We thank the referees for their very helpful and positive comments. We are particularly pleased that referee B finds the work to be well-written, comprehensive, and accessible, in addition to finding the scientific content to be both timely and satisfying. We hope that our responses address referee A's concerns about clarity as well as provide a more direct link to future experiments (which we are actually in the process of conducting!). We believe that the changes made to the manuscript should now render it suitable for publication in PRL.

Below, we respond to each point in turn. The referee's comments are shown in **bold**, our response in plain text, and modifications to the manuscript are shown in *italics*.

Report of Referee A -- LS18194/Hass

The authors investigate a simple one-dimensional model of random walk in random environment (RWRE), namely, a special case of the so-called Beta model. This is an exactly solvable model that was shown by one of the co-authors in 2017. The main focus here is on the extreme statistics of the right-most particle at time t. Using high precision numerical simulations, the authors analyze the behavior of the variance of this particle and reveal several asymptotic regimes which are different from the conventional linear time growth of the variance of a typical particle (like for Brownian motion).

While the results sound interesting, the manuscript requires substantial improvements and clarifications. Moreover, most asymptotic results seem to be reported earlier, as stated by the authors. Finally, even though simple models are often employed in statistical physics to understand the properties of complex systems, the discussion of eventual "experimental confirmations" (as stated in the abstract) sounds a little speculative. In particular, the authors provide no evidence that "...much of the picture presented here will persist beyond the discrete and solvable model setting and be observable in experiments" (from the conclusion section).

The referee raises a number of important points here that we will address sequentially.

As the referee keenly notes, the asymptotic results that are utilized here have, indeed, been reported in previous works. The issue, however, is that those asymptotic results address different asymptotic regimes as t and log(N) go to infinity. However, as the referee says later in their review, for realistic system sizes, log(N) is rather small. Thus, it is not at all clear how to make contact between the asymptotic results and real system sizes that could have a bearing on experiments. The purpose of studying the solvable model and performing numerics is to make this contact and develop a set of theoretical predictions that can be used in subsequent experimental work.

There is, however, also a key theoretical contribution in our work that is confirmed by numerics and should be very useful in experiments. In real world systems it is impossible to measure the

1/N quantile of the random (environment dependent) transition probability distribution, or indeed really any direct measurement of that probability distribution for a single particle in a given environment. However, it is possible to measure the location of tracer particles. We show here that by studying the fluctuations of the maximum of N tracer particles and removing the sampling fluctuations (that can be understood from the standard Einstein diffusion coefficient), it is possible to actually recover the behavior of the environmental fluctuations, namely the fluctuations of the 1/N quantile. In particular, doing this we can recover a new scaling exponent that is typically not seen in diffusions. This approach is tailor made for experimental work.

Regarding the potential for experimental confirmation, we completely understand that this sounds speculative. However, the referee will be happy to hear that we have already designed two experiments and are in the midst of building them. We will say more about this in response to the comment labeled 1) below, and will note how we have changed the text to address this concern.

The final point raised by the referee is about whether this discrete and solvable model will have predictive power. This is a fundamental question and the referee is quite right that we had not sufficiently addressed why we expect a high degree of universality. There is a lot to unpack in this question. For instance, the most basic question of predictive power of studying our model is whether the asymptotic theory results have anything to do with the behavior of the model for finite and even small system sizes and times. This point, we believe, is well addressed in our letter by the numerics and comparisons with theory provided therein.

Of course, the referee is rightly more concerned with the question of whether the results that we develop by studying the Beta RWRE model can be applied in understanding the behavior of (1) other models, and more importantly (2) real world systems. In other words, what about our picture is universal, and why. These are very hard, yet important, questions that deserve a lot of investigation. However, we can provide some evidence here for why we believe there is a high degree of universality in our results.

First, let us clearly lay out (as in the conclusion) what should be universal about our picture. For other models the environmental variance should satisfy the same form of power-laws as uncovered here, up to different (model dependent) coefficients. (This should be compared to the behavior of general independent identically distributed increment random walks where the variance of a walker always shows linearly, up to a diffusion coefficient scaling.) The other part of the picture that should be universal is that the variance of the maximal particle decouples as the sum of the environmental variance and the sampling variance.

Why do we believe that this picture will hold? One piece of evidence for universality is provided by the fact that we can perform the same exact analysis when the U[0,1] distribution (for the random B(t,x)) is replaced with the general Beta(\alpha,\beta) distribution. For general \alpha,\beta>0, the model is still solvable and that structure enables a more general version of the analysis we performed. While the formulas become considerably more complicated, the outcome is that these power-laws and variance decoupling results still hold. Another piece of

evidence comes from a different exactly solvable model, that of sticky Brownian motions studied by Barraquand and Rychnovsky. In fact, that model is a limit of the general \alpha,\beta model, so it is perhaps not so surprising that the phenomena observed in the RWRE setting persists in this limit. On the other hand, this is now a bona-fide continuum model, and thus it becomes clear that what is being observed extends, at least in one case, to the continuum. This is the limit of what has been done (or seems to be doable) using exact solvability.

To move beyond solvable models is often challenging in the setting of complex dynamics such as these. However, there is a saving grace. The environment dependent transition probability p_B(x,t) satisfies a simple recursion relation (given as equation 1 in our manuscript) which can be seen as a discrete version of the multiplicative noise stochastic heat equation. This also provides an interpretation for p_B(x,t) as a directed polymer partition function. For reference, we say that Z(x,t) solves the multiplicative noise stochastic heat equation (SHE) d_t Z(x,t) = ½ d^2_x Z(x,t) + \xi(x,t)Z(x,t) where now x and t are real valued and \xi(x,t) is a space-time white noise field. The logarithm of the SHE h(x,t) = log Z(x,t) solves the KPZ equation d_t h(x,t) = ½ d^2_x h(x,t) + ½ (d_x h(x,t))^2 + \xi(x,t). The point here is that log(p_B(x,t)) satisfies a discrete version of the KPZ equation. The KPZ equation is believed to have a very broad universality class – this has been argued with a host of methods ranging from renormalization group to numerics to real physical experiments (e.g. studying growth models, particle systems, domain wall boundaries, etc). Our expectation for the universality of our picture very much stems from this evidence for the robustness of KPZ universality.

To address this question about predictive power / universality of our results, we have modified our text as indicated below in a manner that hopefully captures the essence of what we have said at further length above. While we agree that it is important to make this clearer, it is not the primary contribution of our work and thus we had to be selective in how much we say here. We have augmented our description in the “Background” section to now read:

“The recursion relation (see (1) below) for RWRE transition probabilities solves a discrete version of the multiplicative noise stochastic heat equation (mSHE) which, in the continuum, is related to the KPZ equation by taking a logarithm. Hence, large deviations for RWREs, in particular beyond the solvable model and even in experimental set- tings, may relate to the KPZ equation and its universality class – especially in light of the rich canon of work on KPZ universality in various contexts using theoretical [59, 76], numerical [77] and experimental [78] methods. The KPZ connection is quite useful since its statistics and power- laws are well studied.”

We have also modified the text in the conclusion to read:

“The link between RWREs and KPZ universality with its wealth of theoretical, numerical and experimental evidence strongly suggests that aspects of the picture presented here will persist beyond discrete and solvable models, even to experiments. When t is of order log(N), variances should be non-universal, depending in a difficult to determine way on the nature of the environment. By contrast, when t >> log(N), we anticipate that the scaling exponents and

functional forms we have identified for the variances of Env^N_t ,Sam^N_t and Max^N_t will be universal, as will the relation (7). The leading coefficients in (5) should be non-universal and hold within them all of the accessible information about the correlation structure of the environment – we call these extreme diffusion coefficients. Further theoretical study, such as for the general α,β Beta RWRE model, should provide a natural first test of this universal picture and an understanding of how the extreme diffusion coefficients relate to the microscopic environment. A continuum model that should provide an even wider testing-ground amenable to numerics involves particles x i (t) for i = 1,2,... satisfying dx_i(t) = F(x_i(t),t)dt + D(x_i(t),t)dB_i(t) where F(x,t) and D(x,t) are random forcing (as in [62]) and diffusivity (generalizing diffusing diffusivity, c.f. [85]) fields common to all particles while B i are Brownian motions independent between different i. Changing the correlation structures of F and D will probe the transition between temporally mixing versus quenched environments, which should have very different behavior (c.f. [86, 87]) and warrants further study."

Overall, it is recommended to incorporate the Supplemental Material (SM) into the main text in order to facilitate reading and resubmit a revised version as a regular article to Physical Review E. The following points should be taken into account.

We appreciate the referee's desire for a single, longer, manuscript. However, we strongly believe that this work contains a clear and impactful point that is of interest to the broad audience of PRL and so have had to work within the confines of this format. Nevertheless, we have pulled in material from the supplement as appropriate and modified the manuscript with an eye towards increased clarity and readability. In particular, the main text is now more self-contained, including a better description of the numerical methods, the key background results that lead to equations (3) and (4), the full description of the function Var^asy(Env^N_t), as well as a clearer explanation for why the KPZ equation and universality class arise in studying the RWRE problem. We have also moved into the main text a very informative figure illustrating the dynamics of our random process.

On top of all of this, the main text now also contains a discussion of potential (and in progress) experimental tests for the predictions developed herein (see below).

1) As said earlier, eventual relations to experiments should be either detailed (which type of experiment or experimental method, what are possible systems, etc), or eliminated. Otherwise they sound as speculations.

The referee is correct that our discussion of future experiments was too brief and hence sounded like speculations. In fact, we are in the early stages of conducting these very experiments. In particular, we are performing two different kinds of measurements. The first is to measure the leading edge of diffusing colloids in a quasi-1d system (a capillary tube). The second is to measure the time of first arrival for photons from a femto-second laser pulse traveling through a scattering media. These complementary measurements on the time of first passage of diffusing objects open the door to experiments done on all manner of diffusing objects, including light or sound diffusing through a scattering medium, dye molecules in a fluid,

or any other object whose first passage can be measured. The present work sets the stage for understanding the outcomes of these experiments by providing a link from asymptotic theory to finite size and time systems, thus providing a detailed description of the way in which space time randomness affects the extreme value statistics of diffusion. In particular, the ½ power-law that we find for the environmental variance as well as the method of accessing this from subtracting the sampling variance from the variance of the observed maximum provides a clear roadmap for how to start to analyze experimental data.

We have included in the conclusion the below text describing in more detail the link to potential experiments:

“One could repeatedly track the motion of the leading edge of diffusing particles in a system of colloids confined to a quasi-1D channel thereby directly measuring $\var \maxnt$ for system sizes ranging from $N ~ 10^2$ to $N ~ 10^{10}$. Further, one can also perform complementary measurements on the time of first passage of diffusing objects, which opens the door to experiments done on all manner of diffusing objects, including light or sound diffusing through a scattering medium, dye molecules in a fluid, or any other object whose first passage can be measured.”

2) It is rather strange that the discussion of RWRE models does not refer at all to the classical book by Barry Hughes entitled "Random Walks and Random Environments". Is there a particular reason to ignore this book?

We apologize for this omission. We were actually unaware of this very nice two volume book on the subject and we appreciate the referee pointing us to it. The book (Volume 2) is primarily concerned with the study of random walks in random environments that are static or unchanging in time (e.g. random walks in percolation clusters). Our own study is focused on temporarily changing environments. However, since we have mentioned both types of models, we have added this book in the background section as reference 42:

“The random walk in random environment (RWRE) model goes back to [36, 37] (see also [38–42])”

3) The authors provide the following argument to justify their choice of one-dimensional setting: "...working in one dimension skirts anisotropy issues arising from the choice of lattice in higher dimensions." Even though this is, of course, true, I suspect that the main reason is that analytical and numerical computations in higher dimensions are much more challenging; in particular, there is no available asymptotic results, while the numerics can be more difficult. As a reader, I would prefer to see this fair argument.

The referee is completely right that our choice of studying this type of one-dimensional discrete model is very much determined by the availability of an exactly solvable model. While our initial version of the text did note this as being important, the emphasis was more on the point that

there is inherent value to studying this type of model. We have rearranged our discussion at the beginning of the “models for diffusion” section to address the referee’s comment. It now reads:

“Although physical diffusion is continuous in time and (typically) occurs in three-dimensional space, here we work with discrete models in one spatial dimension. The principal reason for this choice is that it is the setting for the exactly solvable Beta RWRE [58] (a continuous sticky Brownian motion limit of this model exists [71]) that will enable us to compare numerical results to exact theoretical predictions. Beyond that, discretization is common for numerical simulations and higher dimensions are more challenging numerically due to anisotropy issues arising from the choice of lattice and due to the lack of exactly solvable models, c.f. [62].”

4) Numerical simulations with up to 10^(300) particles must be clearly explained. Even though the authors provide some technical details in the SM, the simulation technique is not presented. In fact, as one speaks about simulations of diffusion processes with particles, the first technique coming to mind is a Monte Carlo simulation. However, a direct version of Monte Carlo simulations (in which the motion of every particle is simulated) is limited to (some) billions of particles. As a consequence, reading about simulations with 10^(300) particles is confusing.

Indeed, an agent-based Monte Carlo simulation of 10^(300) particles would be impossible. Instead, we use the fact (noted on page 2 column 2) that if there are N(x,t) particles at position x and time t, then the number that move to site x+1 at time t+1 will be distributed as a Binomial random variable with N(x,t) samples and with success probability B(x,t). Thus, rather than tracking agents, we track occupation numbers, which equivalently encodes the process. There are some subtleties in sampling Binomials with very large sample sizes which require use of the central limit theorem to approximate the Binomial distribution by the Gaussian for very large occupation numbers. We have added the following text in the main body “Numerical Methods” section in order to make clear that we do not perform an agent-based Monte Carlo simulation.

“We simulate such large systems by tracking occupation variables instead of individual particle trajectories. In particular, if there are N(x,t) particles at site x at time t, then the number that move to site x + 1 is binomially distributed with N(x,t) samples and success probability B(x,t) (the remainder move to site x − 1). We sample these binomial distributions utilizing quadruple-precision floating point numbers and making approximations to the binomial distribution when dealing with sizes beyond our precision limits, as described in [83]. The rightmost particle location (identified by the maximal x with N(x,t) ≥ 1) at each time represents a sample of Max^N_t . By repeatedly sampling new environments along with random walk occupation variables N(x,t) therein we numerically measure Var(Max^N_t).”

We also added, at the end of that paragraph, a statement on the computing resources:
“The data presented in Fig. 2 and 3 took approximately three weeks to run in parallel on 500 cores of the University of Oregon high performance computing cluster, Talapas.”

The SM now also contains an extended description of precisely how our simulation is implemented. The text added reads:

"We expand upon our description of the numerical methods used in our simulations. We consider varying system sizes from N = 10^2 to N = 10^300 with time varying from t = 0 to t = 5000 log(N). We simulate such large systems by utilizing the full range of quadruple-precision oating point numbers and making approximations to the binomial distribution when dealing with sizes beyond our precision limits. These approximations are necessary given the limitations of our precision and do not seem to affect the behavior of the system in any significant manner. In particular, for a site x at time t where the number of particles, N(x; t), is below 2^31 we use the C++ Boost integer implementation of a binomial distributed random number generator to choose the number of particles that will move right versus left. Above 2^31 the integer implementation overflows and we instead approximate the binomial distribution by a Gaussian distribution of mean N(x; t)B(x; t) and variance N(x; t)B(x; t)(1 - B(x; t)), as dictated by the central limit theorem. For sufficiently large N(x; t), the variance itself will fall outside of the precision of a quadruple floating point number. This occurs when N(x; t) > 10^64, above which we approximate the binomial simply by its mean. The number of particles that move right is then N(x; t)B(x; t). Psuodo-code of the simulation is given:

(1) Start N particles at x = 0

(2) For each site:

draw B(x, t) from U[0,1]

if (N (x, t) > 10^64)

N (x + 1, t + 1) = N (x, t)B(x, t)

N (x - 1, t + 1) = (1 - B(x, t))N (x, t)

else if (2^31 < N (x, t) < 10^64)

μ = N (x, t)B(x, t)

σ^2 = N (x, t)B(x, t)(1 - B(x, t))

N (x + 1, t + 1) is drawn from Normal (μ, σ^2)

N (x - 1, t + 1) = N (x, t) - N (x + 1, t + 1)

else if (N (x, t) < 2^31)

N (x + 1, t + 1) is drawn from Bin(N (x, t), B(x, t))

N (x − 1, t + 1) = N (x, t) − N (x + 1, t + 1)

where U[0,1] is the random uniform distribution on the range [0, 1], Normal (μ, σ^2) is the Gaussian distribution with mean μ and variance σ^2, and Bin(n, p) is the Binomial distribution with number of trials n and probability p. Note that our simulation differs from Monte Carlo, or agent based, simulations because we can group all the particles at a site together and iterate over each site. Our code is available at https://github.com/CorwinLab/RWRE-Simulations."

5) What is the purpose of considering systems with 10^(300) particles? In the introduction section, the authors speak about a great number of agents varying from 10^2 to 10^(60). Even if the "upper limit" of 10^(60) does not sound much realistic, the size 10^(300) is fully irrational. What is such a huge system of diffusing one-dimensional particles that the authors have in mind?

This is a great question. The purpose of going to 10^300 is definitely not to try to match an experiment with a similar number of particles as no such systems exist. It comes, instead, from our desire to compare to asymptotic theory over as wide a range as possible and to study how well that theory can be used to fit system sizes varying with N. The theory that we appeal to is all based around taking N and t to infinity such that log(N) and t, or log(N)^2 and t remain at a constant proportion. Even though N=10^300 is quite large, log(N) = 300*log(10)~690 is still relatively small. This is especially true when you are considering power laws involving log(N)^(⅓)~8.83, as we must. Since we are dealing with discrete models, this should be compared to the lattice size 1. Thus, a priori, it was completely unclear to us that anything from the asymptotic theory would be visible even at N=10^(300). It came as a happy surprise that the asymptotic theory remains effective down to very small and physically relevant system sizes. That said, the quality of the fit definitely declines as N shrinks. This much is apparent from Figure 2 and 3 where the numerically observed variance curves match the theory curves for large time, but diverge for small t and small N. As N increases, the small t fit improves considerably. Thus, the real purpose of going to this very wide spread of N is to highlight what aspects of the theory are well-fitted for all N and which are less reliable for small N and small t.

We have included the following discussion in the text. In “Numerical methods” we added the line:

“We consider such large and physically unrealistic system sizes like 10^300 in order to see how asymptotic theory applies for as wide a range as possible of finite system sizes.”

While in “Comparison of Numerical and Theoretical Results” we have added

“Fig. 4 shows the tight matching of the asymptotic theory curves and numerically measured values for the variance of Max^N_t ,Env^N_t and Sam^N_t for a given value of N = 10^7. Notice that for t ≈ log(N) the asymptotic theory and numerical values for the variance of Sam^N_t do not fit as well as for large t. This is likely a result of finite-size effects and quickly goes away at larger values of t or when N increases. The fit for N = 10^300 in Fig. 2 and 3 remains tight over the entire range of t.”

6) On page 3, when discussing SSRW, the authors start with the condition t/log(N) < 1/log(2) and come to N >> 2^t. But the former inequality implies N > 2^t, not much larger?

Indeed, this is confusing notation. In particular, the referee is right that t/\log(N)<1/\log(2) logically implies N>2^t and nothing more. What we really meant, but did not clearly enough say, is that if one assumes that t/\log(N) = \hat{t}<1/\log(2) as t and N tend to infinity, then in that same limit N=e^(t/\hat{t})>>2^t in the sense that N/2^t will tend to infinity (as t and N do). We have included a preface about this language at the beginning of the “Asymptotic Theory Results” section:

“Given a relationship between t and log(N) (e.g. t/log(N)=\hat{t}) we write f(N,t) >> g(N,t) if f(N,t)/g(N,t) tends to infinity as N and t do subject to their relationship.”

In the text that follows, we always specify the scaling relationship between t and \log(N), be it t/\log(N) = \hat{t} or t/\log(N)^2 = \hathat{t}. We hope that this has clarified the meaning of these symbols.

7) As said earlier, the current main text refers too often to the SM that makes it hardly readable. It is therefore suggested to restructure the main text by incorporating the SM, splitting it into proper sections, etc.

We appreciate the referee's concerns. As explained in the letter, there are two main theoretical results that provide the basis for the theoretical asymptotics in the t~log(N) and t~log(N)^2 scaling regimes. These results, in turn, are derived through a highly non-trivial connection between this model and quantum integrable systems (in particular, it is solvable via Bethe ansatz). We have moved the statements of the two main theoretical results into the letter. In particular, below equation (3) we now state:

"As shown in [83], this follows from the result of [58]: For v\in(0,1), \log P_\mathbf{B}(vt,t)=-t I(v) + t^{1/3} \sigma(v)\chi_t where I(v) = 1-\sqrt{1-v^2}, \sigma(v) = (2I(v)^2/(1-I(v)))^{1/3} and \chi_t is random converging to the GUE Tracy-Widom distribution as t goes to infinity."

While below equation (4) we now state:
"where h(0,s) is the height at 0 and time s of the narrow wedge solution to the KPZ equation \partial_s h(y,s) = \frac{1}{2} \partial_y^2 h(y,s) + \frac{1}{2}(\partial_y h(y,s))^2 + \xi(y,s) with \xi space-time white noise. As shown in [83], this follows from the result of [63]: For v\in (0,\infty), \log P_\mathbf{B}(vt^{3/4},t) \approx -\frac{v^2t^{1/2}}{2} -\log(t)/4+\log(v) - v^4/12 + h(0,v^4)."

The calculations deriving what we need from these results are still contained in the SM since they are a bit lengthy when done with proper detail. Still, we think that this modification to the letter better highlights the key theoretical inputs. The theoretical origins (in quantum integrable systems) of these inputs are mentioned in the letter at the end of the first paragraph of the "Asymptotic Theory Results" section.

8) In Eq. (3), there is no \chi (except for \sigma_\chi). Is there something missing?

Thank you for pointing this mistake out. We have now clarified this in the text as

"where $\sigma_{\chi}^2\approx~0.813$ is the variance of the GUE Tracy-Widom distribution [61], [84]"

9) The sentence after Eq. (5) is unclear. Where is the formula with the error function mentioned in this sentence?

Thank you for noting this omission. This formula, that had previously been in the SM, has been moved to the main text immediately after (5). It now reads

For finite N and t these regimes have a gentle crossover that we capture by setting
Var^{asy}(Env^N_t) := I(N,t) V_1(N, t)+(1-I(N,t)) V_2(N, t)
where I(N,t) := \frac{1}{2} (1-erf((t-(\log(N))^{3/2}) / (\log(N))^{4/3})) (with erf(x)= 2/\sqrt{\pi} \int_{0}^{x} e^{-t^2}dt the error function) interpolates from 1 to 0$ over an interval of width (\log(N))^{4/3} around (\log(N))^{3/2}.

10) The short-time ballistic regime has been experimentally observed in various references, earlier than Ref. [28], e.g. by Huang et al., Nat. Phys. 7, 576-580 (2011).

Thank you for catching this oversight. We have added this, as well as other references, to the manuscript as References [28-32].

11) The bibliography style should be homogenized.

Thank you, we will work with the editorial and publication staff to make this happen.

Report of Referee B -- LS18194/Hass

In their manuscript, Hass et al. investigate mathematically and numerically the statistical properties of extremes in an ensemble of N particles diffusing in a common, random environment. Specifically, they present asymptotic results for the largest displacement in an ensemble of non-interacting particles. The authors identify universal power laws in the time dependence and show that the crossover times of the corresponding windows scale as log(N) and log(N)^2. Thus, very large ensembles of particles are needed to separate these time scales and to resolve the different regimes. The authors address this challenge with large-scale simulations and high-precision numerics. From the perspective of physics, a central insight is that the measurement of extremal properties of diffusion can give insight into the environmental fluctuations that is averaged out in experiments probing sample means.

The manuscript is very well written and starts with a comprehensive and accessible introduction. The numerical results are of high quality and the agreement between predictions and stochastic simulations is very satisfying. The topic is timely, both from a probability theory point of view and from the physics demand to better characterize and understand diffusion in complex environments. Yet an open question is whether the identified scaling regimes are indeed relevant for any specific physical phenomenon. The manuscript may become suitable for publication in Physical Review Letters after my following concerns and questions have been addressed.

We thank the referee for their praise and recognition of our work.

How can one obtain Var(Max_t^N) in an experiment? The large ensemble sizes needed exclude the use of single particle tracking. The appeal of the manuscript to physicists would be considerably increased if the authors would describe a concrete experimental setup that could probe their predictions.

We appreciate the referee's suggestion to explain more about the concrete experimental setups that could probe our predictions. In fact, we are in the early stages of conducting these very experiments. In particular, we are performing two different kinds of measurements. The first is to measure the leading edge of diffusing colloids in a quasi-1d system (a capillary tube). The second is to measure the time of first arrival for photons from a femto-second laser pulse traveling through a scattering media. These complementary measurements on the time of first passage of diffusing objects open the door to experiments done on all manner of diffusing objects, including light or sound diffusing through a scattering medium, dye molecules in a fluid, or any other object whose first passage can be measured. The present work sets the stage for understanding the outcomes of these experiments by providing a link from asymptotic theory to finite size and time systems, thus providing a detailed description of the way in which space time randomness affects the extreme value statistics of diffusion. In particular, the ½ power-law that we find for the environmental variance as well as the method of accessing this from subtracting the sampling variance from the variance of the observed maximum provides a clear roadmap for how to start to analyze experimental data.

We have included in the conclusion the below text describing in more detail the link to potential experiments:

"One could repeatedly track the motion of the leading edge of diffusing particles in a system of colloids confined to a quasi-1D channel thereby directly measuring $\var \maxnt$ for system sizes ranging from $N ~ 10^2$ to $N ~ 10^{10}$. Further, one can also perform complementary measurements on the time of first passage of diffusing objects, which opens the door to experiments done on all manner of diffusing objects, including light or sound diffusing through a scattering medium, dye molecules in a fluid, or any other object whose first passage can be measured."

For applications, it seems crucial that the idealization of non-interacting particles is lifted. Is there any perspective that this limitation of the mathematical treatment can be weakened or removed?

This is a great question. While it is true that, conditioned on the environment, particles are non-interacting, if one does not see the environment then the resulting motion of particles appears highly correlated since they are routed into similar paths by the environment. Thus, the RWRE already provides one mechanism for interaction. Of course, we will see in our experiments if this mechanism is sufficient to capture the behavior of extremes in real systems.

We anticipate that for extreme particles the role played by interactions will be diminished since the density is lowest around the extremes. We note this in “Models for diffusion”, saying

“Our model ignores any higher order interactions as we expect them to be less present in the behavior of extreme particles, for which the local density is necessarily low.“

Of course, the referee may be asking about the potential of finding a solvable RWRE model in which, conditioned on the environment, particles interact. This is a very interesting question. For instance, in terms of occupation variables (the N(x,t) in our letter), one can describe a Markov process whereby the mechanism for dividing the particles at x at time t between x+1 and x-1 at time t+1 is specified by a different family of distributions than the binomial (which comes from the independence of the underlying random walkers). In the continuum, this corresponds to having non-solvable choices of the kernel for sticky Brownian motions. While it seems unlikely that there is a solvable version of such a more general model, it may still be possible to connect the transition probability for this model to the stochastic heat equation and hence show that the t^{½} environmental variance long-time scaling exponent persists. This is a great direction to explore both in terms of numerics and stochastic analysis theoretical methods, and we appreciate this suggestion.

Another popular class of models for diffusion in a randomly varying environment is summarized as "diffusing diffusivity" models (see works by Chubinsky/Slater, Chechkin/Metzler, Grebenkov, Jain/Sebastian and others). Can the authors comment on a possible connection to the beta-diffusion model studied here?

Also please comment on a possible extension of the results to the limit of a static random medium, where B(x,t) does not change in time. This case is of interest as it may have application to the diffusion problem on percolation clusters and even the random Lorentz gas.

These are both great questions and we have added some discussion in the “Conclusion” section. The diffusing diffusivity model mentioned here seems to generally involve diffusions in which the diffusion coefficient evolves in a random manner as time progresses. So, in a sense, these models involve a random environment that is the same everywhere in space but changing dynamically in time. Perhaps a closer match to our model is to consider a model in which the diffusivity depends on time and space and evolves itself according to some simple stochastic PDE (or is simply a smoothed version of a Gaussian field). An even more general version of this model would involve both a space-time changing drift field and diffusivity field that represent the common environment in which particles diffuse: i.e., for i=1,2,... we have that the i^th particle x_i(t) satisfies the SDE
dx_i(t) = F(x_i(t),t) + D(x_i(t),t) dB_i(t)
where the F and D are the drift and diffusivity fields that are common to all i, while each B_i(t) represents an independent Brownian motion.
A special case of this model in which D is constant but F is random has been continued in (https://arxiv.org/abs/1705.05159) where it is argued that this should display a similar behavior

to the RWRE model. The most general class of models seems well-worth study, likely through numerics.

The case of a static environment can also be considered from the general perspective of F and D described above, in this case where F(x,t)=F(x) and D(x,t)=D(x). This type of model, e.g. at the level of a discrete RWRE, has been very well studied. The large deviations and first passage behavior of the model has previously received attention (and has rather different behavior with a much slower advance of the maximal particle caused by persistent bottlenecks from the environment), though it seems that an analysis similar to ours – studying how the randomness remains present in large deviations or maximal particle behavior – is lacking in the literature. Thus, we have also mentioned the interest in understanding how the transition between temporally mixing and static environments plays out.

The following is the new text that we have put into the "Conclusion":
"A continuum model that should provide an even wider testing-ground amenable to numerics involves particles x i (t) for i = 1,2,... satisfying dx_i(t) = F(x_i(t),t)dt + D(x_i(t),t)dB_i(t) where F(x,t) and D(x,t) are random forcing (as in [62]) and diffusivity (generalizing diffusing diffusivity, c.f. [85]) fields common to all particles while B i are Brownian motions independent between different i. Changing the correlation structures of F and D will probe the transition between temporally mixing versus quenched environments, which should have very different behavior (c.f. [86, 87]) and warrants further study."

Is there an intuitive picture behind the finding that Max_t = Env_t + Sam_t (see p. 2, col. 1)? If this relation serves as the definition of Sam_t, what is the argument leading to the interpretation that Sam_t describes only the randomness due to sampling?

This is a very good question. We argue in the SM that given the value of Env^N_t (i.e, the location of the environment dependent 1/N quantile), the distribution of Sam_t is approximately Gumbel and independent of Env_t. This is because the profile of the tail probability around the 1/N quantile has two contributions -- the overall decay from Taylor expanding the rate function I in a small neighborhood of Env^N_t, and then the randomness on top of that (stemming from the random environment). We have assumed in our analysis that this randomness has a lower order effect on the sampling, and this is very much in line with our numerical findings. Of course, one can ask from a theoretical perspective why this is the case, i.e., why do we neglect the fluctuations in the taylor expansion on the bottom of page 7 of SM. In KPZ universal class models (such as this) it is generally the case that fluctuations look locally Brownian (or for discrete models, locally like independent increment random walks). Assuming this property transfers to our model, it would justify that the contribution from fluctuations is of lower order than the Taylor expansion of the rate function, as desired. Demonstrating the local Brownianity is something that has not yet been accomplished theoretically for this model though, so we have not gone into detail here. To reflect the above discussion we have added the follow lines on the top of page 8 of the SM:

"This assumption would be true if, for instance, χ_t looked locally like a random walk with independent and identically distributed intervals. This property is believed to be universal to models in the KPZ

class and thus it is reasonable to assume that the same holds here. We do not attempt to justify this theoretically beyond this heuristic, though note that it yields very good agreement with our numerical simulations."

The claimed connection to the KPZ equation is not readily apparent and can only be understood after working through refs. [56, 61] and the supplement. A more detailed explanation in the main text would be appreciated.

Indeed, this came across as a bit of a blackbox and we appreciate the referee's point here. One way to argue this connection is to go to a continuum version of the RWRE such as discussed in https://arxiv.org/abs/1705.05159. Then, the Fokker Planck equation actually produces something quite similar in nature to the KPZ equation. Alternatively, one can view the recursion relation (1) in the letter as a discrete version of the stochastic heat equation which is satisfied by the exponential of the KPZ equation. It is this second heuristic that we note in our text as explanation for why the KPZ equation and universality class arise. Of course, it takes considerably more effort to actually establish the precise links to the KPZ equation and universality class, and this is what the two papers [59] and [63] accomplish for us. Here is the text we added to the last paragraph of the "Background" section:

"*The recursion relation (see (1) below) for RWRE transition probabilities solves a discrete version of the multiplicative noise stochastic heat equation (mSHE) which, in the continuum, is related to the KPZ equation by taking a logarithm."*

Minor points:

Please comment briefly on the fact that one does not need to simulate N independent trajectories, which would be impossible for the huge values for N considered.

We thank the referee for this suggestion and have now added a comment to this effect in the manuscript in our numerical methods section as:

"We simulate such large systems by tracking occupation variables instead of individual particle trajectories. In particular, if there are N(x,t) particles at site x at time t, then the number that move to site x + 1 is binomially distributed with N(x,t) samples and success probability B(x,t) (the remainder move to site x − 1). We sample these binomial distributions utilizing quadruple-precision floating point numbers and making approximations to the binomial distribution when dealing with sizes beyond our precision limits, as described in [83]."

Please add a statement about the required computing resources. It is not obvious to me whether the simulations can be carried out quickly a single workstation or whether they require the use of a large cluster. The authors may further consider making their computer code available as a supplement to the paper.

The computing requirements are relatively modest. One can get meaningful data on a single workstation. However, to achieve the high quality averaged data presented in Figures 2 and 3 took approximately 10,000 cpu-days running on our university's cluster (about two or three weeks running on 500 cores). We have made our code publicly available on github at https://github.com/CorwinLab/RWRE-Simulations (referenced in the SM). At the end of the "Numerical Methods" section in the letter we have added:

"The data presented in Fig. 2 and 3 took approximately three weeks to run in parallel on 500 cores of the University of Oregon high performance computing cluster, Talapas."

p. 1, col. 1: please spell out or define the abbreviation GUE

Done.

figs. 2 and 3: please add straight lines to indicate the power laws t^(1/3) and t^1, for easier reference.

Done.

p. 3/4: I suggest to consistently use the symbol '<<' (\ll) instead of '<' for 'much smaller than'

We have incorporated this suggestion in the text. In particular, we have included a preface about the << language at the beginning of the "Asymptotic Theory Results" section:

"Given a relationship between t and log(N) (e.g. t/log(N)=\hat{t}) we write f(N,t) >> g(N,t) if f(N,t)/g(N,t) tends to infinity as N and t do subject to their relationship."

In the text that follows, we always specify the scaling relationship between t and \log(N), be it t/\log(N) = \hat{t} or t/\log(N)^2 = \hathat{t} and then only use < or > when we compare \hat{t} to a given cutoff value. We hope that this has clarified the meaning of these symbols.

p. 4, col. 1: In the first sentence, it is not clear which coefficient in eq. (5) is referred to, there are two cases.

We appreciate the referee pointing out this confusing language. In fact, we mean that both coefficients should be non-universal and should reveal information about the underlying environment. We have modified the text in the "Conclusion" to now read more clearly:

"By contrast, for t >> log(N), we anticipate that the scaling exponents and functional forms we have identified for the variances of Env^N_t ,Sam^N_t and Max^N_t will be universal, as will the relation (7). The leading coefficients in (5) should be non-universal and hold within them all of the accessible information about the correlation structure of the environment – we call these extreme diffusion coefficients. Further theoretical study, such as for the general α,β Beta RWRE

model, should provide a natural first test of this universal picture and an understanding of how the extreme diffusion coefficients relate to the microscopic environment."

fig. S3, in my view, is a very accessible illustration of the problem and would fit neatly into the main text, if space allows.

We thank the referee for the compliment and have reworked Figure 1 to include this.

Ref. [6]: the correct year is 2016

Fixed.