Comments on Seumas Miller's review of Social Ontology: Collective Intentionality and Group agents in the Notre Dame Philosophical Reviews (April 20, 2014)

Miller's review contains many misunderstandings and other mistakes and requires a critical response. Journal reviews like the present one in general cannot be responded to directly although there sometimes might be the need to have such alternative available to an author. This review is a case in point. Why? I will show that some of the points made clearly indicate that the reviewer has not properly studied the book and therefore is not up to the task in question. I mention some such cases below. (1) The review does not present a description of the book that would be helpful to a potential reader. (2) The reviewer does not properly discuss what is new in the book and does not really evaluate the new theses and results presented in the book. (3) There is a somewhat technical chapter in the book (Chapter 7 using in part game-theoretical tools) that contains important results concerning the conceptual and ontological irreducibility and emergence (with failure of definability and sometimes even supervenience) of we-mode states and actions to I-mode ones but yet is not adequately commented on.

I will next present some detailed criticisms. There are quite a few erroneous claims in the review – many of which I believe are due to the fact that the reviewer has not carefully read the relevant pages in the book and as a consequence attributes views to me that I don't endorse, and he fails to mention many of my central arguments for the views that I do hold.

The reviewer says that "Tuomela positions himself as a nonreductionist collectivist". This claim however needs some qualifications. Rather I speak of the I-mode/we-mode account and occasionally of weak collectivism (see e.g. pp. 29). As to group matters I defend weak conceptual and epistemic collectivism as well as "contingent" weak ontological collectivism that is functionalist with individuals and their interrelations playing a strong role (as groups of course cannot function without their members' activities). In my ontology actual social groups that can act as groups are regarded as real social systems capable e.g. of causally influencing their surroundings through their members. The members have real attitudes in contrast to the attitudes extrinsically attributed to groups (group agents) that are not directly biologically grounded. The groups as social systems and extrinsically attributed attitudes seem to be individualistically irreducible in an ontological sense at least in central cases. Note that group agents in some cases (e.g. in the case of corporations) have fictitious features (e.g. law-based ones) as explained in Section 7 of Chapter 2. Also note that conceptual reducibility and irreducibility

can be conceptually shown here and now, while ontological reducibility or irreducibility are generally a posteriori matters to be decided in casu by empirical research, and this is why I above speak of contingent ontological collectivism. My view of group agents collectivism, however, is not of the strong metaphysical kind that attributes a collective mind and collective consciousness to group agents in the manner of old German idealism (cf. von Gierke).

The reviewer claims on p. 1 that collective intentions in my account "are not ontologically existing, causally operative attitudes". This is not right in the case of joint intentions which are in effect relations between individual group members' we-intentions. The reviewer incorrectly and confusingly speaks of we-intentions and joint intentions in my sense as a group's intentions. To be sure, in Chapter 3 of my book joint intentions, we-intentions, and a group's intentions have been spoken about under the label "collective intention", but in my analytic treatment these three kinds of intention are always clearly distinguished from each other.

Probably the reviewer means by his notion of collective attitude only a group's attitudes. In my account they are quasi-mental attitudes that are *extrinsically* attributed to functional group agents. Yet in the case of correct attribution the resulting predictions have a good chance of being right, and this fact normally supports the reality of the attributed attitudes. This is also a partial answer to the usefulness problem of extrinsic "as if" attitudes in my account: When they causally lead to behavior, they are factually *explanatory* as I show e.g. in Chapter 2, pp. 46-53.

The reviewer does not discuss my view of group agents qua groups that can act as if they were full-blown agents – obviously a central topic in the book. For one thing, a group as an agent cannot function without its members functioning appropriately and, furthermore, it has explanatory power and function in a vicarious sense through its members *acting as group members*. (This is a notion that I have discussed and analyzed in detail in my book *The Philosophy of Sociality* both for we-mode and I-mode groups as indicated clearly in the present book in note 15 to Chapter 3 but the reviewer does not discuss my account.) Let me emphasize that while the group is conceptually the primary agent, it is the individual members that are causally and ontologically primary. Although this is discussed in the book, the reviewer takes no notice of this important matter.

The reviewer says on p. 1 that "Tuomela tends to equate individualism with methodological individualism". But this is not quite right (see pp. 9-13). To be sure, one of my targets in the book is the "strict" individualism incorporated in standard rational choice theory – and through it in those social sciences like economics that seriously rely on (standard) game theory and some of its modern

individualistic modifications. However, I also comment on individualism in other senses. On p. 10 of Chapter 1 I present a version that I call methodological individualism and which is broader than both strict individualism and typical classical methodological individualism. Still that version does not allow groups that are individualistically irreducible. (Cf. my discussion of, and arguments against, the conceptual and ontological individualistic reducibility of we-mode states and actions on pp. 11-12 of Chapter 1, at the end of Chapter 3 and in Chapter 7; my central arguments there are ignored by the reviewer.) Individualism even in this broader sense does not support (conceptually and ontologically) irreducible we-reasoning (understood in terms of identification with the group and a nondistributive "togetherness we", as I say). The presence of agentive groups and of we-mode we-reasoning (based on group identification) are the central elements in the argument of Chapter 7 against the all-encompassing functionality of the mentioned wider sense of individualism (that still accepts groups only in a reductive sense). The reviewer does not mention that result at all (see especially my discussion and interpretative extension of Bacharach's theory and "The Central Theorem", see especially Section 3 of Chapter 7).

My distinction between the (pro-group) I-mode and the we-mode is not based on attitude contents or on types of attitudes. One can therefore e.g. intend a content (e.g. to open the window) both in the I-mode and in the we-mode –but in many cases not at the same time (although one can act for a mixture of I-mode and we-mode reasons). The we-mode case may also be formulated from the intending participant's point of view to result in the intention expression "We will open the window" (that depending on the context can express a person's we-intention or several person's joint intention). This expression is taken to be accepted and the intention had by the *causal-ontological* intending subject "I" and where, so to speak, the primary *intending* subject from a *conceptual* point of view is "we", which also is operative in the *causal* realm.

To the extent the reviewer's question on p. 2 about what a mode is can be taken literally, my short answer is that the notion of mode is clarified on pp. 36-37 where I give an "adverbial " account of mode: Briefly, a mode is the *way* of having the attitude and acting on it. As to I-mode versus we-mode, see the pp. 67-72 for central discussion.

As to my terminology in the case of the mode-discussion, the reviewer is bothered by my writing about "private" mental states instead of using the term "personal". I agree that a better word would be desirable but have not found one so far (but note that "private" is used in a related way in AI). The term "personal" is not feasible here because both I-mode and we-mode mental states are personal

3

in the sense of being applicable to persons, and this creates ambiguity. The content and meaning of the word "private" is yet explained sufficiently clearly in contexts where intentions including pro-group I-mode intentions are analyzed, see especially p. 70 (that the reviewer ignores). Judging from the reviewer's examples, his view may be that there are simple linguistic criteria for distinguishing between individualistic and we-mode states. But that is not a tenable view as I show on pp. 71-72. Consider the reviewer's example sentence, "I intend to do my part of painting the house". This can represent an I-mode intention or a we-mode intention depending on whether I have agreed e.g. to paint the front of the house (*I-mode*) in the context of other participants or whether there is a tight group (with a nondistributive "togetherness we") as a member of which I intend to do my part (*we-mode*), assuming also that in this latter case the central we-mode criteria (group reason, collectivity, collective commitment) are satisfied in normal a case. Here one and the same linguistic phrase can represent either a we-mode action or an I-mode action depending on the situation. Note specifically that the pro-group I-mode is assumed to rely on a *distributive* "we" (in accordance with e.g. Michael Bratman's account in Chapter 6 of his 2014 book) while we-mode thinking and reasoning of individuals in my account relies on a *nondistributive* "we" and a strongly social "togetherness we". Contrary to what the reviewer claims on p. 3, the distinction between pro-group I-mode thinking and we-mode thinking is not one of degree already for this reason (cf. below). I-mode thinking –when not based on individuals' contracts or the like is more autonomous than we-mode thinking. Note, too, that the pro-group I-mode approach cannot accept individualistically irreducible groups – on pain of becoming collectivistic. (The present view of individualism agrees with the account to be found in historical accounts of individualism as used in scientific contexts; see e.g. the review in Udehn 2001.)

In this connection the reviewer also considers my account as applied to standard group situations, where according to my claim the primary intending subject in the we-mode case conceptually is the group (a group agent), "we" for the participants. These participants are taken to causally bring about the intended end as a group – while in the I-mode case both the primary conceptual and ontological intending subjects and the actual subjects operative in the causal realm are individual agents qua a kind of representatives of the group. (As said, the attributed quasi-mental states of the group agents tend to be correct if they tend to be corroborated through what the members do.) According to the reviewer this account involves mysterious collective agents (group agents). However, my discussions in Section 7 on group agents of Chapter 2 and Section 5 on irreducibility (and a kind of interactive emergence) of Chapter 3 make it clear that there is no metaphysical mystery involved in my functionalistic and pragmatic group agents.

In this context the reviewer also makes claims on the basis of some preliminary statements concerning the we-mode notions in the introductory chapter of the book seemingly without having studied the substantial later discussions (especially in Chapter 2). He thus says about my presentation of the basic markers of the we-mode as formulated in a preliminary way on p. 6 of the book that the account makes the I-mode/we-mode distinction one of degree. To the extent that I understand this claim it seems clearly wrong – e.g. the discussion concerning the Central Theorem in Chapter 7 shows that the notions are clearly (rationally and functionally) distinct. Furthermore, the we-mode framework is intrinsically social – partly in view of the nondistributive "we" that it involves, and the mentioned markers of the we-mode based on strong identification with the group that are not as such available to the I-mode approach as shown or at least argued in Chapter 2 (in sections not considered by the reviewer). Also this indicates that the "degree" view is not tenable.

As to collective commitment, the reviewer's critical points are not warranted in view of what I say especially in Chapter 2 (and in my 2007 book). Briefly, in my account commitment concerns a person's binding herself to some item (see my 2007 book, Chapter 1 and the present book, Section 6 of Chapter 2). The case of collective commitment is analyzed by analogy in the general case. One can understand binding oneself in a merely psychological sense that does not involve any normative features or that involves only instrumental normativity. In the group case we have collective commitment that of course can be merely instrumental as well but that anyhow does have some quasi-moral features ("don't let the others down"). As is done in attitude logic (cf. pp. 64-65), I accept that a strong notion of intention such as mine entails commitment and that analogously joint intention entails joint or collective commitment (p. 83). This kind of commitment need not involve stronger than instrumental normativity. However, in the case of group agents the situation is different as the members' collectively binding themselves to an item is grounded by the group that also gives authority backed by group pressure to the collective commitment to the group's ethos (its constitutive ideas and principles). The authority may be based on the instructions (intentions, recommendations, orders, etc.) that the group (e.g. through its leaders) gives. Therefore I speak of group-normativity in this strong case. This normativity is group-based and as such is not entailed by intending but only by intending as a member of a we-mode group in the strong, we-mode sense. I do not operate with moral commitment and with non-revisability, the reviewer's two alternatives. As to the latter, all commitments are revisable in my view (also in the reviewer's example of an intention-in-action, because a commitment is on conceptual ground revisable at any time, irrespective of the intender's real-life psychological incapacities).

I conjecture that I-mode intentions and other I-mode attitudes cannot be atomistically strengthened by the addition of one I-mode condition after another to result in we-mode attitudes (this is in effect also indicated by the Central Theorem in Chapter 7 by its requirement of the presence of irreducible groups and of the members committedly acting as we-reasoners in the we-mode case).

As to the reviewer's comment on what I call implicit agreements, he erroneously claims that I base it on convention, but I do not rely on social conventions in my account and do not say I do. (The account applies also to non-conventional activities: I do refer to Hume's rowing example in a note to Chapter 5 but do not analyze it in terms of convention.) What my account is based on is simply formation of a joint intention by the participants (see (IMPLAGR) on p. 145), and this kind of weak agreement need not as such involve stronger normativity than what joint intentions entail – however, in the group case group-normativity will be involved and works against the participants letting others down by free-riding.

In the section on social reasons, the reviewer discusses many issues that I do not theorize about in my book. He misunderstands my text by seemingly thinking that there is no I-mode acting as a group member (but see the references above). Note, too, that group agents are not a *conceptually* necessary ingredient in my theory – contrary to the reviewer's claim on p. 4 collective acceptance by the members of a group does not entail that the group must be a group agent in a stronger sense that it, if suitably organized, can act. On several occasions I point out that the reasons for postulating functional group agents lie rather in the fact that the use of the notion in a pragmatic and epistemic sense, helps a researcher to gather knowledge of the social world especially in the case of large groups and corporations and also helps to give explanations of some social phenomena as shown in see Section 7 of Chapter 2 on group agents.

The reviewer also comments on my distinction between the action reasons that groups (qua group agents) have and that ground the group members' reasons for performing their parts of the group's actions. His account is confused because he asks about how the reasoning process at the collective level (presumably meaning group level) is connected to the member level reasoning processes. The simple answer to this is that all reasoning is at bottom done by group members: groups qua groups do not think, reason, intend, believe or act except through their members reasoning as group members. This matter is presented and discussed many times in the book (cf. also the irreducibility argument on pp. 11-12).

As to the notion of supervenience, the reviewer says, contrary to my claims, that I think that this correlational relation often correctly describes the abstract connection between the group level and the individual level. I do not in fact claim so for the general case but rather the opposite (see Section 3 of Chapter 3 for my contrary claim), but I do think that it descriptively fits many relevant cases and give clarity to the situation. I agree with the reviewer that supervenience (such as the technical notion of global supervenience) that I briefly consider in the book as such does not give a micro-explanation (or anything of the kind) of the group-level. Supervenience as such is just a descriptive notion that correlates two levels.

The reviewer criticizes my view that in an ideal we-mode case (in the sense of Section 3 of Chapter 7) the members of a we-mode group avoid collective action dilemmas and that thus defection ideally is not possible. But it is worth noting that a kind of defection in the we-mode case (e.g. choice of the individualistically rational choice alternative D in a standard PD that leads the members to the equilibrium outcome DD whereas the group's choice entails the Paretian CC for the members) is possible in the sense of involving temporarily or permanently lapsing to I-mode thinking and acting (and thus giving up the group in a sense). The we-mode group generally views and describes dilemma situation from the group's perspective and its possible choices concern the joint outcomes (such as CC, CD, DC, DD in a two-person PD). In a simple two-person two-choice Hi-Lo game there are four possible outcomes although only the choice between the equilibria HiHi and LoLo here is significant. Classical game theory fails here unless the members can somehow agree on a certain joint outcome. Still as members of a we-mode group they will as default think on the basis of the group's perspective and jointly eliminate other choice possibilities than the Paretian HiHi outcome.

The present kind of group account applies to all collective dilemma situations and is especially useful in the case of situations with a Paretian solution as is the case in the Prisoner's Dilemma. There are many theoretical attempts to solve (or dissolve) Hi-Lo and proper collective action dilemmas but the team-reasoning or we-mode solution can be argued to be overall the best (cf. Bacharach, 1999). Recent experimental results indicate that, depending on the situation a substantial amount of participants do think in the present kind of weakly collectivistic way and reason in terms of what the group should do as a group and what the members should do as group members (see the references in Chapter 7 of my book). Given the difficulties in experimental research and in the interpretation of results (e.g. as mental reasoning processes are hard to study) we-mode thinking can still be argued to take place even when understood in a very idealized sense (e.g. (*) on pp. 196-197). (The reviewer

objects to my way of using idealized notions, but my way of doing philosophy often is to test an approach through its limiting cases –this is what idealization basically involves here.)

The central result of Chapter 7 (and Hakli, Miller, and Tuomela 2010) that we are presently discussing is based on a mathematical theorem by the economist Bacharach. This "Central Theorem", which the reviewer does not even mention, basically says when formulated in my terminology that the we-mode approach and the (pro-group) I-mode approach are functionally different in the specific sense that *they yield different action equilibria* in many tricky cases, and consequently we-mode reasoning and I-mode reasoning in some cases lead to different behaviors. The cases in which such differences appear include not only Hi-Lo but many other interaction situations and especially important collective action dilemma situations (such as the Prisoner's Dilemma). Contrary to Hi-Lo, these include the temptation to free-ride as individualistic game-theoretic reasoning rationally recommends mutual defection (non-cooperation) at least in cases where sufficient assurance that the other will choose C is lacking (note that CC is the group-beneficial joint outcome here). The we-mode account assumes that the members act as a group and adopt utilities on that basis (e.g. in terms of weighted sums of the payoffs). The aforementioned central theorem that involves the irreducibility result – is based on group agency (groups that can act) and group utilities as well as we-reasoning by the members (possibly with some probabilistic uncertainty about the others' action tendencies). It shows the rational-functional difference between the we-mode and the pro-group I-mode accounts at least for some cases. This result is not even mentioned by the reviewer.

Of course, there are individualistic attempts to deal with the PD but basically the free-riding possibility rationally may often lead also pro-group I-moders to mutual defection in a PD. A common idea is to require the transformation of the PD into an Assurance Game (AG), which has CC and DD as its equilibria. Although they prefer CC to DD, the latter is not excluded unless the participants are strongly assured that the other one will choose C.

In the book I try to show also that the we-mode approach at least in some cases does apply to real life contexts and leads to results different from e.g. the pro-group I-mode account as the discussed central result shows.

To rehash, my account indicates that we-mode acting is functionally different from pro-group I-mode acting not only in terms or the smaller set of equilibria that it in many cases (including also many non-Paretian cases) entails in relation to the latter but that this difference at least in some important cases is actualized in real life where people act as group members. The pro-group I-mode

8

solution may sometimes result in mutual cooperation (e.g in a suitably transformed material-payoff PD and e.g. in an AG), but the mutual defection equilibrium is not eliminated, and it remains a powerful alternative in single-shot situations. Note finally concerning the present result against individualism that it may not apply to all kinds of "individualisms" or individualistic activities but it demonstrably works at least against strict traditional game-theoretic and economic individualism. My main aim has been to show that individualism (here primarily the (pro-group) I-mode account) and the we-mode account are functionally and behaviorally different at least in some cases and that thus individualistic theorizing does not suffice as the basis of all theorizing about the social world – a group approach such as the we-mode account is needed to accompany and complement it.

I have above answered practically all the central criticisms presented by the reviewer in the Notre Dame Philosophical Reviews. I have found almost all of them untenable and have tried to argue for my responses briefly and have referred to my 2013 book Social Ontology and in some cases also to my 2007 philosophy of sociality book for more detailed argumentative discussion.*

References

Bacharach, M. 1999. Interactive Team Reasoning: A Contribution to the Theory of Cooperation. *Research in Economics* 53 (2): 117-147.

Bratman, M. 2014. Shared Agency: A Planning Theory of Acting Together, Oxford University Press

Hakli, R., K. Miller, and R. Tuomela. 2010. Two Kinds of We-Reasoning. *Economics and Philosophy* 26 (03): 291-320.

Tuomela, R. 2007. *The Philosophy of Sociality: The Shared Point of View. New York: Oxford University Press.* Paperback ed. 2010

Tuomela, R., 2013. *Social Ontology: Collective Intentionality and Group Agents,* New York: Oxford University Press.

Udehn, L. 2001. *Methodological Individualism: Background, History and Meaning*. London: Routledge.