Cognitive Primitives of Collective Intentions: Linguistic Evidence of Our Mental Ontology

NATALIE GOLD AND DANIEL HARBOUR

Abstract: Theories of collective intentions must distinguish genuinely collective intentions from coincidentally harmonized ones. Two apparently equally apt ways of doing so are the 'neo-reductionism' of Bacharach (2006) and Gold and Sugden (2007a) and the 'non-reductionism' of Searle (1990, 1995). Here, we present findings from theoretical linguistics that show that *we* is not a cognitive primitive, but is composed of notions of *I* and grouphood. The ramifications of this finding on the structure both of grammatical and lexical systems suggests that an understanding of collective intentionality does not require a primitive *we*-intention, but the notion of grouphood implicit in team reasoning, coupled with the individual concept *I*. This, we argue, supports neo-reductionism but poses difficulties for non-reductionism.

1. Introduction

In this article, we address the nature of collective intentions in light of results in generative linguistic theory. To frame the core problem, consider two people making hollandaise with one stirring and the other pouring. Their actions, stirring versus pouring, are simply enough described. However, the nature of their intentions is more complex. The pourer, for instance, does not merely intend to pour but intends, additionally, that a hollandaise sauce will result. The stirrer too must have this second intention. However, since making a hollandaise requires both actions, it may seem that the stirrer intends the pourer's actions and vice versa. What, then, are the individuals' intentions in such cases of collective action?

We thank Jason Alexander, Christian List, David McCarthy, Tom Smith, Robert Sugden and two anonymous reviewers of this journal for helpful input and are further grateful to the organizers and audiences of *Perspectives on We-Thinking* in Trento, *Nudge: Joint Action Workshop* in Warwick, the Choice Group at the London School of Economics, and the members of the 2009-10 Fellows' Reading Group at the Center for History and Philosophy of Science, University of Pittsburgh, where preliminary versions of this material were presented. Gold's research was supported by an award from the Carnegie Trust for the Universities of Scotland. During the revisions to this article, Harbour was supported by the Arts and Humanities Research Council of the United Kingdom, via Project א (Atomic Linguistic Elements of Phi), Sub-Project ו (We And We-intentions), grant AH/G019274/1.

Address for correspondence: Natalie Gold, Department of Philosophy, King's College London, Strand, London, WC2R 2LS, UK; Daniel Harbour, Department of Linguistics (SLLF), Queen Mary, Mile End Road, London E1 4NS, UK.
Email: natalie.gold@rocketmail.com; harbour@alum.mit.edu

Mind & Language, Vol. 27, No. 2 April 2012, pp. 109–134.

Raimo Tuomela and Kaarlo Miller (1988) suggest a three-point account of collective intention that reduces '*we*-intentions' to individual intentions together with a network of mutual beliefs. Michael Bratman (1992, 1993), pursuing a similar intuition, argues that collective action requires 'appropriate attitudes of each individual participant and their interactions' (1993, p. 99). But according to John Searle (1990, p. 404; see also Searle, 1995), 'no such reduction will work'. He presents a counterexample to show that there are things that are not shared intentions that nonetheless satisfy Tuomela and Miller's three criteria. Since then Natalie Gold and Robert Sugden (2007a) and Nick Bardsley (2007) have shown more generally that, according to the three criteria, every Nash equilibrium counts as a *we*-intention. Michael Bacharach (2006) makes the same point about Bratman's account. The key problem identified by these critiques of reductionist accounts is that coincidentally harmonized beliefs are misrepresented as *we*-intentions.

Searle's reaction to such difficulties is to claim that '"we-intentions" are primitive' (Searle, 1990, p. 404; we consider below what kind of primitiveness this might be, e.g., biological, psychological, or conceptual). In Gold and Sugden's account, grouphood is a primitive part of our mental ontology, that is, the collection of concepts humans have evolved to use to explain, plan and rationalize individuals' behaviour. *We*-intentions arise when reasoning about actions assumes a group subject (rather than the speaker reasoning just about him- or herself), the members of which (potentially) subsequently receive subsidiary tasks that accomplish the *we*-intention, that is, the intention of the group. On this view, *we*-intentions are not primitive.

Let us term Bacharach, Gold and Sugden's position 'neo-reductionism' and the contrasting position that could be associated with Searle, that *we*-intentions are primitives of our mental ontology, 'non-reductionism'. (The extent of Searle's commitment to non-reductionism is discussed in Section 5, below.) Different sets of primitives are demanded by or (in)compatible with these two positions. There are five primitives that might be found as evidence for or against neo- and non-reductionism: (a) *I*, (b) *we*, (c) grouphood, (d) team reasoning, (e) *we*-intentions. Neo-reductionism assumes (a) *I* and (c) grouphood as primitives of our mental ontology, together with the logical procedure of (d) team reasoning. Non-reductionism assumes the ontological primitiveness of (e) *we*-intentions. Moreover, where neo-reductionism is committed to the non-primitiveness of (b) *we*, which is viewed as a composite of the concepts of (a) *I* plus (c) grouphood, non-reductionism is not so committed. Therefore, if (e) *we* exists as a primitive of our mental ontology, then this counts as evidence against neo-reductionism and in potential favour of non-reductionism.

In this article, we turn to findings in generative linguistic theory and linguistic typology to assess the plausibility of the commitments and implications of neo- versus non-reductionism. We will consider three varieties of data, concerning the composition of first person pronouns, the syntax of such pronouns, and the structure of the lexicon. These, we argue, directly support the status of (a) *I* and (c) grouphood as linguistic primitives; by contrast (b) primitive *we* enjoys no such support. This means that, whereas the primitives of neo-reductionism enjoy extra

support from pronominal primitives, there are none that threaten neo-reductionism or favour non-reductionism. For (d) team reasoning and (e) *we*-intentions, we find no supporting linguistic evidence. In the case of (d), this appears to be for independent reasons: languages never mark means of logical inference or reasoning. However, for (e), there appears to be no independent reason why shared intentions could not be overtly marked. This is, therefore, a surprising lexical lacuna if *we*-intentions are indeed primitive to our mental ontology. The balance of linguistic evidence thus supports the neo-reductionist position against the non-reductionist.

It should be clear that the case mounted below is one of plausibility and parsimony, rather than being a definitive demonstration of the untenability of one approach to *we*-intentions. Both neo- and non-reductionism appear equally well equipped to differentiate *we*-intentions from coincidentally harmonized ones (and are also equally compatible with the qualia to which Searle affords importance; section 4). Therefore, it is necessary to turn elsewhere for further evidence. We assume that, to the extent that philosophy of mind aims to understand the actual world studied by science (as, e.g., Papineau 2009 argues), it is desirable that our philosophical theories be consistent with our best scientific theories and data. Moreover, we assume that the study of cognition should, insofar as possible, be conducted as a unified whole, with the results of one area being regarded as relevant for the study of others, unless arguments are made to the contrary. So, even if linguistics may seem far from debates about *we*-intentions, we regard it as relevant in principle and in practice to researchers whose concern lies with the theory of collective intentions. And what the linguistic evidence shows is that the primitives of neo-reductionism enjoy independent corroboration, whereas primitives endangering neo-reductionism or supporting non-reductionism do not. Given the simple assumption that a theory with primitives deployable in explanations of a wide range of data is preferable to one that extends only to part of that data and that leads to unsubstantiated expectations elsewhere, we conclude that there is more support for neo-reductionism than non-reductionism.

Our argument is structured as follows. In Section 2, we present the neo- and non-reductionist accounts, as well as the interpretation of Tuomela and Miller in reaction to which they arise. We also demonstrate that both accounts distinguish *we*-intentions from coincidentally harmonized ones. In Section 3, we present the first two types of linguistic evidence—the structure and syntax of pronouns such as *I* and *we*—that pertain to the primitives posited by or (in)compatible with each theory. This presents evidence for the primitives *I* and *group*, and against *we*. Empirically, though, this constitutes quite a narrow slice of language. So, in Section 4, we consider the structure of the lexicon, where languages register many concepts that are irrelevant to the domains of morphology and syntax considered in Section 3. We argue that there are good grounds for believing that absence of primitive *we* and *we*-intentions from this domain entails their absence from our cognitive primitives more generally. Finally, in Section 5, we discuss the implications of these results not just for non-reductionism, but also for the various interpretations that Searle suggests for the notion of primitiveness in his works.

2. A Problem of Primitives

Let us begin by stating the problem and proposed solutions in greater detail. When two people make hollandaise sauce, with one stirring and one pouring, they are said to have a collective intention: that is, in addition to the pourer's intention to pour and the stirrer's intention to stir, both stirrer and pourer must have the intention to produce hollandaise sauce as the joint result of their individual actions. The collective intention is, therefore, more than the mere sum of their individual intentions, i.e. more than the stirrer's intention to stir plus the pourer's intention to pour. These two individual actions can only have the effect of producing a hollandaise if properly coordinated, and not, for instance, if the pourer is by the sink and the stirrer by the hob. Alternatively, two sauciers in neighbouring restaurants might well make batches of hollandaise at the same time, but the sense in which *We are making hollandaise sauce* holds true for them is quite different from that in the stirring-pouring scenario, as each is engaged in a separate activity, the outcome of which does not depend on the other's action: if one saucier fails, the other can still succeed, but if one of the stirrer-pourer pair fails, then so must the other. In sum, coincidentally harmonized actions are not joint actions and coincidentally harmonized intentions are not collective intentions.

Analyses of collective intentions attempt to articulate what differentiates the intentions associated with joint actions from other intentions. A dominant strand of research (e.g. Tuomela and Miller, 1988; Bratman, 1992, 1993) attempts to reduce collective intentions to individual intentions and beliefs, and the relations between them. Tuomela and Miller, for instance, suggest a three-part reduction, the essential features of which can be illustrated with respect to a two-member group, $\{P_1, P_2\}$.[1] Consider some 'joint social action', A, which comprises the subactions A_1 and A_2 for the respective individuals. According to Tuomela and Miller, P_1 has a *we*-intention with respect to A if:

(i) P_1 intends to do A_1,
(ii) P_1 believes that P_2 will do A_2,
(iii) P_1 believes that P_2 believes that P_1 will do A_1, and so on (1988, p. 375).

[1] Tuomela (2009) argues that Searle's and subsequent interpretations of Tuomela and Miller focus too narrowly on their three points and miss the intrinsically cooperative nature of *we*-intentions, which allows the Tuomela-Miller account to avoid the problem of coincidental harmonization. We repeat Searle's interpretation of Tuomela and Miller because it is a part of the historical dialectic and it makes obvious the problem faced by reductionist accounts of *we*-intentions. We use the terms 'non/neo-reductionist' in an historical sense, to refer to accounts that respond to Searle's critique of reductionism and, at this stage, our aim is to outline these two approaches that explicitly avoid the issue of coincidental harmonization. Hence, for the moment, we leave aside the issue of how much of neo-reductionism is implicit or explicit in Tuomela and Miller (1988). However, we return to the issue later, both at the end of this section and in Section 5.

Observe that (i) is an intention of P_1, (ii), a belief of P_1, and (iii), a belief of P_1 about another's belief. As the same beliefs are held, mutatis mutandis, by P_2, the account reduces *we*-intentions to individual intentions and a network of mutual beliefs.

A problem for such reductive accounts, first pointed out by Searle (1990), is that (i)–(iii) may be satisfied in cases where the P_i do not plausibly have any shared intentions. That is, such accounts 'overgenerate'. Searle's own example, concerning a business school, might reasonably be felt to be at the margins of likelihood,[2] thus leading skeptics to wonder whether counterexamples to reductive accounts only arise in similarly arcane cases. However, Gold and Sugden (2007a) show that the problem is quite general: (i)–(iii), or their analogues in other accounts, characterize every Nash equilibrium as a case of collective intentionality. In a Nash equilibrium, each individual's action is a best response to their true beliefs about the others' actions. Since these are intentional actions, this is equivalent to saying that each individual's intention is adapted to their true beliefs about the actions and intentions of the other. Thus the criteria for being a *we*-intention are satisfied. But, in many cases, the Nash equilibrium does not involve a collective intention. Reductionist accounts mischaracterize such cases.

We illustrate this with the stag hunt game (Figure 1). Consider two hunters who can hunt either stag or rabbit. Rabbit provides a small amount of meat, but can be caught by one person. Stag provides more meat, but requires two people to catch one. Thus, if one player hunts stag and the other, rabbit, the stag-hunter will go hungry. There are two pure-strategy Nash equilibria in this game: (*stag*, *stag*) and (*rabbit*, *rabbit*). The players each get a higher payoff in the (*stag*, *stag*) equilibrium. However, the (*rabbit*, *rabbit*) equilibrium has the property of 'risk dominance' (Harsanyi and Selten, 1988); intuitively, hunting rabbit is the safer strategy because, regardless of the other player's action, the rabbit-hunter will never go hungry. In

		Player 1	
		stag	rabbit
Player 2	stag	(10, 10)	(7, 0)
	rabbit	(0, 7)	(7, 7)

Figure 1 *Stag hunt*

[2] Searle (1990, pp. 404-5) shows that Tuomela and Miller's (i)–(iii) are satisfied in the following scenario: 'Suppose a group of businessmen are all educated at a business school where they learn Adam Smith's theory of the hidden hand. Each comes to believe that he can best help humanity by pursuing his own selfish interest, and they each form a separate intention to this effect; that is, each has an intention he would express as ''I intend to do my part toward helping humanity by pursuing my own selfish interest and not cooperating with anybody.'' Let us also suppose that the members of the group have a mutual belief to the effect that each intends to help humanity by pursuing his own selfish interests and that these intentions will probably be carried out with success. That is, we may suppose that each is so well indoctrinated by the business school that each believes that his selfish efforts will be successful in helping humanity.'

evolutionary models, the system generally tends to the risk dominant equilibrium in the long run. Suppose it is common knowledge between P_1 and P_2 that, in stag hunt games, players usually choose *rabbit*. Then the intentions involved in P_1's action are:

(i) P_1 has the intention to choose rabbit.
(ii) P_1 believes that P_2 will choose rabbit.
(iii) P_1 believes that P_2 believes that P_1 will choose rabbit.

Thus, according to Tuomela and Miller's three points, the risk-dominant Nash equilibrium constitutes a collective intention. However, this is wrong: (*rabbit*, *rabbit*) is clearly not a collective action involving a collective intention, as, in the sketched scenario, each player can hunt and catch their own rabbit separately. Moreover, by both choosing *rabbit*, each receives a yield of 7, whereas, if they had truly formulated a collective intention, it would surely have been to play *stag*, for a higher yield.[3]

Tuomela and Miller (1988) qualify the three-point account, saying that it only applies to situations where it can be presumed that the agents are engaged in a 'joint social action'. Tuomela (2009) clarifies how this would exclude Nash equilibria that do not involve collective intentions. Whether or not Tuomela and Miller's account is ultimately successful, examples like the stag hunt show that accounts of collective intentions need to avoid mischaracterizing coincidentally harmonized intentions as collective intentions. (Searle's business school example [note 2] provides a non-game-theoretic illustration of the same point.)[4]

Searle's response to the problem of overgeneration is to claim that no such reduction as Tuomela and Miller's will work. Bacharach (2006) comes to a similar conclusion about reductionism following his critique of Bratman's (1992, 1993) analysis. Instead, Searle regards collective intentions as 'primitive'. Indeed, in later work, he writes that 'Collective intentionality is a *biologically primitive* phenomenon that cannot be reduced to or eliminated in favor of something else' (Searle, 1995, p. 24, our emphasis). Bacharach, by contrast, suggests that the distinctiveness

[3] Whether (*stag*, *stag*) must involve a collective intention is a separate question. Sometimes a given pattern of behaviour can be intended either individually or collectively. For instance, there is a dispute amongst primatologists about whether packs of hunting chimpanzee act collectively or whether each hunter merely places himself in a position that makes him most likely to catch the prey, given where the other hunters are positioned (Tomasello, 2008).

[4] We do not discuss Bratman's account here because he explicitly differentiates it from the analyses provided by Tuomela and Miller and Searle, saying that his shared intentions are states of affairs made up of the interrelated intentions and beliefs of the people who share them; they are not intentions of a special kind, held by individual agents (Bratman 1993, p. 107). Gold and Sugden (2007a) address the relationship between Bratman's account, team reasoning, and Tuomela and Miller's and Searle's approaches. They argue that Bratman's analysis of shared intention can be understood as an account of group agency (a disposition to reason and act as a member of a group), which comes prior to group members' deciding how they will coordinate their actions so as to achieve their goal, that is, prior to the object of Tuomela and Miller's and Searle's analysis.

of collective intentions consists in being the result of a particular method of reasoning. Gold and Sugden (2007a) expand on this point, arguing that collective intentions are those that result from 'team reasoning', where the individual first determines what the group as a whole should achieve ('What should we do?') and then works out their part in the best team plan ('What should I do?').[5] On this view, in addition to team reasoning, reference to the group and to *I* are ineliminable parts of the process of forming a joint intention.[6]

Searle's non-reductionist and Bacharach-Gold-Sugden's neo-reductionist views are, at least prima facie, incompatible. Searle expands on his notion of primitiveness by saying we 'simply have to recognize that there are intentions whose form is: ''We intend to perform act A''; and such an intention can exist in the mind of each individual who is acting as part of the collective' (Searle, 1990, p. 414, double quotation marks added). On the Bacharach-Gold-Sugden view, collective intentions are not primitive parts of our mental ontology, that is, the collection of concepts by which individuals' behaviour is planned, rationalized, and explained; nor are they biologically primitive. (See section 5 for further discussion.)

Despite this incompatibility, both the non-reductionist and the neo-reductionist positions cope equally well with the problem of Nash equilibria. Searle says that *we*-intentions cannot be reduced to *I*-intentions because 'The notion of a *we*-intention, of collective intentionality, implies the notion of *cooperation'* (Searle, 1990, p. 414, italics in original). Given that P_1 does not intend to cooperate with P_2 to achieve (*rabbit*, *rabbit*), there is no collective intention on the non-reductionist account; and given that there is no reasoning by P_1 about what the group consisting of P_1 and P_2 should do, there is no team reasoning and hence no collective intention on the neo-reductionist account. So, for evidence that discriminates between the accounts, we may have to look beyond the domain of the current debate.

One obvious quarter from which such evidence might come is experimental psychology, which might be able to prove or disprove that we use 'team reasoning' as a mental process (see, for instance, Colman, Pulford, and Rose, 2008, and Guala, Mittone, and Ploner, 2009). A related, non-experimental approach would be to identify the cognitive capacities that are required by different theories and to investigate whether the amount of cognitive sophistication required is congruent with that possessed by the agents to whom the theory is supposed to apply (see Pacherie, 2011). However, an alternative to seeking support for the mental process proposed by Gold and Sugden (2007a) is to seek support for their primitives, *I* and

[5] See Gold, in press, for an account of how team reasoning could lead to cooperation in the stag hunt. Gold and Sugden (2007b) show how team reasoning could lead to cooperation in the prisoner's dilemma.

[6] Bacharach, couching his account in terms of rational choice theory, argues that both the group utility function and the individual utility function are primitive. Modulo talk of utility functions, this amounts, we believe, to much the same thing.

group. This is the approach we pursue in remainder of this article, drawing on findings from generative and typological linguistics.[7]

3. Linguistic Evidence of Our Mental Ontology

Our basic assumption, following Donald Davidson (1967, 1973), Richard Montague (1970), and much subsequent work, is that natural language is compositional and that the atoms of such composition cast light on the concepts that form part of our mental ontology (where knowledge of language and our mental ontology are understood internalistically following, for example, Chomsky, 1986). Humans share an inventory of atoms and of algorithms that operate on such atoms and it is the task of generative linguistics, as a branch of cognitive science, to discover what these are and, where possible, to explain why they have such properties as they have. (See Ramchand, 2008 for a recent, highly articulated application of this approach in the domain of verb meaning.)

Amongst these atoms are some that pertain to quantity and number and others that pertain to persons. It is these atoms that, in more complex combinations, yield what we normally call pronouns and, in those languages that have it, agreement. In the domain of person and number, where language after language attests relevant data in the shape of pronouns and/or agreement paradigms, the programme of generative linguistics is to find the set of shared atoms and algorithms that generate all systems that we see attested, whilst avoiding overgeneration, the prediction of unattested systems. Of course, it is only with respect to robustly attested, tightly circumscribed phenomena that we want to rule out generation of unattested systems. Person (to which a putative primitive *I* belongs) and number (to which a putative primitive *group* belongs) are precisely such well attested and circumscribed domains (Corbett, 2000; Cysouw, 2003; Bobaljik, 2008). Given the substantial theoretical investigation of this domain by linguistic theoreticians (e.g. Hale, 1973; Silverstein, 1976; Noyer, 1992; Harley and Ritter, 2002; Harbour, 2007, 2011b), natural language is an ideal testing ground for the primitives of the neo-reductionist account.[8]

[7] It is possible that non- and neo-reductionism engender quite different research agendas with respect to non-human collective action, particularly where such behaviour involves (in contrast to simple pack action) distinct, learned skills contributed by different expert participants, as has been documented in, for instance, hunting by dolphins and killer whales. Collective action as a biological primitive suggests a search for biological homologies across humans, cetaceans, and possibly other animals. Collective action as a form of reasoning, by contrast, suggests a search for the relevant cognitive subcapacities. Whether these research agendas would remain distinct in practice, however, is far from obvious.

[8] Atoms, as the innate concepts that we build our thoughts up out of, are not culture-specific nor are they learned (though their usage may exhibit a maturation phase). As such, distinctions such as those between *I* and others, or between count and mass, differ from those between mugs and jugs or computers and the internet which may involve prototypes and sets of beliefs.

Our main claim will be that there is substantial evidence for neo-reductionist primitives (Sections 3.1–3.2) and none for a primitive *we*-intention. If anything, there is evidence against the latter position (Section 4). The data we adduce in support of the primitives *group* and *I* centres on the form of the word 'we' in diverse languages (Section 3.1) and its behaviour in different grammatical constructions (Section 3.2). Our purpose is to present the data in such a way as to lead to a natural understanding of the primitives that are available to grammatical cognitive systems.[9]

Implicit in our reasoning is the assumption that primitives of grammar are available also to other cognitive systems, such as those implicated in intention sharing and action planning. There is, of course, the possibility of mismatch between the primitives available to grammatical and non-grammatical systems. One instance of this is the neo-reductionist concept of team reasoning: no language that we know of marks propositions that result from team reasoning in any special way. However, this is part of a much broader and, to our knowledge, absolute tendency that languages never mark mode of inference (such as modus ponens, *and*-elimination, etc.). So, this represents a primitive (of the reasoning or planning system) for which there is no linguistic evidence. In contrast, we do not believe that it is plausible to appeal to the possibility of such mismatch in defence of the non-reductionist position as, unlike modes of reasoning, jointness of action and intention is a category that can be linguistically indicated. Assuming *we*-intentions to be a primitive leads to incorrect expectations about the structure of the lexicon (Section 4). (Further possible interpretations of Searle's non-reductionist position are discussed in Section 5.)

3.1 Basic Arguments

Let us begin with languages in which the word for 'we' is obviously not simplex (i.e. is morphologically complex). If we look at how such languages construct their *we*-words, we find (a) that *we* is constructed from *I*, and (b) what is added to *I* in such cases is either a straightforward plural or an element meaning *group* or something similar. Thus, *I* and *group* are the primitives from which *we* is derived.

[9] A reviewer objects that the linguistic notion of *group* merely concerns pluralities, whereas the groups that engage in joint action are conceptually more complex. This disparity is not problematic. On Gold and Sugden's view, the specialness of joint intentions arises from making a plurality the subject of a reasoning process, to the result of which all group members, if they were to reason qua group member, would assent. Of course, a separate debate can be pursued on the notion of assent, asking, for instance, whether all members need to reason or whether it suffices for some to endorse the outcome of others' reasoning or their parcelling out of subsidiary tasks. However, this debate concerns a philosophical notion distinct from that under analysis here.

The pattern of adding a plural to *I* is well attested.[10] We present three examples from two different geographical areas (three distinct language families). In Mandarin Chinese, the word for 'I' is *wǒ*, which is contained in the plural counterpart for 'we', *wǒmen*. The same suffix is found in other plurals, such as *lǎoshǐ* (*men*) 'teacher(s)' and *xuésheng(men)* 'student(s)' (Chappell, 1996). In Vietnamese, formal and informal words for 'I', respectively, *tôi* and *mình*, are again subparts of the formal and informal words for 'we', *chúng tôi* and *chúng mình*. As in Mandarin, the added element, *chúng*, is found in other singular-plural pairs, such as *nó* '(s)he/it (non-adult)', *chúng nó* 'they (non-adults)' (Ngô, 1999). Finally, in Miskitu (a Misumalpan language of Nicaragua), the difference between 'I fell', *yang kauhwiri*, and 'we fell' *yang nani kauhwiri*, is the element *nani*. The same element is used to derive the plural from the singular, as in *aras (nani)* 'horse(s)' (Green, 1992).

Not all languages with constructed *we*-words add the plural to *I* to create *we*. Instead, some add a noun, or similar element, with a meaning like *group*. We give two examples, again from distinct language families and geographical areas. In Thai, *pûag rao* 'we' explicitly contains the word for 'group, party, community'. (The same word may be added to *káo '*(s)he' to create *pûag káo* 'they'; Becker, 2006.) In Japanese, the element added to the words for 'I', for instance, *boku* or *watashi*, to produce 'we' is *tatji*: *bokutatji*, *watashitatji* 'we'. More literally, the meaning of these terms is 'Me and my associates/group'. The same element may attach to proper names, like *John*, to produce *Johntatji* 'John and his associates/group' (Nakanishi and Tomioka, 2004). Thus, *we* is constructed, in some languages, from *I* plus some group-like element.

The evidence just presented creates a strong case for the idea that *we* is constructed out of two more basic notions, namely, *I* and the concept of plurality or grouphood—precisely the Gold-Sugden primitives. However, such constructed *we*'s are noteworthy for a second reason. The languages above add to the singular to create the plural (as English does for common nouns, like *pig(s)*). In other languages, by contrast, one finds nouns for which the plural is the primitive from which the singular is constructed, such as *moch* 'pigs' / *mochyn* 'a pig' in Welsh (Jones, 1991); *áá*

[10] By 'well attested', we mean that it is a linguistically significant minority pattern, not confined to just one or a few regions, languages, or families. A reviewer points out that Daniel (2005) counts 15% of sampled languages as following the pattern of adding a plural to *I*. Although far from a statistical majority, this number is reassuringly high as evidence for the morphological composition in question. (To make a proper assessment of the significance of this number, an issue orthogonal to our concerns, one needs to measure something different—which no typological study has, we believe, ever attempted, probably for reasons of tractability—namely, how frequently separate grammatical categories fuse into one morpheme rather than being realized by separate morphemes. For instance, the Latin dative plural of 'leader', *prīncip-um*, is in no morphologically straightforward sense the addition of a plural morpheme to the dative singular *prīncip-is*. This contrasts with, say, the Turkish dative plural of 'house', *ev-ler-e*, where the plural morpheme, *ler*, is clearly added to the dative singular, *ev-e*. To make sense of the figure of 15%, we would need to know the extent to which languages fuse versus the extent to which they separate person and number with other categories, such as gender and case, in order to establish a general baseline for fusion versus separation of person and number crosslinguistically.)

'trees' / *áádau* 'a tree' in Kiowa, a Kiowa-Tanoan language of Oklahoma (Harbour, 2007); and *sínkɪr* 'fish' / *sínkirr̂* 'a fish' in the Maasai dialect of Maa, a Nilotic language of Kenya (Corbett, 2000).[11] Given the possibility of constructing singulars out of plurals, we must recognize, as a logical possibility, that some languages might construct *I* out of *we* and a 'singularizer' like Welsh *-yn*, Kiowa *-dau*, or Maasai *-r̂*.

We have shown that, if we take the primitives of the neo-reductionist account and treat them as linguistic primitives, we are immediately able to characterize some well-attested linguistic patterns. By contrast, the neo-reductionist position might be threatened if *we* were also a basic entity: primitive *we* might be taken as the pronominal counterpart or entailment of primitive *we*-intentions, thus favouring the non-reductionist position. It is striking, then, that, in the extensive literature on pronominal systems (e.g. Corbett, 2000, 2006; Cysouw, 2003; Siewierska, 2004), and despite the existence of plural-to-singular derivations, no language has been found that derives singular *I* from plural/group *we*. Although some languages 'singularize' common nouns, no languages 'singularize' pronouns.

Now, the claim that *we*-intentions are primitive does not commit one to the claim that the pronoun *we* is primitive and that *I* derives from it. (One might, for instance, claim that *we* and *I* are both primitive, or that *we*-intentions are primitive and that pronouns are quite orthogonal; positions we argue against in Sections 3.2 and 4 respectively.) There is, nonetheless, a notable disparity in terms of how comfortably the two accounts of collective intentions sit with some fairly basic linguistic evidence. More importantly, basic principles of parsimony strongly support any account, the primitives of which can be directly imported into another domain. Given that the neo- and non-reductionist positions are equally able to distinguish coincidentally harmonized intentions from collective intentions, it is precisely with respect to such non-core data that they can most sensibly be evaluated. To defend the non-reductionist account by ignoring such data is contrary to standard scientific practice. (See Section 4 for empirical arguments against the non-reductionist position.)[12]

---

[11] We omit gender prefixes from the Maasai Maa examples, for reasons of simplicity. The phenomenon is further attested in the Uto-Aztecan and Semitic families, and, according to an audience member at LSE, in Dutch. So this appears to be a linguistically significant, statistically minor pattern. A possibly related phenomenon (see, e.g., Acquaviva, 2008) is the existence of morphemes that render non-atomic mass nouns countable, as in Arabic *qamḥ* 'wheat' / *qamḥat* 'grain of wheat', *baqar* 'cattle' / *baqarat* 'cow (head of cattle)'.

[12] A reviewer suggests that, if we talk about 'collective intentions' rather than '*we*-intentions', we would expect the pronominal correlate/entailment of non-reductionism to be the collective, that is, grouphood, rather than primitive *we*. Given that the pronominal evidence does not exhaust the empirical case against non-reductionism, this alternative is not lethal for our case. However, it is a questionable interpretation for independent reasons. First, as discussed in section 4, Searle places significance in the qualia associated with collective intentions. For there to be qualia, there must be an experiencer. Consequently, we are not just dealing with grouphood, but with speaker-inclusive grouphood, that is, with *we*. Second, a different reviewer suggests that it is not open to Searle to assume grouphood as a primitive given his

3.2 Non-Basic Arguments

Anyone reading this article will, of course, be aware that not all languages base their words for *we* on *I*, as this sentence illustrates for English. One might, therefore, think that one type of evidence is being preferentially treated: if the existence of constructed pronouns counts in favour of *we* not being cognitively primitive, then the existence of non-constructed pronouns should count in favour of *we* being primitive after all. As suggested in the previous two paragraphs, this position might be taken as natural pronominal corollary of the claim that *we*-intentions are primitive.

In fact, it is easy to explain this disparity. The phenomenon whereby a complex form (such as *we*) is not merely the pronunciation of its primitive parts (*I* plus plurality/grouphood), is known as *suppletion* (or *fusion*). Besides *we*, it affects plural formation in common nouns, such as *goose/geese* in English, *'išah* 'woman' / *našim* 'women' in Hebrew, and *bič'ni* 'sack corner' / *božˇdo* 'sack corners' in Archi (a Lezgian-Samur language of Dagestan). Furthermore, suppletion occurs in many grammatical domains beyond plurality. For instance, in English, the past tense of *go* is suppletive, *went* (not *goed*), as are the comparative and superlative of *good/better/best* (not *good/gooder/goodest*). The phenomenon of suppletion simply concerns an irregular relation between meaning and pronunciation: the plural of some nouns is not pronounced as noun+plural, but as an irregular, one-off, fused form (similarly for the past tense of some verbs, the comparative of some adjectives, and so on). This does not cause one to revise one's view of what primitives there are, but merely to recognize that meaning and pronunciation are not always perfectly correlated. Therefore the pronunciation of *I* and *we* is sometimes unrevealing of the primitives out of which they are built.

Interestingly, however, even in languages where *we* is not constructed from *I* plus *group*, one can still find evidence for *I* being the more primitive. We present two grammatical phenomena that show this. To do this, some simple concepts from linguistic theory are required. (Readers less concerned with linguistic detail should nonetheless briefly familiarize themselves with the basic theory immediately below, as it is relied on in Section 4.)

3.2.1 Some Basic Theory. The arguments presented below rely on two key concepts of theoretical linguistics: *features* and *underspecification*.

Features are the atomic units out of which pronouns (and other elements) are built. They are usually understood to be predicates, P, that may either be asserted [+P] or denied [−P]. Below, we require reference to two features: [±speaker] and [±singular]. These mean, respectively, 'does (not) contain the speaker' and 'is (not) singular'. Thus, in terms of these features, *I* is [+speaker +singular] and *we* is [+speaker −singular].

methodological individualism (for a discussion of the relation between Searle's methodological individualism and his account of collective intentionality, see Fitzpatrick, 2003).

Underspecification is a means of economizing on information. For instance, the feature matrices for *I* and *we* are oppositely specified only for one feature, [±singular]. So, they can be simplified in one of two ways:

(a) *I* = [+speaker] *we* = [+speaker −singular]
(b) *I* = [+speaker +singular] *we* = [+speaker]

The idea behind underspecified feature matrices is that the unspecified features are understood by default. In (a), for instance, *I* is understood as being [+singular], even though there is no overt feature signalling this.

The data from Mandarin and like languages shows that the underspecification in (b) cannot be correct, for a very simple reason. The pronunciation of *we* in all the cases is larger than that of *I*. As what is pronounced is the features, Mandarin-style *we* must contain more features than *I*. (If (b) were correct, we would expect *I* to be *we* plus something extra, which, as already observed, does not arise for these pronouns.)

Conversely, (a) derives precisely the relations we observe. If *wǒ* meant [+speaker +singular], one of two problems would arise. Either, *wǒ* could not be used to pronounce 'we' [+speaker −singular], as *wǒ* contains [+singular] but 'we' does not. Or else *wǒmen* would be contradictory, on the assumption that *men* means [−singular], for then *wǒmen* would mean [+speaker +singular −singular] = 'contains the speaker and is singular and is not singular'. If we assume, as in (a), that *wǒ* means just [+speaker], and if *men* means [−singular], then the pronunciation of 'we' [+speaker −singular] is *wǒ*, the pronunciation of [+speaker], together with *men*, the pronunciation of [−singular].

With these basics in place, we now turn to two different phenomena. In both cases, our presentation will be the same and will proceed in three stages. (1) We introduce the phenomenon. (2) We develop it as a diagnostic for underspecification, using first versus third person. Above, we considered only underspecification of number (singular versus plural). Applied to person, the notion of underspecification is, simply, this: If the pronouns *I/we/(s)he/it/they* are specified for [±speaker], then first person (*I/we*) must be [+speaker] and third person ((*s*)*he/it/they*) must be [−speaker]. However, we can treat third person as the default interpretation that arises in the absence of specification. It may, in consequence, be underspecified for [±speaker] (cf. Benveniste, 1966 amongst many others). To show that the phenomenon is diagnostic of underspecification means that it distinguishes between fully specified and underspecified pronouns.[13] (3) We apply the diagnostic to singular versus plural, to show that singular is underspecified, plural not. This demonstrates that *I* has a more basic representation than *we* and that *we* is constructed from *I* and other elements.

[13] A background assumption in this discussion is that languages may differ in point of which features they underspecify. In Section 3.2.2, we see that number may be underspecified in Romanian, where it interacts with the person case constraint, though not in French, where it does not. In Section 3.2.3, we see that number may be underspecified in Dhirari, and like languages, where it interacts with ergative marking, though not in Georgian, where it does not.

3.2.2 The Person Case Constraint. (1) The first phenomenon we are concerned with is a well studied and extremely well attested one. We illustrate it first with respect to English. English has two nearly synonymous constructions for 'ditransitive' verbs like *show, give, present, introduce*, and so on: the 'prepositional dative', *She showed them to him*, and the 'double object construction', *She showed him them*. Interestingly, when the direct object (the thing shown) is first or second person, only the prepositional dative is possible: for instance *She showed me to them*, but not *She showed them me*. The person case constraint is the restriction that in the double object construction, when a ditransitive verb takes two pronominal objects, the direct object (in this case, *me*, the thing shown) must be third person.

For certain technical reasons, the effect is somewhat subtle in English: the unacceptable sentences are, for some speakers, only mildly degraded. In most languages where it is attested, however, the aberrant sentences are far more robustly rejected. One such language is French. In sentences such as 'She showed them the book' and 'She showed me to the professor', 'them' and 'me' may be represented by the object pronouns *leur* and *me*. One therefore expects 'She showed them me' (i.e. me to them) to use both *leur* and *me*. However, *Elle me leur a montré* is in fact ungrammatical, and, as in English, a preposition (*à*, and a different form of 'them', *eux*) must be used: *Elle m'a montré à eux*. Although both the languages just discussed are Indo-European, the person case constraint is found across the world (see, e.g., Haspelmath, 2002).

(2) The person case constraint connects to underspecification on a wide variety of analyses (e.g. Anagnostopoulou, 2003; Béjar and Řezáč, 2009; Adger and Harbour, 2007). These essentially argue that the double object construction is able to cope only with a certain quantity of features and, if it is overburdened by the direct object, ungrammaticality results. That is, pronouns that are 'too big', in sense of having too many features, are ungrammatical as the direct object. We explained above the standard view that the first person must be specified [+speaker], but that third person may be underspecified. It follows, therefore, that the person case constraint is a diagnostic of underspecification: fully specified pronouns are unacceptable as direct objects (hence, the difference between the acceptable *She showed them him* and the degraded *She showed them me*).

(3) There is a core set of properties to the person case constraint that is invariant crosslinguistically (for instance, the unacceptability of *me leur*, them me). However, in some more subtle cases, languages do vary as to which combinations of pronouns are ungrammatical. One strand of this variation concerns number. It has been observed that, if speakers find a difference in acceptability between singular and plural pronouns, then it is the singular that is acceptable. So, for instance, Nevins and Săvescu (2008) show that, for some Romanian speakers, 'giving you us' is unacceptable, but 'giving you me' is not. In (2), we said that fully specified pronouns are unacceptable in such configurations. We can straightforwardly account for the

difference between plural and singular in Romanian, given that 'me' [+speaker] is underspecified for number, but that 'us' [+speaker −singular] is not.

Now, recall that we are concerned in this section with languages where *we* is not constructed out of *I* together with some plural or group-like element. Romanian *ne* 'us' is clearly not the pluralization of singular *mă* 'me'. Thus, we have evidence that, even when plural *we* is not overtly constructed out of *I* and a plural or group-like element, it is still non-primitive: it is the pronunciation of [+speaker −singular], which is, self-evidently, a combination of the primitives [+speaker] and [−singular].

3.2.3 Ergativity. (1) Subjects of transitive verbs in many languages receive special marking, a case known as the 'ergative'. The case is found, for instance, in Georgian (a Kartvelian, non-Slavic/non-Indo-European, language of the eponymous country). Compare (a)–(b) with (c):

(a)	*Gogo*	*šemovida*	
	girl	came in	
	'The girl came in'		
(b)	*Me*	*gogo*	*vnaxe*
	I	girl	saw
	'I saw the girl'		
(c)	*Gogo-m*	*portolaxi*	*gaprtskvna*
	girl-ERG	orange	peeled
	'The girl peeled the orange'		

Only in (c) is 'girl' the subject of a transitive verb; in (a), the verb is intransitive, and in (b), the verb is transitive but 'girl' is the object. In the first and second sentences, 'girl' appears in its basic form, *gogo*. However, when the subject of a transitive verb, it appears in the ergative, as *gogo-m*.

The phenomenon of ergative marking is, in many languages, person-dependent. That is, some persons receive it, others do not. In Georgian, for instance, first persons (*me* 'I', *čven* 'we') never receive ergative marking, but third persons (*gogo-m* 'girl-ERG', *gogo-eb-ma* 'girl-s-ERG') do.[14]

(2) Person-dependent ergative marking, although well discussed and documented in the typological literature, has not received as much analytic attention (in terms of features) as the person case constraint. However, one account (Harbour, 2006; Richards, 2010) ties it to underspecification. The idea is that subjects of transitive verbs must be fully specified for person. Consider a language in which third person is underspecified for [±speaker]. When a third person is the subject of a transitive

[14] The ergative can also be restricted to certain tenses or constructions. In Georgian, for instance, it is restricted to (tenses constructed from) the past tense.

verb, it will receive an extra feature, [−speaker]. In contrast, nothing will be added to first persons, as they are already specified as [+speaker]. The pronunciation of such added features yields what is traditionally labelled as the ergative. Hence, ergative marking occurs on the third person in such languages, and never on first. It therefore follows that, when ergative marking occurs only on some (pro)nouns, it is diagnostic of which are underspecified.

(3) Again, as for the person case constraint, the present phenomenon is relevant for our purposes because there are languages where number too is a factor in determining when ergative marking occurs. For instance, in Dhirari (a language of South Australia), the first person singular receives ergative marking, whereas non-singular first persons do not. Similar facts hold for Arabana (related to Dhirari), Gumbaynggir (a language of New South Wales), and Aranda (a language of the Northern Territory). This pattern can be easily captured if we claim, as above, that singular pronouns are underspecified in these languages. In consequence, when they occur as the subject of transitive verbs, the full specification requirement will force them to receive an additional [+singular]. The ergative in these cases is the pronunciation of this extra number feature. This provides a second instance that shows that *we* has a larger feature specification than *I*, from which it follows that *we* must be comprised of several features and so is non-primitive. Again, recall that we are concerned with evidence for the relation of *we* to *I* in languages where the former is not constructed from the latter. To illustrate, briefly, for Aranda (Strehlow *circa* 1944, pp. 91–2, diacritics removed), that ergative marking may reveal this even where the structure of the forms does not, observe that *jinga* 'I' does change in the ergative (to *ata*), that *ilina* 'we (two)' and *(a)nuna* 'we (more than two)' do not, and that neither of the first two (*jinga, ata*) is a sub-element of either of latter two (*ilina, (a)nuna*).

4. Cognitive Exhaustion

The evidence presented above converges on the view that *I* and grouphood are primitive notions and that *we* is constructed out of them. However, if we attempt to use this conclusion to decide between the non-reductionist and neo-reductionist accounts of collective intentions, there is an obvious counterargument to be faced, namely, that there is a primitive *we* concept, or *we*-intention concept, but that the grammatical systems concerned with the phenomena presented above do not have access to it. Absence of linguistic evidence for such a primitive does not evidence its absence from all cognitive systems (cf. the modularity hypothesis of Fodor, 1983).

This counterargument relies, in part, on the view that features used to represent pronouns in grammatical systems do not exhaust all our pronoun-like concepts. We argue in this section that there is strong evidence against this position. That is to say, personal pronouns amount to nothing more than pronunciations of the features manipulated by the syntax, semantics, and morphology. We present two arguments, showing that there is an exact fit between the features posited by linguists and the

pronominal inventories attested in natural languages. To explain the significance of this exact fit, and why it entails that *we* cannot be an independent pronominal primitive, we first discuss the nature of the lexicon, as a store for information that goes beyond purely featural content. In so doing, we articulate a view of the nature of the lexicon (see, e.g., Ramchand, 2008 for more detail), one corollary of which is that *we*-intentions are highly unlikely primitives of our mental ontology. (This is an important aspect of the argument for researchers who, concerned with the ontology of collective intentions, may wonder why they have wandered so deep into the domain of pronouns.)

At first glance, the claim that linguists' features and languages' pronouns match would seem unsurprising, as failure to match would indicate that linguists had not adequately accounted for their data. However, this is to misapprehend what features do in linguistics. If we look at the lexicon of any given language, we find words for many concepts: *cat, dog, fourteen, fifteen, blue, green*. It is, of course, a real task to explain the difference in meaning between these pairs of terms. However, the explanation of these differences does not rely on positing features, such as [±canine], [±even], [±primary]. Rather, features are only posited where there is evidence that a given distinction is made by the syntax or other grammatical systems. Non-featural distinctions, that is, ones that have no impact on the grammar, are said to reside in the lexicon (they are 'encyclopaedic' in the sense of Marantz, 1997; see Fodor, 1977 for an early formulation).

Let us explain this distinction in slightly more detail. To speak English competently, one must know the difference between *cat* and *dog, fourteen* and *fifteen*, and *blue* and *green*. However, the differences between these pairs are entirely irrelevant for syntax, semantics and morphology, the linguistic systems that depend on features, as we now illustrate:

Syntactic phenomenon: passivization. No language is known in which one can passivize verbs done to cats, but not ones done to dogs (that is, in which *The cat has been fed* is grammatical but *The dog has been fed* is not).

Semantic phenomenon: quantifier scope. No language is known in which *fourteen* may have wide-scope and narrow-scope readings, but *fifteen* only narrow-scope. That is, no language is known in which *All the girls know fourteen boys* might mean, in semi-formal notation, either $[\exists_{14} y\colon \mathrm{B}(y)]\ [\forall x\colon \mathrm{G}(x)]\ (\mathrm{K}(x, y))$ 'There are fourteen boys—Andy, Billy, . . . , Neddy, say—and all the girls in question know them'; or $[\forall x\colon \mathrm{G}(x)]\ [\exists_{14} y\ :\ \mathrm{B}(y)]\ (\mathrm{K}(x, y))$ 'All the girls in question know fourteen boys, but each girl's set of fourteen may be distinct'; however, for *fifteen* only the latter type of reading would be available, $[\forall x\colon \mathrm{G}(x)]\ [\exists_{15} y\colon \mathrm{B}(y)]\ (\mathrm{K}(x, y))$ 'All the girls in question know fifteen boys, but each girl's set of fifteen may be distinct'.

Morphological phenomenon: ability to agree (e.g. reflect the singularity/ plurality of a noun). Languages differ in whether, and when, adjectives agree: for instance, in German, the endings of the adjectives are different in *blaues/grünes Papier* 'blue/green

paper' versus *blaue/grüne Papiere* 'blue/green papers'; however, the adjectives are invariant in the English translations. Thus, German adjectives agree, but English ones do not. Such morphological differences are common. However, no language has been found where blue, and related hues, agree but green, and related hues, do not.

Being grammatically inert, the differences between *cat* and *dog, fourteen* and *fifteen*, and *blue* and *green* are not featurally represented. Instead, they are confined to the lexicon.

The counterargument with which we began suggests that pronouns and *we*-intentions might be like animals, numerals, and colours: they might be characterized by differences in meaning that are represented in the lexicon but not in the feature system. If this were the case, then failure to find featural evidence for the primitiveness of *we* might indeed still leave open the possibility that *we*, or *we*-intentions, are primitive in other cognitive systems. We reject this position for the following reasons. (We concentrate first on primitive *we*, from which the argument against primitive *we*-intentions emerges naturally.)

First, the most complex person systems that languages attest comprise four distinctions. We illustrate this with the dual-number pronouns of Tok Pisin (Papua New Guinea; Foley, 1986), which, as an English-lexified creole, makes these differences in meaning particularly apparent to English speakers.

first inclusive	*yu-mi-tu-pela*	(me + you)
first exclusive	*mi-tu-pela*	(me + him/her)
second	*yu-tu-pela*	(you two)
third	*em-tu-pela*	(they two)

There are, however, several pronominal meanings beyond these four that languages could plausibly distinguish in the lexicon. One such is 'you' where all addressees are present versus 'you' where only some are. Notice that this resembles the very frequent crosslinguistic distinction in object deixis: that near you, and that far from you, cf. Scots *that hill*, *yon hill*; and it would be a linguistically practical device for creating group cohesion, a major factor influencing linguistic usage (and the object of study of most of sociolinguistics). Extensive surveys have found no evidence of any such lexicalization (Cysouw, 2003; Bobaljik, 2008). More striking in connection to this study is the absence of a 'choric', or 'mass', *we*. A sentence like *We ran the race* can be true in two quite different senses: if the race was a marathon, then every member of the group ran individually; if the race was a relay, then every member of the group cooperated. Failure of one person in the relay scuppers the race; failure of one person in the marathon does not (cf. the pouring-stirring scenario of Searle, 1990 and the duet-singing of Bratman, 1992). Group supplication (in the form of prayer, or petition writing) is another scenario where the existence of such pronouns is plausible. However, extensive surveys have again found no language in which such choric or mass *we* is specially lexicalized (Cysouw, 2003; Siewierska, 2004). Yet, again, from a sociolinguistic perspective,

one can easily imagine the role such pronouns could have in establishing group identity and cohesion (though some authors dissent; see, e.g., Wechsler, 2010). The fact that such plausible pronouns do not exist argues strongly that the lexicon does not contain any pronouns beyond those that the linguistically relevant features permit.

However, a stronger result obtains, concerning the sets of pronouns that a language may contain. Obviously, not all languages make so many distinctions in their pronouns as Tok Pisin. English makes only three distinctions, conflating first inclusive and first exclusive into a general first person (*we*). Other languages have even more impoverished pronominal systems: for instance, Winnebago conflates first inclusive, first exclusive, and second person (i.e. English *we* and *you*) under *nee* and uses *'ee* for third person only. Let us call the English system a tripartition, and the Winnebago system, a bipartition. Logically, there are 6 possible tripartitions, and 7 possible bipartitions. Of these, only 1 tripartition and 2 bipartitions are attested.[15]

Harbour (2011a) models such variation by proposing two features. Logically, these generate five systems (there being four possible subsets of two features, with two possible orders of semantic composition, if both features are active). Harbour shows that each of these sets is used by some language. This result is important, as it means that the feature inventory generates only attested sets of pronouns—a non-trivial result that has eluded previous researchers (e.g. Noyer, 1992; Halle, 1997; Harley and Ritter, 2002). If, as the counterargument proposes, *we* were a separate, primitive concept, capable of independent lexicalization, then it could be added alongside any feature system. As we already have a feature set that generates all the attested systems, adding an extra primitive is unnecessary. Moreover, given that the feature set generates only the attested systems, adding an extra primitive predicts unattested systems. This is a major problem because, as said above, the avoidance of overgeneration is a principal objective when linguists propose feature systems.

We conclude that, in addition to not being a primitive of the grammatical system, there is no primitive, innate 'we' concept that is represented in the lexicon.

Moreover, the evidence suggests that there is no primitive '*we*-intention' concept in the lexicon. Implicit in the first argument against primitive *we*—that choric *we*

[15] A partition refers to the total set of distinctions that a language makes in the verbal or pronominal domains (for instance, English pronouns and the verbal paradigm for *be* constitute tripartitions). However, languages may *syncretize*, i.e. pronounce identically, distinct partition elements within a given paradigm (for example, the English verb *be* shows a tripartition, 1 | 2 | 3 ≅ *am* | *are* | *is*, but in the plural all partition elements are pronounced identically, {1, 2, 3} ≅ *are*). Some attested syncretisms are not possible partitions, that is, they do not occur across the verbal or the pronominal domain of language as a whole but only in isolated verbal or pronominal paradigms (for instance, {1, 3} | 2 is not a possible partition though the syncretism does occur, for instance, in German for a subset of verbs, viz., modals and preterites, e.g. singular *kann* | *kannst*, plural *können* | *könnt*). Given that syncretisms arise via different mechanisms from partitions (see, e.g., McGinnis, 2005), it is important to bear the distinction in mind, especially when approaching, say, Cysouw, 2003, where the concern is explicitly with syncretisms within paradigms, not simply with partitions.

is never lexicalized—is the assumption that useful concepts have a propensity to be lexicalized. It is striking, then, that literature on pluractionality and on the expression of intention has noted no language in which *we*-intentions receive specialized lexical expression. This certainly cannot be because the concept is not useful (non-useful concepts rarely spawn their own research domains), nor because it is so arcane that we are generally unaware of collective intentions (witness our ready ability to appreciate that the narrow interpretation of Tuomela and Miller's three-point account founders on Nash equilibria). It is genuinely surprising, then, that no language has different means of expressing our having a *we*-intention (to make hollandaise sauce cooperatively, say) versus our all having distinct individual intentions (to make separate batches of hollandaise). Rather, the means that languages employ to express collectivity are not specific to the expression of *we*-intentions; instead, they are more general elements that may be co-opted for such usage. For instance, *together*, though it may express the *we*-intention of *We're making hollandaise (together)* may equally characterize such non-intentional situations as *The shoes are lying together at the top of the stairs* and *I can't put the pieces back together* (similarly, *jointly* occurs in such non-intentional, but nonetheless related, uses as *jointly distributed variables* and *these origins lie jointly in social requirements of human groups and in the fecundity of liminal experiences*).[16]

This argument therefore replicates that made at the end of Section 3.1. It too shows that positing *we*-intentions as conceptual primitives leads to unsubstantiated expectations about the structure of natural languages. However, the current argument is much stronger and more general than the earlier version. The earlier argument concerned only pronominal correlates and compatibilities of non- and neo-reductionism. Here, however, the argument applies to the expression of *we*-intentions in language tout court, not through the narrow lens of pronoun structure, and we find that the purportedly primitive concept is unexpressed even where languages exercise their greatest expressive freedom with respect to concepts, namely, in the lexicon.

By denying that *we*-intentions are primitives or, indeed, entities in our cognitive representations of joint action, the neo-reductionist approach provides a natural account of this lexical lacuna. A non-reductionist might object that team reasoning is also unexpressed in the lexicon, so it seems that we are not treating neo-reductionism and non-reductionism symmetrically (since we do not claim that the absence of any marking of team reasoning is a problem for the neo-reductionist).

---

[16] Although no language distinguishes a special *we* concept, some languages do distinguish between collective readings and non-collective readings of verbs with plural subjects. Strikingly, where languages make this distinction, it is multiple individual intentions, not joint intentions, that receive special expression. A 'distributive' is added in order to indicate that each member of the plurality is acting individually. However, like the English word *together*, the distributive is not confined to expression of intention but may generally apply to non-intentional actions and to actions that are diffuse in ways other than being performed separately by the members of a group (for instance, by being performed at distinct times or locations).

However, as we said above, team reasoning is not expected to receive special lexical expression because, beyond the specialized world of logic and philosophy, means of reasoning are never lexically expressed: languages do not mark whether an assertion has been arrived at by modus ponens, ∃-elimination, and so on.

Finally, we recall an earlier statement, that the lexicon is a core locus for representation of cognitively salient entities. This is worth recalling because a non-reductionist might object that, if the lexicon provides no evidence of *we*-intentions tout court, then it cannot provide evidence of *we*-intentions as primitives of our mental ontology, which might suggest an implicit bias in the methodology pursued above. However, we observe simply that one does not know, before looking, whether languages lexicalize *we*-intentionality or not and so absence of *we*-intentions from non-technical lexicons (as opposed to those of analytic philosophy) is ontologically suggestive, not methodologically detrimental.

5. Implications of the Linguistic Evidence for the Primitiveness of *We*-Intentions

Although we have been arguing against non-reductionist approaches to *we*-intentions, the implications of the foregoing argument for Searle's position depend on what exactly he means when he says that *we*-intentions are primitive.

There is a weak sense in which the Bacharach-Gold-Sugden view is consonant with Searle's (beyond their rejection of joint intentions as networks of mutually shared individual intentions and beliefs). If we take Searle's 'primitive' to mean 'explanatorily prior', then there is no disagreement: before the individual intention (to stir or pour, or to hunt stag or rabbit), comes the collective intention (to make hollandaise, or to go stag-hunting together). Some of Searle's exposition might be seen as supporting the idea that he means 'explanatorily prior':

> The crucial element in collective intentionality is a sense of doing (wanting, believing, etc.) something together, and the individual intentionality that each person has derived *from* the collective intentionality that they share (Searle, 1995, pp. 24–5, his emphasis).

Searle does not specify how *I*-intentions are derived from *we*-intentions. However, both Gold and Sugden (2007a) and Tuomela (2009) discuss how this might be done. Gold and Sugden use schemata of practical reasoning. Tuomela (2009), expanding on the notion of 'joint social action' which supplements the three-point reduction of Tuomela and Miller (1988), explains that agents already have a 'joint intention' between themselves, that they accept the intention expression 'we will do X' (or 'we will do X together'), from which the individual *we*-intentions in their heads are derived, using inference schemas. So, if Searle simply intends 'primitive' as 'explanatorily prior', then his account is compatible with neo-reductionism, with the linguistic evidence and with Tuomela and Miller's account.

However, Searle's use of the term 'biologically primitive', as well as the fact that his is a counterproposal to (one reading of) Tuomela and Miller's account, strongly suggests that he intends a different, innate sense of 'primitive'. Moreover, he expands on his thesis that *we*-intentions are primitive by saying that we 'simply have to recognize that there are intentions whose form is: "We intend to perform act A"; and such an intention can exist in the mind of each individual who is acting as part of the collective' (Searle, 1990, p. 414; double quotation marks added).[17] This, combined with the claim that *we*-intentions are biologically primitive, suggests that Searle intends a sense of primitive that is at odds with the view that collective intentions arise via a reasoning process that presupposes *group* and *I* as primitives.

In *The Construction of Social Reality*, Searle gives his positive argument against analysing *we*-intentions in terms of *I*-intentions and other mental states. He says, 'There is a deep reason why collective intentionality cannot be reduced to individual intentionality': 'The problem ... is that it does not add up to a sense of collectivity' (Searle, 1995, p. 24). To understand this statement fully, one must bear in mind Searle's broader commitments in the philosophy of mind. He argues that mental states are, in part, differentiated by their 'qualia', or qualitative feel (Searle, 1992, pp. 41-3). Thus, *we*-intentions are 'primitive' because of their distinct phenomenology.

So understood, Searle's non-reductionism (in terms of mental states) is not necessarily opposed to Bacharach-Gold-Sugden neo-reductionism (in terms of concepts). Quite simply, we may grant Searle his phenomenology but claim that the peculiar qualia are associated with, for instance, having arrived at a plan via team reasoning or by having intentions involving the linguistic/cognitive primitive group. Whether or not *we*-intentions are analysable in terms of other mental states is a different question from whether or not they are analysable in terms of other, innate concepts and if the linguistic evidence, conjoined with the theory of team reasoning, shows that group is an ineliminable part of the formation of *we*-intentions, then this might be precisely how a sense of collectivity comes in.[18] If, however, Searle intends 'biologically primitive' to indicate parity between *we*-intentions and such cognitive primitives as *I* and *group*, then the evidence from linguistic theory and the structure of lexicon seems strongly to be against him.

---

[17] Similar notions feature in Tuomela's (2009) defence of his earlier account, which he claims has been misunderstood. The result is a position very close to Searle's in that agents 'accept the statement "We will do X" as expressing both their joint intention (when read collectively) and their individual *we*-intentions' and in that the account presupposes an 'irreducible' and 'preanalytic' notion of *we*-intention: 'A minimal intuitive idea here is that the participants are supposed to function as a group or as one agent and to appropriately coordinate or bind together their activities both in their reasoning and acting as group members' (p. 293).

[18] That said, a reviewer questions whether Searle's methodological individualism allows him to posit grouphood as a cognitive primitive. See Fitzpatrick, 2003 for a critique of Searle's position in this regard.

6. Conclusion

Theories of collective intentions must be able to distinguish collective intentions from coincidentally harmonized ones, an issue which we have referred to as the problem of overgeneration. Searle's response to the problem of overgeneration is to claim that *we*-intentions are 'primitive', a non-reductionist position. Bacharach, Gold and Sugden, in contrast, propose a neo-reductionist account, according to which *we*-intentions arise by schemata of reasoning employing the primitives *I* and grouphood. To evaluate the accounts, we have broadened the discussion of mental ontology to include the kinds of primitives that are relevant to linguistic theory. The neo-reductionist primitives find direct and diverse support from linguistic theory, which speaks strongly in favour of such accounts. A primitive pronominal concept *we*, which might constitute potential evidence for non-reductionism or against neo-reductionism, enjoys no corroboration from linguistic theory, and both it and the primitive posited by non-reductionist accounts, namely, *we*-intentions, lead to incorrect expectations concerning the structure and content of the lexicon. If Searle's notion of 'primitive' is intended to afford *we*-intentions the same foundational cognitive status as *I*, *group* and team reasoning, then we conclude that his account is problematic. If, however, his notion of 'primitive' applies only to the qualia associated with having *we*-intentions, then this is not only compatible with the evidence presented here, but might be better understood in its light: the qualia in question are those that attach to intentions derived via team reasoning, and the irreducible collectivity of such intentions (the fact that they are not the mere sum of individuals' *I*-intentions) arises from the irreducible role that grouphood plays their derivation.

Department of Philosophy, King's College London
Department of Linguistics, Queen Mary, University of London

References

Acquaviva, P. 2008: *Lexical Plurals: A Morphosemantic Approach*. Oxford: Oxford University Press.

Adger, D. and Harbour, D. 2007: Syntax and syncretisms of the Person Case Constraint. *Syntax*, 10, 2–37.

Anagnostopoulou, E. 2003: *The Syntax of Ditransitives: Evidence from Clitics*. Berlin: Mouton de Gruyter.

Bacharach, M. 2006: *Beyond Individual Choice*, N. Gold and R. Sugden, eds. Princeton, NJ: Princeton University Press.

Bardsley, N. 2007: On collective intentions: collective action in economics and philosophy. *Synthese*, 157, 141–9.

Becker, B. P. 2006: *Thai-English/English-Thai Dictionary*, 4th edn. Bangkok: Paiboon Publishing.

Béjar, S. and Řezáč, M. 2009: Cyclic agree. *Linguistic Inquiry*, 40, 35–73.

Benveniste, E. 1966: *Problèmes de linguistique générale*. Paris: Gallimard.

Bobaljik, J. 2008: Missing persons: a case study in morphological universals. *The Linguistic Review*, 25, 203–30.

Bratman, M. 1992: Shared cooperative activity. *The Philosophical Review*, 101, 327–41.

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

Chappell, H. 1996: Inalienability and the personal domain in Mandarin Chinese discourse. In H. Chappell and W. McGregor (eds), *The Grammar of Inalienability: A Typological Perspective on Body Parts and the Part–Whole Relation*. Berlin: Mouton de Gruyter, 465–527.

Chomsky, N. 1986: *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger Publications.

Colman, A. M., Pulford, B. D., and Rose, J. 2008: Collective rationality in interactive decisions: evidence for team reasoning. *Acta Psychologica*, 128, 387–97.

Corbett, G. 2000: *Number*. Cambridge: Cambridge University Press.

Corbett, G. 2006: *Agreement*. Cambridge: Cambridge University Press.

Cysouw, M. 2003: *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

Daniel, M. 2005: Plurality in independent personal pronouns. In M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (eds), *World Atlas of Language Structures*. Oxford: Oxford University Press, 146–9.

Davidson, D. 1967: Truth and meaning. *Synthese*, 17, 304–23. Reprinted in D. Davidson, 1984 (ed.), *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 17–36.

Davidson, D. 1973: Radical interpretation. *dialectica* 27, 313–28. Reprinted in D. Davidson, 1984 (ed.), *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 125–39.

Fitzpatrick, D. 2003: Searle and collective intentionality: the self-defeating nature of internalism with respect to social facts. *American Journal of Economics and Sociology*, 62, 45–66.

Fodor, J. 1977: *Semantics: Theories of Meaning in Generative Grammar*. New York: Thomas Y. Crowell Company.

Fodor, J. 1983: *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Foley, W. 1986: *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.

Gold, N. In press: Team reasoning, framing and cooperation. In S. Okasha and K. Binmore (eds), *Evolution and Rationality: Decisions, Cooperation and Strategic Behaviour*, pages to be confirmed, Cambridge: Cambridge University Press.

Gold, N. and Sugden, R. 2007a: Collective intentions and team agency. *Journal of Philosophy*, 104, 109–37.

Gold, N. and Sugden, R. 2007b: Theories of team reasoning. In F. Peter and H. B. Schmid (eds), *Rationality and Commitment*. Oxford: Oxford University Press, 280–312.

Green, T. 1992: Covert clause structure in the Miskitu noun phrase. Ms, Massachusetts Institute of Technology, Cambridge, MA.

Guala, F., Mittone, L. and Ploner, M. 2009: Group membership, team preferences, and expectations, *CEEL Working Paper*, 6–09.

Hale, K. 1973: Person marking in Walbiri. In S. R. Anderson and P. Kiparsky (eds), *A Festschrift for Morris Halle*. New York: Holt, Rinehart, and Winston, 308–44.

Halle, M. 1997: Distributed Morphology: Impoverishment and Fission. In B. Bruening, Y. Kang, and M. McGinnis (eds), *MIT Working Papers in Linguistics 30: PF: Papers at the Interface*. Cambridge, MA: MITWPL, 425–49. Reprinted in J. Lecarme, J. Lowenstamm and U. Shlonsky, 2003 (eds), *Research in Afroasiatic Grammar: Papers from the Third Conference on Afroasiatic Languages, Sophia Antipolis, France 1996*. Amsterdam: Benjamins, 125–50.

Harbour, D. 2006: A feature calculus for Silverstein hierarchies. Talk presented at *Harvard-Leipzig-MIT Workshop on Morphology and Argument Encoding*. Available at http://webspace.qmul.ac.uk/dharbour/.

Harbour, D. 2007: *Morphosemantic Number: From Kiowa Noun Classes to UG Number Features*. Dordrecht: Springer.

Harbour, D. 2011a: The audience: An obituary. Talk presented at New York University. Available at http://webspace.qmul.ac.uk/dharbour/.

Harbour, D. 2011b: Descriptive and explanatory markedness. *Morphology*, 21, 223–40.

Harley, H. and Ritter, E. 2002: Person and number in pronouns: a feature-geometric analysis. *Language*, 78, 482–526.

Harsanyi, J. and Selten, R. 1988: *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Haspelmath, M. 2002: Explaining the ditransitive person-role constraint: a usage-based approach. *Constructions* (www.constructions-online.de).

Jones, T. J. R. 1991: *Welsh: A Complete Course for Beginners*. Teach Yourself Books. Sevenoaks: Hodder and Stoughton.

Marantz, A. 1997: No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, and A. Williams (eds), *PWPL 4.2, Proceedings of the 21st Annual Penn Linguistics Colloquium*. Pennsylvania: University of PennsylvaniaWorking Papers in Linguistics, 201–25.

McGinnis, M. 2005: On markedness asymmetries in Person and Number. *Language*, 699–718.

Montague, R. 1970: Universal grammar. *Theoria*, 36, 373–98. Reprinted in R. H. Thomason (ed.), 1974: *Formal Philosophy: Selected Papers of Richard Montague*. New Haven, CT: Yale University Press, 222–46.

Nakanishi, K. and Tomioka, S. 2004: Japanese plurals are exceptional. *Journal of East Asian Linguistics*, 13, 141–79.

Nevins, A. and Săvescu, O. 2008: An apparent Number-Case Constraint in Romanian: The role of syncretism. Paper presented at the 38th Linguistic Symposium on Romance Languages, Urbana-Champaign, IL.

Ngô, N. B. 1999: *Elementary Vietnamese*. Boston, MA: Tuttle Publishing.

Noyer, R. 1992: *Features, Positions and Affixes in Autonomous Morphological Structure*. Cambridge, MA: MITWPL.

Pacherie, E. 2011: Framing joint action. *Review of Philosophy and Psychology*, 2, 173–92.

Papineau, D. 2009: The poverty of analysis. *Aristotelian Society Supplementary Volume*, 83, 1–30.

Ramchand, G. 2008: *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Richards, N. 2010: *Uttering Trees*. Cambridge, MA: MIT Press.

Searle, J. 1990: Collective intentions and actions. In P. Cohen, J. Morgan, and M. Pollack (eds), *Intentions in Communication*. Cambridge, MA: MIT Press, 401–15.

Searle, J. 1992: *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

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

Siewierska, A. 2004: *Person*. Cambridge: Cambridge University Press.

Silverstein, M. 1976: Hierarchy of features and ergativity. In R. Dixon (ed.), *Grammatical Categories in Australian Languages*. Canberra: Australian Institutes of Aboriginal Studies, 112–71.

Strehlow, T. *circa* 1944: *Aranda Phonetics and Grammar*. Number 7 in The Oceania Monographs, Sydney, NSW: Australian National Research Council.

Tomasello, M. 2008: *Origins of Human Communication*. Cambridge MA: MIT Press.

Tuomela, R. 2009: Collective intentions and Game Theory. *Journal of Philosophy*, 106, 292–300.

Tuomela, R. and Miller, K. 1988: We-intentions. *Philosophical Studies*, 53, 367–89.

Wechsler, S. 2010: What 'you' and 'I' mean to each other: person marking, self-ascription, and theory of mind. *Language*, 86, 332–65.