

A computer simulation of the argument from disagreement

Johan E. Gustafsson · Martin Peterson

Received: 13 January 2010 / Accepted: 27 September 2010 / Published online: 16 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In this paper we shed new light on the Argument from Disagreement by putting it to test in a computer simulation. According to this argument widespread and persistent disagreement on ethical issues indicates that our moral opinions are not influenced by any moral facts, either because no such facts exist or because they are epistemically inaccessible or inefficacious for some other reason. Our simulation shows that if our moral opinions were influenced at least a little bit by moral facts, we would quickly have reached consensus, even if our moral opinions were affected by factors such as false authorities, external political shifts, and random processes. Therefore, since no such consensus has been reached, the simulation gives us increased reason to take seriously the Argument from Disagreement. Our conclusion is however not conclusive; the simulation also indicates what assumptions one has to make in order to reject the Argument from Disagreement. The simulation algorithm we use builds on the work of Hegselmann and Krause (*J Artif Soc Social Simul* 5(3); 2002, *J Artif Soc Social Simul* 9(3), 2006).

Keywords Hegselmann–Krause · Disagreement · Simulation · Meta-ethics · Moral realism · Opinion dynamics

In this paper we shed new light on the Argument from Disagreement by putting it to test in a computer simulation. The Argument from Disagreement seeks to show that

J. E. Gustafsson (✉)
Division of Philosophy, Royal Institute of Technology, Teknikringen 78-B, 100 44 Stockholm, Sweden
e-mail: johangus@kth.se

M. Peterson
Section for Philosophy and Ethics, Eindhoven University of Technology,
P.O Box 513, 5600 Eindhoven, The Netherlands
e-mail: m.peterson@tue.nl

there are no objective and epistemically efficacious moral facts, i.e., moral truths that somehow affect our moral opinions. This is because if objective and epistemically efficacious moral facts were to exist, then there would be no more interpersonal and intercultural disagreement in ethics than in, say, chemistry or mathematics. However, there is in fact much more disagreement in ethics than in other areas, and therefore advocates of the Argument from Disagreement conclude that either no such facts exist, or if they exist, then they are epistemically inaccessible or inefficacious for some other reason.

Sextus Empiricus was the first to offer a detailed account of the Argument from Disagreement.¹ In more recent years, J. L. Mackie's discussion has been particularly influential. According to Mackie, '[r]adical differences between first order moral judgements make it difficult to treat those judgements as apprehensions of objective truths.'² This is because,

the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values.³

Recent discussions of the Argument from Disagreement have revealed at least two major weaknesses. The first is exemplified in Mackie's treatment: He offers no or little support for the claim that variations in moral opinions are more readily explained by the non-realist hypothesis than its realist rival. Why, exactly, is it more reasonable to think that people would stop to disagree if moral facts were to exist? Moral opinions are affected by a large number of factors, including authorities and moral opinions held by our peers and relatives, so advocates of the Argument from Disagreement need to do more to convince us that the existence of moral facts would invariably lead to consensus.

The second weakness, highlighted by several contemporary moral realists, is that it is difficult to reconcile the Argument from Disagreement with the fact that we actually agree on many moral issues, such as the wrongness of torturing innocent children for fun.⁴ More precisely put, it seems that we disagree on *some* moral issues, such as abortion and capital punishment, but not on all. So how can claims about disagreement disprove the existence or epistemic inefficacy of moral facts, given that we agree on many—or even most—moral issues?

The aim of this paper is to show that a computer simulation can overcome both weaknesses described above. If correct, the simulation will thus shed new light on the Argument from Disagreement and, indirectly, the meta-ethical view we call non-sceptic moral realism. Non-sceptic moral realism is the view that moral facts exist and

¹ Against the ethicists, pp. 68–78, *Outlines of pyrrhonism III*, p. 190.

² Mackie (1977, p. 36).

³ Mackie (1977, p. 37).

⁴ See, e.g. Shafer-Landau (2003) and Tersman (2006, p. xii).

that they influence our moral opinions.⁵ The most influential account of non-sceptic moral realism in recent years is Cornell realism, a view that is defended by Richard Boyd, Nicholas Sturgeon, and David Brink.⁶ Cornell realists believe that moral facts are mind-independent facts that cannot be reduced to non-moral facts, and that we under the right circumstances do in fact acquire knowledge about them. In many cases, the existence of moral facts explain why we hold the moral opinions we do in fact hold. Moral facts have a causal influence on our opinions.

Our computer simulation offers a detailed explanation of why the existence of moral facts would lead to consensus on moral issues, and why we do actually agree on some moral issues but not on others. This gives us increased reason to take the Argument from Disagreement seriously. Our results are, however, not conclusive. The simulation also indicates what assumptions one has to make in order to reject the Argument from Disagreement. The simulation algorithm we use builds on the work of [Hegselmann and Krause \(2002, 2006\)](#).

The structure of this article is as follows. In Sect. 1 we explain the basic configuration of the computer model, which is thereafter extended and rendered more realistic in Sect. 2. We model a scenario that includes moral facts and compare it to a scenario without moral facts. Since the scenario comprising moral facts turns out to have no or little resemblance with the world we actually live in, unlike the scenario in which there are no moral facts influencing our moral opinions, we conclude that *either* there is something wrong with the model *or* non-sceptic moral realism is false. In Sect. 3 we offer some reasons for thinking that the last disjunct of the foregoing sentence is the correct conclusion. We do this by showing that the model is very robust. Its outcomes look very much the same even if the initial assumptions are modified quite dramatically. Thereafter, in Sect. 4 and Sect. 5, we explain more in detail how our computer simulation overcomes the weaknesses of the Argument from Disagreement outlined above, and discuss what assumptions one has to make in order to reject the Argument from Disagreement. In these sections we also reply to some possible objections.

1 The basic model

A successful computer model of moral reasoning should be flexible about how moral opinions evolve and change over time. The model has to be able to account for the influence of a number of factors, apart from moral facts, many of which we know relatively little about. In this section we explain the basic configuration of the computer model we use.

To start with, we wish to make room for the possibility that one's moral opinions are influenced, at least to some degree, by opinions held by others. People who think abortions should remain legal may perhaps pay no attention to the most extreme views

⁵ It should be stressed that the meta-ethical position we attack is not an ontological position only, but an ontological plus an epistemological position. For example, a reviewer has pointed out to us that the meta-ethical position defended by Hartmann (1932) might not be vulnerable to our argument. This is because on Hartmann's view moral facts only have a short-lived influence on our opinions.

⁶ [Boyd \(1988\)](#), [Sturgeon \(1985\)](#), and [Brink \(1989\)](#).

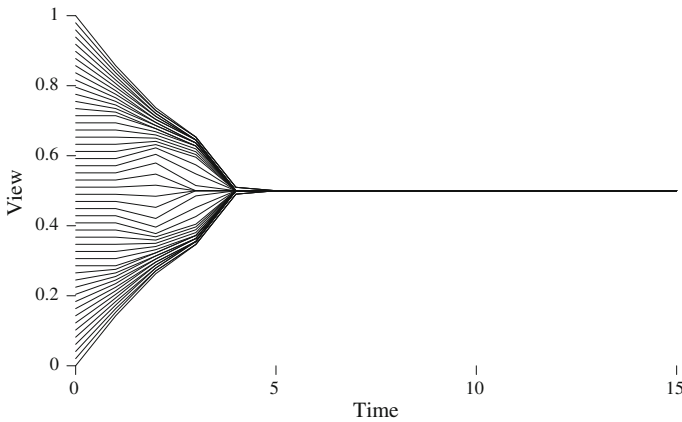


Fig. 1 $\varepsilon = 0.3$

put forwards by their opponents, but they are likely to take into account at least *some* arguments put forward by people holding opinions that are sufficiently similar to their own. We shall model this by adopting a generalized version of a model originally developed by Hegselmann and Krause.⁷ Hegselmann and Krause assume that opinions can vary continuously in an interval of real numbers. It is convenient to let the interval be $[0, 1]$. Since a model is a purely mathematical construction, the outcome of the simulation will not depend on the interpretation of the interval. However, to give the numbers some intuitive meaning, we will assume that they correspond to the agent's opinion about the degree of moral praiseworthiness of the action under consideration. In a discussion of abortion 1 would, for instance, represent maximum praiseworthiness and 0 maximum condemnation of that practice, while 0.5 would represent moral neutrality.

Figure 1 shows a simulation of a moral discussion between a group of 50 individuals, all of which initially hold different and evenly distributed opinions. All individuals are aware of the opinions held by others, and in each round they modify their opinions by taking into account views put forward by others. In this example, each individual considers only opinions that are up to 0.3 units more positive or negative than the opinion he or she held in the previous round. The term *confidence interval* refers to the set of similar views that each agent is prepared to consider. (In the mathematical literature the confidence interval is denoted by ε .) In the original Hegselmann–Krause model, developed for non-moral purposes, it is assumed that disagreeing parties should ‘split the difference’, that is, adopt the opinion that corresponds to the average opinion of the set of opinions within the confidence interval. However, in order to make our model more flexible—and thus less vulnerable to criticism—the extended model presented in the next section assumes that agents are more conservative, in the sense that they put more weight on their own opinion than those of others.

A striking fact about the original Hegselmann–Krause model is that the outcome of the simulation is determined solely by each agent's willingness to adjust his or her

⁷ Hegselmann and Krause (2002).

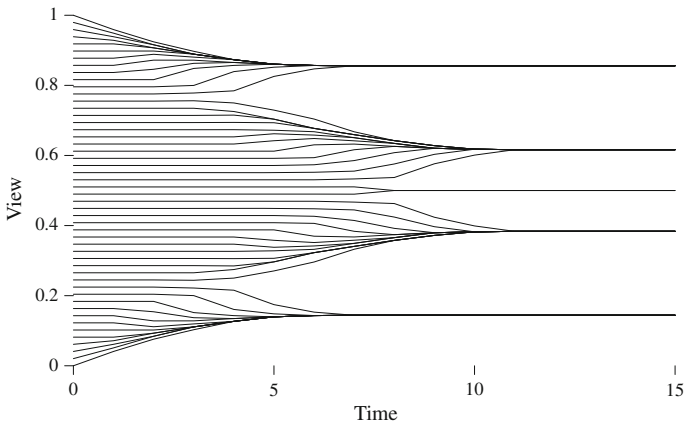


Fig. 2 $\varepsilon = 0.1$

opinion to similar opinions held by others. If each agent's opinion at each round is influenced *only* by the view he held the previous round and other views that are within the confidence interval, i.e., views that are sufficiently close to one's own for being worth considering, then consensus will arise just in case the distance between two neighbouring opinions is never greater than the confidence interval.

In Fig. 2 the population is kept constant while the confidence interval is decreased to 0.1. Hegselmann and Krause refer to the two types of cases illustrated in Fig. 1 and 2 as *consensus* and *plurality*. This strong correlation between the number of surviving opinions and the size of the confidence interval is typical of the Hegselmann–Krause model. The relationship is, however, not monotonic. It has been shown that in some intervals the number of surviving opinions increases as the confidence interval increases.⁸ Furthermore, Hegselmann and Krause show that patterns similar to those in Fig. 1 and 2 arise even if the confidence interval is asymmetric, i.e., if the upper limit of the confidence interval is not equally far from the initial opinion as the lower limit.

A large number of modifications to the original Hegselmann–Krause model have been studied in the literature. A modification that is of particular interest in the present context is that proposed by Hegselmann–Krause in their (2006). In that model, which we will call the *truth-sensitive* Hegselmann–Krause model (as opposed to their original (2002) model) agents are not merely influenced by each other's opinion but also by the true fact of the matter. Hegselmann–Krause assume that only one opinion in the range of opinions is the true one and investigate what happens if either all or some agents let their opinions be affected by the truth (as well as the opinions of others that lie within the confidence interval). The conclusion is that agents will eventually reach agreement on the true opinion, with the exception for the case of an agent whose (i) opinion is not affected at all by the truth, and whose (ii) initial opinion is too far apart from that of others, and is thus not affected by those opinions either. In Fig. 3, the dotted line represents the truth, which is assumed to be stable over time. The influence from other peers is ten times stronger than the influence of the truth. (That is, the 'truth

⁸ Lorenz (2007).

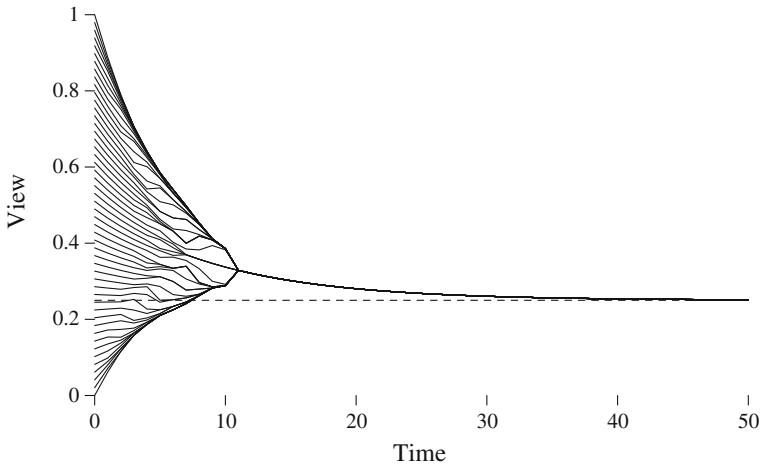


Fig. 3 $\varepsilon = 0.1$ and 10 % truth attraction

attraction' is 10 %.) All other parameters are as in Fig. 2. Note that by introducing a true opinion, a case of plurality is transformed into consensus.

A number of mathematicians have attempted to obtain analytic results for Hegselmann and Krause's models. Krause (2008) proves a convergence theorem for a limited number of cases, and Lorenz (2007) and Lorenz and Lorenz (2008) prove generalized versions of the theorem. The 'thunnel theorem' in Hegselmann–Krause (2006) shows that if all agents are sensitive to the truth, then consensus on the truth will arise in the sense that all agents will eventually approach the truth. Moreover, the 'leading the pack theorem' in the same paper shows that even if only some agents go for the truth this is sufficient for guiding the whole group towards the truth, given that no individual opinion that is unaffected by the truth is at some point left isolated outside the confidence interval. The present paper makes no attempt to add new analytic results to the literature on different versions of the Hegselmann–Krause model.

2 The extended model

It is a striking feature of moral discussions on abortion, capital punishment, and a number of similar issues, that we never seem to reach consensus. It is therefore interesting to study what structural assumptions about moral discussions, if any, are consistent with persistent disagreement. In particular, it is interesting to investigate whether persistent disagreement is compatible with the assumption that moral opinions are affected by moral facts that we somehow perceive or get acquainted with. In this section we will modify the two models for opinion-dynamics introduced by Hegselmann and Krause in their (2002) and (2006), such that they become more relevant for modeling moral discussions.

In the original Hegselmann–Krause model, persistent moral disagreement is possible only if the confidence interval is sufficiently small. Otherwise consensus will

quickly be reached, as shown above. Therefore, in the rest of this paper, we assume that the confidence interval is sufficiently small to enable persistent disagreement, i.e., we will only study cases similar to that in Fig. 2. Although it cannot be excluded *prima facie* that persistent disagreement could arise for larger confidence intervals once more features are added to the model, we have found that this is not the case. (For details, see Sect. 3.) In order to be on the safe side, we also assume that individuals are even more conservative than before. In the extended model, each individual takes his own view to count for several hundred times as much as his moral peer. Although this is hardly rational, this makes the model less vulnerable to criticism, as will be shown below.

Needless to say, it would be implausible to assume that moral opinions are affected *merely* by opinions held by others. Other factors may also play a role. In the extended model three such additional factors are incorporated into the simulation, viz. authorities, external shifts, and random processes. Naturally, the more factors that can be incorporated into the model without affecting the overall result of the simulation, the less vulnerable will its conclusion be to criticism. We do not wish to claim that all moral opinions are always affected by authorities, shifts, and random processes. The point is, on the contrary, that our conclusions will hold water *even if* these factors affect how moral opinions change over time.

An *authority* can be conceived of as an influential person, organization, political party, book, or philosophical movement. Typical examples include the Pope, Greenpeace, the German Nazi party in the 1930s, Peter Singer's book *Animal Liberation*, and Aristotelian virtue ethics. All these entities might affect people's moral opinions, given that the distance between the authority's opinion and that of the agent is within the confidence interval. However, all authorities cannot be right about what they preach, because on many moral issues they advocate inconsistent views. Furthermore, it is reasonable to assume that each authority tends to attract opinions over an extended period of time, but the influence is seldom constant over time. In our model, we assume that each authority, or rather each period in which an authority affects our opinions, extends over a randomized interval of 10–30 rounds. (See Sect. 3.) Of course, some authorities, such as the Church, tend to be influential during several periods. As will be shown in Sect. 4 our conclusions will go through even if our moral opinions are influenced by one or more constant moral authorities, i.e., constant attractors of moral opinions that may differ from the truth. In Figs. 4 and 5 authorities are represented as relatively short, thick horizontal lines. Contrary to what one might have expected, the presence of a large number of authorities lead to more consensus rather than less. Moreover, our simulations also show that the result will be the same no matter whether the strength of each authority is constant or decreases with respect to the distance of the opinion.

Sometimes external *shifts*, such as the political shift towards the left in 1968, affect our opinions quite dramatically during a relatively short period of time. From a mathematical point of view, they can be represented as some subinterval of $[0, 1]$ in which all opinions are suddenly shifted either upwards or all downwards. A major difference between external shifts and authorities is that authorities affect our opinions to a limited degree over a relatively long period of time, while shifts take place within just

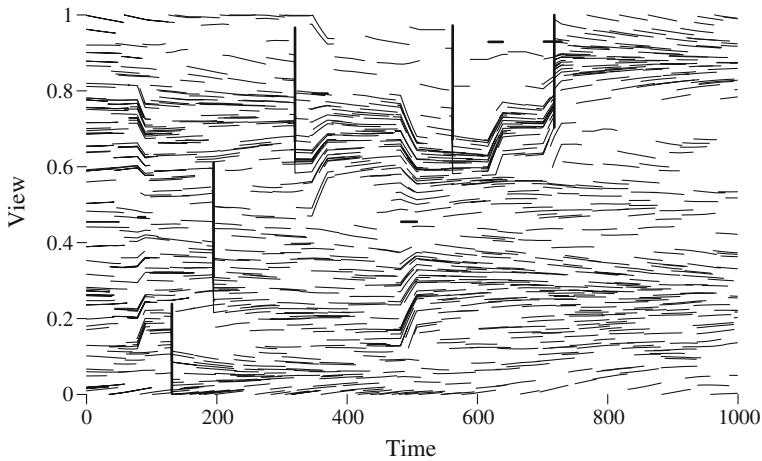


Fig. 4 The extended model without moral facts

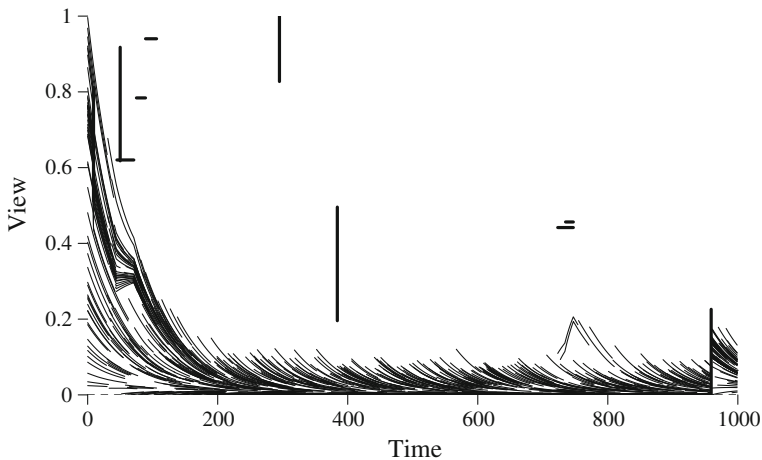


Fig. 5 The extended model with strong moral facts

one round. However, in that round they have a relatively large influence on our moral opinions. In Figs. 4 and 5 external shifts are represented by thick vertical lines.

The third new factor in the extended model is *random processes*. Sometimes opinions change for no good (or bad) reason at all. This may very well be irrational, but since the Argument from Disagreement starts from the assumption that people do *actually* disagree, it seems wise to also take the possibility of random processes into account. In our extended model we have stipulated that opinions change randomly every 30–70 rounds, and that no random change is more than twice the size of the confidence interval. These numerical assumptions are of course arbitrary. However, as will be shown below, not much hinges on them. The robustness analysis presented in Sect. 3 shows that our conclusions do not depend much on what assumptions we make about the numerical values of these variables.

Figure 4 shows a simulation in which the three factors introduced above have been added to the original Hegselmann–Krause model (2002). The most important conclusion to be drawn here is that no consensus is reached, not even after a large number of rounds. In the next section we discuss what assumptions we have made about the values of the underlying variables and the mechanism for aggregating their influence, and we show that the result in Fig. 4 is surprisingly robust. This means that the figure will look roughly the same no matter what one assumes about the size of the population (relative to the confidence interval), the number of authorities, their position in the opinion space, the authorities' extension in time, the number of shifts, the strength of these shifts, and the influence of random processes.

Enter moral facts. We take for granted that if they were to exist, they would somehow affect our moral opinions in much the same way as the truth affects opinions in the truth-sensitive Hegselmann–Krause model. Hence, we assume that if moral facts were to exist, they would be invariant with respect to time—they would always be there.⁹ We assume that moral fact exercise moderate influence on our opinions. For simplicity, we assume that the influence of moral facts is constant over time. Furthermore, we assume that there is only one fact of the matter about each moral issue; hence, it cannot be the case that, say, abortion is both morally praiseworthy and not, everything else being equal.

A moral fact is represented by a dotted line in our plots. In Fig. 5 we assume that the objectively true opinion is to fully reject abortions (i.e., its numerical value is 0), but our simulations show that it is irrelevant to the overall conclusion of the paper what the objectively true opinion happens to be: if the objectively true opinion is a random variable, the total amount of disagreement will be slightly higher, but not dramatically higher. (See Sect. 3 for details.)

To start with, we assume that moral facts affect our moral opinions to the same degree as authorities, i.e., to the same degree as the Church or Peter Singer. Fig. 5 clearly shows that if moral facts were to have such a great influence on our opinions, then consensus would have been reached within a relatively small number of rounds. Suppose that a 'round' in the model is equivalent to, say, one month of discussion. Then consensus on abortion would have been reached within a few years after the discussion began. This is, however, not the case. Abortion is still a very controversial moral issue, on which people hold very different moral views—despite the fact that free and democratic discussions have been going on for at least a century. Therefore, the assumption about the strong influence of moral facts on our opinions made in Fig. 5 is hardly a correct representation of the actual world.

But what if moral facts have far less influence on our moral opinions? In Fig. 6 the influence of moral facts is one third of that of a single authority. Since the number of authorities is large, and there is only one objectively true opinion, and opinions are also affected by many other factors, the total share of the total influence generated by the single objectively true opinion is thus relatively small. (The numerical values of all these variables are listed in Table 1.) However, even if our sensitivity to moral facts is that low, it nevertheless turns out that consensus will be reached within a moderate

⁹ We are aware that this assumption must be interpreted with care. Naturally, similar acts may have different moral status at different points in time due to changes in the external world.

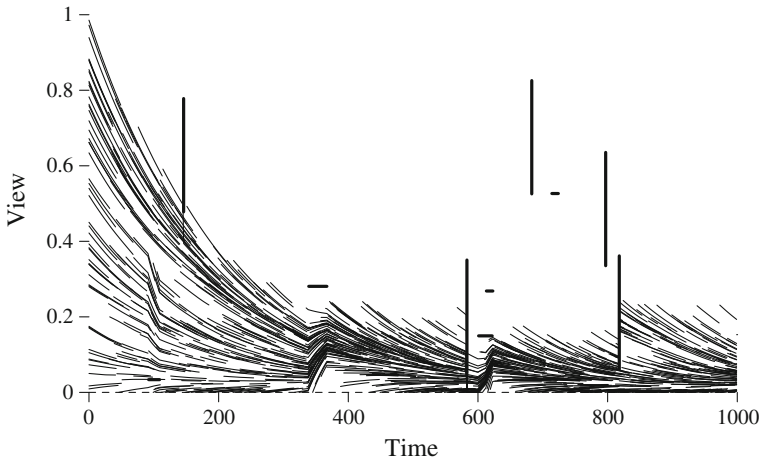


Fig. 6 The extended model with weak moral facts

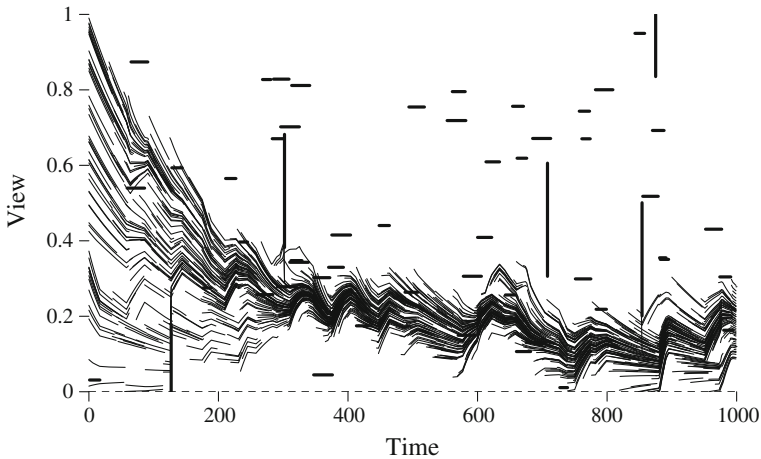


Fig. 7 The extended model with weak moral facts and many authorities

number of rounds. Moreover, this conclusion also holds true if the number of authorities is increased dramatically, (to 50 instead of 5) as can be seen in Fig. 7. Therefore, since no consensus has yet been reached on abortion, our preliminary conclusion is that moral facts about abortion do not exist, or at least do not affect our opinions to any significant degree.

3 Further details about the extended model

In Sect. 1 and Sect. 2 we have tried to give a brief and non-technical overview of how our model works and why it is relevant to the Argument from Disagreement. The main point is that our moral opinions do not seem to be affected by any objective moral

facts, because if they were, we would have reached consensus a long time ago. In this section we offer a detailed presentation of the technical assumptions underlying the model, and discuss their plausibility. The basic message is that the model is surprisingly robust. Our conclusion will not change dramatically if the simulation is repeated a large number of times, or if the initial values in each simulation are altered.

Opinions are represented by real numbers, which are updated in each round depending on the distance to other opinions, authorities, external shifts, random processes, and facts. The initial opinion held by each individual is assigned by an even random distribution. Time is modelled by the number of rounds the process is iterated. The interpretation of time is of course somewhat arbitrary. In order to be on the safe side, we assume that each round lasts no longer than one year, i.e., that individuals are willing to reconsider their moral opinions at least once a year. Arguably, this is a very conservative assumption. If opinions are actually reconsidered more frequently, say once a month, our case becomes even stronger. This is because consensus would then be reached much quicker. If each round were to last much longer, then people would be too conservative for the model to be realistic. Therefore, it is not a problem that the unit of time is to some extent arbitrary, as long as the assumptions we base the simulations on are sufficiently conservative.

The numerical values of all parameters are listed in Table 1. The simulation lasts for R rounds and in each round there are n peers each holding an opinion between 0 and 1. The views of all peers within ε affects the view of a peer. Each peer has a certain lifespan that may be as low as $peerMinAge$ and as high as $peerMaxAge$. When a peer has reached the end of its lifespan it is replaced by a peer with a randomized view within $mutRnge$ of the view of its predecessor. There are $authNr$ authorities that may last from $authMinDur$ to $authMaxDur$ rounds. An authority affects the views of all peers with a view within $authRnge$ of the authority. Furthermore, there are $shftNr$ shifts that in a single round changes the views of all peers holding a view within $shftRnge$ from a certain view. The shift either increases or decreases the views of the affected peers by $shftPull$. The truth affects all peers with a certain strength in all rounds. The relative strength of the influence from a peer's earlier view, other peers, authorities, and the truth are $consInfl$, $peerInfl$, $authInfl$, $TInfl$. Let $x_i(r)$ be the view of peer i at round r . Let $authView_i$ be the view of authority i and let $authDur_i$ be the duration of i and let $authBeg_i$ be the start round for i . Let $shftCntr_i$ be the centre of range of views affected by shift i and let $shftDir_i$ be the direction i pulls, and let $shftRnd_i$ be the round i occurs.

We are now in a position to describe the simulation algorithm for the extended model. See Fig. 8 for an accompanying flow chart.

The Simulation Algorithm for the Extended Model: Given the parameters (given in Table 1) simulate the changes in views for n persons over R rounds.

1. [Initialize] Set r to 0. For $i = 1$ to n , set $x_i(0)$ to a random real number between 0 and 1 and set $rToLive_i$ to a random integer between $peerMinAge$ and $peerMaxAge$. For $i = 1$ to $authNr$, set $authDur_i$ to a random integer between $authMinDur$ and $authMaxDur$ and set $authBeg_i$ to a random integer between 1 and R and set $authView_i$ to a random real number between 0 and 1. For $i = 1$ to $shftNr$, set $shftRnd_i$ to a random integer between 1 and R and set $shftDir_i$ randomly

to either -1 or 1 . For $i = 1$ to $shftNr$, set $shftCntr_i$ to a random real number between 0 and 1 such that $shftCntr_i + (shftRnge + shftPull) * shftDir_i \geq 0$ and $shftCntr_i + (shftRnge + shftPull) * shftDir_i \leq 1$.

2. [Increase round] Increase r by 1 .
3. [Conservatism] For $i = 1$ to n , set $consPoint_i$ to $x_i(r - 1)$.
4. [Peers] For $i = 1$ to n , set $peerPoint_i$ to

$$\begin{cases} x_i(r - 1) & \text{if } I(i, r - 1) = \emptyset \\ |I(i, r - 1)|^{-1} \sum_{j \in I(i, r - 1)} x_j(r - 1) & \text{otherwise,} \end{cases}$$

where $I(i, r - 1) = \{k \in \{1, \dots, n\} : |x_i(r - 1) - x_k(r - 1)| \leq \varepsilon \wedge k \neq i\}$.

5. [Authorities] For $i = 1$ to n , set $authPoint_i$ to

$$\begin{cases} x_i(r - 1) & \text{if } A(i, r - 1) = \emptyset \\ |A(i, r - 1)|^{-1} \sum_{j \in A(i, r - 1)} authView_j & \text{otherwise,} \end{cases}$$

and set $authTotInfl_i$ to $|A(i, r - 1)| * authInfl$, where $A(i, r - 1) = \{k \in \{1, \dots, authNr\} : |x_i(r - 1) - authView_k| \leq authRnge \wedge authBeg_k \leq r \leq authBeg_k + authDur_k\}$.

6. [Age] For $i = 1$ to n , decrease $rToLive_i$ by 1 .
7. [New views] For $i = 1$ to n , set $x_i(r)$ to

$$\frac{consInfl * consPoint_i + peerInfl * peerPoint_i + authTotInfl_i * authPoint_i + TInfl * T}{consInfl + peerInfl + authTotInfl_i + TInfl}$$

8. [Shifts] For $i = 1$ to n , if $|S(i, r - 1)| > 0$ then set $x_i(r)$ to $x_i(r) + shftPull * shftDir_j$ where $j = \min(\{m : m \in S(i, r - 1)\})$ and $S(i, r - 1) = \{k \in \{1, \dots, shftNr\} : |x_i(r - 1) - shftCntr_k| \leq shftRnge \wedge shftRnd_k = r\}$.
9. [Successors] For $i = 1$ to n , if $rToLive_i = 0$ then set $x_i(r)$ to a random real number between $x_i(r - 1) - mutRnge$ and $x_i(r - 1) + mutRnge$ such that $0 \leq x_i(r) \leq 1$ and set $rToLive_i$ to a random integer between $peerMinAge$ and $peerMaxAge$.
10. [Done?] If $r < R$ then go back to step 2 otherwise terminate. ■

As emphasized above, the values of the parameters are to some extent arbitrary. Therefore, in order to neutralize the objection that our results depend too heavily on these values, we have performed a series of simulations that help to assess the robustness of our results. Briefly put, it turns out that even if the values of the parameters are significantly altered, the outcome of the simulations will be more or less the same. We iterated the simulation 10,000 times, under the assumptions that moral facts exist, each time with new randomly chosen values that were up to fifty percent higher or lower than the initial values described above. (The maximum and minimum values we used are listed in the two rightmost columns of Table 1.)¹⁰ At the end of each simulation, the distance between the opinions being furthest away was measured. Fig. 9 shows the relative number of times the distance between the opinions being furthest away were between 0 and 0.1 units, and between 0.1 and 0.2 units, and so on. Fig. 10 shows the

¹⁰ In case $peerMaxAge$ turns out be lower than $peerMinAge$, the former is set to $peerMinAge$.

Table 1 Parameters

	Fig. 4	Fig. 5	Fig. 6	Fig. 7	Fig. 12	Fig. 13	Fig. 14	Min	Max
<i>R</i>	1000	1000	1000	1000	1000	1000	1000	1000	1000
<i>n</i>	50	50	50	50	50	50	50	25	75
ϵ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.05	0.15
<i>peerMinAge</i>	30	30	30	30	30	30	30	15	45
<i>peerMaxAge</i>	70	70	70	70	70	70	70	35	105
<i>mutRnge</i>	0.15	0.15	0.15	0.15	0.15	0.08	0.15	0.075	0.225
<i>authNr</i>	5	5	5	50	5	5	5	2	8
<i>authMinDur</i>	10	10	10	10	10	10	10	5	15
<i>authMaxDur</i>	30	30	30	30	30	30	30	15	45
<i>authRnge</i>	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.175	0.525
<i>shftNr</i>	5	5	5	5	5	5	5	2	8
<i>shftRnge</i>	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.075	0.225
<i>shftPull</i>	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.05	0.15
<i>consInfl</i>	200	200	200	200	200	200	200	100	300
<i>peerInfl</i>	1	1	1	1	1	1	1	0.5	1.5
<i>authInfl</i>	3	3	3	3	3	3	3	1.5	4.5
<i>TInfl</i>	0	3	1	1	1	1	1	0.5	1.5
<i>T</i>	0	0	0	0	0	0	0	0	1
<i>constAuthNr</i>					1	7	7		
<i>constAuthInfl</i>					10	100	20		
<i>constAuthRnge</i>					0.35	0.08	0.08		

same plot under the assumption that moral facts do not exist. It seems to be relatively difficult to resist the conclusion that the non-existence of moral facts is a necessary condition for persistent moral disagreement.

Fig. 11 shows what happens if the location of the moral fact is determined by an evenly distributed random variable. As can be seen in the figure, this yields slightly more disagreement compared to the situation in which the value of the fact is 0, but the difference is not dramatic. We have not been able to find any clear explanation of this effect. However, a possible conclusion one could draw is that if moral facts exist they are more likely to be located somewhere in the middle of the interval rather than at the extremes.

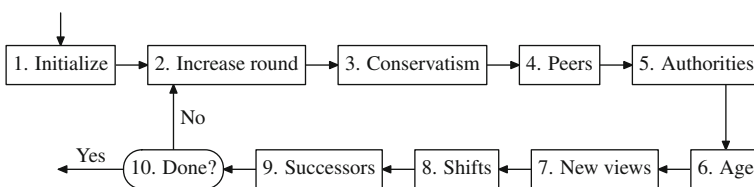


Fig. 8 Flow chart for the simulation algorithm

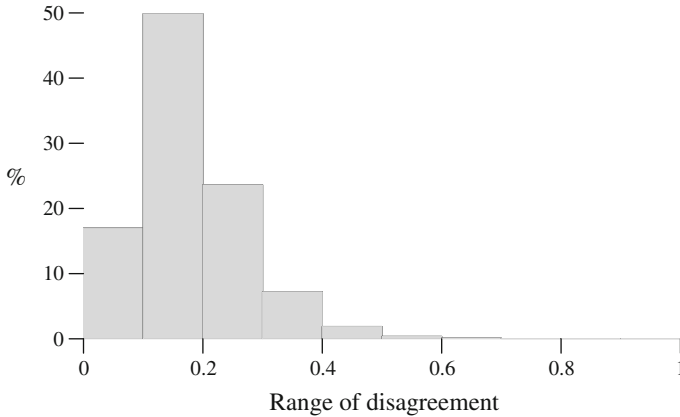


Fig. 9 Range of disagreement after 1000 rounds in 10,000 randomized simulations with moral facts at $T = 0$

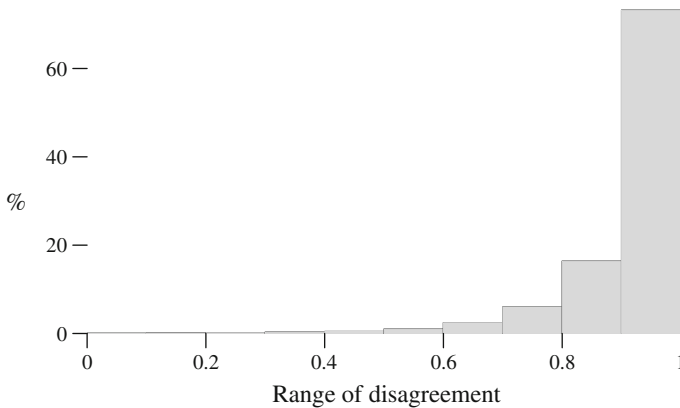


Fig. 10 Range of disagreement after 1000 rounds in 10,000 randomized simulations without moral facts

4 Constant authorities

In Sect. 2 we briefly mentioned that some authorities, such as the Church, may influence our moral opinions over very long periods of time. From a modelling perspective, the Church can be conceived as a *constant* authority, i.e., as an authority that affects our opinions in all rounds. How does the presence of one or more constant authorities affect the robustness of our conclusion?

Figure 12 shows a simulation with a moral fact at 0 and a constant authority near 1. A possible example could be a situation in which it is, say, wrong not to allow gay marriages although the Church holds the opposite view. Naturally, at most two opinions will survive in the long run, no matter the relative strength of the fact and the constant authority. (In this particular example, the second surviving opinion is just below 1. This is due to the relative strength of the moral fact and the opinion preached

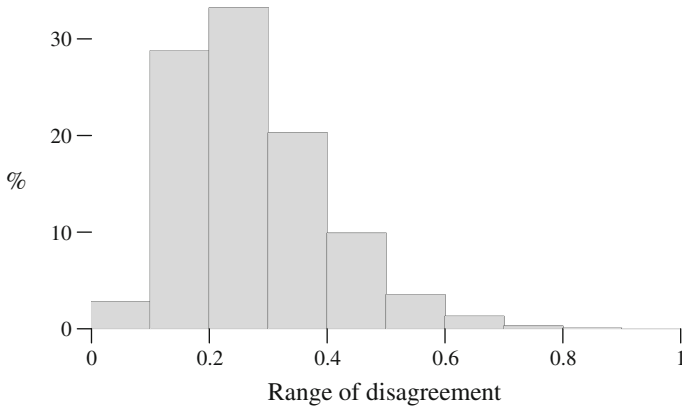


Fig. 11 Range of disagreement after 1000 rounds in 10,000 randomized simulations with randomized moral facts

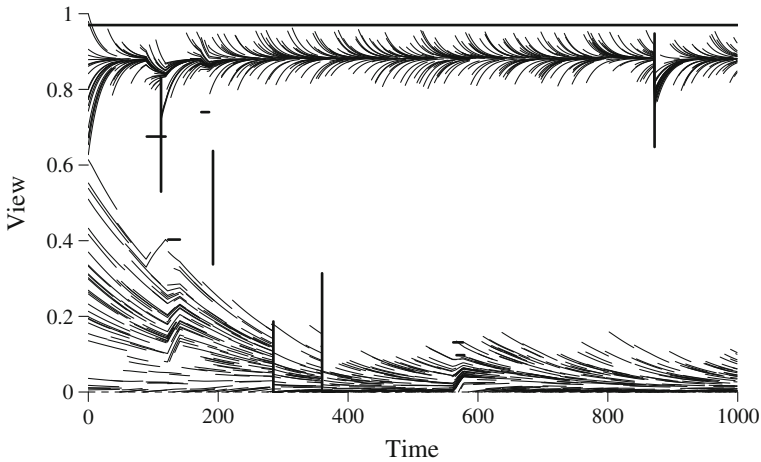


Fig. 12 Disagreement and facts with a constant authority

by the Church.) However, this is clearly not a correct account of how moral opinions evolve over time. Our opinions do not tend to converge to two extreme views. On the contrary, nearly *all* opinions between the two extremes tend to be represented when people disagree on moral issues. Moreover, in Fig. 12 all change is change towards an extreme view, and once that view has been reached the individual never departs from it in the future. This is implausible. Therefore, Fig. 12 is not a convincing picture of how people do actually reason about moral issues. This indicates that the non-sceptic moral realist has little reason for thinking that the disagreement we observe is due to the existence of some (possibly incorrect) constant authorities, such as the Church.

At this point it might be objected that non-sceptic moral realists could try to make their claim about moral facts compatible with the prevalence of widespread moral disagreement by simply increasing the number of constant authorities. Fig. 13 shows a

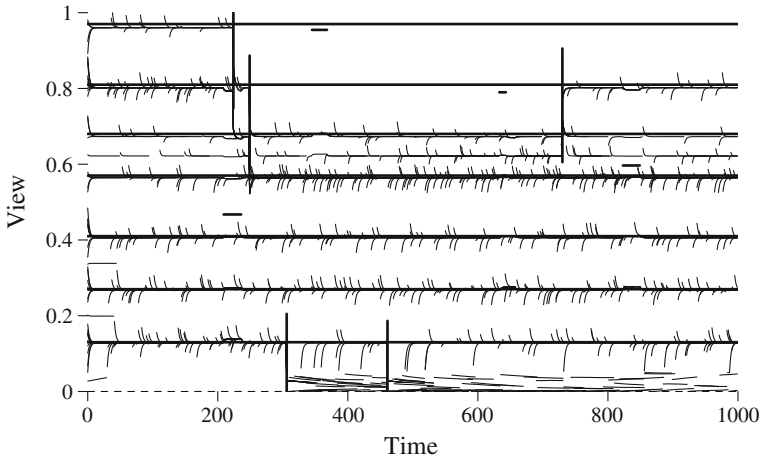


Fig. 13 Many constant authorities and a single fact

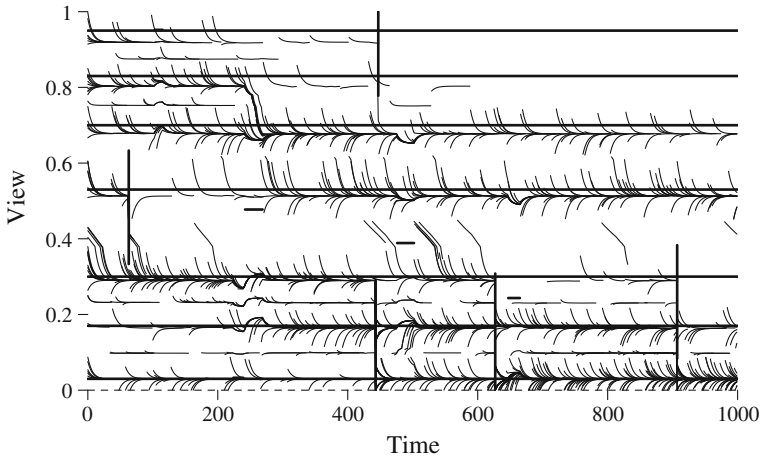


Fig. 14 Less influential constant authorities and a single fact enables more change

scenario with seven constant authorities and a single fact. Unfortunately, this figure also fails to give a convincing account of how moral opinions evolve over time. Although we can now observe considerable disagreement, people in this simulation simply do not change their opinions as much as they sometimes do in real life. Moreover, in Fig. 13 all change is change towards a constant authority, and once the individual holds the same view as the authority, he or she will never depart from it in the future. This is implausible. We therefore conclude that this is not a convincing non-sceptic moral realist picture of moral disagreement.

However, let us consider yet another way in which the non-sceptic moral realist could try to rescue his position. This time we assume that constant authorities are much less influential than before and that people are more likely to randomly change their opinions; see Fig. 14 and Table 1 for details. During the first couple of rounds

this figure appears to give a plausible account of how disagreement could be possible despite the existence of an objective moral fact that influence people's opinions. As can be seen in the figure, people do actually change their opinions over time. However, a potential problem for the non-sceptic moral realist is that after a few hundred rounds some parts of the opinion space is empty, or almost empty. No one holds views that are close to those of the two uppermost authorities. That said, the figure suggests that there are some assumptions under which epistemically efficacious moral facts are compatible with persistent disagreement. Thus, a possible way to defend non-sceptic moral realism is to argue that some assumptions that generate this type of scenario are indeed the correct ones.

Finally, another approach that would guarantee persistent disagreement, regardless of any facts, is to significantly increase the effect of randomness on the peers. If *mutRnge* is sufficiently high and *peerMinAge* sufficiently low, then a large number of peers in each round will have a random view independently of any peers in earlier rounds and any other factors. This move will trivially ensure persistent disagreement. However, this would result in a peculiar version of moral realism; while moral facts would have some influence on our moral opinions this influence would nevertheless be drowned out by the influence of chance.

5 A better foundation for the Argument from Disagreement?

In the introduction we identified two weaknesses of traditional treatments of the Argument from Disagreement, namely, (i) the absence of an explanation of why the non-existence of moral facts leads to moral disagreement, and (ii) the absence of an explanation of why we disagree on some moral issues but not on all.

The results presented in this paper offer new insights into both these issues. We have shown that under a wide range of empirical assumptions the absence of moral facts is a necessary although not a sufficient condition for persistent disagreement (of the right structure). If moral facts were to exist, and if our opinions have long been affected by them to at least some non-negligible degree, then consensus would have emerged quickly. Moreover, if our moral opinions were also affected by some constant authorities the resulting disagreement would in many cases not be of the right structure, as explained in Sect. 4. This means that the basic premise of the Argument from Disagreement has gained some additional support: If moral facts existed we would not disagree widely on controversial issues, such as abortion and capital punishment. However, we do in fact disagree widely on many such issues.

The observation that the absence of moral facts (under a wide range of empirical assumptions) merely is a necessary and not a sufficient condition for persistent disagreement also helps to explain why we disagree on some moral issues but not on all. The point is that consensus can arise even if no moral facts exist. The simulations have revealed several such consensus-inducing mechanisms apart from moral facts. As shown above, the challenge is to show how persistent disagreement is possible, not to explain consensus.

Having said all this, it might be objected that our model proves too much. There is also considerable disagreement about whether there is life in other galaxies, whether

there is a god, and on exactly what Aristotle tried to tell us about the virtues. So does not our model predict that agreement should have emerged a long time ago on all these controversial issues? Our reply is that our model merely predicts that we would quickly reach consensus *if* our opinions were affected at least to some degree by the fact of the matter (towards the fact, not in the other direction). In the examples mentioned above most of our opinions are presumably not positively affected by any fact of the matter. Of course, this does not imply that there cannot be any objectively true answer to these questions. All we are entitled to conclude from our simulations is that whenever there is persistent disagreement the truth has not positively affected our opinions.

One might also object that our computer simulations are self-defeating.¹¹ This is because we seem to disagree not just on issues related to normative ethics, but also on many issues in meta-ethics, including the question whether non-sceptic moral realism is true. So if our model is correct, then there cannot be any fact about whether non-sceptic moral realism is correct that influences our meta-ethical opinions, since such a fact would have led to agreement. One might therefore wonder how our model could support this meta-ethical fact. The answer is that our model merely implies that the fact about whether non-sceptic moral realism is correct has not affected our opinions in the past. This does not rule out that our ability to let our future opinions be affected by the relevant facts might increase as we adopt new methodological approaches to meta-ethics. One such new approach is the computer simulation presented in this paper. Therefore, our simulation is not self-defeating.

The same move is of course open to the non-sceptic moral realist. While our simulation might have shown that there does not exist any moral facts that have influenced our moral opinions in the past, there might exist moral facts that will do so in the future. Perhaps some new methods in normative ethics will enable us to access the truth. It seems unlikely that any argument from past disagreement can rule out this possibility.

6 Conclusion

By showing how the Argument from Disagreement can be modelled in a computer simulation, we have shed new light on this influential argument. As explained above, our findings seem to lend some additional support to the rejection of non-sceptic moral realism, at least under a wide range of empirical assumptions about how moral opinions evolve over time. However, the most important conclusion is perhaps that the *methodology* we use appears to be fruitful for moral philosophers wishing to discuss the meta-ethical significance of moral disagreement. By giving up some of the seemingly plausible assumptions we use in our model, it may very well be possible to construct *some* version of moral realism that is immune to the Argument from Disagreement—if so, this would then tell us something important about the structure of a plausible form of moral realism. We thus conclude that computer simulations provide us with a new tool for assessing meta-ethical debates about moral disagreement.

¹¹ This objection was suggested to us by Jonas Olson.

Acknowledgements Early versions of this paper were presented to audiences at conferences and seminars at the University of Colorado at Boulder, Stockholm University, Uppsala University, Lund University, Luleå University of Technology, Delft University of Technology, and the Buenos Aires Metaethics Workshop. We would like to thank the participants for very helpful comments. We also wish to thank two anonymous reviewers for equally helpful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Boyd, R. N. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Ithaca, NY: Cornell University Press.
- Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Hartmann, N. (1932). *Ethics* (Vol. 1–3). London: George Allen & Unwin Ltd.
- Hegselmann, R., & Krause, U. (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 1–33.
- Hegselmann, R., & Krause, U. (2006) Truth and cognitive division of labour: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation*, 9(3), 1–28.
- Krause, U. (2008). Compromise, consensus and the iteration of means. *Elemente der Mathematik*, 63, 1–8.
- Lorenz, J. (2007). *Repeated averaging and bounded confidence modeling, analysis and simulation of continuous opinion dynamics*. PhD thesis, Universität Bremen.
- Lorenz, J., & Lorenz, D. A. (2008) On conditions for convergence to consensus. Arxiv preprint arXiv:08032211.
- Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.
- Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford: Oxford University Press.
- Sturgeon, N. L. (1985). Moral explanations. In D. Copp & D. Zimmerman (Eds.), *Morality, reason, and truth: New essays on the foundations of ethics* (pp. 49–78). Totowa: Rowman & Allanheld.
- Tersman, F. (2006). *Moral disagreement*. Cambridge: Cambridge University Press.