Forthcoming in The British Journal for the Philosophy of Science

Putting theory in its place: The relationship between universality arguments and empirical constraints

Grace E. Field

ABSTRACT

In light of its empirical undetectability, physicists have attempted to establish Hawking radiation as universal—as a phenomenon that should appear regardless of the possible details of quantum gravity, whatever those details might be. But, as pointed out in a recent article by Gryb, Palacios, and Thébault (2019), these universality arguments for Hawking radiation seem broadly unconvincing compared to the Wilsonian renormalization-group universality arguments for condensed matter physics.

Motivated by their apparent failure, compared with the overwhelming success of universality arguments in so many other contexts, I address the question: in which situations should we expect to be able to construct successful universality arguments? In other words, which situations are 'universality-argument-apt'?

I distinguish between two notions of success for a universality argument: 'strength' and 'relevance'. I argue that we should only expect to be able to construct universality arguments that are successful in the sense of being significantly relevant to a given domain if (1) we know enough about how that domain's micro-physics is structured, or (2) we are able to empirically test the domain's macro-behaviour, or if we are in both situations at once. These conditions are useful, most obviously, as a clarification of what universality arguments are capable of. But I argue that they are also useful for two less direct reasons: they clarify the status of analogue experimentation, and thereby show us where we stand in our search for empirical confirmation of Hawking radiation.

1 Introduction

Universality arguments for Hawking radiation have recently been employed in the philosophy literature as a tool to provide analogue black hole experiments with confirmatory power (Dardashti, Hartmann et al. [2019]). And yet, as shown by Gryb et al. ([2020]), none of the existing proposed universality arguments for Hawking radiation seem nearly as convincing as the well-established Wilsonian renormalization-group arguments in condensed matter physics.

Their apparent weakness, contrasted with the clear success of universality arguments in so many other fields, highlights the need for a more detailed analysis of whether we should even expect universality arguments to navigate empirically uncharted territory. This paper aims to fill that gap, by articulating in-principle limits on the domains for which we should expect to be capable of constructing successful universality arguments.

I begin in Section 2 by assessing the connection between universality arguments and analogue experiments. Section 2.1 summarizes the existing literature on universality arguments and analogue experiments, both in general (Section 2.1.1) and for Hawking radiation in particular (Section 2.1.2). Section 2.2 attempts to articulate the ingredients—both formal and practical—that we need to have at our disposal in order to gain confirmatory power from analogue experiments using the Bayesian framework established in Dardashti, Hartmann et al. ([2019]).

I distinguish between two interpretations of this framework—one in which it requires us to have a particular universality argument in mind (Section 2.2.1), and one in which it does not (Section 2.2.2)—and I argue that only the first is viable. Therefore, we at least need to have such an argument at hand in order to gain some degree of confirmation from an analogue experiment. But, I suggest, further conditions need to hold for that confirmation to be significant: in particular, we need to be confident that our universality argument is significantly positively relevant to both the source and target systems. This translates into formal probability constraints on the variables in the Bayesian framework.

Section 3 goes one step further to ask: in which situations can we have that confidence? After distinguishing between two components of success for a universality argument—strength and relevance—in Section 3.1, and comparing our knowledge of condensed matter physics with our knowledge of quantum gravity in Section 3.2, I suggest in Section 3.3 that we should only expect to be able to construct a significantly relevant universality argument within a given domain if either

(1) we know that the micro-physics of the physical systems in question—i.e. the physical systems that supposedly all exhibit the same macro-phenomenon—are of the same type in key respects; or

(2) we are able to empirically access those systems' macro-behaviour,

or if we are in both of those situations at once.

These conditions are useful, most obviously, as a clarification of what universality arguments are capable of. But I will argue in Section 4 that they support various other

morals. They clarify the status of analogue experimentation—in particular, by elucidating the underlying conditions that are required for an analogue experiment to provide significant (Sections 4.2.1 and 4.2.2) rather than negligible (Sections 4.1.1 and 4.1.2) confirmatory power. More broadly, they are essential as a reminder of how our theoretical tools are shaped by our empirical capabilities (Section 4.3). If we cannot empirically access the macro-behaviour of the physical systems in question, and if we do not already know that they share the relevant micro-structure, then no theoretical argument—not even a universality argument—can fill the empirical gap.

2 Confirmation via Analogue Experimentation

This Section will begin by summarizing the existing literature on universality arguments and analogue experimentation, both in general (Section 2.1.1) and for Hawking radiation in particular (Section 2.1.2). I will then assess the conditions required for an analogue experiment to provide confirmatory power—both to some degree and to a significant degree—after distinguishing between two possible interpretations of Dardashti, Hartmann et al. ([2019]) (Sections 2.2.1 and 2.2.2).

It will become clear that the confirmatory power of an analogue experiment depends on our having access to a universality argument that is significantly positively relevant to both the source and target systems. This will motivate a closer analysis of the relationship between universality arguments and empirical constraints in the second half of the paper (Sections 3 and 4).

2.1 Background

2.1.1 Analogue experiments and universality arguments

The physics literature

Physicist Bill Unruh was the first to suggest analogue experimentation as a solution to the empirical undetectability of Hawking radiation. He realized that any fluid, if moving at an ever-faster rate of flow in some direction, will develop a sonic horizon analogous to the event horizon surrounding a black hole singularity.

This intuitive analogy can be made mathematically rigorous (Unruh [1981], pp. 1351–2). We can view the sound waves in our fluid as a massless scalar field propagating in a (3+1)-dimensional Lorentzian geometry with metric

$$g_{\mu\nu} = \begin{bmatrix} -(c_{sound}^2 - v_0^2) & \vdots & -v_o^j \\ \cdots & . & \cdots \\ -v_0^i & \vdots & \delta_{ij} \end{bmatrix}. \tag{1}$$

Comparing $g_{\mu\nu}$ with the linearized Schwarzchild metric $g_{\mu\nu}^{Schwarzschild}$ in Painlevé-Gullstrand coordinates, the isomorphism should be clear:

$$g_{\mu\nu}^{Schwarzschild} = \begin{bmatrix} -(c^2 - \frac{2GM}{r}) & \vdots & -\frac{2GM}{r}x^j \\ \cdots & \cdot & \cdots \\ -\frac{2GM}{r}x^i & \vdots & \delta_{ij} \end{bmatrix}. \quad (2)$$

The mathematics used to derive the emission of Hawking radiation from a Schwarzschild black hole horizon can therefore be applied equally well to derive the emission of analogue Hawking radiation from an acoustic horizon. We are left not only with an intuitive analogy between the two systems—fluid and astrophysical—but with an isomorphism between the mathematical models by which we believe they can be adequately described.

This isomorphism is the basis for the reasoning behind analogue experimentation, both for black holes and in general. Given two systems described by isomorphic mathematics, advocates of analogue experimentation suggest that we should be able to draw conclusions about one based on experimental data collected on the other.

Research groups at the University of British Columbia, the Technion, the University of St Andrews, the University of Nottingham, and the French National Centre for Scientific Research have led work towards building and completing these analogue experiments; and so far, their results have been positive.[1] Steinhauer's group, for example, claims to have observed spontaneous analogue Hawking radiation in a Bose-Einstein Condensate analogue black hole (Steinhauer [2016]; de Nova et al. [2019]; Kolobov et al. [2019]).[2]

Dardashti, Thébault et al. ([2017]) and Dardashti, Hartmann et al. ([2019])

Dardashti, Thébault et al. ([2017]) and Dardashti, Hartmann et al. ([2019]) were the first to articulate the reasoning behind these analogue gravity experiments in the philosophy literature.[3] Dardashti, Thébault et al. ([2017], pp. 67, 73) suggest, based on a five-step inductive argument, that such analogue experiments should be able to provide confirmation for hypotheses about their empirically inaccessible 'target' systems under two conditions: as long as our mathematical models of the two systems are isomorphic, and as long as we have Model-External and Empirically Grounded Arguments ('MEEGA') to support the strength of that isomorphism.

Dardashti, Hartmann et al. ([2019]) identify the 'MEEGA' as universality arguments and formalize this reasoning in a Bayesian framework. The framework introduces four binary variables, which I label $\mathcal{E}$, $\mathcal{M}^S$, $\mathcal{M}^T$, and $\mathcal{X}$.[4] Following Dardashti, Hartmann

1 See e.g. Unruh ([2014]), Steinhauer ([2016]), Jacquet and König ([2020]), Weinfurtner et al. ([2011]) and Rousseaux et al. ([2008]). The group at UBC is led by Unruh himself, and works collaboratively with the group at Nottingham. For more information, see https://phsites.technion.ac.il/atomlab/ and https://www.gravitylaboratory.com.

2 Steinhauer ([2016]) observed the correlations between positive and negative modes that are thought to be a distinctive feature of Hawking radiation, de Nova et al. ([2019]) extracted the Hawking spectrum, and Kolobov et al. ([2019]) studied the radiation produced by an evolving effective black hole horizon.

3 They were certainly not the first to discuss analogical reasoning in more general terms—see e.g. Mill ([1843/1930]), Keynes ([1921]), Hesse ([1964]), Hesse ([1966]), Hesse ([1973]), Hesse ([1974]), Hesse ([1988]), and Bartha ([2010]). Neither were they the first to discuss the issues associated with similarity and analogy in modelling and simulation—see e.g. work by Sterrett ([2006]) and Weisberg ([2013]), among many others.

et al. ([2019], pp. 4–6), $\mathcal{E}$ takes the positive value E if evidence in support of our mathematical model $\mathrm{M^S}$ of the source system obtains.[5] It takes the negative value $\sim E$ if such evidence does not obtain. $\mathcal{M}^{\mathcal{S}}$ takes the positive value M^S if our mathematical model $\mathrm{M^S}$ of the source system (i.e. our hypothesis about the source system) is indeed an adequate description of the source system, and it takes the negative value $\sim M^S$ if not. Similarly, $\mathcal{M}^{\mathcal{T}}$ takes the positive value M^T if our mathematical model $\mathrm{M^T}$of the target system is an adequate description of the target system, and the negative value $\sim M^T$ otherwise. Finally, and most importantly for our discussion, the variable $\mathcal{X}$ takes the positive value X if a universality argument that supports the adequacy of both $\mathrm{M^S}$ and $\mathrm{M^T}$ obtains.

Dardashti, Hartmann et al. ([2019], p. 9) rigorously show that $P(M^T|E) > P(M^T)$—that evidence collected on the source system confirms our hypothesis about the target system—as long as the following four inequalities hold:

$$0 < P(X) < 1; \tag{3}$$

$$P(M^S|X) > P(M^S| \sim X); \tag{4}$$

$$P(M^T|X) > P(M^T| \sim X); \tag{5}$$

$$P(E|M^S) > P(E| \sim M^S). \tag{6}$$

Those inequalities, and the corresponding transfer of evidence from source to target, are represented by the Bayesian network in Figure 1.[6] In particular, adopting a more streamlined notation on which $y := P(Y)$, $\overline{y} := P(\sim Y)$ and $y_z := P(Y|Z)$ for any Y and Z,[7] the degree of confirmation provided by an analogue experiment satisfying these inequalities is given by:[8]

$$\Delta_C := m_e^T - m^T = \frac{x\overline{x}}{e}(m_x^S - m_{\overline{x}}^S)(m_x^T - m_{\overline{x}}^T)(e_{m^S} - e_{\overline{m}^S}). \tag{7}$$

The logic, stripped of its Bayesian details, is remarkably intuitive. If we have a universality argument that is a theoretical argument for why S and T should be adequately described by models of the same isomorphism class 'M', and should therefore exhibit analogous phenomena $\mathrm{P^S}$ and $\mathrm{P^T}$, then empirical results that show S exhibiting $\mathrm{P^S}$ are simultaneously empirical support for our universality argument. But if we have empirical support for our universality argument, and if that universality argument applies to both S and T, then we simultaneously have empirical support for what our universality argument says about T: namely, that T should exhibit $\mathrm{P^T}$.

4 I use a slightly different notation than Dardashti, Hartmann et al. ([2019]): my $\mathcal{M}^{\mathcal{S}}$ is their $\mathcal{A}$ and my $\mathcal{M}^{\mathcal{T}}$ is their $\mathcal{M}$.

5 Note the font conventions that are adopted here and throughout. *CALLIGRAPHIC* font is reserved for binary variables; *ITALIC* font for the possible values of those variables; and PLAIN TEXT font for other entities/concepts invoked by those variables.

6 For a discussion of how this Bayesian network might apply to analogical reasoning more generally, see Feldbacher-Escamilla and Gebharter ([2020]).

7 m_e^T, for example, represents $P(M^T|E)$; $e_{\overline{m}^S}$, similarly, represents $P(E| \sim M^S)$.

8 The equation below combines equations (14) and (18) in Dardashti, Hartmann et al. ([2019], p. 9).

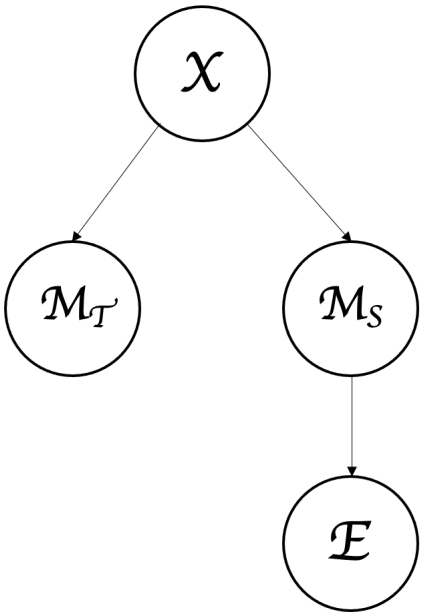

Figure 1: Analogue experiments supplemented with universality arguments, based on Figure 2 in Dardashti, Hartmann et al. ([2019], p. 6).

Crowther et al. ([2019]) and Evans and Thébault ([2020])

Yet, not everyone agrees that the reasoning behind analogue experimentation is reliable. Steinhauer writes of his 2016 experiment that '[t]he measurement reported here verifies Hawking's calculation, which is viewed as a milestone in the quest for quantum gravity' (Steinhauer [2016], p. 964). But Harlow, for example, dismisses the same measurement as 'an amusing feat of engineering' that 'won't tell us anything about black holes' (Wolchover [2016]).

In the philosophy literature, Crowther et al. ([2019], pp. 16–23) have argued that analogue simulation suffers from a severe circularity problem, whereby we are only warranted in using data collected on the source system to make inferences about the target system under the assumption that our prior beliefs about the source and target systems are correct. In Bayesian terms, their concern is about whether we can even draw the arrow between the variables $\mathcal{X}$ and $\mathcal{M}^{\mathcal{T}}$—or, equivalently, about whether we can assume that the inequality (5) holds without taking our hypothesis about the target system for granted.

They write:

> it was thus already assumed that black holes at least probably fall under the relevant universality class when the network in Fig. 2 [our Figure 1] was drawn: by connecting the variables $\mathcal{X}$ and $\mathcal{M}$ [our $\mathcal{M}^{\mathcal{T}}$], we thereby stipulated a correlation between black hole physics and the universality class. It is this step which presupposes (probabilistically) what we would like to establish, viz., that black holes fall into the relevant universality class [...] What we are interested in is precisely whether black holes are appropriately modelled by the model summarised in Sect. 3.1 [the model M^{T}]. In

> the Bayesian framework, this translates into the question of whether an edge should be drawn between $\mathcal{X}$ and $\mathcal{M}$ [$\mathcal{M}^{\mathcal{T}}$] in the Bayesian network in Fig. 2 at all, and not just to what the priors are. (Crowther et al. [2019], p. 22)

Evans and Thébault ([2020]) respond that this objection collapses into inductive scepticism. By comparing analogue experimentation with other well-trusted inference processes, they argue that the confirmatory power of evidence is not determined by the accessibility (or inaccessibility) of the target system of interest. Instead, they suggest, it is determined by the extent to which 'inductive triangulation' is possible: the extent to which we are able to appeal to some distinct mode(s) of inductive reasoning to support the assumed similarity between the system being experimented on and the system to which we want to extrapolate.

In their words,

> Assuming some minimal adequacy of the modelling framework as a description of a target system, or a non-zero probability for the relevant hypothesis, is a *necessary pre-condition of all inductive reasoning* [...] since Crowther et al. are not assuming the position of inductive sceptics, and so are not offering an argument based upon unreasonable doubt, they surely cannot rule out strategies for inductive triangulation that lead to a termination of their justificatory demands. (Evans and Thébault [2020], pp. 15–6)

Thus, even if we accept the formal result in Dardashti, Hartmann et al. ([2019], p. 9)—a result which would be hard to deny—questions remain about how it should be interpreted, and what it means for analogue experiments in practical terms.

The first question is: what ingredients do we need to have at our disposal in order to claim that the inequalities (3) through (6) hold? Are Crowther et al. ([2019]) correct that we cannot access those ingredients without collapsing into circularity, or are Evans and Thébault ([2020]) correct that Crowther et al. ([2019])'s objection collapses into inductive scepticism? The second question is: even when we have those minimal ingredients, when will the corresponding degree of confirmation, Δ_C, be significant? This question needs to be assessed both formally and informally: we need to establish the formal relations that must hold for Δ_C to be significant, and we need to outline the underlying practical situations that will yield those formal relations. I will attempt to answer both questions in what follows, after looking more closely at our epistemic relationship with Hawking radiation in Section 2.1.2.

2.1.2 Hawking radiation and the trans-Planckian problem

Hawking radiation is a pillar of black hole thermodynamics (see e.g. Wall ([2018])). It is what allows black holes to thermally interact with their environment; without Hawking radiation, we would not have any understanding of how such interaction might be possible.

And yet, the theoretical basis for Hawking radiation is not entirely secure. It suffers from two well-known problems: empirical undetectability and the trans-Planckian problem.

Empirical Undetectability: Hawking radiation, if it exists, is far too weak a signal to detect empirically. This is a simple consequence of the predicted temperature of Hawking radiation, $T = \frac{\hbar\kappa}{2\pi k}$, and the fact that black hole surface gravity κ is inversely proportional to mass. A black hole of the Sun's mass, for example, should radiate at a temperature of approximately 10^{-7}K—a plainly undetectable signal compared to the 2.7K cosmic microwave background temperature. Furthermore, we are not yet technologically able to produce smaller black holes (which would radiate at a much higher temperature) in the laboratory.

The Trans-Planckian Problem: Hawking radiation is produced in the high-frequency (trans-Planckian) regime; this is exactly the domain of conditions in which we should expect genuine quantum gravity effects to become relevant. Any Hawking radiation that we receive close to $\mathscr{J}^+$ has been exponentially redshifted on its long journey away from the horizon, and therefore must have emerged from the black hole as an extremely short-wavelength, high-energy signal. Yet, Hawking radiation is a prediction of quantum field theory on curved spacetime, a framework which is only thought to be reliable at sub-Planckian frequencies.[9]

Astrophysical Hawking radiation is therefore exactly the kind of phenomenon that an analogue experiment, if reliable, could be used to illuminate. It is empirically inaccessible, and we have good reason to doubt its existence on theoretical grounds—so we need to test whether it exists, but we cannot use a conventional experiment to perform that test. Furthermore, as outlined in Section 2.1.1, we do have access to source systems that can test for the presence of analogue Hawking radiation.

But recall that even the advocates of analogue experimentation agree that it needs to be supplemented with appropriate universality arguments to gain confirmatory power. On this front, Hawking radiation comes up short: Gryb et al. ([2020]) have shown that the existing universality arguments for Hawking radiation do not seem especially convincing when compared with the classic Wilsonian renormalization-group arguments applied to condensed matter physics.

According to their extensive review of the physics literature, the main candidate arguments can be grouped into three categories:

(a) arguments from the equivalence principle,

(b) arguments from horizon symmetries, and

(c) arguments from modified dispersion relations.

Each attempts to establish, on theoretical grounds, that the existence of Hawking radiation should be independent of the details of quantum gravity. But not one has been unanimously accepted by the physics community as sufficient grounds to dismiss the trans-Planckian problem—and, according to Gryb et al. ([2020]), for good reason. I will summarize each in turn.

9 For more literature on the trans-Planckian problem, see the references in Gryb et al. ([2020]), footnote 5: Gibbons ([1977]), Unruh ([1981]), Jacobson ([1991]), Jacobson ([1993]), Unruh ([1995]), Brout et al. ([1995]), Helfer ([2003]), Jacobson ([2004]), and Harlow ([2016]).

(a) arguments from the equivalence principle

The arguments from the equivalence principle suggest that the Hawking radiation produced by a black hole must be equivalent to the thermal Unruh spectrum that would be viewed locally by an observer accelerating through flat spacetime at a rate corresponding to the black hole spacetime's curvature (Gryb et al. [2020], pp. 16–9). However, Gryb et al. ([2020], p. 18) emphasize that the equivalence principle only holds for local observations, whereas Hawking radiation depends on global properties of the spacetime (Gryb et al. [2020], p. 18). And, even worse, the Unruh spectrum does not in fact seem to be universal against quantum gravity effects (Agulló et al. [2009]; Alkofer et al. [2016]; Gryb et al. [2020], p. 18).

(b) arguments from horizon symmetries

The arguments from horizon symmetries identify the production of Hawking radiation with the production of Goldstone bosons at the horizon, and proceed to show that the production of Goldstone bosons should be unaffected by any changes in micro-physics that preserve the horizon's symmetries and topological structure. However, even if we grant the identification between Hawking radiation and Goldstone bosons (which is a non-trivial identification), these arguments are so idealized that they lack physical plausibility. They only clearly apply to nonevaporating black holes, and even then, only when back-reaction effects are ignored (Gryb et al. [2020], pp. 19–22).

(c) arguments from modified dispersion relations.

Gryb et al. ([2020], p. 24) take the third category, the arguments from modified dispersion relations, to provide 'the most physically plausible derivation of Hawking radiation available'. These arguments incorporate modifications to the dispersion relation into Hawking's original derivation in an attempt to model the impact of possible quantum gravity effects. They are 'readily applicable to a huge range of physical systems', including 'a wide variety of analogue black hole systems' (Gryb et al. [2020], p. 24). But even as the most promising contenders, they are not entirely reliable as universality arguments for astrophysical Hawking radiation. In a 2005 paper, Unruh and Schützhold show that although the Hawking spectrum remains unaltered by some changes to the dispersion relation at high frequencies, it is altered dramatically by other such changes 'which do not appear to be unphysical or artificial' (Unruh and Schützhold [2005], p. 11). Furthermore, one can reasonably doubt 'whether sub-luminal modified dispersion relations really are sufficient to model quantum gravity effects' (Gryb et al. [2020], p. 25).

Thus Gryb et al. ([2020]) have warned that none of the existing universality arguments for Hawking radiation are reliably successful. They identify the weakness of these arguments as a lack of 'integration' between the insensitivity that they provide towards 'token-level' variations on the one hand, and 'type-level' variations on the other: even the arguments that seem to convincingly protect Hawking radiation against high-frequency effects in analogue black hole systems seem unconvincing as universality arguments for astrophysical Hawking radiation (Gryb et al. [2020], p. 25).

The question then becomes: should these arguments be able to support the confirmation of astrophysical Hawking radiation via analogue experimentation based on the Bayesian reasoning outlined by Dardashti, Hartmann et al. ([2019])? If not, why not? And if so, to what extent? Will the degree of confirmation provided be significant?

Furthermore, why are the existing universality arguments for astrophysical Hawking radiation unconvincing? Do they seem limited as solutions to the trans-Planckian problem precisely because of our inability to empirically probe the physics underlying the effect whose presence they are designed to secure? And, more generally, in which situations should we expect to be able to construct universality arguments that do seem convincing? These questions, alongside the questions outlined at the end of Section 2.1.1, will become the focus of the next Section and the rest of the paper.

2.2 Dardashti et al. (2019): Two interpretations

One major barrier to interpreting Dardashti, Hartmann et al. ([2019]) concerns an ambiguity in the definition of the variable $\mathcal{X}$. $\mathcal{X}$ is defined to take the positive value X when a universality argument in support of both $\mathrm{M^S}$ and $\mathrm{M^T}$ obtains. But that definition can be interpreted in two different ways.

We might take it to imply that we already have a specific universality argument A in mind, which we know to support both $\mathrm{M^S}$ and $\mathrm{M^T}$—in formal terms, for which the inequalities $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$ both hold. The possible values of the variable $\mathcal{X}$ would then be given by:

X : A obtains (i.e. its conclusions do indeed follow from its premises).

$\sim X$: A does not obtain.

Notice that, on this interpretation, $\mathcal{X}$ is really just the same as the variable $\mathcal{A}$ defined by the argument A: $\mathcal{X}$ takes the positive value X exactly when A obtains, and $\mathcal{X}$ takes the negative value $\sim X$ exactly when A does not obtain.[10] I will call this Interpretation 1.

We might, instead, take the variable $\mathcal{X}$ to remain agnostic about whether a universality argument that supports both $\mathrm{M^S}$ and $\mathrm{M^T}$ exists, in which case the positive value of $\mathcal{X}$ would be given by the following conjunction:

X : $\exists$ some argument A such that $\big(P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)\big) \wedge \big(\text{A obtains}\big)$.

The negative value of $\mathcal{X}$ would be given by the following disjunction:

$\sim X$: $\forall$ arguments A, $\big($the inequality $P(M^S|A) > P(M^S|\sim A)$ does not hold and/or the inequality $P(M^T|A) > P(M^T|\sim A)$ does not hold$\big) \vee \big(\text{A does not obtain}\big)$.

I will call this Interpretation 2.

At the end of Section 2.1.1, I set out two tasks. First, to identify the ingredients that we need to have at our disposal to claim that the inequalities (3) through (6) hold—and

[10] By 'the variable $\mathcal{A}$ defined by A', I mean the variable that would take a positive value A whenever the argument A obtains, and the negative value $\sim A$ otherwise.

therefore, to be able to claim that an analogue experiment is able to provide some degree of confirmation for hypotheses about its target system. And second, to identify the conditions—formal and informal—under which the corresponding degree of confirmation provided will be significant. I will begin by addressing these tasks from the perspective of Interpretation 1 (Section 2.2.1). Then I will explain why Interpretation 1 is the only viable interpretation (Section 2.2.2).

2.2.1 Interpretation 1

I begin with the first question: from the perspective of Interpretation 1, under what circumstances will the inequalities (3) through (6) hold? In other words, under what circumstances will analogue experiments be able to offer some degree of confirmatory power?

On Interpretation 1, inequality (3) has to do with the probability that the conclusion of our argument A does indeed follow from its premises. It is satisfied if and only if that probability falls between zero and one. This is a very weak condition—all we need is an argument whose validity remains somewhat unknown.

What about the next two inequalities, (4) and (5)? On Interpretation 1, we are starting out with a particular universality argument in mind, which we already know to satisfy the inequalities $P(M^S|A) > P(M^S| \sim A)$ and $P(M^T|A) > P(M^T| \sim A)$. Thus, based on the equivalence of the variables $\mathcal{X}$ and $\mathcal{A}$, (4) and (5) are taken care of from the outset. But, of course, we are not even able to get our reasoning off the ground unless we do in fact possess such an argument.

What kind of argument will satisfy those inequalities? Should we expect such arguments to be in short supply, leaving Dardashti, Hartmann et al. ([2019])'s proposal formally sound but practically unhelpful? The answer, surely, is no: given any candidate universality argument and any two systems S and T, it seems reasonable to allow some nonzero chance that our argument will prove positively relevant to our hypotheses about the systems. This is all that is required for the inequalities to hold. So almost any universality argument should give us what we need to satisfy (4) and (5).

The final condition, $P(E|M^S) > P(E| \sim M^S)$, is satisfied whenever we know of a piece of evidence that would be more likely to obtain if our hypothesis about the source system were correct. On its own, this inequality should be extremely easy to satisfy: in a sense, all it requires is that we understand what our hypothesis about the source system predicts. But, of course, for our reasoning to be practically meaningful, we need to have access to an experimental setup that should, in principle, be able to show the evidence E of interest.

Thus, we have established, from the perspective of Interpretation 1, the practical circumstances in which analogue experiments will be able to provide some degree of confirmatory power when supplemented with universality arguments. We need to have access to a universality argument whose validity remains somewhat unknown—in the sense that it has not yet been definitively confirmed or disconfirmed—and which has at least some small chance of being positively relevant to our hypotheses about the source and target systems. And we need to have access to an experiment that can provide evidence which is somewhat more likely to obtain if our hypothesis about the source system is correct.

These requirements may seem surprisingly weak; but they are meant to be so, because so far we have not said anything about the degree of confirmation that our experiment will be able to provide. That is the next question: on Intepretation 1, in what circumstances, both formal and practical, will the degree of confirmation provided by an analogue experiment be significant?

Here we have to refer back to the expression for confirmatory power established in Dardashti, Hartmann et al. ([2019], p. 9) (our equation (7)). Once again, in the notation introduced in Section 2.1.1, that expression is given by:

$$\Delta_C := m_e^T - m^T = \frac{x\overline{x}}{e}(m_x^S - m_{\overline{x}}^S)(m_x^T - m_{\overline{x}}^T)(e_{m^S} - e_{\overline{m}^S}). \tag{8}$$

So the degree of confirmation provided depends on the values of four factors. Even if an analogue experiment satisfies the minimum practical conditions just outlined, and is therefore able to provide some degree of confirmatory power, that power will be negligible unless the product of these four factors is significant. I label them γ, Δ_S, Δ_T, and Δ_E:

$$\begin{aligned} \gamma &:= \frac{x\overline{x}}{e} \\ \Delta_S &:= m_x^S - m_{\overline{x}}^S \\ \Delta_T &:= m_x^T - m_{\overline{x}}^T \\ \Delta_E &:= e_{m^S} - e_{\overline{m}^S}. \end{aligned}$$

Δ_C will, of course, be at its greatest when each of these factors is as large as possible, but it is not immediately clear what that means—and in particular, whether any of the factors γ, Δ_S, Δ_T, and Δ_E can become arbitrarily large and thereby yield a large value of Δ_C even when the other factors are negligibly small. The limits on the values of γ, Δ_S, Δ_T, and Δ_E should therefore be investigated.

I claim that all of these factors are in fact bounded above, given (3) through (6), and provided that the value of e is not too small. This means that from a purely formal perspective, each of γ, Δ_S, Δ_T, and Δ_E must be significant for Δ_C to be significant—and conversely, if any of those factors is negligibly small, then Δ_C will be negligibly small.

The boundedness is easy to see for Δ_S, Δ_T, and Δ_E. Each of these factors is greater than zero, by (4), (5), and (6). But at the same time, each is equal to a difference between probabilities which are themselves bounded between zero and one. So we have $0 < \Delta_S \leqslant 1$, $0 < \Delta_T \leqslant 1$, and $0 < \Delta_E \leqslant 1$. The boundedness of γ is only slightly less straightforward, requiring the extra assumption that e is not too close to zero.

We know that $\gamma > 0$, since $0 < x < 1$ (from (3)), and since $x + \overline{x} = 1$. So all we need to show is that γ is bounded by above under reasonable conditions. Using $x + \overline{x} = 1$, the expression for γ can be simplified to:

$$\gamma = \frac{x(1-x)}{e}, \tag{9}$$

where $0 < x < 1$ (from (3)) and $0 \leqslant e \leqslant 1$ (because e is a probability). The numerator is certainly bounded above: given $0 < x < 1$, $x(1-x)$ reaches a maximum value of 0.25 when $x = 0.5$. So we only need to worry about the denominator. But e would have to be

smaller than 0.25 for the value of γ to be greater than 1, and e would have to be smaller than 0.025 for the value of γ to be greater than 10. Thus, as long as e is not too close to zero, the value of γ will not be too great.

In fact this seems like a reasonable requirement to impose. The value of e represents the prior probability $P(E)$ that evidence E will be observed in the source system. We would have to be incredibly unsure of our experimental setup and incredibly unsure of our theory of the source system for this value to be very small.

In purely formal terms, we therefore have an answer to our question: for the degree of confirmation provided by an analogue experiment to be significant, each of the factors γ, Δ_S, Δ_T, and Δ_E must be significant (subject to the reasonable assumption that the prior $P(E)$ is not prohibitively close to zero).

What does this mean practically? The value of γ will depend on the details of the case. But the meanings of Δ_E, Δ_S, and Δ_T are easy to interpret. Δ_E encapsulates the relevance of our hypothesis about the source system to the presence of the evidence E: it will be significant whenever we are able to probe a phenomenon in the source system that would be much more likely to exist if our hypothesis about the source system were correct. Practically speaking, this is fairly straightforward: it boils down to our having access to a reliable experimental test of the source, or 'analogue', phenomenon of interest.

Δ_S and Δ_T are the factors that I believe deserve the most attention: they concern the relevance of our universality argument to our hypotheses about the source and target systems. Δ_S encapsulates the degree to which the argument invoked by the variable $\mathcal{X}$ is positively relevant to our hypothesis about the source system, and Δ_T encapsulates the degree to which that argument is positively relevant to our hypothesis about the target system.

If we are not confident that our universality argument is relevant to our hypothesis about the source system, then Δ_S will be small, leaving Δ_C small. Similarly for the target system. But if we are very confident that the argument is positively relevant to both hypotheses, then both Δ_S and Δ_T will be significant, leaving room for Δ_C to be significant. Whether it is in fact significant will depend, in addition, on the values of γ and Δ_E.

Thus we have made substantial headway towards answering our two questions, from the perspective of Interpretation 1. If $\mathcal{X}$ is taken to refer to a particular universality argument that is known to satisfy the conditions $P(M^S|A) > P(M^S| \sim A)$ and $P(M^T|A) > P(M^T| \sim A)$ from the outset—which, as I've shown, is in fact a very weak requirement—we have articulated the practical circumstances in which an analogue experiment will be able to provide some degree of confirmatory power, and we have established the formal conditions under which that confirmatory power will be significant. The practical circumstances which support those formal conditions remain so far unknown, leaving room to assess that final question in Section 3.

Next, I will explain why the other interpretation of Dardashti, Hartmann et al. ([2019]), in which $\mathcal{X}$ remains agnostic about whether a particular universality argument exists, is not substantially different from Interpretation 1 and does not offer meaningful advantages. The answers established above are therefore definitive.

2.2.2 Interpretation 2

On Intepretation 2, the possible values of $\mathcal{X}$ are defined as follows:

X : $\exists$ some argument A such that $\big(P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)\big) \wedge \big($A obtains$\big)$.

$\sim X$: $\forall$ arguments A, $\big($the inequality $P(M^S|A) > P(M^S|\sim A)$ does not hold and/or the inequality $P(M^T|A) > P(M^T|\sim A)$ does not hold$\big) \vee \big($A does not obtain$\big)$.

This interpretation might initially seem to offer some advantages over Interpretation 1. It automatically satisfies inequalities (4) and (5), since if $\mathcal{X}$ takes the value X, then some argument obtains which satisfies $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$. And so surely $P(M^S|X) > P(M^S|\sim X)$ and $P(M^T|X) > P(M^T|\sim X)$ hold.

Furthermore, Interpretation 2 does not force us to commit ourselves to examining a particular universality argument A. Indeed, it does not even force us to commit ourselves to the existence of an argument A that satisfies $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$. So it might seem to give us more for less: namely, non-zero confirmatory power given fewer assumptions.

However, while it is possible to represent the problem this way, it is unhelpful for various reasons. And furthermore, Interpretation 2 does not in fact offer the advantages that it might seem to offer at first sight.

First of all, the prior of X, $P(X)$, now depends not only on the probability that a suitable universality argument holds, but on the probability that such an argument (i.e. an argument satisfying the inequalities $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$) exists. This leads to a complicated and almost self-defeating nesting of probabilities, in which the $P(X)$ that we are feeding into a Bayesian network that requires $P(M^S|X) > P(M^S|\sim X)$ and $P(M^T|X) > P(M^T|\sim X)$ depends, in turn, on the probability that $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$ hold for some particular argument. To define that probability, we would need to have some particular candidate argument(s) in mind. So we have not really abstracted away from particular arguments after all.

Secondly, the values of $P(M^T|X) - P(M^T|\sim X) := \Delta_T$ and $P(M^S|X) - P(M^S|\sim X) := \Delta_S$ will surely depend on the values of $P(M^T|A) - P(M^T|\sim A)$ and $P(M^S|A) - P(M^S|\sim A)$ for particular arguments—or, more precisely, on the probability that $P(M^T|A) - P(M^T|\sim A)$ and $P(M^S|A) - P(M^S|\sim A)$ will take on certain values for particular candidate universality arguments. Again we are required to consider particular arguments, and again we are presented with a nesting of probabilities.

Third, and finally, there is really no need to shift from Interpretation 1 to Interpretation 2. As I emphasized in Section 2.2.1, it should be very easy to find a particular argument that satisfies the inequalities $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$. The challenge is to find arguments that come along with significant values of Δ_S and Δ_T, a challenge that is not eased by adopting Interpretation 2.

Thus, Interpretation 1 is the clear way forward: to be able to define the probabilities in Dardashti, Hartmann et al. ([2019])'s Bayesian network, we must have a particular

universality argument in mind which we know to satisfy the (weak) conditions $P(M^S|A) > P(M^S|\sim A)$ and $P(M^T|A) > P(M^T|\sim A)$. And for the corresponding degree of confirmation to be significant, that argument must produce significant values of Δ_S and Δ_T.

Having established these formal constraints, the final question is: in which practical situations should we expect to be able to construct such a universality argument? That will be the focus of Section 3.

3 When Can a Universality Argument Be Successful?

I begin by distinguishing between two components of success for a universality argument: 'strength' and 'relevance' (Section 3.1). Success as relevance will be my focus here. Motivated by a comparison of the renormalization-group universality arguments in condensed matter physics and Unruh and Schützhold ([2005])'s modified dispersion relation universality argument for Hawking radiation (Section 3.2), it will become clear that there are two such situations, which I will denote (1) and (2) (Section 3.3). Section 4 will then assess the implications of this result, both for analogue experimentation in general and for Hawking radiation in particular.

3.1 The two components of success

There are two sides to the success of a universality argument, which I will label 'strength' and 'relevance'. The strength of a universality argument concerns the probability that its conclusions do in fact follow from its premises. In our Bayesian framework, this corresponds to $P(X)$, the probability that the universality argument at hand obtains. The role of $P(X)$ for analogue experiments has already been examined in detail in Dardashti, Hartmann et al. ([2019]) (see, for example, their Section 4.2 on the role of $P(X)$ in defining the saturation point of confirmatory power for multiple-source analogue experiments).

What has not been examined in detail is the role of relevance—the extent to which a universality argument is known to be significantly positively relevant to the systems of interest. It is this sense of success that corresponds to the factors Δ_S and Δ_T identified in the previous Section, because the positive relevance of any variable $\mathcal{A}$ to the value of any variable $\mathcal{B}$ is captured precisely by the difference between $P(B|A)$ and $P(B|\sim A)$. The practical underlying conditions required to support the construction of significantly relevant universality arguments are what I will attempt to elucidate here.

3.2 Condensed matter vs. quantum gravity

It is well-known that renormalization-group arguments provide highly convincing explanations for why microscopically diverse condensed matter systems like liquids and ferromagnets exhibit the same critical behaviour in the region of a continuous phase transition. By comparison—as argued by Gryb et al. ([2020]), and as reviewed here in Section 2.1.2—even the most promising universality arguments for Hawking radiation seem unconvincing.

Before moving on to more abstract analysis, it will be helpful to examine these two examples in slightly more detail, to assess why the Wilsonian renormalization-group arguments in condensed matter physics are able to succeed and why even the most promising candidate universality argument for Hawking radiation appears to fail.

The renormalization-group arguments in condensed matter physics use an iterative procedure to show that for a wide array of condensed matter systems (which are composed of discrete and approximately evenly spaced micro-particles, and can therefore often be modelled as lattice structures), all micro-features other than their spatial dimension d and their order parameter dimension n are irrelevant to the values of the exponents in the power laws that describe their critical behaviour. This explains why systems with different short-range details, including the short-range details that distinguish a d-dimensional-n-order-ferromagnet-lattice from a d-dimensional-n-order-liquid-lattice, can and should exhibit the same large-scale power-law behaviour in the region near a continuous phase transition.

The argument has a high degree of strength: it is based on convincing mathematical argumentation, and so—allowing, of course, for certain approximations and assumptions—we can be fairly confident that its conclusions follow from its premises. But that is not the only reason why it succeeds. It succeeds, also, because we know enough about the microscopic structure underlying condensed matter physics to be confident that the systems of interest do share the features that the argument invokes. In particular, we can be confident that a vast array of systems in condensed matter physics, which superficially seem to be very different systems, do share the feature that they can all be described as lattices of a particular dimension d and order parameter n at the microscopic level. In the terminology that I have adopted, we can be confident that the universality argument is successful in the sense of being relevant to the systems of interest.

But, in fact, the argument would have been able to succeed in this sense even if we did not have that confidence. For we could have guessed that our systems were lattice structures at the microscopic level, and we could have constructed a renormalization-group argument based on that guess. Since we are able to empirically access the macroscopic behaviour of condensed matter systems, we would have then been able to test our guess by testing whether the argument's predictions about universality held true.

Hawking radiation presents an entirely different situation. Since we do not have a full theory of quantum gravity—or even agreed constraints on what such a theory might look like—and since Hawking radiation is currently empirically undetectable, we have no choice but to guess about whether the possible details of quantum gravity are of the type invoked by a given universality argument. And at the same time, we are left unable to empirically test the accuracy of what that guess predicts.

Indeed, even what Gryb et al. ([2020]) consider to be the most promising universality argument for Hawking radiation, namely Unruh and Schützhold ([2005])'s argument from modified dispersion relations, falls into this category. It makes a guess about the possible micro-physics of quantum gravity—a reasonable guess, but nevertheless a guess—whose predictions cannot be empirically tested due to the empirical undetectability of astrophysical Hawking radiation.

In particular, Unruh and Schützhold ([2005])'s argument only establishes Hawking

radiation as universal assuming three conditions on the structure of quantum gravity, (i), (ii), and (iii):

(i) a preferred reference frame—which does not exist at low frequencies, where all reference frames are equivalent—both exists at high frequencies, and is the freely falling reference frame,

(ii) high frequency excitations begin in their ground state, and

(iii) whatever effects dominate the high frequency domain evolve much more quickly than the effects that dominate at lower frequencies.

(Unruh and Schützhold 2005, 9)

Even the authors themselves admit that they are not at all sure whether the possible micro-physics of quantum gravity do indeed exhibit these features. They write,

> [t]he Hawking effect is not *a priori* independent of the laws of physics at the Planck scale, but it can be made so by imposing the three assumptions [(i), (ii), (iii) above] [...] However, we have also demonstrated counter-examples, which do not appear to be unphysical or artificial, displaying deviations from Hawking's result. Therefore, whether real black holes emit Hawking radiation or not remains an open question. (10-11)

These two examples, and the differences between them, suggest that the success as relevance of a universality argument depends directly on the degree to which we are able to empirically access the microscopic and macroscopic domains of physics that relate to the supposedly universal phenomenon in question. In the next Section, I will argue in more precise terms that this is, indeed, the case.

3.3 The two 'universality-argument-apt' situations

A universality argument aims to explain why we should see macroscopic uniformity despite microscopic diversity. The simplest such explanation would surely be an argument about systems that look as if they are microscopically diverse, but whose micro-physics can be described by the same mathematical model.

Yet this simplest possible formulation must, in fact, capture the spirit of all universality arguments. The more complex versions begin with physical systems whose micro-physics are genuinely different. But, to be convincing—i.e. convincingly relevant to the systems of interest—they must proceed to show that in certain key respects, the systems' microscopic models actually have features in common.

To draw support from Batterman, 'explanations of universal (multiply realized) behavior proceed by finding principled reasons *in the physical theory of the realizers* for the irrelevancy of certain details that otherwise are quite important for answering certain explanatory questions' (Batterman [2002], p. 119). A successful universality argument must convincingly argue that the systems' micro-models differ only in those irrelevant details.

	Our State of Knowledge		
Micro-structure (of the physical systems)	*Known* Type-A	*Unknown* But we can guess it is type-A	*Known* Type-A
Macro-structure (of the physical systems)	*Unknown* But we can predict it is type-M based on our knowledge that the micro-structure is type-A	*Known* Empirical tests are able to show whether it is type-M	*Known* Empirical tests are able to show whether it is type-M, and we can also predict it is type-M based on our knowledge that the micro-structure is type-A

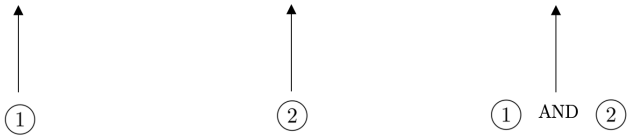

Figure 2: The universality-argument-apt situations.

	Our State of Knowledge
Micro-structure (of the physical systems)	*Unknown* But we can guess it is type-A
Macro-structure (of the physical systems)	*Unknown* We can guess it is type-M based on our guess that the micro-structure is type-A, but empirical tests are unable to show us whether either of our guesses are correct

Figure 3: General characterization of situations that are not universality-argument-apt.

Thus, for a universality argument to achieve its aim, we must be confident that the micro-physics of the physical systems at hand are indeed of the same type—call it type-A— with respect to the key structural features invoked by the argument.

A universality argument must therefore be of the following form:

> Given that the micro-physics of physical system S has type-A structure, changes in its micro-features $\{f_1, f_2, \ldots, f_n\}$—which would make it a different physical system, but still with type-A micro-structure—would not undermine the adequacy of a type-M model as its macro-description;

where a type-M model is the type of model that predicts the supposedly universal phenomenon. The antecedent, 'given that the micro-physics of *S* has type-A structure', is essential.

There are therefore only two 'universality-argument-apt' situations, in which a universality argument can succeed in the sense of being significantly relevant to the systems at hand (for a schematic outline, see Figures 2 and 3).

Either (1): we have micro-models of our physical systems that we are very confident about, which we can show are all type-A.

Or (2): we can guess micro-models for our systems, which we can show are all type-A, run arguments on those models that make predictions about the universality of certain phenomena, and then compensate for our initial guessing by empirically testing whether their predictions are correct (whether the macro-phenomena predicted to be universal are indeed universal).

These results provide further insight on our earlier comparison of the successful renormalization-group universality arguments in condensed matter physics and the

unsuccessful modified dispersion relation universality arguments for Hawking radiation. Namely, we know enough about the micro-structure of condensed matter physics, and have enough empirical access to its macro-behaviour, to be in both situations (1) and (2). But for Hawking radiation, we are not in either situation. Our state of knowledge with respect to Hawking radiation is not currently universality-argument-apt.

By acknowledging this state of affairs, along with the two allowable situations (1) and (2) from which it stems, we are not only able to better understand the status of our search for Hawking radiation. Also, given any area of physics, we are able to quickly and simply judge whether that area is able to support universality arguments that are successful in the sense of being significantly positively relevant to the systems of interest.

In the next Section, I will connect these conclusions back to the formal and informal conditions required to support the presence and significance of confirmatory power for analogue experiments supplemented with universality arguments.

4 Morals

The results of Section 3 are significant for an obvious reason: if we know which situations can support the construction of successful universality arguments, we can better decide where to direct our energy, and we can better explain the successes and failures among existing efforts.

But those are not the only morals. The conditions (1) and (2) connect simply and clearly with the formal and informal conditions underlying the confirmatory power of analogue experimentation that were established in Section 2.2, thereby clarifying the status of analogue experimentation both in general and for Hawking radiation in particular (Sections 4.1.1, 4.1.2, 4.2.1, 4.2.2). Additionally, and perhaps most importantly, they emphasize the priority of empirical over theoretical justification—while at the same time revealing that direct empirical access is not always an essential precondition for confirmation (Section 4.3).

4.1 On the presence of confirmatory power

4.1.1 In general

Based on Section 2.2, three practical ingredients are required for an analogue experiment to be able to provide some degree of confirmatory power. We need to have access to a universality argument whose validity remains somewhat unknown. We need that argument to have at least some small hope of being positively relevant to our hypotheses about the source and target systems. And we need to have access to an experiment that can provide evidence which is somewhat more likely to obtain if our hypothesis about the source system is correct.

The second requirement might initially seem problematic—indeed, it is where Crowther et al. ([2019]) seem to take issue with Dardashti, Hartmann et al. ([2019]), as outlined in Section 2.1. Crowther et al. ([2019]) suggest that, by requiring us to assume the existence of a universality argument that is positively relevant to both our source and target hypotheses, Dardashti, Hartmann et al. ([2019]) fall into circular reasoning.

And yet, as I have argued in Section 2.2, the requirement of positive relevance is extremely weak and should be satisfied by almost any candidate universality argument. As an objection to the presence of confirmatory power, the argument in Crowther et al. ([2019]) therefore cannot stand. Here I agree with Evans and Thébault ([2020]), who suggest that Crowther et al. ([2019])'s argument collapses into inductive scepticism.

Where an objection might be convincingly raised is on the question of whether the confirmatory power provided is significant. On that question I would agree with Crowther et al. ([2019]), as far as analogue gravity is concerned: the existing analogue black hole experiments do not seem to be able to confirm the existence of astrophysical Hawking radiation to any significant degree, as I will elaborate further in Section 4.2.2. But, as I have shown, this question is not about whether we can assume $P(M^S|X) > P(M^S| \sim X)$ and $P(M^T|X) > P(M^T| \sim X)$. Instead it is about the magnitude of the differences $P(M^S|X) - P(M^S| \sim X)$ and $P(M^T|X) - P(M^T| \sim X)$, and, correspondingly, about whether we are in situations (1) or (2) with respect to the systems of interest.

Crowther et al. ([2019], p. 22) do seem, to some extent, to anticipate this response. They write:

> One may retort to this charge that all the theorem discussed above [the theorem which secures the confirmatory power of analogue experiments] requires is that $P(M^T) \neq 0$ and that $P(M^T|X)$ is ever so slightly greater than $P(M^T| \sim X)$. Thus, it may be insisted, the initially required commitment to M^T and to its probabilistic dependence on X is minimal and so epistemically responsible. However, that these assumptions are not as innocent as they appear can be seen from the following straightforward consequence of the theorems in DHTW. Particularly, by moving to multi-source confirmation (Dardashti et al. (2018, §4.2)), but in principle already in the simple case considered here (and in their §4.1), we can perform a large number of terrestrial analogue experiments and thereby achieve an ever increasing (but in general bounded, cf. their §4.2 and Appendix 2) degree of confirmation.[11]

However, they need not worry. As shown by Dardashti et al. (2019, 9-10),

$$lim_{n\to\infty}\Delta^{(n)} = \overline{x}(m_x^T - m_{\overline{x}}^T) = \overline{x}\Delta_T, \tag{10}$$

where $\Delta^{(n)}$ denotes the degree of confirmation provided by an n-source analogue experiment. So even for multiple-source experiments, the value of Δ_T is a limiting factor. By assuming the inequality $P(M^T|X) > P(M^T| \sim X)$, we do not risk attributing a high degree of confirmatory power to multiple-source analogue experiments unless the value of Δ_T is significant.

4.1.2 For Hawking radiation

In particular, the minimal conditions for the presence of confirmatory power are satisfied by the existing analogue experiments for Hawking radiation. The validity of those arguments certainly remains somewhat unknown. And they are certainly positively relevant to the source and target systems to at least some very small extent—there is

[11] I have replaced their M's with M^T's to match the notation I have been using throughout.

surely some chance, for example, that both the physics underlying analogue black holes and the physics underlying astrophysical black holes do both satisfy conditions (i)-(iii) in Unruh and Schützhold ([2005]).

Finally, since at least 2016 we have had access to analogue experiments that are able to show analogue Hawking radiation, an effect which coincides directly with whether our hypothesis about the source system is correct. So the existing analogue experiments for Hawking radiation are certainly able to provide some degree of confirmation for the hypothesis that astrophysical Hawking radiation exists.

4.2 On the significance of confirmatory power

4.2.1 In general

Based on Section 2.2, the confirmatory power offered by an analogue experiment will only be significant if four factors are significant: $\gamma = \frac{x\overline{x}}{e}$, $\Delta_S = m_x^S - m_{\overline{x}}^S$, $\Delta_T = m_x^T - m_{\overline{x}}^T$, and $\Delta_E = e_{m^S} - e_{\overline{m}^S}$.

Δ_S and Δ_T, I have argued, are the factors that require the most interpretation. They will be significant whenever we are confident that our universality argument is positively relevant to both the source and target systems. Based on Section 3, we will have this confidence exactly when one or both of conditions (1) and (2) hold for both systems: in short, if we are confident (or able to make ourselves so through empirical investigation) that the micro-physics of the systems have the features invoked by our universality argument.

In other words, given a universality argument whose antecedent invokes type-A micro-structure, the universality argument can only transfer confirmation of the adequacy of a type-M macro-model from the source system to the target system to any significant extent on the condition that the source and target systems are known to have type-A micro-structure.

This verdict need not be interpreted as a negative outcome for the confirmatory power of analogue simulation. On the contrary, it suggests that analogue experiments, when supplemented with universality arguments, can provide a significant degree of confirmation for hypothesized features of their target systems, just as Dardashti, Hartmann et al. ([2019]) suggest. It only narrows down the situations in which that significant transfer of confirmation is possible.

4.2.2 For Hawking radiation

Following Gryb et al. ([2020]) (see Section 2.1.2), none of the existing universality arguments for Hawking radiation seem as convincing as their Wilsonian counterparts. Gryb et al. ([2020]) diagnose the problem as a lack of 'integration': the arguments that seem convincing for token-level variation do not seem convincing for type-level variation, and vice versa.

This lack of integration, I suggest, amounts to a mismatch between the values of Δ_S and Δ_T: the arguments that are known to be significantly relevant to our hypothesis about the source system are not known to be significantly relevant to our hypothesis about the target system. If I am correct that Δ_S and Δ_T will only both be significant if we are in

situation (1) or (2) with respect to both systems, then there is an underlying explanation for this mismatch. Namely, whenever we are confident that our analogue system is of the type invoked by one of our universality arguments, we do not currently know enough about quantum gravity—and in particular, how quantum gravity affects parameters that we are able to model in the currently available analogue systems—to be able to claim that the target system's microphysics are of that same type. And therefore, Δ_T is bound to be low even when Δ_S is high.

The existing analogue experiments for Hawking radiation are therefore currently unable to confirm the existence of astrophysical Hawking radiation to any significant extent. This is consistent with what Gryb et al. ([2020]) suggest. Out of the three candidate arguments that they examine in detail, the only one that has significant robustness with respect to analogue black holes—namely, Unruh and Schützhold's modified dispersion relations argument—has extremely limited robustness with respect to astrophysical black holes.

Indeed, given how very ignorant we are about the micro-structure of quantum gravity, it seems unlikely that any universality argument will be able to play a significant role in confirming the existence of astrophysical Hawking radiation until our knowledge of quantum gravity improves. Our ignorance of quantum gravity is not only an ignorance of its structural details. Currently, we cannot even confidently outline constraints on its structure, beyond the fact that it must be consistent with quantum field theory on curved spacetime in the limit of large wavelengths and weak spacetime curvature. Therefore, it is hard to envisage how the antecedent of any candidate universality argument could currently be known to hold of the target, astrophysical, system.

This situation, however, could change. It could change if we were to learn more about the physics underlying astrophysical black holes—and, in particular, about how that physics affects parameters that we are currently able to model in analogue black holes—such that one of the existing universality arguments would yield significant values for both Δ_S and Δ_T. Or it could change if we were able to develop some new universality argument for which conditions (1) or (2) would be satisfied for both systems; although, as I just mentioned, it is difficult to imagine how such an argument could currently be constructed. Or, finally, it could change if we were able to develop new analogue systems that could more closely model quantum gravity effects. Such systems would allow Δ_S to be significant for a wider range of universality arguments—including, possibly, arguments for which Δ_T is also significant.

4.3 On the status of empirical access

It would be a shame not to consider what the above arguments have to say, more generally, about the status of empirical accessibility in scientific investigation.

They certainly suggest that manipulability is not paramount: we do not necessarily need to be able to empirically manipulate a phenomenon of interest in order to confirm hypotheses about its behaviour. But what about accessibility?

On the one hand, they suggest that we do not need to be able to directly access the physics underpinning that phenomenon, as long as we can access enough other features of the system to make ourselves confident that it falls into a universality class whose behaviour we understand (or can come to understand through further empirical

investigation). But on the other hand, conditions (1) and (2) emphasize that we do need exactly enough access to be able to gain that confidence.

On the first two points—that neither manipulability nor direct access are strictly required to gain significant confirmation about a phenomenon's behaviour—I agree with Evans and Thébault ([2020]). Indeed their first case study, stellar nucleosynthesis, is plausibly a case in point. The nuclear processes that govern energy production in the stellar core are unmanipulable and inaccessible, and nevertheless, Evans and Thébault ([2020], pp. 10–12) point out, it seems reasonable to claim that our model of those processes is well-established and well-confirmed.

But, as Evans and Thébault themselves note, the system of interest here is not entirely inaccessible. In particular, even though we cannot observe the nuclear processes themselves, we have enough access to other stellar signals to gain confidence that such processes do have the features invoked by the uniformity principles on which our model relies. I would suggest—and this corresponds to my third point above—that the confidence conferred by this access is exactly what makes our model of stellar nucleosynthesis so convincing.

At the same time, my claim that direct empirical access to the phenomenon of interest is not always essential captures my deepest disagreement with Crowther et al. ([2019]).[12] Crowther et al. ([2019]) seem to argue that in order to gain confirmation from an analogue experiment, we would need to be able to directly access the phenomenon of interest in the target system—an obviously problematic requirement, since such access would undermine our motivation for performing the analogue experiment in the first place. I argue, instead, that we need to be able to access the target system just enough to make ourselves confident that it has the features invoked by an appropriate universality argument.

This is a much weaker requirement. It allows us to move away from dismissing analogue experimentation as viciously circular, and towards accepting analogue experimentation as a promising basis for empirical confirmation of target phenomena.

5 Conclusion

When successful, universality arguments are able to establish the irrelevance of various features to a whole class of systems' behaviour. They are powerful predictive and explanatory tools—for predicting as-yet-unobserved universality in empirically accessible systems, and for explaining observed universality in systems that have already been investigated.

Because they claim to establish the irrelevance of so many features to the macroscopic behaviour of so many systems, they can easily seem to override the need for empirical investigation of those systems' underlying structure. And yet, motivated by Gryb et al. ([2020])'s comparison between the successful renormalization-group argument for condensed matter physics and the much less successful candidate universality arguments for Hawking radiation, I have argued that empirical constraints limit the situations in which we should expect to be able to construct successful

12 That is, aside from the fact that I would ascribe nonzero but negligible confirmatory power to the existing analogue black hole experiments, whereas they would assign exactly zero.

universality arguments.

We have seen that a universality argument is necessarily of the form,

> Given that the micro-physics of physical system S has type-A structure, changes in its micro-features $\{f_1, f_2, \ldots, f_n\}$—which would make it a different physical system, but still with type-A micro-structure—would not undermine the adequacy of a type-M mathematical model as its macro-description;

so that it can only tell us, to any significant extent, about whether the same macro-phenomenon really will appear in microscopically diverse systems if we are confident that its antecedent, 'given that the micro-physics has type-A mathematical structure', holds of each system.

We can only ever be confident that this antecedent holds of each system if

(1) we have micro-models of our systems that we are very confident about, which we can show are all of the same type in the respects invoked by the argument; and/or

(2) we can empirically test the macro-behaviour of our systems, to compensate for our initial uncertainty about whether their micro-physics are indeed of the same type in those respects.

I have shown that this conclusion connects in a very clear and simple way with the confirmatory power of analogue experiments supplemented with universality arguments. As Dardashti, Hartmann et al. ([2019]) themselves show, the confirmatory power Δ_C of an analogue experiment supplemented by a universality argument is given by (cf. Equation (7)):

$$\begin{aligned}\Delta_C &:= P(M^T|E) - P(M^T) \\ &= m_e^T - m^T \\ &= \frac{x\bar{x}}{e}(m_x^S - m_{\bar{x}}^S)(m_x^T - m_{\bar{x}}^T)(e_{m^S} - e_{\overline{m}^S}).\end{aligned}$$

So the confirmatory power of an analogue experiment supplemented with a universality argument is directly proportional to both $\Delta_S := P(M^S|X) - P(M^S|\sim X)$ and $\Delta_T := P(M^T|X) - P(M^T|\sim X)$.

If Δ_S and Δ_T are small, which will happen exactly when my conditions (1) and (2) fail to hold, then the degree of confirmation Δ_C provided by our analogue experiment will necessarily be small (provided that the other factors involved are bounded above, which I have shown to be true of Δ_E, Δ_S and Δ_T generically, and of γ under reasonable conditions). If Δ_S and Δ_T are significant, then the degree of confirmation Δ_C provided by our analogue experiment will be (or at least can be) significant.

These results are noteworthy for various reasons. Most obviously, they tell us exactly when we should not expect to be able to construct a successful universality argument: namely, when we are not in situations (1) or (2). At the same time, they clarify the status of analogue experimentation and the status of existing universality arguments for Hawking radiation.

Only (2) explicitly relies on our ability to empirically access the system at hand. But both (1) and (2) go back to empirical access in the end, because we will only find ourselves in situation (1)—with confidence in our systems' micro-models—if we have been able to perform enough empirical tests to gain that confidence. Like all theoretical arguments, universality arguments are therefore grounded in empirical knowledge. Their success is shaped, and ultimately limited, by the availability of empirical evidence.

Acknowledgements

I am indebted to two anonymous referees for their comments and suggestions, and to Hasok Chang, Erik Curiel, Karim Thébault, Sean Gryb, and Pete Evans for many helpful conversations on this topic. I received valuable feedback from audiences at the 2019 Hanneke Janssen Memorial Prize Ceremony, the 2020 Oxford Philosophy of Physics Graduate Conference, and the 17ème Journées de la Matière Condensée Conference. I would like to especially thank Jeremy Butterfield for comments and guidance on many previous versions of the paper. This work was supported in part by the University of Cambridge Harding Distinguished Postgraduate Scholars Programme, and in part by funding from the Social Sciences and Humanities Research Council of Canada.

Department of History and Philosophy of Science
University of Cambridge
Cambridge, UK
gef30@cam.ac.uk

References

Agulló, I., J. Navarro-Salas, G. J. Olmo and L. Parker [2009]: 'Insensitivity of Hawking radiation to an invariant Planck-scale cutoff', *Physical Review D*, **80**.

Alkofer, N., G. D'Odorico, F. Saueressig and F. Versteegen [2016]: 'Quantum gravity signatures in the Unruh effect', *Physical Review D*, **94**.

Bartha, P. [2010]: *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*, Oxford: Oxford University Press.

Batterman, R. W. [2002]: *The Devil in the Details*, Oxford: Oxford University Press.

Brout, R., S. Massar, R. Parentani and P. Spindel [Oct. 1995]: 'Hawking radiation without trans-Planckian frequencies', *Physical Review D*, **52**, pp. 4559–68.

Crowther, K., N. Linnemann and C. Wüthrich [May 2019]: 'What we cannot learn from analogue experiments', *Synthese*.

Dardashti, R., S. Hartmann, K. Thébault and E. Winsberg [2019]: 'Hawking Radiation and Analogue Experiments: A Bayesian Analysis', *Studies in History and Philosophy of Modern Physics*, **67**, pp. 1–11.

Dardashti, R., K. P. Y. Thébault and E. Winsberg [2017]: 'Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity', *British Journal for the Philosophy of Science*, **68**, pp. 55–89.

de Nova, J. R. M., K. Golubkov, V. I. Kolobov and J. Steinhauer [May 2019]: 'Observation of thermal Hawking radiation and its temperature in an analogue black hole', *Nature*, **569**, pp. 688–91.

Evans, P. W. and K. P. Y. Thébault [2020]: 'On the limits of experimental knowledge', *Philosophical Transactions of the Royal Society A*, **378**.
Feldbacher-Escamilla, C. J. and A. Gebharter [2020]: 'Confirmation Based on Analogical Inference: Bayes Meets Jeffrey', *Canadian Journal of Philosophy*, **50**, pp. 174–94.
Gibbons, G. [1977]: 'Quantum processing near black holes', in R. Ruffini (*ed.*), *Proceedings of the First Marcel Grossman Meeting on General Relativity*, Amsterdam: North-Holland, pp. 449–58.
Gryb, S., P. Palacios and K. P. Y. Thébault [2020]: 'On the Universality of Hawking Radiation', *British Journal for the Philosophy of Science*, **0**, pp. 1–32.
Harlow, D. [2016]: 'Jerusalem lectures on black holes and quantum information', *Reviews of Modern Physics*, **88**.
Helfer, A. D. [May 2003]: 'Do black holes radiate?', *Reports on Progress in Physics*, **66**, pp. 943–1008.
Hesse, M. B. [1964]: 'Analogy and Confirmation Theory', *Philosophy of Science*, **31**, pp. 319–27.
Hesse, M. B. [1966]: *Models and Analogies in Science*, Notre Dame: University of Notre Dame Press.
Hesse, M. B. [1973]: 'Logic of discovery in Maxwell's electromagnetic theory', in R. Giere and R. Westfall (*eds.*), *Foundations of Scientific Method: The Nineteenth Century*, Bloomington: University of Indiana Press, pp. 86–114.
Hesse, M. B. [1974]: *The Structure of Scientific Inference*, Berkeley: University of California Press.
Hesse, M. B. [1988]: 'Theories, Family Resemblances and Analogy', in D. H. Helman (*ed.*), *Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy*, Dordrecht: Kluwer Academic Publishers, pp. 317–40.
Jacobson, T. [Sept. 1991]: 'Black-hole evaporation and ultrashort distances', *Physical Review D*, **44**, pp. 1731–9.
Jacobson, T. [July 1993]: 'Black hole radiation in the presence of a short distance cutoff', *Physical Review D*, **48**, pp. 728–41.
Jacobson, T. [Apr. 2004]: 'Introduction to Quantum Fields in Curved Spacetime and the Hawking Effect', *Preprint*.
Jacquet, M. J. and F. König [June 2020]: 'The influence of spacetime curvature on quantum emission in optical analogues to gravity'.
Keynes, J. M. [1921]: *A Treatise on Probability*, London: Macmillan.
Kolobov, V. I., K. Golubkov, J. R. M. de Nova and J. Steinhauer [Oct. 2019]: 'Spontaneous Hawking radiation and beyond: Observing the time evolution of an analogue black hole'.
Mill, J. S. [1843/1930]: *A Domain of Logic*, London: Longmans-Green.
Rousseaux, G., C. Mathis, P. Massa, T. G. Philbin and U. Leonhardt [2008]: 'Observation of negative-frequency waves in a water tank: a classical analogue to the Hawking effect?', *New Journal of Physics*, **10**.
Steinhauer, J. [Oct. 2016]: 'Observation of quantum Hawking radiation and its entanglement in an analogue black hole', *Nature Physics*, **12**, pp. 959–65.
Sterrett, S. G. [2006]: 'Models of Machines and Models of Phenomena', *International Studies in the Philosophy of Science*, **20**, pp. 69–80.

Unruh, W. G. [May 1981]: 'Experimental Black-Hole Evaporation?', *Physical Review Letters*, **46**, pp. 1351–3.
Unruh, W. G. [Mar. 1995]: 'Sonic analogue of black holes and the effects of high frequencies on black hole evaporation', *Physical Review D*, **51**, pp. 2827–38.
Unruh, W. G. [2014]: 'Has Hawking Radiation Been Measured?', *Foundations of Physics*, **44**, pp. 532–45.
Unruh, W. G. and R. Schützhold [2005]: 'Universality of the Hawking effect', *Physical Review D*, **71**, pp. 1–11.
Wall, A. C. [July 2018]: 'A Survey of Black Hole Thermodynamics'.
Weinfurtner, S., E. W. Tedford, M. C. J. Penrice, W. G. Unruh and G. A. Lawrence [2011]: 'Measurement of stimulated Hawking emission in an analogue system', *Physical Review Letters*, **106**.
Weisberg, M. [2013]: *Simulation and Similarity*, Oxford: Oxford University Press.
Wolchover, N. [Nov. 2016]: 'What Sonic Black Holes Say About Real Ones', *Quanta Magazine*.