“Let’s build an Anscombe box”: assessing Anscombe’s rebuttal of the statistics objection against indeterminism-based free agency

Thomas Müller¹

Received: 16 March 2021 / Accepted: 26 December 2021 / Published online: 7 March 2022
© The Author(s) 2022

Abstract

Towards the end of her famous 1971 paper “Causality and Determination”, Elizabeth Anscombe discusses the controversial idea that “‘physical haphazard’ could be the only physical correlate of human freedom of action”. In order to illustrate how the high-level freedom of human action can go together with micro-indeterminism without creating a problem for micro-statistics, she provides the analogy of a glass box filled with minute coloured particles whose micro-dynamics is subject to statistical laws, while its outside reliably displays a recognisable variation of a certain image created by an external cause. Anscombe’s example is somewhat enigmatic, and she provides no details to support its viability. In this paper we discuss the context of Anscombe’s analogy, and we provide basics of the theory and practice of building an Anscombe box.

Keywords Indeterminism · Free agency · Top-down causation · Anscombe · Incompatibilism · Agent causation · Wild coincidences · Pereboom

1 Introduction

One of the main strands of Anscombe’s famous inaugural lecture at Cambridge, “Causality and Determination” (Anscombe, 1971),¹ is to make sense of causality

¹ Page references without further indication of a source are to this paper, in the pagination of the 1981 *Collected Philosophical Papers* reprint edition.

This article belongs to the topical collection “Causality and Determination, Powers and Agency: Anscombean Perspectives”, edited by Jesse M. Mulder, Dawa Ometto, Niels van Miltenburg, and Thomas Müller. Thanks to Jesse Mulder and to two anonymous referees for helpful comments on an earlier draft, and to Dawa Ometto for help with refactoring the program code.

✉ Thomas Müller
Thomas.Mueller@uni-konstanz.de

¹ Department of Philosophy, University of Konstanz, Box 17, 78457 Konstanz, Germany

without the assumption of determinism. Along the way, she argues that we should view the laws of nature, in so far as they are required to make sense of causality, as constraining rather than fully determining what happens. The laws of planetary motion, which provided a strong but ultimately disastrous template for laws of nature since the 17th century, may well be deterministic, “like the rules of an infantile card game”. Not so, however, for those laws that pertain, e.g., to the motion of us humans and other animals: “[I]n relation to what happens on and inside a planet the laws are, rather, like the rules of chess; the play is seldom determined, though nobody breaks the rules” (143).

Assuming that the notion of an indeterministic law is thus vindicated, and given that we have good evidence that such indeterministic laws are at play in our world, the question arises whether we thereby have in our hands also “what was wanted for defending the freedom of the will” (145), as some have thought. Anscombe points out that two objections are commonly raised against an affirmative answer. The first is what today is mostly discussed under the heading of *the luck objection* (see, e.g., Widerker & Schnall, 2015): it cannot be that “this ‘mere hap’ is [...] the physical correlate of ‘man’s ethical behaviour’” (145), because acting based on mere hap would be even worse than being determined. Accepting this objection pushes a defender of free will to a compatibilist reconciliation of physical determinism and freedom. Anscombe makes it clear that she has no sympathy for compatibilism: “The reconciliations have always seemed to me to be either so much gobbledegook, or to make the alleged freedom of action quite unreal” (145f.). Her reason is that “actions are mostly physical movements”, and “if these physical movements are physically predetermined by processes which I do not control, then my freedom is perfectly illusory” (146)—basically, this is a form of Van Inwagen’s consequence argument (Van Inwagen, 1983, p. 56). That incompatibilist argument by itself, as Anscombe notices explicitly, is not sufficient to establish a sensible idea of freedom, but “[t]he truth of physical indeterminism is [...] indispensable if we are to make anything of the claim of freedom”, and “there is nothing unacceptable about the idea that ‘physical haphazard’ should be the only physical correlate of human freedom of action; and perhaps also of the voluntariness and intentionality in the conduct of other animals” (146).

Thus, Anscombe rejects compatibilism and proclaims that physical indeterminism is acceptable as the sole physical correlate of free agency. Anscombe does not go on to develop a positive view of indeterminism-based freedom in her paper in any detail.² She does go on, however, to discuss and reject a second objection against the view that indeterminism can be freedom-friendly, and that objection, she says, is “more to the point” (146), i.e., harder to deal with than the luck objection. Despite its alleged seriousness, this second objection, which I will call *the statistics objection*, does not play a prominent role in the current free will discussion.³ Anscombe describes the

² See Kane (1998), Palmer (2014), and Müller and Briegel (2018) for different attempts at providing more details.

³ The statistics objection bears certain interesting similarities to Pereboom’s “wild coincidences” objection against agent-causal libertarianism (Pereboom, 2001, 2014, pp. 65–69), which has been the focus of several recent papers (e.g., Baker, 2017; Runyan, 2018; Taggart, 2021). See notes 7 and 8 for some details about Pereboom’s argument and its relation to Anscombe’s text. I provide a brief comment about the relation in the conclusions section (Sect. 5).

objection clearly, and she sketches a possible reply to the objection towards the end of her paper, providing a complex analogy as her last move before wrapping up. As far as I can see, Anscombe's analogy, the "Anscombe box" as I will call it, has not been thoroughly discussed,⁴ and it certainly does seem a bit enigmatic. It is my main aim in this paper to describe the statistics objection and to analyse to which extent Anscombe's box analogy helps to counter the objection.

2 The statistics objection

Anscombe's first exposition of the statistics objection runs as follows:

[Q]uantum laws predict statistics of events when situations are repeated; interference with these, by the *will's* determining individual events which the laws of nature leave undetermined, would be as much a violation of natural law as would have been interference which falsified a deterministic mechanical law. (145; italics in original)

The background of this challenge is that the actual quantum laws, which Anscombe mentions here as proof that "physics went indeterministic" (145), are unlike the rules of chess in an important respect. The chess rules only say which moves are possible given a certain configuration of pieces, but they do not say *how possible, to which degree* the various moves are possible; they do not prescribe any statistics on the moves. In the opening configuration, White has 20 possible moves. Nothing in the rules of chess says anything about the distribution among these moves. Let us assume that 44% of all first moves are 1. e4.⁵ *Nothing at all in the rules* stands against that percentage in 2022 dropping to 5%, or rising to 95%. If all players of chess choose to cooperate accordingly, then these would be the statistics, and this would leave the rules of the game completely untouched.

In quantum physics, however, the fundamental Born rule for measurement outcomes is a *statistical law*: it describes *both* the *possible outcomes* $e_1, \dots, e_N$ of a measurement of an observable A on a system in state $|\psi\rangle$, and it provides a precise numerical *probability* p_i for each possible outcome e_i .⁶ Such a statistical law seems to be open to a coordinated attack, which Anscombe describes as follows:

Certainly if we have a statistical law, but undetermined individual events, and then enough of these are supposed to be pushed by will in one direction to falsify

⁴ For some references to texts in which Anscombe's analogy is mentioned, see note 13 below.

⁵ According to chessree.net, accessed 17 January 2021.

⁶ Mathematically, for the case of a finite-dimensional Hilbert space $\mathcal{H}$ of dimension N , with a self-adjoint operator A representing the measurement of an observable and $|a_1\rangle, \dots, |a_N\rangle$ a basis for $\mathcal{H}$ consisting of eigenstates of A with non-degenerate eigenvalues $e_1, \dots, e_N$ (so that $A|a_i\rangle = e_i|a_i\rangle$), the possible measurement outcomes are exactly the N eigenvalues $e_1, \dots, e_N$, and the probability of measuring value e_i , given that the system considered is in the quantum state $|\psi\rangle$, is $p_i = |\langle a_i|\psi\rangle|^2$. In this paper I will not engage with the view that it may be possible to interpret quantum mechanics in a deterministic way. I acknowledge that there are interesting aspects to that discussion (see, e.g., Adlam, 2018). Some of my own views for holding quantum physics to be indeterministic are given in Müller and Briegel (2018, p. 231f.). For the purposes of discussing Anscombe's paper, it is enough to acknowledge that she mentions quantum physics as proof of the claim that "physics went indeterministic" (145).

the statistical law, we have again a supposition that puts will into conflict with natural laws. (146)

If the rules of chess were statistical, and they contained a clause to the effect that 44% of all games shall be opened by the move 1. e4, then a coordinated effort by the world's chess players (declaring 2022 to be “the year of 1. d4”, say) would be able to invalidate the rules. Similarly, if my free choice between having tea or coffee in the morning corresponded to a physical haphazard of the form “measure observable A on a system in state $|\psi\rangle$ ”, where outcome e_1 corresponded to tea and outcome e_2 to coffee, and the Born rule fixed the statistics to be 50:50, then I could, by freely choosing to abandon coffee forever, invalidate the Born rule, breaking a law of quantum physics by my free actions.

Anscombe is quick to point out that the assumptions behind the objection are unrealistic:

But it is not at all clear that the same train of minute physical events should have to be the regular correlate of the same action; in fact, that suggestion looks immensely implausible. It is, however, required by the objection. (146)

It seems clear, indeed, that my having tea, as well as my having coffee, is multiply realizable, and is in fact realized by different states of my body (not just my brain) on different mornings. This may or may not be enough to reject the statistics objection.⁷ It appears that Anscombe feels that more should be said, as she offers an “analogy to illustrate this point” (146).⁸

3 The analogy of the Anscombe box

The context in which Anscombe presents her analogy has just been given: she is about to provide an illustration that the same action can have different microphysical

⁷ Runyan (2018) provides many pertinent references to empirical studies attesting to the variability of human behaviour. He does not relate these studies to Anscombe; the context of his discussion is, rather, Pereboom's already mentioned “wild coincidences” objection (see note 3; see also note 8).

⁸ As mentioned above in note 3, the statistics objection has parallels to Pereboom's “wild coincidences” objection. In order to keep this paper focused on Anscombe, I do not elaborate on the parallels, except for a brief remark in the conclusions section (Sect. 5). To provide some background: Pereboom's argument is directed against agent-causal libertarianism. He argues that already the conceptual coherence of agent-causal libertarianism is in doubt (Pereboom, 2014, p. 69), but the stronger point against the position is that it is “doubtful for empirical reasons that derive from our best physical theories” (Pereboom, 2014, p. 69f.). Briefly, the objection is as follows: Consider a given class of free human actions and their respective physical components. Assuming indeterminism (as the libertarian does), each such physical component will have a precise probability assigned, e.g., 32%. Now for the single instances, such a probability assignment never amounts to a verdict of necessity or impossibility (unlike in the deterministic case). But in the long run, “the force of the statistical law is that for a large number of instances it is correct to *expect* physical components in this class to be realized close to 32 percent of the time” (Pereboom, 2014, p. 66f.; italics in original). The problem is that it is “incredible” that this should come about through agents' free choices: “if the occurrence of these physical components were settled by the choices of agent-causes, then their actually being chosen close to 32 percent would amount to a [wild] coincidence” (Pereboom, 2014, p. 67). This problem is strikingly similar to Anscombe's description of a possible coordinated attack against a statistical law; see the passage “Certainly if we have ...” from p. 146 quoted in the main text above.

correlates. It is worth quoting her analogy in full. In her text, it follows as a new paragraph directly after the passage “But it is not [...] by the objection” quoted above.

Let me construct an analogy to illustrate this point. Suppose that we have a large glass box full of millions of extremely minute coloured particles, and the box is constantly shaken. Study of the box and particles leads to statistical laws, including laws for the random generation of small unit patches of uniform colour. Now the box is remarkable for also presenting the following phenomenon: the word “Coca-Cola” formed like a mosaic, can always be read when one looks at one of its sides. It is not always the same shape in the formation of letters, not always the same size or in the same position, it varies in its colours; but there it always is. It is not at all clear that those statistical laws concerning the random motion of the particles and their formation of small unit patches of colour would have to be supposed violated by the operation of a cause for this phenomenon which did not derive it from the statistical laws. (146)

This, then, is our task: Can we build such a box? I will show that we can. When describing an implementation in Sect. 4 below, I will just take some liberty with the example, most prominently in replacing the name of the Atlanta company⁹ by the graphically simpler and more appealing peace sign.

Before embarking on this task, it is necessary to analyse the analogy and to dismiss readings that threaten to trivialise it.

3.1 Making the analogy explicit

The context in which Anscombe provides her example makes it clear that “this point”, the point illustrated by the analogy, is the implausibility of the claim that “the same train of minute physical events should have to be the regular correlate of the same action” (146). The action, therefore, must correspond to the writing discernible on the sides of the box. Translated to my tea vs. coffee predicament, always seeing the same writing then corresponds to always taking tea, and never taking coffee. The possible variations of the writing (shape and formation of letters, size, position, colours) provide an analogy for the different ways of executing (the different possible tokenings of) that same action. Thus, starting with external factors, I can have my tea at 7:15 or at 7:30, I can fill the kettle with 200ml of water or with 250ml, I can use my otter cup or my fish cup to drink from, and so on. Facing the choice between tea and coffee, I can act straight away or hesitate for a while, I can reason that I had tea yesterday and it was good, or that I should drink less coffee, or make my tea on autopilot as it were. I can pour the water by raising my arm to an angle of 90° or 100°; I can wait for the kettle to shut off automatically or shut it off manually when I hear the water boil. All these different ways of having my tea will clearly correspond to different “train[s] of minute physical events”, involving different events in my frontopolar cortex, my

⁹ I wonder whether Anscombe chose the name to rhyme with the perhaps more fitting “mene tekell” as reported in the Book of Daniel 5:25. There, the proverbial writing on the wall is by the hand of God, certainly “a cause for this phenomenon which did not derive it from the statistical laws”. Be that as it may, Anscombe does not explicitly consider divine intervention in her paper, and neither will I.

motor cortex, my auditory cortex, and whatnot. And certainly, anyway, each day when I get up, my body is in a different state from the day before.

The complexity of the box, which contains “millions of extremely minute coloured particles”, is small when compared to the complexity of our bodies—the human brain alone contains on the order of 10^{11} neurons, plus a lot of other stuff, so Anscombe’s example is at least a million times less complex than our bodies. But fittingly, her example (the writing on the sides of the box) is also vastly less complex than, say, having a cup of tea. So overall that may be a reasonable down-scaling of complexity, both of the box and of the phenomenon of agency, in the interest of tractability.

While these aspects of the analogy are fairly straightforward, it is harder to get the other parts straight. The analogy to the indeterministic but statistical laws of physics would seem to be the statistical laws for the “random motion” of the coloured particles in the box—but Anscombe also mentions such laws for the particles’ “formation of small unit patches of colour”. On the physics side of things, Anscombe does not really mention a mesoscopic level that would correspond to the colour patches, but then she does not go into any detail about the “minute physical events” under consideration at all, and so we may assume that these also come in different, perhaps not very neatly separated layers of complexity.¹⁰

One thing that the text leaves somewhat open for interpretation, as far as I can see, is what corresponds to the shaking of the box. It seems that at least *prima facie*, there are two possible readings. One is that the shaking of the box is what drives the indeterminism in the dynamics of the particles (and, derivatively, of the patches). The underlying idea could be that the tiny particles, in a realistic implementation, will have to be relatively large when compared to quantum objects, so that their basic dynamics is mechanical and thus, perhaps even deterministic. The shaking of the box would then be what leads to the behaviour of the box having to be characterized by statistical laws. A somewhat different, perhaps more interesting reading is that the particles are subject to an underlying indeterministic dynamics to begin with, but *in addition*, the box is subject to uncontrollable outside forces. This latter reading may be more in line with the systematic role that the analogy plays. For one, Anscombe’s term “particles” suggests that the indeterministic laws of physics should correspond to the laws of interaction inside the box, prior to any external shaking. And second, the larger context of discussion here is an indeterministic notion of human or animal agency. All biological entities are physical entities, whose ultimate constituents obey the indeterministic laws of quantum mechanics, *and* all biological entities live in mostly unpredictable and largely uncontrollable environments. Animals *are*, literally, “constantly shaken”, subject to outside influences beyond their control.¹¹ In my morning routine, for example, I may be subject to the news on the radio, to the smell of the flowers on my table, to the changing view out my window, or to the noise

¹⁰ If one were to be serious about a mesoscopic layer, one might, for example, think of quantum dots, whose properties can be computed based on more fundamental quantum laws, but whose technological use shows that they can be considered as separate units. The laws governing these mesoscopic objects will inherit some of the indeterminism of their more fundamental base, often in the form of noticeable fluctuations around an average.

¹¹ Internal causes may also play a “shaking” role, e.g., when a conscious thought comes up through association. I thank an anonymous referee for pointing out this additional reading.

and vibrations of a passing train. So I will proceed on the assumption that Anscombe construes her box as subject to an indeterministic inner, inter-particle dynamics *plus* an additional outer “shaking” influence, which may correspond to the external cause that is responsible for the patterns on the sides.

The last part of the analogy we still have to come to terms with concerns the lesson to be drawn from the example of the box. Anscombe singles out as potentially problematic “a cause for this phenomenon” (the writing on the box, my choosing tea) “which did not derive it from the statistical laws”: it is *that type* of cause that stands behind the statistics objection whose failure the box is meant to illustrate. By way of contrast, we can read Anscombe as acknowledging that a cause for a high-level phenomenon that *is* derived from the low-level statistical laws does not pose a problem for those laws, and therefore cannot give rise to the statistics objection to begin with. This makes good sense: if a high-level law is derived in the sense of being a mere aggregate (e.g., a coarse-graining) of the lower-level dynamics, then of course a cause conforming to these aggregate laws can also operate in conformity with the lower-level laws. Thus, Anscombe must hold that the box illustrates the fact that there can be a high-level cause for a high-level phenomenon that operates in some relevant sense independently.¹² The operation of such a cause will lead to dynamics that may or may not still conform to given low-level statistical laws. In presenting her analogy as a case in which the low-level statistical laws remain intact, Anscombe aims to show that the statistics objection need not be taken seriously. Assume that a person’s will—my choice of having tea rather than coffee each morning—is such a cause: then Anscombe says that the minute physical events constituting the different tokenings of that same action of choosing tea will not threaten the statistics of the indeterministic micro-dynamics that constrain what goes on in my body and its surroundings.

3.2 A systematic lesson from the analogy, or easy ways out?

Anscombe offers no details about the supposed workings of her box, and she treads carefully when expressing her claims: she speaks of an “analogy”, and her conclusion is in the form of a double negation that contains two hedges (“[i]t is not at all clear that those statistical laws [...] would have to be supposed violated”, 146). As far as I can see, there has been no thorough discussion of the systematic force of her analogy in the literature either.¹³ In order to evaluate Anscombe’s rebuttal of the statistics objection,

¹² Anscombe does not use the terms “top-down causation” or “downward causation”, and I do not think that this indicates a defect, as these terms are often handled in a problematic way. Ellis (2016, Ch. 4) provides a thorough discussion of different kinds of inter-level dependencies in terms of top-down causation (see also Ellis, 2018). He states: “Because of the existence of random processes at the bottom, there is sufficient causal slack to allow this kind of top-down effect to occur without violation of physical causation” (Ellis, 2016, p. 236). This point is parallel to the one Anscombe is making. Ellis does not, however, explicitly address the statistics objection, and Anscombe is not mentioned in his book at all.

¹³ An early reference to Anscombe’s box appears in Kenny (1973, p. 101), who notes that the mere imaginability of the example doesn’t prove much; I agree. Kenny, however, seems to misclassify the example as an argument for incompatibilism, while Anscombe makes it clear that the example is meant merely to defend the incompatibilist against the statistics objection. Sheeler (1983, p. 296f.) presents several aspects of Anscombe’s analogy. He states that he “fail[s] to see how the will can be introduced into this analogy” because a box subject to the will should also allow for the word “Guinness” on its sides. Indeed, part of our

therefore, more work is needed in order to substantiate the analogy. That work will be done in Sect. 4.

There are, however, also two ways of reading Anscombe's analogy which turn it into something like a triviality. These readings can be assumed to miss the message that Anscombe is trying to convey, given that Anscombe takes the statistics objection seriously and must assume that she is really providing a helpful and nontrivial analogy. But I believe that it is important to discuss the easy ways out explicitly, as this will also help to sharpen our view for the real contribution. The two readings correspond to taking the analogy for granted without further ado, and to denying its possibility, respectively.

Easy 1: The box as a gift. In presenting her analogy, Anscombe just *gives* us the box. She says that she “construct[s] an analogy”, but there is no real construction going on; we are simply invited to “[s]uppose that we *have*” (146, my italics) such a box, and to find out about its features. What we find out, according to Anscombe's story, are facts about the box's micro-dynamics (“laws concerning the random motion of the particles and their formation of small unit patches of colour”) and about the macro-phenomenon of the writing on the sides of the box. Anscombe says that we find out about the micro-dynamics through the “[s]tudy of the box and particles”, so that the statistical laws are established from a study of the actual dynamics; and we simply *see* the writing when looking at the box from the outside. Nothing is said about the cause of the writing, but in her conclusion, Anscombe considers a cause that “did not derive it [i.e., the writing] from the statistical laws”. If we assume that this is also a feature of her box—i.e., if it is part of her illustration that the writing is caused independently from the micro-laws—then the box indeed provides a strong rebuttal of the statistics objection. An independent cause operates, and yet the statistical laws remain intact.

There are two problems with this easy way out. First, how can we be sure that Anscombe is really describing a scenario that is possible? How can we be sure that such a writing on the walls can really be due to an independent cause? The story can stipulate that this is so, of course, but this stipulation is unsupported.¹⁴ Second, if we agree with the stipulation of an independent cause, but go on to derive the statistical laws from the actual happenings inside the box, then it is trivial that that cause does not go against these laws, simply because the independent cause is operative when we

Footnote 13 continued

work in Sect. 4 will be devoted to allowing for some flexibility in the operating high-level cause. Ometto (2016, p. 200) addresses the example in its relation to the multiple realizability of human actions. He holds that “Anscombe is not saying that such a box could actually exist”; and indeed, Anscombe does hedge her presentation of the analogy, as we have seen. Mulder (2014, p. 277f.) follows the example in more detail and points out its importance as an illustration of the relative independence of higher-level phenomena (e.g., life processes) from their lower-level (e.g., chemical-material) basis, connecting back to Anscombe's “rules of chess” characterization of indeterministic laws of nature. Anscombe's analogy is also mentioned in the following works (no claim of completeness is implied): Corcoran (2013, p. 273) refers to the example briefly and credits Runyan with bringing it to his attention (p. 281, n. 8). Runyan (2014, p. 255, n. 39) merely reports that Anscombe has successfully rejected the statistics objection. Zeis (2016, p. 255f.) also cites Anscombe's example approvingly without further analysis.

¹⁴ This appears to be Kenny's point; see note 13.

study the box and its constituents to establish these laws.¹⁵ Proceeding in this way, it is rather that the statistical laws are derived from the external cause! So all we are left with is an unsupported stipulation leading to a trivial conclusion. That cannot be what is meant.

Easy 2: Causal closure. The second easy way out is even simpler: If we subscribe to the idea of the causal closure of the micro-physical, then we can be sure that there *cannot be* a cause that does not derive the phenomenon of the writing on the sides of the box from the micro-dynamics. On that view, the question of whether the operation of such a high-level cause can violate any laws simply does not arise. If the box is real, then all of its features must be based on the micro-dynamics, and that's all.

This may be a consistent position (even though I doubt it), but holding it means that the point of the analogy is lost, and that the illustration has no force against the statistics objection. That objection starts by assuming that a person's will can be viewed as an independent cause, and claims that a coordinated operation of that cause can threaten the low-level statistics. Saying that the will can be no such cause may be a way out for some indeterminist friends of freedom, but not for Anscombe, who holds that "freedom at least involves the power of acting according to an idea" (146), which she says is nowhere to be found in indeterministic physics.

Not so easy: Drawing a systematic lesson. Rejecting these two easy ways of dealing with Anscombe's illustration helps to get clear on what would need to be done in order to strengthen her case against the statistics objection. It will not do to reject the idea of an external cause out of hand: that would mean to reject a premise of the statistics objection that Anscombe accepts. And neither will it do to have that external cause always operative when we establish the micro-laws: that would mean to trivialize the fact that the micro-laws are vindicated. So, in order for Anscombe's analogy to have force against the statistics objection, we have to have an external cause that can be present or absent (writing or no writing; tea or coffee¹⁶), and we have to study the box in some average regime where the cause sometimes operates and sometimes doesn't. If we derive statistical micro-laws in this way, the statistics objection has real bite: when we go on to study the micro-dynamics *in a situation in which the external cause always operates* (when we study the box in the regime where the writing is always there, as described in Anscombe's illustration; when I always choose to have tea), it could very well be that the previously established micro-statistics are violated. Anscombe's claim in presenting her analogy is that *this need not happen*.

Anscombe's point is, then, that one can *construct* (her wording on p. 146) a box that illustrates a failure of the statistics objection despite being a reasonable analogy of free agency. Her claim is not that there cannot be other boxes that do fall prey to the statistics objection. Indeed, it is relatively easy to construct a box on which an external cause operates in such a way that the micro-statistics do change, even though that box "is constantly shaken" (see Sect. 4.2). The real work thus lies in actually constructing

¹⁵ Baker, following the integrative approach of Clarke (2003), seems to be happy to endorse this option, saying that on the agent causalist's view, "the agent-cause is *already part of* the observational evidence" (Baker, 2017, p. 3092, italics in original).

¹⁶ Or: "Coca-Cola" or "Guinness"; see Sheeler's point discussed in note 13.

a box of the type that Anscombe targets with her illustration. I will attempt such a construction in Sect. 4.3.

4 “Let’s build an Anscombe box”

Having set out the task above, I go on to describe a model of an Anscombe box that is implemented as a fairly simple Python program.¹⁷ I start with the core features of the model, which underlie all extensions and which thus provide the basis for the micro-statistics absent any external causes (Sect. 4.1). In Sect. 4.2 I describe an external cause that leads to trouble vis-à-vis the statistics objection. Finally, in Sect. 4.3 I describe a pattern-generating external cause operating on the box in the way envisaged by Anscombe, i.e., one whose operation does not fall prey to the statistics objection. I keep the discussion at a fairly abstract level, relegating a fuller description of the model, including commented pseudo code, to Appendix A.

4.1 Basics of the Anscombe box

The basic framework for the model of Anscombe’s analogy to be discussed here is a box consisting of $N \times N \times N$ cells, where each cell is considered to be the locus of one of Anscombe’s coloured particles. For simplicity’s sake, my illustration will consider two colours only. Initially, the box is filled with a random distribution of the coloured particles, thus assigning one colour to each cell in the box. (In all the examples described here, $N = 50$, and the ratio of the colours is roughly 1 to 4.) The stochastic micro-dynamics is modeled in discrete time steps. At each time step, two neighbouring cells are picked at random, and then it is decided by a stochastic rule whether the particles in these two cells are exchanged or not. At the next time step, again two cells are picked at random, uncorrelated to the previous pair, it is again stochastically decided whether the corresponding particles are exchanged or not, and so on.¹⁸ Figure 1 shows a typical view of the six sides of the box (top) and a histogram of the number of cells that have witnessed a given number of exchanges. The dynamics is purely internal: no external cause is active, and while “small unit patches of uniform colour” (146) are indeed formed randomly, the dynamics typically creates no discernible patterns.¹⁹

What are the statistical laws? With a view to studying alterations of the basic dynamics via the action of an additional external cause, it is necessary to specify the success criterion for Anscombe’s analogy, i.e., to spell out what it means that the “statistical laws” remain unaltered. What exactly are those laws? Anscombe, as we have seen, mentions at least three levels of complexity: (1) the lowest level of the random motion

¹⁷ Code available from the author upon request.

¹⁸ This model thus belongs to the class of probabilistic cellular automata as described, e.g., by Deutsch and Dormann (2017, Ch. 5).

¹⁹ Of course, *any* pattern may form by chance—it is just that the probability for that to happen is very small. See note 30 for a rough estimate according to which the probability for a large discernible peace sign forming on a side of the box by chance alone is of the order of 10^{-60} , very close to zero indeed.

Fig. 1 Basic Anscombe box: final view of the six sides (top) and histogram showing on the y axis how many of the $50 \times 50 \times 50$ cells in the box have witnessed the number of exchanges given on the x-axis during the 10^7 time steps. Thus, e.g., around 6500 cells in the box witnessed 60 exchanges. The total number of exchanges was 7,524,081 (75.2%).

of the individual particles, (2) an intermediate level of the formation of small patches of uniform colour, and (3) the highest level of the patterns visible on the outside of the box. The basic dynamics just sketched specifies (1) via a uniform rule for exchanging the particles in randomly chosen adjacent cells. (The basic rule is, roughly, to draw a

random number between 0 and 1 and to do the exchange if that number is less than 0.75.²⁰) No further specifications are given for levels 2 or 3, so that what happens at these levels is just a coarse-graining of what happens at level 1.

The effect of the level 1 rule can be conveniently summarized in two ways (see Fig. 1): first, via the total number of exchanges that occur during the simulation, and second, via the distribution of the number of cells that witness a given number of exchanges. I have chosen a histogram representation of that distribution for easy visual inspection, as this helps to highlight the differences between the sorting cause of Sect. 4.2 and the pattern-generating cause of Sect. 4.3. The success criterion I will handle in the comparison is that (a) the total number of exchanges and (b) the distribution of these exchanges over the individual cells of the box remain intact.

At this point, however, an important objection can be raised.²¹ Isn't the case for retaining the statistical level 1 laws hopeless, on conceptual grounds? After all, an external cause that makes a difference to what happens at level 3—reliably creating a discernible pattern—must do this by somehow influencing how the particles move: a difference at level 3 must be grounded in some difference in level 1. The particles and their motion are, after all, all there is; the simulated box consists of nothing but the particles. And if these particles show patterns that they do not show under the basic level 1 dynamics, this proves that their level 1 dynamics must be different from the basic case. There must, at a minimum, be novel level 1 correlations that are not present in the wholly uncorrelated basic dynamics.

This is true: In the simulation, if there is no difference at level 1, there can be no difference at level 3.²² The open question is when such level 1 differences count as a violation of the “statistical laws”. On the criterion offered above, as long as the exchange probability and the distribution of exchanges remain intact, the laws are preserved. But I admit that there must be other differences at level 1, otherwise, there could be no differences at level 3. I offer two considerations that, I think, are relevant and which support my criterion. First, if the only way of arguing for a difference in the level 1 dynamics is indirect, from the observed level 3 differences, then I hold that no differences in level 1 laws have yet been found. Such differences should be attested at the level of the motion of the particles directly.²³ Second, the abstractness and purity of the model, which simulates the box as a closed system, actually makes it harder to retain the low-level laws. A real box, as envisaged by Anscombe, must have parts in addition to the particles—for example, the glass walls, and some physical mechanism

²⁰ See Appendix A for pseudo code providing more details.

²¹ I thank an anonymous referee for pressing this objection. The referee also provides a helpful illustration in terms of a 1-dimensional model of a bit string subject to random coin-flip dynamics overlaid by a pattern-generating dynamics that keeps the marginal probabilities (the proper analogue of the number and distribution of exchanges in my model) intact. For that bit string model, indeed, it would be correct to say that the pattern-generating interventions amount to a change in the micro-dynamics. The difference with my case is that in the bit string model, one can easily express the resulting correlations at the micro-level, while I argue that here, such correlations can only be inferred indirectly.

²² While I am not convinced that the notion is always helpful, here we have a clear case in which one can say that the occurrence of the level 3 patterns *supervenes* on the level 1 dynamics.

²³ As shown in Sect. 4.3, in the model offered here, the motion of the particles *is* affected by pattern-building locally (after all, otherwise the patterns would not appear), but as the pattern-building sites are distributed randomly, in the long run, any such effects on the particles' motion wash out.

for moving the particles about. (Likewise, human beings consist of much more than neurons and their interconnections; we are real agents with sensors and actuators.) Any real system is an open system, coupled to an unpredictable and uncontrollable environment. (Recall Anscombe's point about the box being "constantly shaken".) In an open system, it is impossible to keep track of all the correlations, as these dissipate into the environment. So dialectically, it is admissible to use a less stringent criterion of sameness of statistical laws for a closed system than for an open system, and this provides additional support for my success criterion of the constancy of the overall number and distribution of exchanges.

4.2 The sorting box: victim to the statistics objection

The first external cause to be introduced here influences the exchange of neighbouring particles in the following fashion. For a given pair of adjacent particles of different colours, if one of them has a higher x coordinate, then the exchange goes through exactly if it leads to the darker colour coming out on top.²⁴ The overall effect of that external cause is, therefore, to sort particles by colour along the x direction, so that the initially randomly mixed box tends towards two fractions of different colours whose boundary becomes sharper over time. Figure 2 shows a typical long-term view of the six sides of the box (top) and a histogram of the number of exchanges that a given cell has been subject to. The fundamental dynamics and the dynamics of the creation of "small unit patches of uniform colour" has obviously been changed: these patches become much larger and more cohesive. Also, the rate of exchanges is different from the base case. Inspection of the histogram reveals another difference: the histogram is asymmetric, showing a stronger prevalence of smaller exchange rates. These are due to the fact that at the boundary, over time, there is a fairly stable sorting of colours along the x direction that prevents exchanges via the rules of the external cause.

This external cause is of the type targeted by the statistics objection: its operation leads to an infringement of the box's micro-dynamical laws, as it alters the frequency and distribution of exchanges of neighbouring particles. If the will were such an external cause on our bodies, then we would be able to violate the micro-statistical laws of physics by long-term coordinated acts of the will.

4.3 The peace box: vindicating Anscombe's analogy

In order to vindicate Anscombe's analogy, I introduce an external cause that generates patterns at a number of sites on the sides of the box.²⁵ That cause interferes with the dynamics locally at the sites at which the patterns are built, but since these sites vary randomly over time (they are selected by profiting from stochastic fluctuations that produce partial patterns at random locations), in the overall statistics there is no discernible deviation from the base case. This can be checked by inspecting the histogram and comparing it with that of the base case. Figure 3 shows a typical view

²⁴ See Appendix A for details including pseudo code for this rule.

²⁵ See Appendix A for details including pseudo code for the pattern-generating rule and for the selection of the pattern-building sites.

Fig. 2 Anscombe box with sorting cause active: final view of the six sides (top) and histogram showing on the y axis how many of the $50 \times 50 \times 50$ cells in the box have witnessed the number of exchanges given on the x -axis during the 10^7 time steps. Thus, e.g., around 6500 cells in the box witnessed 50 exchanges, and around 3500 cells witnessed 60 exchanges. The total number of exchanges was 6,506,518 (65.1%).

of the six sides of the box (top) and a histogram of the number of exchanges that a given cell has been subject to. During the simulation, peace sign patterns (or other patterns that can be selected in the program) are visible practically all the time.

The six faces of the Anscombe Box

Fig. 3 Anscombe box with pattern-generating cause active: final view of the six sides (top) and histogram showing on the y axis how many of the $50 \times 50 \times 50$ cells in the box have witnessed the number of exchanges given on the x -axis during the 10^7 time steps. Thus, e.g., around 6500 cells in the box witnessed 60 exchanges. The total number of exchanges was 7,522,410 (75.2%). A total of 8 patterns were completed during the simulation.

This example shows that it is possible to construct a box of the type Anscombe envisages in her analogy: a pattern-generating external cause influences the motion of the particles, which remain subject to stochastic micro-dynamics, in such a way that a

discernible pattern is created on the sides of the box while the micro-statistics remain intact. The stochastic micro-dynamics provide the leeway for the external cause to build macroscopic patterns without thereby violating the laws at the micro-level. The multiple realizability of the macroscopic patterns, together with the random nature of the actual realizations, has the effect that the minute interventions of the external cause in the generation of the patterns are washed out over time, leaving no discernible trace in the basic micro-statistics.

5 Conclusion

Anscombe's classic "Causality and Determination" offers manifold inspiration. In this paper the focus has been on the Anscombe box analogy of free agency, which has inspired me to simulate such a box via a computer program. It turns out that Anscombe's point behind her analogy can be substantiated by concrete simulation results: Given underlying stochastic dynamics subject to statistical laws, it is possible to implement a separate, external cause (analogous, perhaps, to the human will) whose effect leads to the creation of discernible, sensible patterns while at the same time not violating the constraints of the basic micro-statistics.

Such an external cause has to strike a fine balance in using the underlying micro-indeterminism in the creation of the patterns, as external causes are not guaranteed to act in accordance with the micro-statistics. This is shown by the example of the sorting box of Sect. 4.2, where an external cause generates a pattern in a way that violates the micro-statistics. The statistics objection, that is, has real bite. It is therefore significant that it is possible to construct a model in which an external cause maintains the fine balance between engaging enough of the micro-indeterministic events for the benefit of pattern creation while maintaining the micro-statistics. This is shown by the example of the peace box of Sect. 4.3.

The simulation results reported here may also shed some light on Pereboom's "wild coincidences" objection to agent-causal libertarianism mentioned in notes 3, 7, and 8 (Pereboom, 2001, 2014, pp. 65–69) and the ensuing recent discussion (e.g., Baker, 2017; Runyan, 2018; Taggart, 2021), which unfortunately does not refer to Anscombe's box analogy. At least *prima facie*, the stability of the simulated micro-statistics under the action of an external pattern-generating cause suggests that it may be possible to avoid Pereboom's objection. A thorough evaluation must, however, be deferred to another paper.

Relatedly, the simulations do not show in any way that the human will has to be understood as an external cause in Anscombe's sense. There are various views of how our causal ability to act on the basis of reasons can be understood, and many empirical and conceptual questions remain open. The picture of a rational will operating on the body from the outside, as a wholly separate, external cause (suggestive perhaps of certain variants of dualism) may not be the best way to understand the independent and separate status of the will vis-à-vis our bodies' microphysical base. If, for example, our distinct causal power of having a rational will emerges from the underlying micro-dynamics, then the statistics objection is avoided on conceptual grounds, as an agent-cause is then integrated into the physical happenings. In terms of the box analogy,

this would mean that the stochastic dynamics in the box themselves give rise to the patterns. This would, however, require the box to have a structured inside, rather than the homogeneous, unstructured inside of the boxes discussed here.

Certainly a lot needs to be done to explore these other options for understanding free agency under indeterminism. I believe that for these other approaches, too, concrete modeling will be beneficial, but this must be left for another occasion. The dialectical point of the exercise here has only been to show that even if one subscribes to an extreme view of the will as an external cause operating on the body from the outside, which is at least suggested as an option in Anscombe's text, then statistical laws at the micro-level can still leave room for our free agency.

Acknowledgements Support by DFG Grant MU1816/8-1 and by VolkswagenStiftung Grant Az:97721 is gratefully acknowledged.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Details of the Anscombe box implementation

In this appendix I provide technical details that develop and substantiate claims made above in Sect. 4.²⁶

The box consists of $N \times N \times N$ discrete cells, each containing one particle of a specific colour. In the illustrative examples in the main text, $N = 50$, resulting in a total number of cells $N_{Cells} = 50^3 = 125,000$, and there are just two colours, purple (0) and yellow (1). Simulations run for $N_{Steps} = 10^7$ discrete time steps. At each step, the occupants of one pair of neighbouring cells can be exchanged. We consider three scenarios that differ in the rule for whether that exchange takes place: the basic dynamics (D1; Sect. 4.1), the sorting dynamics showing that the statistics objection has real bite (D2; Sect. 4.2), and the peace sign dynamics that vindicate Anscombe's analogy (D3; Sect. 4.3).

In order to keep things simple, we deviate from Anscombe's example in that we employ only a single pattern of fixed size $n \times n$ (in our examples, $n = 21$), orientation (no rotation), and colour (yellow). It is easy to extend the simulations to a number of patterns of different sizes, orientations, and colours by using a list of bitmaps in place of a single bitmap p as described here.²⁷

²⁶ I thank an anonymous referee for prompting me to provide more details via pseudo code.

²⁷ The sizes and orientations as well as the colours can be chosen randomly. The size n has to be reasonable for the given side length of the box N . From experience, n should be between 7 and $N/2$. Sizes of $n < 7$ are useless in that no pattern can be discerned, and patterns larger than $N/2$ imply that only a single pattern can fit on any surface of the box without overlap.

The pattern initialization step is independent of the chosen dynamics, and goes as follows:²⁸

Let p be an $n \times n$ bitmap showing a peace sign

Let r be the fraction of 1s in p , i.e., $r = \frac{\# \text{ 1s in } p}{n \cdot n}$

Given r , we now put 1s into the appropriate number of randomly chosen²⁹ cells to initialize the box to match the colour distribution of the pattern:

Initialize the $N \times N \times N$ box as an array $B[\cdot, \cdot, \cdot]$ of all 0s

Let $i = 0$

While $i < r \cdot N^3$

 Choose random x, y and z (independently) from $\{0, \dots, N - 1\}$

 If $B[x, y, z] == 0$

 Let $B[x, y, z] = 1$

 Let $i = i + 1$

This set-up results in a box in which 0s and 1s are randomly distributed, showing no correlations, and such that the fraction of 1s corresponds to the fraction of 1s in the peace sign pattern, which is roughly $r = 0.2$.

Now it is possible, but extremely unlikely, that such a randomly filled box will already show a peace-sign pattern on one of its sides. While that pattern has the typical ratio of 0s and 1s, it is atypical in showing spatial correlations among the colour patches—in fact, these correlations are what the pattern *is*, and by which we recognize it.³⁰ It is, however, to be expected that certain patches on the sides of the randomly initialized box will go some way towards resembling the pattern. Those that provide the best fit will be used as the initial attractors for the pattern-building external cause. As a very simple measure of goodness of fit, we can compute the goodness of an $n \times n$ bitmap b from one of the sides of the box for the $n \times n$ bitmap p to be

$$\text{goodness}(b, p) = \frac{\# \text{ of matching 0s and 1s}}{n \cdot n}.$$

²⁸ In case the dynamics D1 or D2 are selected, the pattern p will not be used, but it is important to fix the ratio of 0s and 1s to be the same in all cases for comparison. This could, in these cases, of course also be done by hand.

²⁹ A brief remark on terminology: by a random choice, in accordance with established technical usage, I mean a sample drawn from a uniform distribution. Thus, to choose a number randomly from $\{1, 2, 3\}$ means to choose one of the numbers 1, 2, and 3, with equal probability $1/3$ for each. In philosophy, “random” is sometimes contrasted with “probabilistic” with the meaning that a random choice has no probabilities assigned. According to the technical usage I follow in this paper, random choices are specific probabilistic choices, viz., with uniform probabilities.

³⁰ To give a very rough calculation for the probability of finding the 21×21 peace sign pattern on the outside of a $50 \times 50 \times 50$ box: The pattern involves 93 yellow patches, the probability for any given cell in the box to be yellow is around 0.2, so the probability for 93 given cells to be yellow is $0.2^{93} \approx 10^{-64}$. Now there are 6 sides each offering $29 \cdot 29$ positions for placing a 21×21 pattern, so there are approximately 10^4 ways of placing a pattern. This results in a probability of roughly 10^{-60} for finding the peace sign pattern on the sides of the randomly initialized box. (Here we ignore the fact that the other $21^2 - 93 = 348$ cells in the pattern have to be purple. This is roughly compensated for by the fact that we recognize imperfect patterns and that we ignore rotations, so that there are many more ways of producing a recognizable peace sign.)

Fig. 4 **a** The peace sign pattern; **b** a random distribution of colours selected as a site for growing a pattern ($goodness = 0.728$); **c** a pattern that is considered good enough ($goodness = 0.982$ exceeds threshold of 0.98)

In our simulations, this goodness-of-fit measure is used for two purposes: to select sites for pattern-building, and to determine when a pattern is good enough to be recognizable. Figure 4 shows, for $n = 21$, the peace sign pattern (a), a typical first site for pattern building from the randomly initialized box (b; $goodness = 0.728$), and an exemplary pattern that in our simulations is considered good enough to be considered complete (c; $goodness = 0.982$; our threshold is 0.98).

The list of initial sites for pattern building is constructed as follows (in our simulations, $N_{Sites} = 6$ and $N_{Trials} = 100$):

```

Let Sites be the empty list
For i from 0 to  $N_{Sites}$ 
  Let  $best\_site = None$ ,  $best\_goodness = 0$ 
  For j from 0 to  $N_{Trials}$ 
    Pick a random site s on one of the sides of the box
    If s collides with a site in Sites, let  $g = 0$ 
    Else let  $g = goodness(B[s], p)$ 31
    If  $g > best\_goodness$ 
      Let  $best\_goodness = g$ ,  $best\_site = s$ 
  Append  $best\_site$  to Sites

```

This step (which is, of course, only needed for the pattern-producing external cause) results in a list of sites at which some of the work towards building the wanted pattern has already been taken care of by random fluctuations in the initialization step.

The simulation main loop then is the following, for all the three dynamics *dyn* (D1, D2, or D3):

```

Initialize the pattern p, box B, list Sites, and the interactive plot
For i from 0 to  $N_{Steps}$ 
  Pick a random cell  $\langle x_1, y_1, z_1 \rangle$  and a direct neighbour  $\langle x_2, y_2, z_2 \rangle$ 
  Invoke  $try\_exchange(dyn, \langle x_1, y_1, z_1 \rangle, \langle x_2, y_2, z_2 \rangle)$ 
  At every 50,000th step, update the interactive plot
Output the final configuration and statistics

```

³¹ Here, $B[s]$ is the $n \times n$ bitmap from the side of the box specified by *s* and from the location on that side that is also specified by *s*.

The different dynamics give rise to a case distinction in the function *try_exchange*, which determines whether to exchange the given particles. There are two global parameters that influence the exchange: *prob_exchange* (set to a fairly high value of 0.75 in our simulations) is the probability that particles will be exchanged in the basic dynamics, and *prob_anyway* (set to a low value of 0.01 in our simulations) is the probability that particles will be exchanged come what may, even, so to speak, against the advice of the external cause. (Setting these probabilities to values different from the extremal values, 1 and 0, respectively, is our way of acknowledging that the box is “constantly shaken”.) Irrespective of the chosen dynamics, the function *try_exchange* starts as follows:

Let a be randomly selected from the interval $[0, 1]$
 If $a < \textit{prob_anyway}$ then do the exchange and return

For $\textit{dyn} = \text{D1}$ (the basic dynamics without an external cause), the function *try_exchange* then continues simply like this:

Let b be randomly selected from the interval $[0, 1]$
 If $b < \textit{prob_exchange}$ then do the exchange
 Else don't do the exchange

For $\textit{dyn} = \text{D2}$ (the sorting cause), the function *try_exchange* is a bit more involved:

Let c_1, c_2 be the colours of the particles in the given cells
 Let b be randomly selected from the interval $[0, 1]$
 If $c_1 == c_2$ and $b < \textit{prob_exchange}$ then do the exchange and return
 If $c_1 == 0$ and coordinate $x_1 \leq x_2$ then do the exchange
 Else don't do the exchange

This function thus prescribes that for cells of the same colour, the particles are exchanged with the probability *prob_exchange* of the basic dynamics. Otherwise, the particles are exchanged exactly if that leads to a particle with colour 0 percolating (weakly) upwards in x -direction. As can be seen from Fig. 2, this results in the yellow (colour 1) particles getting stuck near the $x = 0$ surface.

Finally, for $\textit{dyn} = \text{D3}$ (the pattern-generating external cause), we need to differentiate what happens at the surface of the box from what happens inside, and we need to handle some bookkeeping about the sites at which patterns are grown, so the function *try_exchange* becomes a little more complex:

Let b be randomly selected from the interval $[0, 1]$
 If both cells are in the interior of the box then
 If $b < \textit{prob_exchange}$ then do the exchange
 Else don't do the exchange
 return
 If none of $\langle x_1, y_1, z_1 \rangle$ and $\langle x_2, y_2, z_2 \rangle$ fall in a patch s in *Sites*
 If $b < \textit{prob_exchange}$ then do the exchange³²
 Else don't do the exchange

³² Here a different probability can be chosen if—unlike in our case—the pattern building leads to a different exchange rate at the surface, thus compensating for the action of the external cause.

```

return
# Continuing, we know a pattern at site  $s$  is involved
If the pattern at  $s$  will be improved by the exchange then
  Do the exchange
  If the goodness of the pattern at  $s$  is now good enough
    Delete  $s$  from the list Sites
    Find a new site  $s'$  to grow a pattern and add  $s'$  to Sites
  Else don't do the exchange

```

As this pseudo code shows, there is only one occasion on which this pattern-building external cause influences whether an exchange takes place: when two particles are such that exchanging them makes a difference to a patch on the surface of the box that is listed as one of the *Site* areas at which patterns are being built. In that case, the particles are exchanged exactly if that improves the fit of the pattern in the area with the target pattern. Once a pattern has been built to be sufficiently good, it is abandoned to the basic dynamics, meaning that it will dissolve over time. As a replacement, a new site for pattern generation is chosen, taking advantage of already existing, randomly occurring similarities with the target pattern, exactly as in the initialization step above. The sites for pattern generation thus fluctuate randomly across the surface of the box, so that in the long run, no site on the outside of the box is preferred. Also, note that the pattern-building external cause only influences particle exchanges that involve the surface layer of the box, blocking those that negatively influence a pattern that is growing there. For all other exchanges, the basic dynamics remains completely unchanged.

References

- Adlam, E. C. (2018). Quantum mechanics and global determinism. *Quanta*, 7(1), 40–53.
- Anscombe, G. E. M. (1971). *Causality and Determination*. Cambridge University Press, Cambridge. Page references are to the reprint in *Collected Philosophical Papers*, Vol. 2, Oxford 1981, pp 133–147.
- Baker, J. (2017). Rejecting Pereboom's empirical objection to agent-causal libertarianism. *Synthese*, 194(8), 3085–3100.
- Clarke, R. (2003). *Libertarian accounts of free will*. Oxford: Oxford University Press.
- Corcoran, K. (2013). Human persons are material only. In J. P. Moreland, C. Meister, & K. A. Sweis (Eds.), *Debating Christian Theism* (pp. 270–282). Oxford: Oxford University Press.
- Deutsch, A., & Dormann, S. (2017). *Cellular automaton modeling of biological pattern formation. Characterization, examples, and analysis* (2nd ed.). Birkhäuser.
- Ellis, G. F. R. (2016). *How can physics underlie the mind? Top-down causation in the human context*. Berlin: Springer.
- Ellis, G. F. R. (2018). Commentary: Top-down causation and quantum physics. *Proceedings of the National Academy of Sciences of the USA*, 115(46), 11661–11663.
- Kane, R. (1998). *The significance of free will*. Oxford: Oxford University Press.
- Kenny, A. (1973). Freedom, spontaneity, and indifference. In T. Honderich (Ed.), *Essays on freedom of action* (pp. 89–104). Routledge.
- Mulder, J. M. (2014). Conceptual Realism. The Structure of Metaphysical Thought. Ph.D. thesis, Utrecht University. Department of Philosophy, Quaestiones Infnitae, Vol. 81.
- Müller, T., & Briegel, H. J. (2018). A stochastic process model for free agency under indeterminism. *Dialectica*, 72(2), 219–252.
- Ometto, D. (2016). Freedom and Self-Knowledge. Ph.D. thesis, Utrecht University. Department of Philosophy, Quaestiones Infnitae, Vol. 95.

- Palmer, D. (Ed.). (2014). *Libertarian free will: Contemporary debates*. Oxford University Press.
- Pereboom, D. (2001). *Living without free will*. Cambridge University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford University Press.
- Runyan, J. D. (2014). *Human agency and neural causes*. Palgrave Macmillan.
- Runyan, J. D. (2018). Agent-causal libertarianism, statistical neural laws and wild coincidences. *Synthese*, 195, 4563–4580.
- Sheeler, R. (1983). Essentialism and human nature. Ph.D. thesis, Bedford College, The University of London.
- Taggart, C. P. (2021). A response to the problem of wild coincidences. *Synthese*, 198, 11421–11435.
- Van Inwagen, P. (1983). *An essay on free will*. Oxford: Oxford University Press.
- Widerker, D., & Schnall, I. (2015). On the luck objection to libertarianism. In A. Buckareff, C. Moya, & S. Rosell (Eds.), *Agency, freedom, and moral responsibility* (pp. 94–115). Palgrave Macmillan.
- Zeis, J. (2016). Anscombe and the metaphysics of human action. *American Catholic Philosophical Quarterly*, 90(2), 249–262.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.