock. It has been my contention in this section that there is a type of reasoning that can be valid if it is *theoretical* reasoning, but that is invalid if it is *practical* reasoning. This may appear so obviously true as to be hardly worth saying. But it is, I submit, worth emphasizing because in our practical thinking we are inclined to confuse *quasi*-normative reasoning with genuine normative reasoning. Not only philosophical laymen are commonly unaware that sentences such as '*S* ought to ϕ', '*S* must not ϕ', or '*S* has an obligation to ϕ' can be simple statements of fact. I think that not distinguishing between the descriptive and the prescriptive interpretation of such utterances has also caused mistakes in the philosophical literature.[14]

[12]The correct symbolization of hypothetical norms is a notorious problem. Different symbolizations have been proposed, but none has firmly established itself. I have here chosen '$s \rightarrow \mathrm{O}c$', which is only one possibility.

[13]In deontic logic, the distinction between a descriptive and a prescriptive interpretation of formulas such as Oϕ (it is obligatory to ϕ) or Pϕ (it is permitted to ϕ) has been commonly ignored. Even as careful a writer as von Wright (1983) – one of the founding fathers of modern deontic logic – had to admit that in his earlier writings he had conflated a "logic of norms" (i. e., a prescriptive interpretation) with a "logic of norm-propositions" (i. e., a descriptive interpretation).

[14]See, for instance, Gensler (2002) who holds that "thinking that A is wrong and at the same time acting with the intention of doing A" is inconsistent (p. 217). This is not correct if 'A is wrong' is descriptively interpreted.The importance of *quasi*-normative reasoning can also be seen from the fact that many of our "hypothetical imperatives" seem to be best understood as elliptical pieces of this type of reasoning (see von Wright 1983, 199-200). Suppose you believe that Jones wants to make a favourable impression at his job interview, and you believe that he can only make such an impression if he gets a haircut. On the basis of these premises you hold that Jones ought to get a haircut.

In passing, I wish to mention that also so-called ethical principles admit of a descriptive analysis. The content of a physician's ethical belief that doctors ought to act in the best interest of their patients may be expressible by the normative statement that medical codes of ethics require physicians to act that way. If a physician only believes this proposition but does not value it, he is *not inconsistent* if he holds this normative belief but does not act in the best interest of his patients.

2 Genuine Normative Reasoning

The aim of this section is to discuss, from a logical point of view, two kinds of normative reasoning that are genuinely practical. That is to say, their conclusion is valuational rather than a mere belief.

2.1 Purely normative reasoning

The type of reasoning I propose to call purely normative reasoning represents the opposite of *quasi*-normative reasoning. If you believe that you ought to ϕ and your reasoning is *quasi*-normative, then you believe something about ϕ-ing but you do not value it. On the other hand, if your reasoning is purely normative, your conclusion-state is a desire, not a belief. That is to say, if you express your conclusion by saying 'I ought to ϕ' you express your valuation of ϕ-ing but you do not make a statement about ϕ-ing.[15] For instance, by holding that you should lose weight, you may only express your desire to reduce without making any statement about weight loss.

Purely normative reasoning comes in two main versions. (i) In the more prevalent form, the reasoner values the *content* of a norm. Let a simple example serve to illustrate this. Jones holds that he must stop smoking. Schematically, we can write this as follows: Jones holds that (it must be the case that (Jones stops smoking)). What Jones values is the proposition that he stops smoking, not the proposition that it must be the case that Jones stops smoking. By saying 'I must stop smoking', Jones may only express his pro-attitude towards the proposition that he stops smoking, without stating, say, that it can cause cancer.

(ii) Sometimes, however, a reasoner values the normative status of a proposition (i. e., that it is forbidden, permitted, or forbidden). Suppose I hold that so-called partial birth abortions ought to be forbidden. Writing this schematically, we get: I believe that (it ought to be forbidden that (partial birth abortions are performed)). In this example, I value the fact that this type of abortion is *forbidden*. (My desire may be due to what I believe about partial birth abortions. But this *reason* for my valuation is not part of the content of my normative belief.) That is to say, by holding that partial birth abortions ought to be forbidden, I may only express my pro-attitude towards *forbidding* it, without making a statement about it. Thus, reasoning based on such a type of normative belief is purely normative.

To assess this type of normative reasoning from a logical viewpoint, let us analyse a concrete example. Suppose you put forward this argument: "Organ transplants should not only be available to the rich. But if human organs are allowed to be sold, only the rich will be able to afford them. The selling of human organs should therefore not be allowed." This is a linguistic expression of your reasoning. To assess it, we need to go from the linguistic level to the mental

However, the content of your conclusion-belief may be the proposition that Jones will only attain his aim of making a favourable impression if he gets a haircut (which you express by saying 'Jones ought to get a haircut'). Hence, the reasoning may be *quasi*-normative reasoning without any normative force.

[15] This article has been written on the basis of my expressivist meta-ethical position. But it is not tied to it. Upon changing what needs to be changed, my exposition is understandable on the background of other meta-ethical views, too.

level. After all, reasoning is a mental process and the premises and the conclusion are mental states.

We can plausibly assume that the sentence 'Organ transplants should not only be available to the rich' is not a normative statement but expresses your disvaluing of the proposition that organ transplants are only available to the rich. Obviously, if you disvalue this proposition, you can express yourself by uttering the sentence 'Organ transplants should not only be available to the rich'. Schematically, we can therefore state the first premise of your reasoning as follows: Des (Organ transplants are not only available to the rich) – writing 'Des' for 'I desire that'.

The second premise expresses your descriptive belief that only the rich will be able to afford organ transplants if human organs are allowed to be sold. We can write this as B (If human organs are allowed to be sold only the rich will be able to afford them).

Let it be assumed that the conclusion of your argument is not a statement but a normative sentence that expresses your disvaluation of the proposition that the selling of organs is allowed. (As will become apparent below, only on this assumption is the reasoning under consideration valid.) This disvaluation can be *expressed* by uttering the sentence 'The selling of human organs should not be allowed'. Again, we can write the conclusion more exactly as Des (The selling of human organs is not allowed). We can describe now your reasoning as follows:

(1) Des (Organ transplants are not only available to the rich).

(2) B (If organs are allowed to be sold, only the rich will be able to afford them).

Thus, (3) Des (The selling of human organs is not allowed).

The reasoning is valid. I think this is intuitively obvious, but it will be helpful to have a closer look at it by applying the principle of valid practical reasoning that I have stated in Section 1. I said that a piece of practical reasoning is valid *iff* the set consisting of the premises and the negation of the conclusion is *inconsistent*.

Let us assume that you hold the premises, but you negate the conclusion. That is to say, you *desire* (i) that organ transplants are not only available to the rich, you *believe* (ii) that they are available only to the rich if human organs are allowed to be sold, and at the same time (iii) you negate the conclusion (which is equivalent to either being indifferent between selling and not selling them or desiring that their selling is allowed). Let us take the second possibility (the reasoning for the first is analogous). I think it is clear that desiring (i), believing (ii) and at the same time desiring that human organs are allowed to be sold, is being in an incoherent state of mind that is tantamount to inconsistency. This is not to say that it is impossible that a person may entertain all of them. But it is a logical mistake because if you negate the conclusion the content of this negated conclusion (i. e. the proposition that selling of human organs is allowed) logically entails that your desire that organ transplants are not only be available to the rich is frustrated; and, on the other hand, if you assent to the conclusion, its content (i. e. the proposition that selling of human organs is not allowed) logically entails that this desire is satisfied. Given the premises, denying the conclusion is, to adopt a term from I. Kant, an "inconsistency in will".

For definiteness and to make the notion of valid practical reasoning clearer, let me symbolize the reasoning using '*r*' for 'organ transplants are only available to the rich' and '*s*' for 'the selling of human organs is allowed'. On a suggested interpretation we get $Des(\neg r), B(s \leftrightarrow r) \models Des(\neg s)$. (Notice that we need to formalize the second premise as a *biconditional*; a conditional does not make the reasoning valid.) This reasoning is valid because the set consisting of the premises and the negated conclusion is *inconsistent*. That is, the set $\{Des(\neg r), B(s \leftrightarrow r), Des(s)\}$ is inconsistent. It is, however, not inconsistent because it is impossible that a person holds all

members of this set at the same time. It is rather inconsistent because there is a special logical relationship between the *contents* of the premise-states and the *content* of the conclusion-state. In short, this relationship is such that (i) the content of the negated conclusion (in our example '*s*') entails, together with the content of the descriptive premise, the proposition *r* (that transplants are only available to the rich). That is, the negated conclusion logically leads to a *frustration* of your desire that transplants are *not* only available to the rich. This can easily be seen when we consider the argument consisting of these contents: $\neg r, s \leftrightarrow r \models s$. Obviously, *s*, together with $s \leftrightarrow r$, entails *r*. (ii) On the other hand, the content of the conclusion (i. e., $\neg s$) entails, again together with $s \leftrightarrow r, \neg r$. That is to say, the conclusion logically implies that your desire (that transplants are not only available to the rich) is *satisfied*. A piece of practical reasoning is valid if both conditions are satisfied.

I have said that a reasoner sometimes values the normative status of a proposition – e. g., the proposition that partial birth abortions are *forbidden*. It is not difficult to see that such reasoning can also be valid. For instance, if the sentence 'Partial birth abortions should be forbidden' expresses the reasoner's desire that they are forbidden, then the reasoning is analogous to the type we have discussed above. To avoid tedium, I shall not discuss here the details of this kind of reasoning.

It should be easy to see at this point that in purely normative reasoning, validity does not depend on normative beliefs. They are, in fact, redundant. Returning to our example, whether you *believe* that organ transplants should not only be available to the rich is logically immaterial. What matters is that you *disvalue* their being available only to them. The conclusion was entailed by this disvaluation and a *descriptive* belief. But in the next subsection we shall see that sometimes normative beliefs *are* relevant to the validity of normative reasoning.

2.2 Hybrid normative reasoning

In Section 1, it will be remembered, I have said that the expression of a normative belief can be a statement and that such beliefs do not entail a practical conclusion. In the previous subsection, on the other hand, I have shown that normative sentences are sometimes not statements but only expressions of the reasoner's valuation. In other words, deontic sentences can be descriptively or prescriptively interpreted, and since the 'or' is here inclusive, there is the possibility that a deontic expression does both, stating a fact and expressing a valuation. Strictly observing the *distinction* between descriptive and prescriptive discourse does not imply holding that deontic expressions cannot be given "both a prescriptive and a descriptive interpretation" (von Wright 1983, 201). That is to say, the sentence '*S* must not ϕ' can express the (true or false) belief that ϕ-ing is forbidden and at the same time it can express the reasoner's disvaluing of *S*'s ϕ-ing.[16]

To make this clear, let us consider an example. John believes that he ought to divide the money equally. He can express this belief by saying 'I ought to divide the money equally'. Now, by saying this, he can make a statement (e. g., the statement that there exists a norm

[16]That a sentence can be at the same time a factual statement and a normative expression is not obvious. Some further points will be helpful to make this view more plausible. This doctrine was (among others) defended by Hare (1952) who holds that a normative sentence like 'I ought to do X' can be a mixture of one (or more) statements *and* a value judgement (p. 167). We may use one of his examples (which I have slightly adapted) to make this view clearer (see p. 122). The sentence 'This is a good egg' can be used for expressing the belief that it is not decomposed (its descriptive use) and at the same time it can express a favourable attitude towards the egg (its evaluative use). A similar view has been held by Hampshire (1949) who argues that the normative and the descriptive element of moral judgements are often almost inextricably combined (p. 480).

which requires him to divide the money equally)[17] and at the same time he can express his pro-attitude towards dividing the money equally. I shall call reasoning of this type *hybrid* normative reasoning.

In hybrid reasoning, we have always a least one premise or a conclusion that is both descriptive and valuational. Notice, however, that these two aspects need not be equally salient. Sometimes the descriptive component is dominant (e. g., when a doctor tells his patient that he must take his blood pressure medication), sometimes the valuational component is central (e. g., when the patient holds that he must take his blood pressure medication).[18]

I move now on to discussing the logical assessment of hybrid reasoning, a topic largely unexplored. As we shall see, in a logical evaluation we need to employ different logical systems, and the result of our assessment may not be straightforward because this type of reasoning can be both valid and invalid, albeit in a different sense.

To see when hybrid reasoning is valid, let us analyse a simple piece of reasoning. Jones argues "I should not drive my car because drunken driving is wrong and I have been drinking". It will be helpful to restate it schematically as follows:

(1) Drunken driving is wrong.
(2) I have been drinking.
Thus, (3) I should not drive my car.

This argument is an expression of Jones' reasoning. We assume that this reasoning is hybrid. Its assessment requires then that we separate the *quasi*-normative reasoning from the purely normative reasoning and evaluate them individually.

(i) *Ex hypothesi*, Jones holds the statements 'Drunken driving is wrong' and 'I have been drinking'. The conclusion follows only if 'I should not drive my car' is descriptively understood – for instance, as an expression of the belief that it is forbidden that he drives his car. To be a bit more precise, we can state the descriptive strand of Jones' reasoning as follows:

(1') B (Drunken driving is forbidden).[19]
(2') B (I have been drinking).
Thus, (3') B (It is forbidden that I drive my car).

This reasoning is valid. I think this is clear on an intuitive level. It should be easy to see at this point that we have to give the deontic operator 'it is forbidden that' a descriptive interpretation because the reasoning is, as I have already mentioned in Section 1, *theoretical* reasoning.

(ii) We assume that Jones does not only make a statement when he holds that he should not drive his car. By 'I should not drive my car' he can express his disvaluation of the proposition that he drives the car. His reasoning is valid, if 'Drunken driving is wrong' is also a normative

[17] Of course, this is not the only possible descriptive meaning. The sentence 'I ought to divide the money equally' can, for instance, also express the belief "If I don't divide the money equally I am acting unjustly" (which John does not want to do). What statement is being made by a deontic sentence, if any, always depends on the context.

[18] There is one further point which I should like to mention in this connexion. The descriptive and the prescriptive aspect of a normative expression need not concur. You can *state* that ϕ-ing is obligatory and at the same time express your *disvaluing* of ϕ-ing. There is no contradiction in stating that I have an obligation to pay property tax and at the same time express my disapproval of having this obligation. It is no news that we can have different attitudes (e. g., beliefs and valuations) towards one and the same proposition, and it is also clear that these attitudes may not concur.

[19] Clearly, if Jones *believes* that drunken driving is forbidden, he can *express* this by saying that it is wrong. The statement 'Drunken driving is wrong' is then true if there exists a norm to the effect that drunken driving is not allowed.

utterance and not only a statement. In short, the purely normative strand of the reasoning is valid if we interpret it as follows:

(1") Des (People do not drive when they have been drinking).
(2") B (I have been drinking).
Thus, (3") Des (I do not drive my car).

This purely normative reasoning is also valid. But the reason for its validity is, as I have already outlined, different. It is valid because the set consisting of the premises and the negated conclusion is inconsistent. If you desire that people do not drive their car when they have been drinking, you believe that you have been drinking, and at the same time you desire to drive your car, you are practically inconsistent (which does not rule out that you are in this motivational state). This set is inconsistent because the content of the negated conclusion-state (i. e., that Jones drives his car) entails (together with the content of the descriptive premise) that his desire that there is no drunken driving is frustrated; while, on the other hand, the content of the conclusion-state (together with the content of the descriptive premise) facilitates the satisfaction of this desire.[20]

To be clear, it is my contention that the argument (1) – (3) is an expression of *both* the descriptive reasoning (1') – (3') and the purely normative reasoning (1") – (3"). That is to say, if Jones believes that drunken driving is forbidden (1') and desires that people do not drive when they have been drinking (1"), then he *can* express these two propositional attitudes by saying 'Drunken driving is wrong' (1). Likewise, if he believes that it is forbidden that he drives his car (3') and desires that he does not drive it (3"), then he *can* express this by saying 'I should not drive my car' (3).

In our example, both the descriptive and the normative reasoning are valid, and the conclusions of both strands of reasoning concur. Jones validly infers that it is forbidden that he drives a car and his reasoning to the desire of not driving the car is also conclusive. But this need not be the case. In hybrid reasoning, the descriptive part can be valid when the purely normative part is invalid (and vice versa); and the conclusions may conflict. To make the nature of hybrid reasoning and its logical assessment clearer, I shall now illustrate this by briefly discussing two examples.

(i) Conflicting conclusions: Suppose you believe that you must give a lecture on Monday mornings, you also believe that it is Monday morning and therefore you believe you must give now a lecture. By symbolizing your reasoning (using '*m*' for 'it is Monday morning' and '*l*' for 'I give a lecture') we obtain $B(m \rightarrow \mathrm{O}l), Bm \models B(\mathrm{O}l)$. On a descriptive interpretation, the reasoning is valid because its content (i. e., $m \rightarrow \mathrm{O}l, m \models \mathrm{O}l$) is a valid argument.

Suppose, however, that you hate lecturing on Monday mornings.[21] Your valuation $Des\neg(m \rightarrow l)$ and your belief *Bm* entail $Des(\neg l)$ – i. e., your desire not to give a lecture.[22] (Desiring not to lecture on Monday mornings, at the same time believing that it is Monday morning and desiring to give (now) a lecture is *inconsistent*.)

You have now validly inferred that you have an obligation to give a lecture; and your desire not to lecture is also implied by your premises. To be sure, this does not only show the triviality that you can have a reason for holding that you have an obligation to do something and at the

[20] As I have already mentioned, I am glossing here over the details of practical validity. But I think our example is so simple that its validity can be intuitively appraised.

[21] The sentence 'I must give a lecture on Monday mornings' can be used to express your devaluation of the proposition that you lecture on Monday morning (e. g., by emphasizing the phrase 'Monday mornings').

[22] Be it noted that *Des* ($\neg l$), i. e. your desire not to lecture, can – in an appropriate context – be expressed by saying 'I must give a lecture' (which is the conclusion of your argument).

same time have a reason for disvaluing it. It rather illustrates the interesting fact that by putting forward one argument (in our example, "I must give a lecture because I have to lecture on Monday mornings and it is Monday morning"), you can express two pieces of reasoning that have conflicting conclusions.

(ii) Different logical status: Suppose you believe that you must support your old parents and you think that you can only support them if you visit them regularly. On a descriptive interpretation (where 'I must support my old parents' is a normative statement, stating, say, that you have this legal obligation), these premises do *not* entail that you must visit your parents regularly (where 'I must visit my old parents regularly' is also a statement).[23]

However, if 'I must support my old parents' expresses your desire to support them, and you therefore desire to visit them regularly, your purely normative reasoning *is* valid. (In short, visiting them entails that your desire to support them is satisfied, while not visiting them entails that this desire is frustrated.)

This result shows that one argument (e. g., "I must visit my old parents regularly because I must support them and I can support them only by visiting them regularly") can express two pieces of reasoning. One of them is valid, the other invalid.

3 Conclusion

The result of this investigation may be summarized as follows: Contrary to how it may appear to some philosophers, there is not only one kind of normative reasoning. There are at least three types. One of them, which I have dubbed '*quasi*-normative', is in fact theoretical reasoning. At first sight, it may appear to be practical because a superficial understanding makes it easy to overlook the difference between normative statements and linguistic expressions of valuations. Clearly demarcating *quasi*-normative reasoning from genuinely normative reasoning seems to be particularly important. We are inclined to conflate the two forms of reasoning because in expressing their premises and their conclusion, we can use the same deontic terms, and we tend to ignore that these terms can have a purely descriptive use.

With regard to *genuine* normative reasoning, we have seen that it can be valid. I am aware that I was only able to present a brief sketch of its logical assessment. It should, however, be noted that it is not normative *beliefs* that render this type of reasoning valid but *valuations*. It is not the fact that you believe you ought to ϕ, but the fact that you value ϕ–ing which allows you to draw a practical conclusion. In this type of reasoning, normative beliefs as such are logically redundant.

Of particularly importance seems to me *hybrid* reasoning. Its nature is not well understood and its logical assessment has, as far as I know, never been explored in detail. My exposition, incomplete though it is, should have shed some light on this common form of normative reasoning.

References

Anscombe, G. E. M. (1963) *Intention* (2nd ed.), Oxford: Blackwell.

Audi, R. (2001) *The architecture of reason: The structure and substance of rationality*, Oxford-New York: Oxford University Press.

Barnes, G. W. (1983) 'The conclusion of practical reasoning', *Analysis, 43*(4), 193–199.

Black, M. (1964) 'The gap between 'is' and 'should", *The Philosophical Review, 73*, 165–181.

[23] The argument $\mathrm{O}s, s \leftrightarrow v \models \mathrm{O}v$ (where 's' stands for 'I support my old parents' and 'v' for 'I visit my old parents regularly') is *invalid* if we assess it by applying the system that has come to be called the *standard deontic logic*.

Broome, J. (2001) 'Normative practical reasoning', *Proceedings of the Aristotelian Society (Supplementary Volume), 75*(1), 175–193.

Clarke, D. S. (1985) *Practical inferences*, London: Routledge & Kegan Paul.

Edwards, P. (1955) *The logic of moral discourse*, New York: The Free Press.

Genova, A. C. (1973) 'Searle's use of 'ought", *Philosophical Studies, 24*, 183–191.

Gensler, H. J. (1996) *Formal ethics*, London: Routledge.

Gensler, H. J. (2002) *Introduction to logic*, London: Routledge.

Hampshire, S. (1949) 'Fallacies in moral philosophy', *Mind, 58*, 466–482.

Hansson, S. O. (2006) Preferences. Available at: http://plato.stanford.edu/entries/preferences.

Hare, R. M. (1952) *The language of morals*, Oxford: Clarendon.

Kenny, A. J. (1978) Practical reasoning and rational appetite, In J. Raz (ed.), *Practical reasoning*, Oxford: Oxford University Press, pp. 63–80.

Mackie, J. L. (1977) *Ethics: Inventing right and wrong*, London. Penguin Books.

Mitchell, D. (1990) 'Validity and practical reasoning', *Philosophy, 65*, 477–500.

Rabinowicz, W. & Rønnow-Rasmussen, T. (2004) 'The strike of the demon: On fitting pro-attitudes and value', *Ethics, 114*, 391–423.

Raz, J. (1999) Practical reason and norms, Oxford: Oxford University Press.

Searle, J. R. (2001) *Rationality in action*, Cambridge/Mass.: MIT Press.

Spielthenner, G. (2007) 'A Logic of Practical Reasoning', *Acta Analytica, 22(2)*, 139–153.

Spielthenner, G. (2008) 'The Logical Assessment of Practical Reasoning', *SATS – Nordic Journal of Philosophy, 9(1)*, 91–108.

Stevenson, C. L. (1944) *Ethics and language*, New Haven and London: Yale University Press.

Thornton, M. T. (1982) 'Aristotelian practical reason', *Mind, 91*, 57–76.

Wright, G. H. v. (1963) *The varieties of goodness*, London: Routledge & Kegan Paul.

Wright, G. H. v. (1983) *Practical reason: Philosophical papers (vol. 1)*, Oxford: Basil Blackwell.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 85–93, 2013

Joseph Raz on the Problem of the Amoralist

Terence Rajivan Edward

School of Social Sciences
The University of Manchester
t.r.edward@manchester.ac.uk

Abstract
Joseph Raz has argued that the problem of the amoralist is misconceived. In this paper, I present three interpretations of what his argument is. None of these interpretations yields an argument that we are in a position to accept.

1 Introduction

The problem of the amoralist is the problem of providing a person who is not moral with a reason to be moral. In his paper 'The Amoralist', Joseph Raz concerns himself with this problem, but he does not attempt to solve it. Rather the problem itself is his target (1999: 273-4). Raz implies that what is called the problem of the amoralist is not a genuine philosophical problem. He describes the problem as misconceived, as involving an illusory distinction and as arising from a false assumption. This paper offers three different interpretations of Raz's argument against the supposed problem being an actual problem. The argument is objectionable on all of these interpretations. Before introducing any of the interpretations, it will be useful to state Raz's conception of the problem.

An amoralist is a person who is not moral. They are without morality. Someone can be an amoralist while not even understanding what it is to be moral. Someone might also understand but see no reason to be moral, or so it seems at first blush. 'Why should I be moral?' this amoralist asks. The problem of the amoralist, as Raz conceives it, would be solved if we can give a good argument for being moral to the amoralist who poses this question.

Raz's conception of the problem also includes a proposition about what the significance would be of there not being any good argument. Morality would not be rationally defensible – would not be rational, for short:

> If one can be an amoralist then the validity of morality is undermined unless one can be amoral only because of ignorance or irrationality. Morality, the underlying thought is, is rationally defensible only if it can marshal arguments in its support which an amoralist must rationally accept. (1999: 273)

This quotation gives some inkling of what it is for morality to be rational. Raz adopts a provisional criterion for when a person has morality that enables a clarification of this matter. On this criterion, to be moral is to believe that each person is valuable in themselves, simply in virtue of being a person. For morality to be rational is for there to be a good argument for the truth of this belief (1999: 276). We will consider a proposal for the requirements that an argument must meet in order to be good later in this paper.

Raz's provisional criterion is questionable. Can one be a moral person just by having a belief? Does not having this quality also involve being disposed to act in certain ways? Towards

the end of 'The Amoralist', Raz himself subjects his provisional criterion to scrutiny and finds it wanting (1999: 299-301). It was introduced so that he could proceed with evaluating the problem of the amoralist with a greater degree of precision than if nothing had been said about what it is to have morality. Raz reassures us that what he has to say can be adapted for other understandings (1999: 274).

2 An Interpretation of Raz's Argument

In this section, I will present an interpretation of Raz's argument based on some of the statements he makes regarding how to interpret it. One of these statements also concerns another argument: an argument for why the rationality of morality depends on being able to provide an amoralist with a reason for being moral. After introducing this other argument, Raz tells us that he will attack a fundamental presupposition of it:

> True, this argument is too simple. It disguises many ambiguities, and it begs many questions. I will not, however, try to challenge it directly. Rather I will undermine its most fundamental presupposition. It sees morality as a separate domain. The amoralist stands outside it and refuses to go in. (1999: 276)

Raz thinks of the presupposition that he identifies as fundamental because it is impossible for there to be an amoralist unless the presupposition is true and the problem of the amoralist depends on this possibility in order to be a problem at all. He will therefore attack the problem by attacking the presupposition.

We can reconstruct Raz's argument as three premises and two inferences from these premises:

(1) The problem of the amoralist is only a genuine philosophical problem if it is possible for there to be an amoralist.

(2) A necessary condition for the possibility of an amoralist is that morality is a separate domain.

(3) Morality is not a separate domain.

From (2) and (3):

(4) It is impossible for there to be an amoralist.

From (1) and (4):

(5) The problem of the amoralist is not a genuine philosophical problem.

I have formulated this argument without clarifying what it means for morality to be a separate domain. I have also not explained Raz's grounds for denying that morality is a separate domain. These two tasks will occupy the rest of this section.

Raz describes the amoralist as standing outside of morality. This is a metaphorical description because morality is not a region of space or a physical entity that occupies a region of space, outside of which the amoralist stands. Raz's articulation of the presupposition that he attacks is also metaphorical: morality is a separate domain. If one thinks of the amoralist as standing outside of morality, then it seems that an attack on this presupposition will remove the possibility of an amoralist. But this will only be the case if the presupposition, when understood in less metaphorical terms, still captures a necessary condition for the possibility of an amoralist. I do not think that Raz's understanding of it fulfils this requirement. This objection will be made in the next section. At this point, I will present his understanding.

For morality to be a separate domain, on Raz's understanding, is for there to be moral interests and for there to be other practical interests, all of which are not only something other than moral interests but can also be pursued without being moral. Raz does not spell out that

this is his understanding, but it is suggested by the fact that much of his essay is spent arguing that there is a practical interest which, though not itself a moral interest, cannot be pursued without having morality. The closest Raz comes to spelling out his understanding is in the following passage, in which he compares his provisional criterion for having morality with an alternative criterion, according to which to have morality is to engage in a distinctively moral form of reasoning:

> What is common to the view that the mark of morality is acceptance of the principle that people are of value in themselves and to the suggestion that it is marked by the deployment of a special method of argument is a conception of morality as an autonomous area, distinct from other practical concerns. This assumption, seen in operation in Nagel's argument, and essential to all contractarian approaches to morality, though not only to them, explains how the amoralist is possible: he is someone standing outside morality and denying that there is a route, a rationally compelling route, which could lead him in. (1999: 302)

The word 'practical' is being used here in a very broad sense, in which it contrasts with theoretical and aesthetic. When Raz criticizes others for assuming that morality is distinct from other practical concerns, he means something more than that these others are assuming that there are practical interests that are not moral interests. He means that they are assuming that all of these practical interests also do not require having morality in order to be pursued.

Having considered Raz's understanding of what it is for morality to be a separate domain, let us turn now to what Raz has to say against it having this quality, in other words his justification for (3). Raz focuses on the interest of friendship. We can divide what Raz has to say about friendship into two parts. First, there is his reason for thinking that friendship is not a moral interest. Second, there is his reason for thinking that friendship nevertheless requires morality in order to be pursued. Regarding the first part, Raz is very brief. He tells us:

> Volunteering to work for Oxfam in one's spare time may be good for one because it is a morally good thing to do. But there is nothing specifically morally good about having friends. People bereft of friends may have a lonely and impoverished life, but they are not morally at fault. (1999: 295)

Raz talks about having friends in this quotation, but when he considers why friendship requires morality, he focuses on being a friend (1999: 285). I believe his thinking is best formulated throughout in terms of being friends. Adapting what is said in the quotation, his view might be that there is nothing morally good about being a friend, so being a friend is not a moral interest. Alternatively, it might be that a person who decides not to be friends with anyone cannot be morally faulted purely because of this decision, so being a friend is not a moral interest. Raz's reason for saying that being a friend is not a moral interest is less than fully clear. I will not dwell on this issue, however, as it is does not matter greatly within this paper.

Raz has much more to say on why being a friend, even if it does not count as a moral interest itself, nevertheless requires some sort of morality. Raz claims that to be a friend to another one must have concern for the well-being of the other, independently of what one can gain from their well-being, and that to have this is to treat the other as having value in themselves (1999: 287). Raz seems to associate this treatment with believing that the other has value in themselves, such that there cannot be this way of treating the other without the belief. He considers various objections from hypothetical amoralists to the claim that being a friend therefore involves having morality. I will present two of these objections. Firstly, could not someone be friends with another yet value them purely as a person who brings certain good

things into their life, without regarding the other as having value in themselves? Raz denies that this is possible:

> The more extreme amoralist... may say that he cares about his friend, treats him as a person of value, only when doing so serves his own interests in the friendship. This amoralist simply is not a friend to this other person. (1999: 288)

Another objection that Raz considers, from a more moderate amoralist, is that one can believe that one's friends have value in themselves without believing that each person has value in themselves. One can believe that one's friends have some quality which not everyone has – that they are all artists, for instance – and that it is this quality which gives them value in themselves. By Raz's provisional criterion for having morality, one can therefore be a friend without being moral. To counter this objection, Raz turns against his provisional criterion and subjects it to criticism. He says that we do not have sufficient grounds for judging that the supposed amoralist who values their friends in themselves because of some less-than-universal quality is actually outside of morality:

> The failure to identify a position which marks the moralist off from the (reformed) amoralist was a failure to find a way of reading 'people have value in themselves' which renders it both true and appropriate to be the mark of morality. (1999: 301)

I will pass over the readings that Raz considers of the claim that people have value in themselves, because there is an obvious objection to his argument, as interpreted in this section.

3 An Objection to the Argument

The argument ascribed to Raz in the previous section requires that we understand what it is for morality to be a separate domain in such a way that its being a separate domain is a necessary condition for the possibility of an amoralist. Otherwise (2) will be false. But Raz's understanding does not appear to fulfil this requirement. On that understanding, for morality to be a separate domain is for there to be moral interests and for there to be other practical interests, all of which are not only something other than moral interests but can also be pursued without being moral. However, if that is our understanding, then just one instance of a practical interest which is not a moral interest, yet cannot be pursued without having morality, would mean that morality is not a separate domain. However, in the case of a single exception, it seems that there could be an amoralist still. They would not be able to pursue that practical interest, but they would be able to pursue other practical interests. Even if there are many practical interests that presuppose morality, so long as a person does not need to pursue a morality-presupposing interest in order to live at all, there can be an amoralist.

Part of the difficulty with evaluating Raz's understanding of what it is for morality to be a separate domain is that the concept of a practical interest, or 'practical concern', to use his term, has not been sufficiently clarified. Raz does not clarify it in his paper. I am not sure how to clarify it myself, beyond giving examples of practical interests, such as maintaining one's health and having a fulfilling working life. What we can say is that there needs to be a response to the concern articulated in the previous paragraph, or else we cannot accept premise (2) of the argument. In fact, in the absence of a response, we ought to reject it, because the concern looks as if it will survive any plausible clarification of what a practical interest is.

4 A Second Interpretation and an Objection

There are ways of interpreting Raz's argument which render it immune to the objection in the previous section. A second interpretation begins with the observation that Raz, at certain

points in his paper, does not sound as if he is denying that there can be amoralists at all. Rather he sounds as if he is denying that there can be amoralists who are relevant to the problem of the amoralist. To be relevant to the problem of the amoralist, as Raz conceives it, a possible amoralist must have qualities which enable them to challenge the rationality of morality. Raz denies that all possible amoralists have these qualities:

> With Williams we can leave on one side the amoralist who has no concern for people, no friendships, no people he likes or is fond of, and who has no desire for such feelings, attitudes, and relationships... [His] life is so severely limited that—for reasons similar to those explained above concerning the person who denies any values—he poses no challenge to morality. The challenge is posed by an amoralist who can have a rich and rewarding life, while denying the value of people. Such an amoralist is like us in valuing friendship and companionship. (1999: 283)

On Raz's conception of the problem of the amoralist, it includes the proposition that if there is no satisfactory answer we can give to the amoralist who asks, 'Why should I be moral?' then morality is not rational. But in light of this proposition, the problem entails that there can be an amoralist whose life is not severely limited, or so Raz maintains. For only this kind of amoralist, he thinks, can challenge the rationality of morality.

In the quotation above, Raz suggests that if amorality cannot be combined with being a friend, then the amoralist's life is too severely limited to challenge the rationality of morality. For this view to have any plausibility, then being a friend has to cover other close relationship roles. In line with this point, Raz says that he is counting other close personal relationships as friendships (1999: fn. 22). With this expanded usage in place, we arrive at a second interpretation of his argument:

(1) The problem of the amoralist is only a genuine philosophical problem if it is possible for there to be an amoralist who challenges the rationality of morality.

(2) A necessary condition for the possibility of such an amoralist is that there can be an amoral life which is not severely limited.

(3) An amoral life must be severely limited.

From (2) and (3):

(4) It is impossible for there to be an amoralist who challenges the rationality of morality.

From (1) and (4):

(5) The problem of the amoralist is not a genuine philosophical problem.

The third premise of this argument is supported by the following sub-argument: if being an amoralist cannot be combined with being a friend, then an amoral life must be severely limited; being an amoralist cannot be combined with being a friend; therefore an amoral life must be severely limited. In the rest of this section, I shall object to the significance attached to being friends, on this interpretation of the argument.

There is a conception of what it is for morality to be rational according to which the fact that amorality cannot be combined with being a friend is irrelevant. On Raz's provisional criterion, having morality is a matter of having a certain belief: the belief that each person is of value in themselves, in virtue of being a person. One might think then that, since the belief that defines morality is not self-evident, for morality to be rational is for there to be a good argument for this belief. A good argument, one might further think, is an argument for the truth of this belief which is sound, which does not require having the belief in order to be understood and which cannot be reasonably objected to, whatever one's stance towards the belief is. If that is the case, then the rationality of morality depends on being able to provide

a good argument to the amoralist who asks, 'Why should I be moral?' But if we accept this line of thought, then the fact that being an amoralist cannot be combined with being a friend, supposing it is a fact, is irrelevant for setting aside the problem of the amoralist. We can still proceed along the same line of thought for answering the question.

It is precisely this line of thought which Raz considers as grounding the philosophical need to answer the amoralist:

> If one can be an amoralist then the validity of morality is undermined unless one can be amoral only through ignorance or irrationality. Why should one think so? The simple argument runs somewhat as follows. If morality is valid, that is, if people are valuable in themselves, then it is possible for people to come to know that. Moreover, it is possible for people who are amoral to realize that there are rationally compelling reasons to accept that people are valuable in themselves. (1999: 276)

If this is why the amoralist poses a challenge to the rationality of morality, it is altogether unclear why there would no longer be this challenge should being a friend presuppose morality. Raz questions the line of thought just quoted by questioning whether having morality consists in having the belief that people are valuable in themselves. But his questioning does not help overcome the objection that it is irrelevant whether or not the amoralist can have friends. He implies that to be moral is to believe, not that each person is of value in themselves, rather that each person potentially has this value (1999: 300). We can adapt the conception of morality being rational presented in the previous paragraph to cope with this position. For morality to be rational is for there to be a good argument, addressed to the amoralist as well, for the truth of this belief. The challenge to provide an argument thus remains whether or not the amoralist can have friends.

5 A Third Interpretation and Objections

Raz thinks that if an amoral life must be severely limited, then the amoralist cannot challenge the rationality of morality. The third interpretation of Raz's argument attributes the same main argument to him as the previous interpretation, but differs from it over the sub-argument for the third premise: that an amoral life must be severely limited. It does not say that the incompatibility between amorality and being a friend is enough, on its own, to entail that an amoral life is limited to this extent. Raz, however, suggests that the points he makes about friendship can be adapted for other things (1999: 283-284). There are many other potential aspects of one's life which can only be accessed by having morality. They too presuppose morality. It is in light of this fact, according to the argument on the third interpretation, that the only possible amoralist is one whose life is too severely limited to challenge the rationality of morality. Hence the problem of the amoralist is not a genuine problem.

Raz indicates that his argument is not to be understood as appealing exclusively to the amoralist's lack of the opportunity to be friends in the following passage:

> By examining the amoralist who has at his disposal the full range of goods by which his life can be enriched, and investigating the evaluative presuppositions of these goods we can—I will argue—demonstrate that there is no gulf between the moralist and the amoralist, and we can do so more securely and in a more far-reaching way than if we disregard these value-presuppositions in trying to extend the amoralist's sympathies and motivations. (1999: 284)

Raz's thought here is that if we try to conceive of an amoralist who has the opportunity to choose various potentially rewarding options in their life, while remaining amoral, our conception is incoherent because all or virtually all of these options presuppose having morality. The only possible amoralist is one whose life is too severely limited to challenge the rationality of morality, while the closest thing to an amoralist whose life is not that severely limited is a person within the horizon of morality. Raz puts his conclusion, somewhat obscurely, in terms of there being no gulf between the moralist and the amoralist. What he means is that only an amoralist whose life is not that severely limited is relevant to the problem of the amoralist, but a person with such a life is actually within morality.

Since Raz writes primarily about the relationship between having morality and being friends, in order for the argument on the third interpretation to work, we need to be able to apply his points about friendship to other potential aspects of a person's life. The result of this application needs to be that the only possible amoralist is one whose life is too severely limited to be relevant to the problem of the amoralist. I do not think that we can apply Raz's points about friendship to yield this result. When arguing that being a friend presupposes morality, because it requires concern for the well-being of one's friend independent of one's own gain, Raz sets aside something that might appear to be friendship but is actually something else. The person whose behaviour is superficially friend-like is nevertheless not a friend if they do not have unselfish concern for the well-being of the other. The setting aside of this person as not a true friend was noted earlier in this paper. It fits with how people talk. For instance, someone in a difficult situation may find that certain people, whom they previously regarded as friends, suddenly keep their distance, despite not being in any great danger. They might react by saying, 'These people were not really my friends.' Raz implies that to be a friend to the person would be to show concern for that person's well-being, in a way that is not calculated for one's own gain, and this implication seems to capture the thinking behind the reaction. Somewhat confusingly, he refers to the relationship which these people had with the person in difficulty as limited friendship (1999: 288). A more appropriate term, granting Raz's outlook, is pseudo-friendship. They are pseudo-friends with the person. (The question of whether this person was only pseudo-friends in return need not detain us.) Now being a pseudo-friend does not presuppose morality and so can be pursued by the amoralist. So although the amoralist misses out on being a friend, there is something that resembles it which is available to them. This leads to a worry about Raz's effort to establish that the amoralist's life is severely limited.

Raz's main example of something that presupposes morality is being a friend, but he thinks we can apply his point about this to other things. However, since with this example, there is something that resembles the morality-presupposing good which is available to the amoralist, we are left with the thought that the same will be true of other things that presuppose morality. For each potential aspect of a person's life that presupposes morality, or many of them, there will be something that resembles it which does not carry this presupposition and which is therefore available to the amoralist. For example, if philosophy presupposes morality, the amoralist cannot be a philosopher but they can be something resembling it, namely a sophist (Nussbaum 1985: 129-131; Nussbaum 1999: II). And so, while it may be that the amoralist's life strikes us as worse because of what cannot be accessed, we will not be able to dismiss this figure as having a life that is too severely limited to be of relevance to the problem of the amoralist, owing to such options being available to them. This is the worry.

Raz does not clarify the notion of a severely limited life. But at least one kind of amoral life that he describes obviously deserves to be thought of as severely limited:

> The amoralist does not believe in morality, either because he doubts its validity, or because he is not aware of it, or does not comprehend it. This does not mean, of course, that he does not believe in any values, in anything being valuable. That would reduce the amoralist to the level of an animal able to pursue its bodily imperatives only, a creature driven by hunger for food or sex, by the need to discharge bodily functions, and to protect itself from extremes of heat or cold. Such creatures pose no challenge to moral philosophy. (1999: 274)

According to Raz, the role that value beliefs play in action entails that one must believe that some things are valuable in order to lead a life that goes beyond the pursuit of bodily imperatives. An amoralist who does not believe this, and therefore only has these imperatives to guide them, obviously leads a severely limited life. Raz thinks that any person outside morality also leads a life that is too severely limited for them to challenge the rationality of morality, even if they do have some value beliefs. But the amoralist conceived above does not have a life that is so obviously severely limited. I think it is as close to the life which Raz describes as rich and rewarding as it is to the severely limited life he presents in this quotation. Furthermore, merely asserting that most other goods are like friendship in what they presuppose should not convince us that an amoral life is without its distinctive rewards. We cannot simply assume, as Raz invites us to, that there are no valuable life options that in turn presuppose amorality.

So far I have cast doubt on Raz's argument that the amoralist's life is too severely limited to challenge the rationality of morality by casting doubt on his claim that an amoral life must be severely limited. Raz does not do enough to secure this claim. In any case, one can grant the claim yet still object to Raz's argument on the third interpretation. The objection that was made in the previous section can be adapted for it. I imagine an amoralist articulating the objection like so: "To be moral is to believe that each person is valuable in themselves, in virtue of being a person. For morality to be rational, in the sense in which I am interested, there needs to be an argument that establishes the truth of this belief and this argument must be addressed not just to people with the belief but also to someone like me. Perhaps not holding this belief means that my life is severely limited. But that does not entitle you to dismiss the challenge I have set: either provide a good argument to me for the belief that defines morality or concede that morality is not rationally defensible. The only way to dismiss the challenge is to show that my criterion for morality being rational is mistaken and there is no reason to think that you can show this just by arguing that my life must be severely limited." Raz does not respond to the amoralist who thinks in this way, despite initially motivating the problem of the amoralist by appealing to their criterion of rationality. Without a response, his argument on the third interpretation also does not work.

6 Solving Versus Dissolving the Problem

It is tempting to ignore Raz's criticism of the problem of the amoralist and simply treat the material in his paper as an effort to solve the problem. The only other evaluation of his paper that I have encountered adopts this approach (Pedersen 2006: 59-60). According to it, Raz tries to provide the amoralist with a reason to be moral by saying that it is in their interests, so that they can enrich their lives with goods that presuppose morality. But, whatever the merits of this way of treating the material in his paper, Raz himself is opposed to it. Apart from insisting that the problem of the amoralist is his target, he also objects to this way of solving the problem.

The objection is made while operating with his provisional criterion for having morality. But Raz gives no sign of losing his investment in the objection when he subjects that criterion

to criticism, and it can be varied to fit with his final position. The objection is that establishing that it is in the interests of the amoralist to be moral may not establish that the belief which defines morality is true, when this is what needs to be shown:

> Some may try to show that since amoralists have to give up many goods which it is in their own interest to pursue, they can be true to their beliefs only at the cost of harming their self-interest. Perhaps it can even be shown that because of this anyone has a self-interested reason to believe in the value of people, though this step falls foul of my objection to Nagel. It does not prove the amoralist wrong. (1999: 282)

When the problem of the amoralist is clarified using Raz's provisional criterion, there is a kind of amoralist who regards the belief that defines morality as false. To provide that amoralist with a reason to be moral, one must argue that the belief is true, not or not merely that the amoralist's life will be worse off in the absence of the belief, otherwise one has failed to engage with their position (1999: 278).

Given what Raz says about appeals to self-interest, his effort to establish that the genuine amoralist must have a severely limited life, that anyone with a rich and rewarding life is actually within morality, is puzzling. For what difference does establishing this make? The principal value of establishing this point appears to be that it gives the amoralist a self-interested reason to be moral. It may be proposed that if the limitations of a life without morality are extreme enough, one cannot be an amoralist without having a severe mental disorder; then there would be no obligation to ensure that an argument for being moral is accessible to the amoralist. Raz does not explicitly say that this is his strategy against the problem of the amoralist, and what he does say is not sufficient for him to adopt this strategy. Certainly, there is a case to be made that an amoralist who can only pursue bodily imperatives is subject to a severe mental disorder. But we do not have reason to think that an amoral life must be limited to this extent nor do we have reason think that other amoralists are all subject to some kind of mental disorder, some kind which means that there is no need to justify morality to them.

References

Nussbaum, M. (1985) 'Sophistry about Conventions', *New Literary History*, 17, 129-139.

Nussbaum, M. (1999) 'The Professor of Parody', *The New Republic*, 22nd February 1999.

Pedersen, J. R. R. (2006) *"Why should I be moral?" A critical assessment of three contemporary attempts to give an extra-moral justification for moral conduct*, M. Phil dissertation, University of St. Andrews.

Raz, J. (1999) 'The Amoralist', in *Engaging Reason: On the Theory of Value and Action*, Oxford: Oxford University Press.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Volume 7, Number 1, pp. 95–109, 2013

The Rise and Fall of Disjunctivism*

Walter Horn

calhorn@rcn.com

Abstract

In the direct realist tradition of Reid and Austin, disjunctivism has joined its precursors in proudly trumpeting its allegiance with naïve realism. And the theory gains plausibility, particularly as compared with adverbialism, if one considers a Wittgensteinian line of argument regarding the use of sensation words. But 'no common factor' doctrines can be shown to be inconsistent with the naïve realism that has served as their main support. This does not mean that either disjunctivism or the Wittgensteinian perspective on language acquisition that informed it must be false. It does indicate, however, that linguistic arguments against private or internal meanings do not imply perceptual directness and that the espousal of direct realism—naïve or not—does not require adherence to disjunctivism.

Disjunctivism[1] is often associated with the proposition that for a subject S to have (or 'enjoy') a perceptual experience intensionally involving (or 'as of') some external object O, S must either actually perceive O or be in a condition in which it is 'merely to S' as if he were perceiving O, with the 'merely' indicating that both disjuncts cannot be true.[2] Of course, it is hard to imagine a philosopher of any description denying that someone who thinks he is seeing or hearing something is either really doing so or merely seems to be doing so, but is never (barring some Gettier-inspired scenario) both. As the point of the 'merely' is to require that the disjuncts are mutually exclusive, the entire assertion is nearly vacuous—a proposition that could be agreed to by the (hostile) sense-data theorist and the (sympathetic) direct realist alike. What *is* of consequence about the disjunctivist position is the claim that there need be no epistemically relevant[3] 'common factors' shared by both the veridical and merely ostensible perceptual experiences described by those disjuncts, that, e. g., there is no X such that X must be apprehended both by those who are hearing bells and those only seeming to do so (McDowell, 1982, 2002; Martin, 2009a; Dancy, 2009).

A variety of virtues, both epistemic and metaphysical, have been claimed for disjunctivism (Soteriou, 2010), among the most important being its alleged consonance with common sense and ordinary language. The view has been thought to allow philosophers and psychologists to join with plumbers, ballerinas, and civil engineers in entirely dispensing with sense-data,

*Thanks are due to Larry Tapper, Gerald Vision, and an anonymous referee.

[1]The position was pioneered by J.M. Hinton (1967: 217–27) and developed in Hinton, 1973. It was subsequently popularized by Paul Snowdon and John McDowell among others. For a diverse collection of works on the subject, see Byrne and Logue, 2009. See also Putnam, 1999.

[2]As will be seen as we continue, in spite of my use of 'intensionally,' I want 'seems' and 'perceive' to be taken in a quite broad sense. If there are entirely non-cognitive forms of perception (say that of infants or animals), disjunctivism, as I understand the position, is to apply to all of them. It is not similarity of belief, however tepid or fervent, that is important here, but similarity of what might be called (if it did not beg any questions) 'perceptual state.'

[3]Just what is meant by 'epistemically relevant' will be clarified as we continue. For the present, the main thing to understand is that the existence of an entirely causal commonality is not sufficient to produce epistemic relevancy.

where these are alleged to be entities that are (i) distinct from all physical items external to the perceiver (and from all surfaces of such items, if surfaces are not themselves considered physical); and (ii) such that the perception or other apprehension of them is indispensible to the perception of physical objects.[4] Thus, disjunctivism is in the tradition not only of Thomas Reid's 18th Century critique of the Locke/Berkeley/Hume 'way of ideas' and the early 20th Century 'multiple relations critique of sense-data (Moore, 1917, Dawes Hicks, 1917), but is also aligned with the later adverbial theory[5] Like the contemporary disjunctivist, these earlier realists were anxious to deny that when someone mistakes a blue door for a green one, there must be something (not necessarily mental, but certainly not the door) which is actually green. Reid and these later deniers of sense-data argued that any strictly philosophical inference[6] from an item's looking green to somebody to the existence of a (perhaps different) item that must really *be* green, is fallacious.[7] Partly because of this heritage, disjunctivism is commonly seen as akin to naïve realism: they both attempt to show that direct (i.e., with no epistemic intermediaries) access to garden variety physical objects can be consistent with the occasional manifestation of confounding perceptual simulacra. In fact, claims have been made by disjunctivists that *all* forms of direct realism (including naïve versions) entail the non-existence of any common factors to veridical and illusory perceptual situations (Martin, 2009a; Sturgeon, 2006). I hope to show in this paper, however, that, in spite of their shared antipathy to sense-data theories, disjunctivism and naïve realism are actually incompatible. This inconsistency results from the latter theory's acceptance of exactly the sort of common factor that disjunctivism denies can ever occur.

As I shall be assigning various views to the 'naïve realism' genus throughout this paper, it will be helpful to specify more precisely what I mean by that term:

> *Naïve Realism* = df. A theory of perception according to which the vast majority of our every-day perceptual judgments are correct because, on its view, physical objects external to and independent of perceivers generally have the perceptual properties (colors, feels, shapes, etc.) that they are perceived to have (although they may have many others as well).[8]

[4]Sense-data are also regularly claimed to be items with which we have 'direct acquaintance,' a relationship which is said to allow some of our beliefs to be absolutely certain or incorrigible. This characteristic is not obviously required for the interposition of sense-data in perception of physical objects to make all such perceptual experience indirect or representative, however, so the concept of 'acquaintance' will not be discussed in this paper.

[5]For accounts of adverbialism, see, generally, the works of C.J. Ducasse and Roderick Chisholm on perception. Perhaps the most comprehensive statement of the position may be found in Chisholm, 1957.

[6]Obviously, inferences can sometimes be made from non-green doors to, e. g., green light or green lenses, but not without additional empirical information, perhaps involving the scientific principles of optics. I take such inferences not to be strictly philosophical. A recent discussion of Reid's philosophy of perception and his manner of dealing with perceptual illusions may be found in Horn, 2010.

[7]These philosophers, such as Moore (1917: 245–52), sometimes added the claim that 'looks statements' are unanalyzable.

[8]A similar account was recently provided by Charles Travis and was attributed by him to Putnam as well. According to Travis, NR is, 'roughly, just the view that perception is awareness of one's surroundings; so that the objects of perception are, at least typically, what does in fact surround us – notably, objects, such as pigs and Marmite, and facts of things being ways they are, such as that pig's staring at one through the railings of its sty' (Travis 2005: 53). See also Devitt 1996. I don't wish to claim that there are no competing definitions of naïve realism about, only that mine is a traditional even orthodox definition that lacks the obvious defects of some modern competitors. For example, Snowdon (2005: 138) allows a naïve realist to consistently hold that he and everyone else is and has always been dreaming, since his definition doesn't require that any perception has ever been (to use his term) 'genuine.' Another theorist requires naïve realism to be a mereological theory according to which perceptual experiences would seem to require tops, bottoms and weights in addition to beginnings, endings and durations (Martin, 2009a, 2009b). One benefit of my definition, I think, is that it is consistent with both 'engulfment-type' theories like Martin's and intentional theories not requiring mereological analyses. Furthermore, while it does not explicitly require 'directness'

Since common sense is thought to take perception not to require such intermediaries as sense-data (Austin, 1962), naïve realism is considered a species of direct realism and, as such, a foe of phenomenalism and indirect realism, views according to which physical objects are either 'constructed out of' or inferred from sense-data, sensory states, or some other sort of non-physical items. One mid-20th Century champion of common-sense realism, Everett Hall, put the naïve theory this way: we must not infer from the 'undeniable commonsensible fact that we perceive tables and chairs out in the room, not in our heads' that there is some sort of 'law of projection' at work, somehow launching our interior ideas or images out into the world. When looking at a sheet of paper in front of him, the naïve realist will simply make 'the bold assumption that the only thing possessing the congeries of properties [we perceive] is the sheet of paper [rather than] look into our brains for them, or invent some unobservable mental events that display them' (Hall, 1959: 81). I think it is safe to say, then, that naïve-realism, like the disjunctivism that claims to be allied with it, is hostile to any theory according to which the perception of green doors generally requires the perception or other epistemic apprehension of any non-physical particulars. I hope to show, however, that the two views are mismatched in spite of this shared animosity.

From the time of Reid until the flowering of adverbialism in the mid-1950s, direct realist forces—including those troops relying almost exclusively on common sense and ordinary language for their arguments—had difficulty explaining hallucinatory or dream events. This is because, unlike ordinary errors or distorting illusions, when dreams or hallucinations of green doors are in play, there may not be anything in the vicinity of the ersatz perceptual experience even to *look* green.[9]

This is where both adverbialism and disjunctivism came to the rescue. S's perceptual experience may have no object whatever, yet remain sufficient for him to be truly said to have some such property as *being appeared to green doorly*. Similarly, it being merely to S as if he were seeing a green door does not (at least explicitly) require anything (green or not) to be present to anyone for S to be in such a perceptual state. So both techniques seemed to handle what might be called 'the problem of the unavailable particular.'[10] The crucial difference between the adverbial and disjunctive doctrines is that the latter does not allow even that there need be some epistemic state of the (ostensible) perceiver that obtains both when someone sees a green door and when he merely thinks he does.[11] That is, while the two schools agree that there need be nothing that even looks green when S has a hallucination of a green door, the adverbial theory suggests that S will at least be in the identical perceptual state (where this is construed as a psychological or epistemic rather than neural condition) when he sees a green door and

of perception (more on this later), the definition's externality and independence criteria would seem to make other suppositions more difficult to support.

[9]Dawes Hicks, 1924 offers one example of a pre-adverbial attempt to deal with hallucinatory experiences without recourse to sense-data. Alston (1990, 1999) is an attempt at a revival of a pre-adverbial 'theory of appearing.' I would classify Alston's position as a type of naïve realism, but one which, by seeming to 'particularize' content, handles hallucinations quite awkwardly–just as Dawes Hicks' earlier attempts did. For a good discussion of how the inference to sensa requires a misunderstanding of nature of intentionality, see Harman, 1990.

[10]Another approach to this problem has been to interpret perceptions largely in terms of the apprehension or (mis)attribution of universals. So, for example, it was held by Hall (1961: 29–37), that when a green door is either seen or hallucinated, various physical properties are taken as being exemplified. When it is a hallucinatory experience these properties aren't really exemplified, but only 'objectively present' or 'ascribed'. For a discussion of Hall's views on perception, see Horn, 2010.

[11]The term 'epistemic' is included here because the disjunctivist need not assert that there may be no brain or other physical state regularly, or even *always* common to S's two experiences. She can concede that, being a scientific matter, claims regarding the existence of those sorts of common elements are outside the scope of philosophy proper. What the disjunctivist denies is that there is any common perceivable (or 'apprehendable').

when he hallucinates/dreams/misperceives one. Disjunctivism thus goes a step further than adverbialism: by its lights, neither a green door, a possibly non-green other perceptual object, nor even a shared sensory state may be inferred from the existence of a hallucinatory occurrence—at least without the intervention of science. Put another way, for the Chisholmian, not only can both 'I am seeing a green door' and 'I am appeared to green doorly' be truly asserted by the same person at the same time, the being-appeared-to conjunct *must* obtain if the seeing conjunct is to do so (at least in that sense of 'seeing a green door' which requires the perceptual object to actually look like a green door). The experience that one enjoys when perceiving is thus being claimed to be in some sense identical in kind to what can occur when one is not. For the disjunctivist, of course, the mutual exclusivity of the veridical and merely ostensible disjuncts is taken to imply the impossibility of someone being in any psychologico-epistemic condition that is necessary to both the perceiving and non-perceiving states.

This purported advantage over adverbialism may seem like a solution in search of a problem if we consider the elimination of sense-data and the associated problem of the unavailable particular the only *desiderata* for the disjunctivist. In fact, however, there were other problems in the air when Hinton first wrote on this subject. These included several philosophical problems surrounding the meaning of sensation words that adverbialism not only seemed ill-equipped to address, but to which that theory seemed particularly vulnerable. This clue to disjunctivism's origin is more easily discerned if we consider the fact that the theory was developed at Oxford at a time when that school was very much under the spell of (the later) Wittgenstein. As is well known, for the Wittgenstein of the *Investigations*, the notion that ordinary judgments about the world—perceptual or otherwise—are synthesized from private mental states is as absurd and pernicious as any philosophical idea ever could be. Not only did Wittgenstein deride the idea of perceptual words having non-public origins, he also heaped scorn on the notion of any but nominal essences tying together classes of propositional attitudes. Since the adverbialists claimed that seeing a green door requires the obtaining of at least one of some particular class of mental states denoted by phrases like 'being appeared to green doorly,' a Wittgensteinian can be expected to demur based on the tenet that even if any such strange properties exist, neither any single example nor any particular group of them can be necessarily involved in the perception of a green door. On this view, seeing green doors involves a 'game' with rules as infinitely malleable as those involving *being honorable* or *acting silly*. We can, perhaps, capture all of some 'type' of perceptual experience under some phrase via stipulation, but there will be no 'real essence' referred to by general terms of that sort (Wittgenstein, 1953).

It can thus be seen why mid-20th Century Oxonians sought to toss such alleged properties as *being appeared to greenly* into the same dustpan into which green sense-data had earlier been thrown: neither batch of purported entities was thought to be of any use in the (to them, absurd) task of 'getting us to' physical objects. On their view, inferences to the external world neither are nor can be made from 'sensory states' any more than they are or could be made from sense-data. The Cartesian hope of getting from certainties within us to a physical world outside us, whether manifested by indirect realism or phenomenalism, was declared to be as false as the prior Anselmian hope of getting from a concept to a deity. But the Oxonians brought good news too: they offered a novel way to support the old direct realist claim that no inferences from inner states to outer things were ever needed in the first place, the claim that green doors are available to us directly.

If we climb the ladder of semantic ascent, we will be able to see how this new support was constructed. The Wittgensteinian argument (Wittgenstein, 1953: §293and *passim*) involved several related contentions, at least one being clearly empirical: the claim that acquisition of

perception-related 'mental words' involves a prior understanding of words for public perceivables (like green doors).[12] The assertion that descriptions of sensory states as well as those of sense-data are entirely 'parasitic' upon descriptions of physical objects was thought to be fatal to both phenomenalism and indirect realism, since it was believed that the order (in the senses both of classification and chronology) of being must parallel the order of understanding. This diagnosis of parasitism was sometimes thought to have been achievable by conceptual analysis—discovered solely through contemplation of the meaning (or uses) of phrases like 'green afterimage' and 'green door, and was sometimes held to follow from the truth of the empirical claim regarding language acquisition already alluded to. In either case, a linguistic argument had now come to the fore in the argument for directness.[13]

This perspective was not limited to Oxford. Wittgensteinian empirico-conceptual theses regarding the connection of language and perception were among the central tenets of some of the most important Anglo-American philosophers of the 20th Century. I here provide three examples—only one by an Oxford philosopher. First, here is Wilfrid Sellars:

> [T]he concept of *looking green*, the ability to recognize that something *looks green*, presupposes the concept of *being green*, and... the latter concept involves the ability to tell what colors objects have by looking at them – which, in turn, involves knowing in what circumstances to place an object if one wishes to ascertain its color by looking at it.

(Sellars, 1956: 274). W. V. O. Quine urged the following on the opening pages of *Word and Object*:

> Linguistically, and hence conceptually, the things in sharpest focus are the things that are public enough to be talked of publicly, common and conspicuous enough to be talked of often, and near enough to sense to be quickly identified and learned by name; it is to these that words apply first and foremost. Talk of subjective sense qualities come mainly as a derivative idiom... [I]mmediate experience simply will not, of itself, cohere as an autonomous domain. References to physical things are largely what hold it together. These references are not just inessential vestiges of the initially intersubjective character of language, capable of being weeded out by devising an artificially subjective language for sense data. Rather they give us our main continuing access to past sense data themselves...

(Quine, 1960: 1–2). Similarly, Strawson (1979: 43–4) claimed that any account of sensible experience which attempts to eliminate commitments to physical objects in favor of observers' 'subjective episodes' would nevertheless...

> embody or reflect a certain view of the world, as containing objects, variously propertied, located in a common space and continuing in their existence independently of our interrupted and relatively fleeting perceptions of them. Our making of such judgments implies our possession and application of concepts of such objects... [O]ur sensible experience itself is thoroughly permeated with those concepts of objects which figure in such judgments.

All three of these philosophers, in spite of their diverse positions on many issues, can thus be seen to agree on the Wittgensteinian thesis that the entire world of mental entities—whether

[12]There was no epistemic circularity involved in this assertion, since the words for publicly available items (like 'green' or 'door') were taken *not* to require any prior understanding of private, 'mentalese' terms.

[13]It should not be taken from this or what follows that I do not consider G.E. Moore's earlier paradigm case argument for common sense also to have been an important precursor of the disjunctivist defense of direct realism. And, of course, that argument also has a linguistic form.

considered as objects, contents, qualia, or acts—is describable only in a kind of pidgin physical object language.[14] Hinton and the other early disjunctivists may thus be seen as creators of a shorthand statement of the Wittgensteinian argument that progresses from premises regarding:

(i) the denial that real essences tie terms for various 'kinds' of perceptual experiences together; and

(ii) the necessity of the intersubjectivity of primary linguistic referents to the acquisition of any natural language

to a conclusion according to which even adverbialism yields too much to the forces of phenomenalism and indirect realism. As hallucinations are the derivative (or 'parasitic') items here, there can be nothing 'within' them which is not only found in veridical perceptual experiences as well, but whose apprehension is necessary for the latter experiences to occur.

With this linguistic argument added to the anti-sensa arsenal, we may, I think, characterize the main features of the current dispute over disjunctivism as follows: On one side much is made of: (i) powerful intuitions regarding the existence and 'incorrigibility' of a certain type of perceptual qualia—whether viewed as states, contents or objects—that are extremely hard to ignore or dismiss; (ii) the indisputable fact that most of us have been taken in by illusions or dreams at one time or another; and (iii) the reasonableness of an expectation that if the important proximate causes (the neurological conditions) of two perceptual events have identical characteristics, the effects will be identical in all important respects as well (Broad, 1914; Robinson, 1994: 151–62). On the other side we find: (i) a claimed consistency with common sense and ordinary language; (ii) an Occamist desire to avoid a category of entities that it is argued a correct theory of the world is better off without (especially where their main support is said to be Hume's fallacious argument from perceptual relativity); and (iii) a hypothesis regarding the derivative meaning of 'mental words,' backed by empirical claims regarding the acquisition of such words.

Although disjunctivism wasn't formally proposed until the late 1960s, if the above historical gloss is correct, the basic linguistic features of the quarrel between direct realists and their adversaries has not changed much since the publication of *Philosophical Investigations* or even, perhaps, since the circulation of the *Blue Book*. That being the case, is there no hope of any progress at this late date? I believe the answer here is 'No and Yes.' With respect to the main issue of the philosophy of perception, it is my view that there is unlikely to be anything resembling a conclusive resolution of any such basic 'categorial' dispute as that which has persisted for hundreds of years between the supporters of direct realism on one side and the supporters of phenomenalism and indirectness on the other. Certainly, direct realists have by this point made quite plain that, no matter what erroneous, dream, or psychedelic experience they are presented with, they will deny that it requires the existence of any non-physical entities, items whose apprehension on other occasions make it possible for us to veridically perceive such things as green doors. And the foes of directness have been similarly immovable. As with most age-old philosophical issues, further multiplication of cases here is unlikely to be of much use, in my opinion. [15]

[14]For the classic statement of this view, see, generally, Wittgenstein (1953), especially Part II. It has been suggested by Glock, (2003: 21) that Quine's views on this matter were actually derived entirely from Skinner. For a discussion of the similarities of a number of Skinnerian and Wittgensteinian views regarding language acquisition, see Day (1969).

[15]For a contrasting view, one according to which—in spite of hundreds of years of non-dispositive sparring—direct realism may by 'conclusive argument' finally be laid to rest and indirect realism crowned the final victor, see Coates. (2007: 62–98) and Fumerton (2006). For the contrary claim, that direct realism may be definitively demonstrated, see Armstrong (1961).

But with respect to the more limited disjunctivist assertion regarding the mutually exclusive natures of veridical and non-veridical perceptual experiences, the situation may not be quite so dire. Because disjunctivism occupies only the forward-most position in the dispute over the directness of perception, and because it relies on its close ties to naïve realism (Martin, 2009b; Snowdon, 1992; Campbell: 2002) it seems to me to leave open a legitimate chance of its own refutation. Surely, if one of the main supports of disjunctivism is its claimed consonance with our every-day picture of the world, a showing of inconsistency between the two viewpoints must at least be a serious blow. I believe, in any case, that it can be demonstrated that there are perfectly ordinary situations in which naïve realism is committed to common factors of a type that are antithetical to disjunctivism, and, there being so little else to support the doctrine, I think this inconsistency with common sense makes quite clear that disjunctivism can no longer be considered a serious contender in the philosophy of perception. The structure of my approach, however, is such that even if all of its claimed consequences are correctly drawn, it will not enable us to infer either that disjunctivism is certainly false or that the Wittgensteinian claims regarding the derivativeness of 'mentalese' or the implausibility of essentialism are inaccurate. We can get no further than the conclusion that if the linguistic premises are correct, either some additional premise was inserted by the disjunctivist or their inference to 'no common factors' is invalid. Certainly, the linguistic tenets alone seem obviously consistent with naïve realism.[16] This result it seems to me, is sufficient to allege that disjunctivism has fallen or should fall from any remaining philosophical grace.

I have claimed that disjunctivism is inconsistent with naïve realism and have defined the latter viewpoint, but I have not as yet attempted a formal definition of 'disjunctivism.' As I indicated above, simply repeating that one of the *perceiving/merely as if perceiving* disjuncts must always be false is not terribly helpful, since (given our understanding of 'merely') that fact would seem to be consistent with every theory of perception. We might, then, try to define the position by referring to the prohibition of required common factors between veridical and non-veridical perceptions. As we take this tack, we should remember that when critics have suggested such factors as neurological states or the property of *being indistinguishable by S from the state of affairs expressed by the other disjunct* as possible counter-examples, adherents of disjunctivism have been unmoved. They have simply responded with something like, 'Those aren't the kind of common factors that are dangerous to our theory' (Putnam, 1999). Clearly, then, our definition must indicate just what sorts of states of affairs *are* the dangerous ones, the ones whose existence could make the theory false.

The key here is first to recall the original *desideratum* of disjunctivism: dispensing not only with sense-data, but with sensory states or any other items claimed both to be epistemically (rather than only causally) required for the perception of some physical object, and also to be distinct from that object or any part of it. That is, the view was to be an example of a direct realist position that denies that there are any items E such that, in order for any physical object O to be perceived by a perceiver S at t, E would itself/themselves have to be perceived or otherwise apprehended by S at t, in spite of being neither individually nor jointly identical to O or any part of O. Such Es are the dangerous items, then, the entities the proof or discovery of which would be fatal to the position. So, when constructing our definition, we must remember that the distinctive feature of disjunctivism is not its (quite mainstream) disapproval of sensa, but its no-common-factor claim. That was Hinton's great contribution, the advance he made

[16]It may be instructive to consider that neither Sellars, Quine, nor Strawson took their common perspective on the derivativeness of mental terms to imply a no-common-factor theory of perception.

upon all his predecessors.[17] Thus, when we say that disjunctivism is a type of perceptual realism that will be false if there are 'dangerous entities' of the type explained above, we cannot restrict our antipathy to mental items; it is a view according to which there may be no required entity *of any type* which is 'relevantly common' to both actual and merely ostensible perceptual experiences as of the same physical object. But what is *relevant commonality*? I think it is this:

> An entity (or process) E is *relevantly common* both to some perceiver S's actual perception of some (intentional) object O and to his merely ostensible perceptual experience as of O = *df*. E is such that both (i) it is only in virtue of S's perceiving or otherwise apprehending E that S actually perceives O; and (ii) S's perceiving or otherwise apprehending E is sufficient for S to be in a condition that is merely to him as if he were perceiving O.[18]

The use of *sufficiency* in the second disjunct illustrates a remark made earlier, that the uses of 'perception' and 'apprehension' are intended to be quite minimalistic and broad: no advanced cognition is required for perceptual 'takings' as here understood. Apprehension of no more than the minimum deemed necessary for 'ostensible perception' is taken to be sufficient to make the second disjunct true. Such a tack allows for the only "difference" between the two experiences to be *veridicality*. If more than simple apprehension of E—some cognitive or attitudinal extra—were required to produce veridicality, the two disjunctive experiences might be internally distinguishable.

With this characterization in hand we can define 'disjunctivism' in such a way that we may actually be able to test its truth in various perceptual situations.

> *Disjunctivism* = df. A species of direct realism according to which for all ostensible perceptual experiences by S of O, there is no entity (or process) that is relevantly common to both (i) S's (actually) perceiving O and (ii) it merely being to S as if he were perceiving O.[19]

It is worth noting that neither *causality* nor *indistinguishability* seems to create any difficulty here. The definitions do not rule out the existence of one or more neurological states that might be common to the seeing of a green door and the hallucinating of one, since such states, not being apprehended, are not epistemically relevant. However important optic nerves might be to vision, we don't actually perceive them when we see something. Similarly, such properties as *being indistinguishable by S from the psychological state S is in when he is actually perceiving O* will not be relevantly common even if such properties are of necessity exemplified in illusory as well as veridical perceptual experiences. Again, while (correctly) perceiving my front door as purple and mis-perceiving it as blue may both require the existence and seeing of my door, such common element is neither necessary nor sufficient for any merely ostensible perceptual experiences of it as being one color or the other.

---

[17] Unless Husserl actually preceded Hinton in this matter (Smith, 2008). I am not competent to judge that claim.

[18] The 'merely' again indicates that S's perceptual experience is not veridical. It may be noted by some readers that I make no attempt to distinguish between hallucinations and non-hallucinatory illusions here, even though, for example, (correctly) perceiving my front door as purple and mis-perceiving it as blue may both require the existence and apprehension of my door. It is my view that the important dichotomy here is between veridical and merely ostensible. Since having the 'right object' available for perception is neither necessary nor sufficient to 'as if' experiences according to the disjunctivist, such objects cannot be accounted 'relevantly common'—in spite of the importance of their apprehension to the truth value of the first disjunct.

[19] Although disjunctivists have sometimes talked as though perception and hallucination involve two fundamentally antithetical and necessarily non-interacting worlds, in the spirit of charity, I will take the above definition to make only the most narrow interdiction: at any given time t, there can be nothing relevantly common between (i) S's having a veridical perceptual experience as of O at t, and (ii) S's only seeming to have such a perception at t.

It can be seen that the above definition of 'disjunctivism' includes a denial of the necessity of 'relevant commonalities' whatever their claimed ontological status, rather than simply denying the existence of sense-data or other mental items to which we are believed by some to have special perceptual access. It thus allows disjunctivists to plausibly deny that their theory is no more than a restatement of antipathy to sense-data and sensation-apprehension. That is, the definition has the virtue of allowing the possibility for the theory to make a legitimate contribution to the philosophy of perception, by going beyond all of its anti-sensa predecessors. I believe, in sum, that the disjunctivist has no grounds for complaint against this definition. With its use, however, I think it can be shown that disjunctivism cannot be true if naïve realism is.

It has been noted by both disjunctivists and non-disjunctivists alike (Alston, 1999; Martin, 2006; Brewer, 2008) that not all non-veridical perceptual experiences involve hallucinations or (we may devoutly hope) evil demons, realistic dreams, or the machinations of neuro-scientists, mad or otherwise. It may well be, in fact, that relatively few of them do. Let us, therefore, see how disjunctivism, as defined above, handles the following case, which, like Grice's well-known example of the reflected pillar (Grice, 1961; see also Tye, 2007) , involves both vision and a mirror, but is simpler and more commonplace, since it does not require a Gettier-type defeasor. Suppose that someone ('S') is in a restaurant he has never been in before and, while waiting to order (time t1), thinks he sees a waitress walk by in a room in front of him. What he is actually seeing, however, is an extremely life-like reflection of this waitress on the mirrored wall before him: she is really walking directly behind S.

Let us first describe this (I hope not terribly artificial) case disjunctively.

(1) S is either seeing the waitress walk in front of him or it is merely to S as if he were seeing the waitress walk in front of him.[20]

As we are concerned with the naïve realist take on this situation, let us next consider what a typical member of this clan (I'll call him 'Wesley') might say about these disjuncts after having S's dining experience described to him. Presumably we can expect Wesley to respond with something like this: 'Since the waitress is actually behind S, the first disjunct must be false, which would make the second one true.' As we have seen, this isn't too helpful, the key matter really being whether there are relevantly common factors at work here—but let us defer that question for the moment while we consider a subsequent perceptual experience S enjoys along with his meal.

Suppose that by the time dessert is served, a time (t2) by which S has come to realize that there is a mirror in front of him, he sees a reflection of the same waitress making another pass, and let us suppose that this second reflection is exactly similar to the one he observed when he first sat down. This time, however, because of his new understanding of the restaurant's decorative scheme, what S enjoys is a perceptual experience as of the waitress walking behind him. The disjunctive restatement thus now yields:

(2) S is either seeing the waitress walk behind him or it is merely to S as if he were seeing the waitress walk behind him.

[20]While it may seem odd to utilize neither an illusion nor a hallucination in this example, we should remember that what is important for our purposes is the hunt for common epistemic factors as between veridical and merely ostensible perception, not the particular manner in which the events in the second disjunct are classified. While there may be important differences between the predicaments faced by brains in vats, dreamers, LSD partakers, jaundice sufferers, hall-of-mirror visitors, evil demon victims, etc., the consistent disjunctivist must keep her eye on the prize—a denial of the necessary interposition of epistemically relevant common factors between accurate perceptions and merely 'as if' ones. Obviously, there is a continuum of types perceptual error, whether involving drugs, demons or reflective devices, and disjunctivism should be expected to handle all of them.

What would Wesley say about the truth values of the disjuncts in this case? Obviously, naïve realists cannot generally take the position that mirrors, windows, lenses, screens, fog, etc., prevent perception of physical objects from taking place: intervening media are regular facts of our perceptual lives. For philosophical proponents of common sense like Wesley, it is not the imposition of the mirror, but only S's ignorance of its placement in front of him that prevented him from seeing that the waitress was walking behind him at t1. In fact, most realists of any stripe would agree that S was already seeing the waitress *simpliciter* at that time; he was simply mistaken about her location with respect to his table. Thomas Reid, among the wariest opponents of claims regarding indirect perception in the history of philosophy, conceded that 'even a child gets the better of [the deception produced by mirror images] and knows that he sees himself only,' and added that, for those who understand optics, mirrors 'give just and true information' (Reid, 1785: I, i and II, xxii). Presumably, Wesley will join with Reid on this matter, taking the first disjunct to be true, and the second to be false, since the latter requires the non-veridicality of S's perceptual experience.[21]

I have postponed the central questions for Wesley here, those involving the existence of epistemically relevant common factors, because of a complexity involved in my example that I believe is better off avoided. Given the exact similarity stipulated to hold between the reflected images exploited by S at t1 and t2, a critic of the disjunctive paraphrases given above might deride the fact that they seem to imply that at t1, S has an experience as of the waitress walking in front of him that is indistinguishable from his actually seeing the waitress walk behind him at t2. I believe the disjunctivist would respond to any such imputation of oddness in something like the following manner: 'Of course it is not the case that it being to S as if the waitress is walking in front of him and it being to S as if the waitress is walking behind him are indistinguishable experiences! They are palpably different, and, in fact, this confused complaint does not even really depend in any important way on knowledge of mirror placement. For example, the scenario would work as well if S came to understand, not where the mirrors are, but the concept of *being behind* between his two viewings.' This disjunctivist might conclude by reminding us that, on her view, background knowledge—what perceivers believe and understand—is far from irrelevant to the perceptual process. And as we have seen, the position seems to have been created partly to reflect various insights expressed by a philosopher who was extremely sensitive to issues surrounding the concept of *seeing as* (Wittgenstein, 1953: Part II).

This is, no doubt, an interesting topic, but I believe it can be avoided for our present purposes. To escape these complexities, we must simply sift out the 'seen as' (or attributional) portion of S's experiences both at t1 and t2, and concentrate on his perception (or non-perception) of the waitress *simpliciter*, that is, without concern to anything he happens to perceive about her (Dretske, 1969). In a word, does S see her or doesn't he? Adopting this stripped down approach will also simplify consideration of what the naïve realist may be expected to say about S's experiences at t1 and t2, not of the waitress herself, but of her reflected images—for those, too, would seem to be ostensible objects of external perception.

Removing any aspect of (2) that involves attribution of the waitress's position relative to S, we get

(3) S is either seeing the waitress or it is merely to S as if he were seeing the waitress.

[21]On a stricter and less common-sense oriented view, one according to which a veridical perception of some object O can never depend on the aid of an intervening reflection, the truth-values of the disjuncts would, of course, be reversed.

Again, (3) is undeniable. But deciding which disjunct will be considered true and which false at each of the two times in question is a little trickier. As noted above, for common sense advocates like Reid, our knowledge of how mirrors work allows for the possibility of their involvement in veridical perceptions of what they reflect, at least in favorable situations. It may be, however, that while some naive realists (including Wesley) will hold that S therefore sees the waitress at t1, others (like Wesley's sister, Eve) may claim that veridical perception involving mirroring cannot occur without awareness of the interposition of any mirrors in play at the time. For the latitudinarian innocent like Wesley, then, the first disjunct of (3) will be true and the second false. For stricter naïfs like Eve, these truth values will be reversed. Presumably, at t2, both siblings will agree that the first disjunct is true and the second false.[22]

Let us now consider what may be said about S's perceptual relation, not to the waitress, but to the mirror images he either knowingly or unknowingly uses to be in a perceptual situation as of seeing his waitress (whether he's thought to really see her or not).

(4) S is either seeing the mirror images of his waitress or it is merely to S as if he were seeing them.

What will Wesley and Eve say about S's perceptual experience at t1? Does S, because of his ignorance of the mirror, fail to perceive these images or does he successfully perceive them without realizing it? Let us suppose that Wesley takes the position that S *does* see the mirror images, that the first disjunct is true and the second false. Eve, however, insists that, because of S's ignorance of what's going on around him, he is not really perceiving either the waitress or her reflection—though she likely will not deny that S is *somehow* apprehending these images. On Eve's view, then, both disjuncts of (4) are false at t1 because she takes it to be incorrect to say that S is enjoying a perceptual experience as of the reflections in the first place. On her view, S is not strictly in a perceptual situation with respect to the mirror images at all at t1, though he *is* in a non-veridical one with respect to the waitress.

At t2, the results can be expected to be different: both siblings will agree that the first disjunct of (4) is true and the second false. Since Eve's (perhaps unorthodox) view regarding 'perceiving unawares,' is no longer relevant once S understands the lay of the land, there will be no obvious realist account according to which S should not even be described as having an ostensible perception of the reflection.

I hope it is becoming clear that the responses of Wesley and Eve to the either-or statements presented to them are a prelude to a crucial failure of the disjunctivist approach—at least from the siblings' common-sense vantage points. Let's review. According to Wesley, at both t1 and t2, S sees both the waitress and her reflected images.[23] Eve agrees with Wesley about what is happening at t2 (S sees both the waitress and her reflection), but insists that at t1 S doesn't really perceive either the waitress or her reflection. Eve's take is that at the earlier time S merely had a perceptual experience as of the waitress being in front of him, and had no perceptual experience as of the reflections. She would not, presumably, insist that S has no epistemic relationship with the images whatever at t1, for if S did not apprehend them in *some* fashion at that time, his ability to recognize that, for example, the waitress is wearing the same clothes at t2 that she was wearing earlier, would be quite difficult to explain.

[22] As noted above, there may be even stricter theorists who would balk at the first disjunct even at t2, whether or not S knows of the mirrored wall at that time.

[23] We should not infer from this that Wesley's latitudinarian instincts result in him thinking that S was never wrong about the waitress or her reflection that night, however. Remember, he has already said that he believes that S had a non-veridical perceptual experience as of the waitress being in front of him at t1.

We are now ready to ask our prototypical naïve realists the key question: Is apprehension (whether by perceiving or in some other manner) of the reflected images of the waitress necessary both to S's seeing her (at those times at which they believe S does in fact see her) and to it being merely to him as if he were seeing her? Wesley and Eve will certainly both respond in the affirmative. As to the veridical disjunct, how else but with such images as those used by S could one with eyes in the front of his head see someone behind him? And, with respect to the 'merely as if' disjunct, all realists can be expected to hold that in order to be confounded by a mirror image, one must somehow 'take it in.' Surely, if one utilizes these mirror images in the way S does (in the aid of veridical perception or error, as the case may be), one must be apprehending them somehow: there cannot just be a causal relationship at work. The siblings may not agree about whether it is strictly true at each time that S *sees* the images, but they will certainly agree that he must have *some* kind of epistemic access to them without which he couldn't see the waitress at either time. For again, one thing common sense is absolutely clear about is that people may sometimes be aided in perception of physical objects by the interposition of undistorted mirror images (which is why periscopes are sometimes useful on golf courses, and automobiles are equipped the way they are), and, unlike, say, corrective lenses or ambient light, the reflected images are *cognitively available*. And their apprehension is sufficient to produce (indeed *is*) an ostensible perceptual experience.

Taking the single case of S's seeing or merely seeming to see the waitress at t1, it is clear that whether we believe that he sees her or is rather in the throes of an error-producing illusion at that time, S's apprehension of the self-same items (reflected images of the waitress) are required. This is inconsistent with disjunctivism as we have defined it. And, as indicated above, I don't think the definition can be altered in a manner that would eliminate this problem without either its failure to correctly represent the views of its supporters or its degeneration into a simple and old-fashioned insistence that sense-data (and affiliated) theories are false.

That mirror images are relevantly common to certain veridical and non-veridical perceptual experiences will likely give comfort to sense-data supporters, in spite of the physicality of mirrors and their reflections. 'Why,' they may ask, 'if mirror images may be both required and relevantly common to certain veridical perceptions and illusions, could not other things be so as well? As there are cases of indirect perception in our every-day lives, can anything ensure that there will be no circumstances in which we (perhaps unwittingly) use the sort of *mental* common factors that *all* direct realists object to? Of course, it is open to the direct realist to simply continue to deny the existence of any perceived, perceivable, or otherwise epistemically accessible non-physical entities in the perception of physical objects, in spite of examples involving mirrors, just as he formerly did when proffered examples involving dreams and hallucinations. Certainly, direct realism can be defined in such a manner that its truth or falsity will not depend on the fate of disjunctivism.

But where, exactly, has the disjunctivist gone wrong here? Why have her appeals to common sense and Wittgensteinian tenets regarding language acquisition not settled this matter in her favor? The problem, I believe, is that the disjunctivist has put upon these linguistic premises a burden other than that which they are designed to support. As nothing prevents language from being learned by children brought up in a hall of mirrors (or Platonic cave), it cannot be *indirectness* that must be proscribed according to the Wittgensteinian argument: it can only be *privacy/internality*. But disjunctivism is patently a promise of directness. The theory's proponents may lean on claims about how we learn observation sentences and mental predicates, but those claims are extraneous to its central doctrine, which is the categorical denial of indirectness in perception, based on the no-relevant-common-factor proposition.

It is also worth pointing out that no truly 'naïve' theory ought to have ever been expected to take an unequivocal position on what must always be true in perceptual experience in the first place. A common sense approach is unlikely to make any lead pipe guarantees of directness (whether or not as a result of some sort of claimed 'engulfment' of external objects by our perceptual experiences)[24] and should not be depended upon to do so by any more formal doctrine.

But if disjunctivism has been discredited, where does that leave direct realism? As indicated above, the non-disjunctive direct realist may decide to countenance relevantly common factors even between perceptions of green doors and hallucinations of them, so long as those factors are not non-physical items. Situations like those occurring at S's dinner may force a direct realist to concede that ordinary perception is not always direct, but as mirror images are physical, scientifically measurable items, it is open to him to consistently deny such ghostly stuff as sense-data and apprehended sensations, either through the use of adverbialism, the apprehension of universals, or through some other non-disjunctivist theory.

References

Alston, W. 1990: Externalist theories of perception. *Philosophy and Phenomenological Research,* 50, 73–97.

Alston, W. 1999: Back to the theory of appearing. *Philosophical Perspectives*, 13, 181–203.

Armstrong, D.M., 1961: *Perception and the Physical World.* New York: Humanities Press.

Austin, J. 1962: *Sense and Sensibilia*. London: Oxford University Press.

Brewer, B. 2008: How to account for illusion. In F. MacPherson, F. and A. Haddock (eds.) *Disjunctivism: Perception, Action and Knowledge*. New York: Oxford University Press.

Broad, C. 1914: *Perception, Physics and Reality*. London: Cambridge University Press.

Byrne, A. and Logue, H. (eds.) 2009: *Disjunctivism: Contemporary Readings*. Cambridge, MA.: MIT Press ['Byrne and Logue'].

Campbell, J. 2002: *Reference and Consciousness*. Oxford: Clarendon Press.

Chisholm, R. 1957: *Perceiving: A Philosophical Study*. Ithaca: Cornell University Press.

Coates, P. 2007: *The Metaphysics of Perception*. New York: Routledge.

Dancy, J. 2009: Arguments from illusion. In Byrne and Logue (eds.), *Disjunctivism: Contemporary Readings*. Cambridge, MA: MIT Press.

Dawes Hicks, G. 1917, reprinted 1938: The basis of critical realism. *Critical Realism: Studies in the Philosophy of Mind and Nature*. London: MacMillan.

Dawes Hicks, G. 1924, reprinted 1938: On the nature of images. *Critical Realism: Studies in the Philosophy of Mind and Nature*. London: MacMillan.

Day, W. 1969: On certain similarities beween the philosophical investigations of Ludwig Wittgenstein and the operationism of B.F. Skinner. *Journal of the Experimental Analysis of Behavior,* 12, 489–506.

Devitt, M. 1996: *Realism and Truth.* Princeton University Press.

Dretske, F. 1969: *Seeing and Knowing*. Chicago: University of Chicago Press.

Fumerton, R. 2006: Direct realism, introspection, and cognitive science. *Philosophy and Phenomenological Research,* 73, 680–695.

[24]'[T]he actual objects of perception, the external things such as trees tables and rainbows… partly constitute one's conscious experience, and hence determine the phenomenal character of one's experiences.' And, he admonishes, 'This talk of constitution and determination should be taken literally' (Martin (b), 2009: 93).

Glock, H. 2003: *Quine and Davidson on Language, Thought and Reality*. Cambridge: Cambridge University Press.

Grice, H. 1961: The casual theory of perception. *Proceedings of the Aristotelian Society*, xxxv, 121–52.

Hall, E. 1959: The Adequacy of a neurological theory of perception. *Philosophy and Phenomenological Research*, 20, 75–84.

Hall, E. 1960: *Philosophical Systems: A Categorial Analysis*. Chicago, IL.: University of Chicago Press.

Hall, E. 1961: *Our Knowledge of Fact and Value*. Chapel Hill, N.C.: North Carolina Press.

Harman, G 1990: The intrinsic quality of experience. *Philosophical Perspectives*, 4, 31–52.

Hinton, J. 1967: Visual experiences. *Mind*, 76, 217–27.

Hinton, J. 1973: *Experiences: An Inquiry Into Some Ambiguities*. Oxford: Oxford University Press.

Horn, W. 2010: Reid and Hall on perceptual relativity and error. *The Journal of Scottish Philosophy*, 8, 115–45.

Martin, M. 2006: On being alienated. In T. Gendler, and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press.

Martin, M. 2009a: The reality of appearances. In Byrne and Logue (eds.), *Disjunctivism: Contemporary Readings*. Cambridge, MA.: MIT Press.

Martin, M. 2009b: The limits of self-awareness. In Byrne and Logue (eds.), *Disjunctivism: Contemporary Readings*. Cambridge, MA.: MIT Press.

McDowell, J. 1982: Criteria, defeasibility and knowledge. *Proceedings of the British Academy*, 68, 455–79.

McDowell, J. 2002: Knowledge and the internal revisited. *Philosophy and Phenomenological Research*, 64, 97–105.

Moore, G. 1918–19, reprinted 1922: Some judgments of perception. *Philosophical Studies*. London: Routledge and Kegan Paul.

Putnam, H. 1999: *The Threefold Cord*. New York: Columbia University Press.

Quine, W. V. O. 1960: *Word and Object*. Cambridge, MA.: MIT Press.

Reid, T. 1785: *Essay on the Intellectual Powers of Man*.

Robinson, H. 1994: *Perception*. London: Routledge.

Sellars, W. 1956: Empiricism and the philosophy of mind. In H. Feigl and M. Scriven (eds.), *Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*.

Smith, D. 2008: Husserl and externalism. *Synthese*, 160, 313–333.

Snowdon, P. 1992: How to interpret direct perception. In T. Crane (ed.) *The Contents of Experience: Essays on Perception*. Cambridge: Cambridge University Press.

Snowdon, P. 2005: The formulation of disjunctivism: a response to Fish. *Proceedings of the Aristotelian Society,* 105, 119–127.

Soteriou, M. 2010: The disjunctive theory of perception. In *The Stanford Encyclopedia of Philosophy*. Winter, available from http://plato.stanford.edu/archives/win2010/entries/perception-disjunctive/.

Strawson, P. 1979: Perception and its objects. In G. Macdonald (ed.) *Perception and Identity: Essays Presented to A.J. Ayer with His Replies*. Ithaca: Cornell University Press.

Sturgeon, S. 2006: Reflective disjunctivism. *Proceedings of the Aristotelian Society*, 80, 185–216.

Travis, C. 2005: The face of perception. In Y. Ben-Menahem (ed.) *Hilary Putnam*, Cambridge: Cambridge University.

Tye, M. 2007: Intentionalism and the argument from no common content. *Philosophical Perspectives*, 21, 589–613.

Wittgenstein, L. 1953: *Philosophical Investigations*. New York: MacMillan.