On Robustness in Cosmological Simulations

Marie Gueguen*†

The Cold Dark Matter model faces many controversies at small scales, as simulations fail to reproduce the observed properties of dark matter (DM) halos. Since rival DM models differ on their predictions about the structure of DM halos, understanding their mass distribution is crucial for determining the nature of DM. However, only numerical approaches to determining what a hypothesis implies for the mass distribution are possible. Hence, simulations are a crucial part of evaluating DM rival models, as is assessing their reliability. I argue that robustness analysis is not a sufficient criterion for the trustworthiness of cosmological simulations.

1. Introduction. The standard cosmological model describes a nearly homogeneous early universe, where small density inhomogeneities evolve with time through gravitational collapse to form the large- and small-scale structures we now observe. An essential component of this cosmological model is a mysterious kind of matter called 'dark matter' (DM) that only interacts gravitationally. The Cold Dark Matter (CDM) model predicts with great accuracy important large-scale structure properties but does not fare as well on small scales, where simulations fail to reproduce the observed abundance and demographics of DM halos' structure.[1] Since rival models such as cold,

*To contact the author, please write to: Center for the Philosophy of Science, University of Pittsburgh, 1117 Cathedral of Learning, 4200 Fifth Ave., Pittsburgh, PA 15260; e-mail: mgueguen@uwo.ca.

†I am very grateful to Chris Smeenk for his support and advice. I would also like to thank Frank van den Bosch, Eric Winsberg, Craig Fox, the participants of the Rotman Summer Institute of Cosmology, and an anonymous referee for very helpful feedback and discussion. This article is based on work done while I was funded as a graduate student researcher under John Templeton Foundation grant 61048, New Directions in Philosophy of Cosmology.

1. In Cold Dark Matter, 'Cold' means that particles move at nonrelativistic speed.

Philosophy of Science, 87 (December 2020) pp. 1197–1208. 0031-8248/2020/8705-0036$10.00
Copyright 2020 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

warm, or self-interacting DM agree on a large scale, but differ on their predictions about the structure of DM halos, understanding how mass is distributed in these halos is crucial for determining the nature of DM. At such a scale, though, only numerical approaches to determining what a hypothesis implies for the halos' mass distribution are possible. Nonlinear effects related to star formation and gas dynamics make it impossible to determine the mass distribution analytically. Hence, numerical simulations are needed to determine the mass distribution, and these simulations are a crucial part of evaluating the CDM model and various rival hypotheses.

Understanding in which case a simulation can succeed in (dis)confirming a model is, yet, still a challenge in cosmology. At small scale, indeed, the CDM seems to do worse than its rivals: simulations predict, for instance, much more satellite galaxies within DM halos than is actually observed. Before 1998, however, the problem was the exact inverse: simulated DM halos did not present enough substructure compared to observations. When a model is so sensitive to modeling assumptions, what is the conclusion that should be drawn from a mismatch between simulation outcomes and observations? How can we assess whether this 'missing satellite' problem stems from numerical artifacts or constitutes a genuine failed prediction?

In biology and climate science, evaluation of when numerical evidence confirms a model and in what sense this confirmation must be understood has been based on robustness analysis (Levins 1966; Weisberg 2012; Wimsatt 2012).[2] In cosmology, astrophysicists have been relying on a similar methodology, according to which results that resist a change of values of the numerical parameters are considered trustworthy. In this article, I argue that robustness is not a sufficient criterion for determining when a prediction is reliable in *N*-body simulations, for equally robust but mutually exclusive predictions can obtain in *N*-body simulations. Even more worrying, robustness is sometimes a direct consequence of numerical artifacts.

2. Robustness and Convergence Studies. The CDM model faces a number of problems at small scale. For example, it predicts way more substructure in a DM halo of the size of the Milky Way than is actually observed. Only 59 satellite galaxies seem to orbit our Milky Way, whereas several thousand are predicted by this model. Likewise, the density profile drawn from this model by Navarro, Frenk, and White (1997) predicts a steep, cuspy profile in the central region of the DM halo, with infinite density at the center. Yet, observations favor, especially for dwarf galaxies and low-surface-brightness galaxies, a 'cored' profile, with a flatter density profile as the radius tends toward zero. Observations for dwarf and low-surface-brightness

2. Although, see Parker (2011) for an argument to the effect that robust predictions in climate science models are not sufficient for trustworthiness.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

galaxies are especially problematic because these galaxies are mostly made of DM, which means that this discrepancy cannot be washed away by adding baryonic physics to the simulations.[3]

Given that the predictions of the CDM model are not drawn from first principles but from analytical fits to DM-only simulations, assessing the extent to which these two problems challenge this model is as difficult as it is urgent. How can we assess whether the discrepancy between the simulated systems and those observed stems from the physical model or from an erroneous code, whether the simulated outcome is altered by numerical artifacts or constitutes a genuine failed prediction? Astrophysicists rely on robustness analysis to decide when the outcome of a simulation is trustworthy. Robustness analysis was first introduced by Levins (1966) as a way to assess the reliability of models in population biology in the absence of a background theory providing analytically soluble equations. According to Levins, a method is needed to evaluate the impact of the simplifications on which models are based and to determine "whether a result depends on the essentials of the model or on the details of the simplifying assumptions" (423). This role is played by robustness analysis: by addressing the same problem with a diversity of models based on different 'lies', one can test whether these models agree on their predictions. Such agreement is taken to confirm the independence of the models from their characteristic simplifications: "Hence our truth is the intersection of independent lies" (23). Levins's account relies on a form of eliminative reasoning, according to which each model must exclude a given possibility. However, Orzack and Sober (1993) have shown that this reasoning only yields valid inferences when an exhaustive set of all possible models is examined. Such an exhaustivity is yet excessively difficult to achieve.

Wimsatt further fleshed out this methodology by suggesting a four-step procedure for robustness analysis:

1. Analyze a *variety* of *independent* derivation or measurement processes.
2. Look for things that are *invariant* over the results of these processes.
3. Determine the *scope* of the processes across which they are invariant and the *conditions* on which their invariance depends.
4. Analyze and explain any relevant *failures of invariance* (Wimsatt 2012, 62).

For Wimsatt, the purpose of robustness analysis is, first, to distinguish the "reliable from the unreliable"; second, to show the invariance of that which reliability (e.g., a prediction) is scrutinized over different models, in order to

3. A more exhaustive review of all the controversies arising at small scales for the CDM model can be found in Weinberg et al. (2015).

https://doi.org/10.1086/710839 Published online by Cambridge University Press

build confidence in their independence from characteristic simplifications: and, finally, to determine the scope of this invariance. Wimsatt does not justify the robustness of a property via eliminative reasoning but through its overdetermination by independent models. Supposing that this independence can be qualified, such an account does not require us to consider all possible models but only models independent 'in an appropriate way'. In what follows, I consider any procedure satisfying these Wimsattian features an instance of robustness analysis.[4]

Robustness analysis, in astrophysics, takes the form of 'convergence studies'. The idea is to determine whether unconstrained numerical parameters affect the outcome of simulations, by systematically varying their value and defining the value range under which the structure of the simulated halo remains unaffected by such variations. In that case, the halo is deemed 'appropriately resolved'. Such a procedure satisfies the characteristic features of robustness analysis listed above and will thus be considered as such throughout this article.

The influential convergence study of Power et al. (2003) has contributed to set up the parametrization of *N*-body simulations for the last 15 years. Their methodology consists in, first, simulating a large cosmological volume with low resolution, tracking the structure growth seeded by primordial density fluctuations. Then, they zoomed in on some targeted halos and resimulated them at higher resolution. Using this sample of halos, several hundred simulations (typically with a resolution of 32^3 particles) were run, allowing them to survey the parameter space by varying the numerical parameters and drawing preliminary convergence results. Finally, the convergence criteria were confirmed with another series of simulations of higher mass resolution—a series run with 64^3 particles and a few (given how expensive they are) with 128^3 and 256^3 particles.[5] If, in the region of the parameter space defined by the convergence criteria, the predictions remain the same despite the increased resolution, then one can trust that they are independent of the numerical parameters' values. The convergence of the simulated mass profile is supposed to warrant its independence from numerical parameters and that it is not affected by artifacts.

4. More recently, Weisberg (2012) has suggested a procedure to establish robust theorems that consists in examining a group of models, searching for a 'robust property', then finding the core structure giving rise to this property. However, the lack of modularity of simulations in cosmology undermines attempts to find the common structure responsible for the robust property and the subsequent formulations of a robust theorem. I therefore leave aside Weisberg's proposal.

5. That convergence should be reached in time $t \leq 1.7\tau_r$, with τ_r the relaxation time, is an example of such a convergence criterion, since the influence of artificial collisions between particles seems to produce a core profile after $1.7\tau_r$.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

3. Against Convergence

3.1. Convergence Is Not Sufficient. Confidence in predictions about the abundance of DM subhalos extracted is usually explained by the fact that they seem unaffected by an increase of resolution above 50–100 particles per subhalo (see, e.g., Springel et al. 2008). As mentioned above, however, this prediction is very sensitive to modeling assumptions. Up to the close of the last century, simulations were suffering from an 'overmerging' problem, in that not enough substructure was predicted to match the observations. Several culprits had been proposed back then, with no consensus on the cause of the subhalos disruption but with an agreement that it was a numerical problem. Moore, Katz, and Lake (1996) blamed inadequate force softening.[6] Carlberg (1994), however, argued that a low mass resolution could cause two-body heating and artificially enhance matter disruption. Since this problem was superseded by a 'missing satellite' problem as the resolution of simulations increased, no definitive conclusion was drawn about the cause of the overmerging. Yet, as shown by van den Bosch et al., subhalos disruption is still "extremely prevalent in modern simulations, with . . . ~ 65 percent of all subhaloes accreted around $z = 1$ . . . disrupted by $z = 0$" (2017, 3044). The question thus arises: Are *N*-body simulations still suffering from overmerging, or can they be considered reliable, on the basis of convergence studies? Is the subhalos disruption a physical mechanism or the result of numerical artifacts?

This question is addressed by van den Bosch and Ogiya (2018), in a paper also aiming at gaining a better understanding of the nonlinear effects of tidal stripping on subhalos.[7] Tidal processes are very difficult to describe, as the stripping of matter causes the subhalo remnant to fall out of virial equilibrium and then to revirialize by expanding, thereby provoking more stripping of matter, and to fall out of equilibrium once again.[8] No analytical theory is available to describe such complicated processes of de- and revirialization: only simulations can tell how the density distribution of the subhalo is affected

6. Because of limits in computational power, real DM particles are substituted in simulations by heavy particles. Hence, the gravitational force can generate very large, unphysical accelerations when two particles get very close to each other. Force softening is used to smooth the gravitational potential and suppress these accelerations below a typical distance—the 'softening length'.

7. 'Tidal stripping' refers to the escape of matter due to the tidal forces exerted by the host halo on the subhalo. Beyond some limit, the tidal forces exerted by the host overcome the gravitational force bounding the subhalo together, resulting in its dislocation.

8. The virial theorem applied to celestial bodies states that the total energy of a system is equal to half its gravitational potential energy. If this equality does not hold, the system is either collapsing—the gravitational potential energy exceeds the kinetic energy—or expanding. A system for which this equality applies is considered in *virial equilibrium*.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

by tidal stripping. It is thus crucial to disentangle what pertains to this stripping mechanism and what artificially results from an inadequate parametrization in simulations.

In order to do so, the authors came up with an idealized scenario in which the physical hypothesis of tidal stripping could be tested against the numerical hypothesis that matter disruption is mostly caused by inadequate force softening. They focused on the tidal evolution of an isolated subhalo on a circular orbit, given that, in such a scenario, only the host halo's tidal field and the numerical parameters can affect the subhalo disruption. Its isolation forbids high-speed encounters with other subhalos causing more matter to disrupt, and the circular orbit excludes a fast pericentric passage that would deform the orbiting body and cause internal heating and thus more mass loss. The goal is to isolate two possible causes of the disruption and find the culprit by looking at the evolution of the bound mass of subhalos when varying the strength of the tidal field and the force softening.

Figure 4 in van den Bosch and Ogiya (2018, 4071) suggests that increasing the strength of the tidal field by decreasing the orbital radius does not lead to the disruption of subhalos, except in the extreme case in which the orbital radius is chosen so that $r_{\rm orb}/r_{{\rm vir},h} = 0.1$.[9] However, at $r_{\rm orb}/r_{{\rm vir},h} = 0.1$, the subhalo totally disrupts after ~ 13 Gyr. This results concords with the results of the Millennium and the Bolshoi simulations according to which $r_{\rm orb}/r_{{\rm vir},h} = 0.1$ is the typical radius at which subhalos undergo disruption. If this prediction is a physical prediction, as it is taken to be by most astrophysicists, it should not be significantly affected by variations in numerical parameters.

Now, look at the impact of varying force softening and mass resolution (van den Bosch and Ogiya 2018, fig. 7, 4074). According to Power et al. (2003), for instance, the optimal value for the force softening $\varepsilon_{\rm opt}$ ranges between 0.02 and 0.06. Since a force softening ε that is too large results in smaller density distribution at small radii, and since less dense systems are more exposed to tidal stripping, for $\varepsilon > \varepsilon_{\rm opt}$ one expects enhanced stripping and matter disruption. Likewise, a small softening is more exposed to two-body relaxation effects that flatten the central density profile and enhance disruption. Thus, force softening should have no other consequence than increasing the mass loss. Surprisingly however, the simulations show the opposite: for $\varepsilon < \varepsilon_{\rm opt}$, the bound remnants are larger and survive longer. One can suspect then that the physical disruption is very sensitive to the value assigned to ε.

9. Here, $r_{\rm orb}/r_{{\rm vir},h} = 0.1$ is the orbital radius of the subhalo expressed in units of the host halo's virial radius; $r_{\rm orb}$, the distance between the centers of mass of the host and subhalo; and $r_{{\rm vir},h}$, the virial radius, defined as the radius inside of which the average density is $\Delta_{\rm vir} = 97$ times the critical density for closure.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

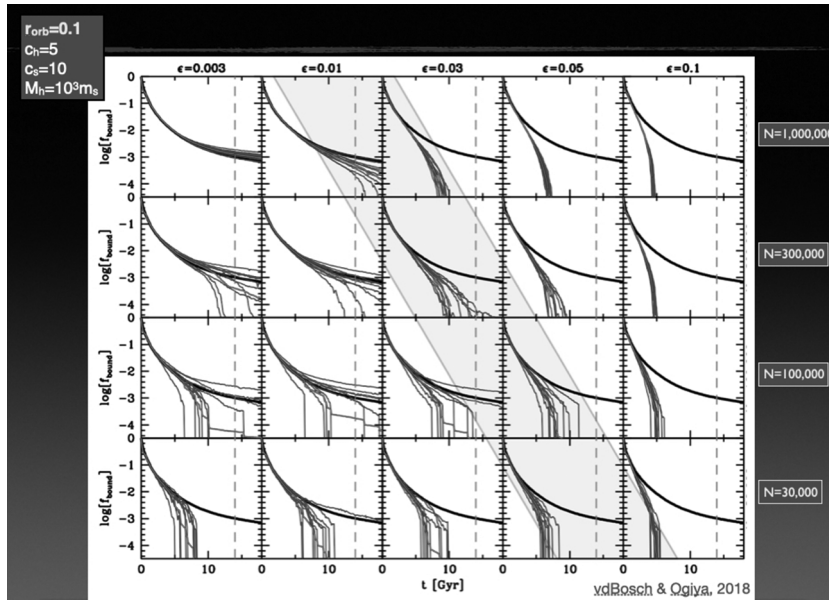


Figure 1. Bound fraction of mass as function of time for simulations with different mass resolution N_p and different force softening ε. (Source: van den Bosch and Ogiya [2018], fig. 10, 4077.) Color version available as an online enhancement.

Consider now simulations at the orbital radius $r_{\rm orb}/r_{{\rm vir},h} = 0.1$ when the mass resolution and the force softening are increased together (see fig. 1). If you look for a region in the parameter space where the instability caused by mass resolution is kept under control and a 'converged' fraction of bound mass $f_{\rm bound}(t)$ can be found, the upper-left corner of figure 1 gives you the closest result to such an ideal.[10] But there, subhalos do not disrupt after 13 Gyrs. On the contrary: a large fraction of bound mass survives.

Note the contrast between this result at large N_p and small ε and those observed along the shaded band, which corresponds to the scaling between mass resolution and force softening defined by Power et al. (2003). This scaling is obeyed by most of the state-of-the-art simulations since 2003. If one focuses on the averaged results of simulations, the bound fraction of mass appears converged: the prediction stays more or less the same despite increasing the mass and the force resolution, which is precisely why it is considered as robust by Power et al.'s standards. Yet, this indicates a full disruption of the

10. The sensitivity to discreteness noise is characterized by van den Bosch and Ogiya (2018) using the variance $\sigma_{\log f}$ in $\log(f_{\rm bound})$. 'Converged' here means that "(i) no significant changes occur when N_p is increased further, and (ii) the standard deviation in $f_{\rm bound}$ after one Hubble time is sufficiently small (i.e., $\sigma_{\log f} \leq 0.05$)" (4076).

https://doi.org/10.1086/710839 Published online by Cambridge University Press

subhalo between 5 and 8 Gyrs. Thus, simulations converge on predicting that subhalos fully disrupt after 8 Gyrs but also on predicting that they survive after 13 Gyrs.

What should alert the philosopher here is that convergence alone will not tell which of these results is the correct one or whether one of them is. Given the range of the optimal softening defined by Power et al. (2003), increasing the mass resolution and the force softening gives good agreement among simulations and confidence in the prediction that subhalos at $r_{\mathrm{orb}}/r_{\mathrm{vir},h} = 0.1$ will not survive after one Hubble time. But if another region of the parameter space is scrutinized, convergence is found on a prediction that contradicts the former. As summarized by van den Bosch and Ogiya, convergence is "not a sufficient condition to guarantee that the results are reliable" (2018, 4067).

3.2. Pseudoconvergence or Convergence? In the previous subsection, I argued that robustness analysis in the form of convergence studies is not sufficient to exclude numerical artifacts and that robust predictions cannot be considered reliable on such grounds. Here, I push this thought further and argue that convergence can result from artifacts, following the work of Anton Baushev on the cusp-core problem.

In *N*-body simulations, real DM is represented by a limited number of heavy test bodies, to make the computational task tractable while preserving the averaged density of the system. However, whereas DM is collisionless, heavy test bodies undergo collisional effects, which in turn affect the density profile of DM halos. These collisional effects are characterized by the relaxation time $\tau_r = N(r)/(8 \ln \Lambda \,.\, \tau_d)$, with $N(r)$ the number of test bodies inside a sphere of radius r, $\ln\Lambda$ the Coulomb logarithm, and τ_d the characteristic dynamical time of the system at radius r. This is why Power et al. (2003) recommend, as one of their convergence criteria, that convergence be reached in a time $t \leq 1.7\tau_r$ (see n. 5).

Baushev, however, has shown that cuspy profiles are generated by an intensive energy relaxation. This fact raises important questions. DM halos undergo violent relaxation when they collapse, as density inhomogeneities create small-scale gravitational fields mediating the exchange of energy among DM particles. But they do so only at the moment of the collapse. The halo, once formed, has a stationary gravitational field. Why is it, then, that *N*-body simulations based on the CDM model predict cuspy density profiles for formed halos? If the cuspy profile stems from collisionality, where does it come from? Is it artificial or physical?

Baushev et al. (2017) address this question with a methodology very similar to that of van den Bosch and Ogiya (2018). First, they propose an idealized scenario in which all sources of collisionality are turned off except the one under scrutiny. Thus, they simulate an isolated halo, to avoid the

https://doi.org/10.1086/710839 Published online by Cambridge University Press

tidal influence of nearby halos and the gravitational capture of more mass from the surrounding, potentially leading to a secondary violent relaxation. They chose a halo with a Hernquist density profile that behaves like the Navarro et al. (1997) profile in the central region but is fully stable. Given that such a halo has a constant gravitational potential field $\phi(r)$, the density and velocity profiles should remain the same and all (the implicit functions of) the integrals of motion such as the specific energy $\omega = \phi(r) + (v^2/2)$, the specific angular momentum $\vec{K}$, and the apocenter distance r_0 should be conserved.[11] Thus, any variation of them must be due to numerical effects. Such symptoms of artificial collisions can be tracked in their simplified scenario. Figure 2 of Baushev et al. (2017, 6) shows the variations of the integrals of motion as a function of the radius over 200 snapshots and confirms that they vary significantly even in a single time step. Thus, the convergence criterion of Power et al. (2003) by no means guarantees that all sources of numerical artifacts have been excluded.

This preliminary conclusion raises another interesting question: Why does the density profile remain stable if the integrals of motion vary? In other words, is the stability of the cuspy profile a mere coincidence or the result of the numerical effects observed above? If the system is collisional, then it is better modeled by the Fokker-Planck equation, which models dynamical friction and diffusion, than by the collisionless Boltzmann equation usually appealed to. In that case, the diffusion streams created by the particle interactions could contribute in forming a stable density profile. And, as it turns out, the Fokker-Planck equation does have a stationary solution close to Navarro et al.'s (1997). If the cusp is a product of this Fokker-Planck diffusion, then artificial collisions should form a downward and an upward stream of particles, with increasing and decreasing r_0 respectively, that complement each other and thereby explain the stability of the profile. This prediction can be tested by taking two adjacent snapshots S_1 and S_2 and calculating, for an array of radii r, the number of particles $\Delta N_+(r)$ that had $r_0 < r$ at S_1 and $r_0 > r$ at S_2 and the number of particles $\Delta N_-(r)$ that had $r_0 > r$ at S_1 and $r_0 < r$ at S_2. Figure 4 in Baushev et al. (2017, 9) confirms that the Fokker-Planck streams are present, complement each other very well outside a given radius, and are important enough to shape the density profile.

In sum, the variations of the integrals of motion unambiguously demonstrate that converged N-body simulations still suffer from artificial collisionality. Convergence fails to identify results independent of artifacts. Moreover, confidence in the density profile is not warranted by its stability since this stability is itself produced by virtue of numerical artifacts.

11. Here, r_0 is the maximum distance that the particle can move off the center and depends only on the integrals of motion: $\omega = \phi(r_0) + K^2/2r_0$.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

3.3. Discussion. Although convergence failed in the two cases discussed above, in diagnosing artifacts, one could defend robustness by arguing that Baushev et al.'s (2017) and van den Bosch and Ogiya's (2018) studies exemplify the fourth step of Wimsatt's procedure—that they demonstrated the unreliability of simulations based on the breaking down of their robustness. These two papers cannot, however, be considered instances of robustness analysis, for at least two reasons. Let us consider first the target of van den Bosch and Ogiya (2018). What the authors examined is not the trustworthiness of the simulations outcomes but that of the convergence criterion itself. Their main conclusion is that "most, if not all, disruption of substructure in *N*-body simulations is numerical in origin," a conclusion that "questions whether the fact that subhalo mass functions appear to be converged down to 50–100 particles per subhalo implies that results are reliable" (4084). In other words, the authors showed that, even though convergence was reached, numerical artifacts had not been excluded and that the results backed up through convergence were still not reliable. What is at stake here is not whether the amount of substructure predicted through *N*-body simulations is a reliable prediction but whether the reason it is taken to be a reliable prediction is sound.

Still, one could insist that this study makes a negative use of robustness analysis, that is, that it uses robustness to show that predictions made about the number of satellite galaxies that should be observed in halos are not reliable, because their apparent robustness breaks down when the resolution is increased further. However, this reconstruction of their argument is misleading: what we see in their paper is not that the appearance of convergence of the results based on Power et al. (2003) breaks down while 'true' convergence is reached for higher resolution. What we see instead is a first instance of convergence for a given range of values and a second one for another range of values. Asserting that the robustness of the first breaks down when increasing the resolution is assuming that we have more reasons to trust the second set of results than the first. However, on the basis of convergence only, we have equally good reasons to accept one result or the other—and, conversely, equally good reasons to reject them. Convergence alone will not tell us which result to trust. But here is where the problem becomes particularly sharp: in the astrophysical context at least, no other criterion is available to back up or supplement the results of robustness analysis. Simulations are used to determine what rival DM models tell us about the structure of the universe precisely because nonlinear effects related to star formation and gas dynamics make it impossible to determine the distribution of matter in DM halos analytically. Thus, there is no analytical solution against which the results of simulations can be tested. Furthermore, a comparison between observations of 'real' DM halos and simulations is of no help, for at least two reasons. While the similarity between the model and the real world is often used

https://doi.org/10.1086/710839 Published online by Cambridge University Press

to validate the former, what is under scrutiny here is precisely the pertinence of such a comparison given that we do not know whether the simulated outcome reliably tracks the predictions of the CDM model or crucially depends on the parametrization process. Second, parametrization—or 'calibration', or 'tuning'—is also the process through which one improves the agreement with real world data when other processes of validation are not feasible (Oberkampf 2014, 33). How could one then appeal to the agreement between observations and simulations when trying to circumscribe the impact of calibration on the simulation outcome? But in the absence of any other criterion to assess reliability, how could one conclude anything from the fact that two mutually exclusive but converged results can be found in two different regions of the parameter space?

Even more undoubtedly, robustness analysis has no role to play in Baushev et al.'s result. Like that of van den Bosch and Ogiya (2018), the aim of Baushev et al. (2017) is not to test the reliability of the Navarro et al. (1997) density profile per se but to test whether its robustness says anything about its reliability. This goal does not involve any attempt to make the stability of the profile break down. On the contrary, their argument shows that its robustness will not break down because it results from numerical artifacts. Simulations converge because the artificial diffusion streams generated by numerical artifacts complement each other. The lesson to be drawn from this article is that robustness by no means says anything about the physical nature of the prediction made and even less about its reliability.

4. Conclusion. I have argued that robustness analysis, in the form of convergence studies, fails to exclude numerical artifacts in *N*-body simulations and thus to warrant their reliability. Simulations in astrophysics constitute a very exciting opportunity for philosophers to explore the limits of robustness and to suggest possible rivals better suited for cosmological simulations. 'Crucial' simulations, where a numerical hypothesis about the origin of a prediction is tested against a physical explanation, used both by van den Bosch and Ogiya (2018) and Baushev et al. (2017), constitute a possible candidate for this task.

REFERENCES

Baushev, A., L. del Valle, L. Campusano, A. Escala, R. Muñoz, and G. Palma. 2017. "Cusps in the Center of Galaxies: A Real Conflict with Observations or a Numerical Artefact of Cosmological Simulations?" *Journal of Cosmology and Astroparticle Physics*. https://iopscience.iop.org/article/10.1088/1475-7516/2017/05/042.

Carlberg, R. 1994. "Velocity Bias in Clusters." *Astrophysical Journal* 433:468–78.

Levins, R. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54 (4): 421–31.

https://doi.org/10.1086/710839 Published online by Cambridge University Press

Moore, B., N. Katz, and G. Lake. 1996. "On the Destruction and Over-Merging of Dark Halos in Dissipationless N-Body Simulations." *Astrophysical Journal* 457:455–59.
Navarro, J. F., C. S. Frenk, and S. D. White. 1997. "A Universal Density Profile from Hierarchical Clustering." *Astrophysical Journal* 490 (2): 493–508.
Oberkampf, W. L. 2014. "Verification and Validation in Computational Simulation." Unpublished manuscript, Plasma Science and Fusion Center, MIT.
Orzack, S. H., and E. Sober. 1993. "A Critical Assessment of Levins's *The Strategy of Model Building in Population Biology* (1966)." *Quarterly Review of Biology* 68 (4): 533–46.
Parker, W. S. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78 (4): 579–600.
Power, C., J. Navarro, A. Jenkins, C. Frenk, S. D. White, V. Springel, J. Stadel, and T. Quinn. 2003. "The Inner Structure of ΛCDM Haloes." Pt. 1, "A Numerical Convergence Study." *Monthly Notices of the Royal Astronomical Society* 338 (1): 14–34.
Springel, V., J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. White. 2008. "The Aquarius Project: The Subhaloes of Galactic Haloes." *Monthly Notices of the Royal Astronomical Society* 391 (4): 1685–711.
van den Bosch, F. C., and G. Ogiya. 2018. "Dark Matter Substructure in Numerical Simulations: A Tale of Discreteness Noise, Runaway Instabilities, and Artificial Disruption." *Monthly Notices of the Royal Astronomical Society* 475 (3): 4066–87.
van den Bosch, F. C., G. Ogiya, O. Hahn, and A. Burkert. 2017. "Disruption of Dark Matter Substructure: Fact or Fiction?" *Monthly Notices of the Royal Astronomical Society* 474 (3): 3043–66.
Weinberg, D. H., J. S. Bullock, F. Governato, R. Kuzio de Naray, and A. H. G. Peter. 2015. "Cold Dark Matter: Controversies on Small Scales." *Proceedings of the National Academy of Sciences* 112 (40): 12249–55.
Weisberg, M. 2012. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
Wimsatt, W. C. 2012. "Robustness, Reliability, and Overdetermination (1981)." In *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*, ed. Léna Soler et al., 61–87. Dordrecht: Springer.

https://doi.org/10.1086/710839 Published online by Cambridge University Press