Causal Interpretations of Probability[1]

Wolfgang Pietsch, Munich Center for Technology in Society, Technische Universität München, Arcisstr. 21, 80333 München, Germany

The prospects of a causal interpretation of probability are examined. Various accounts both from the history of scientific method and from recent developments in the tradition of the method of arbitrary functions, in particular by Strevens, Rosenthal, and Abrams, are briefly introduced and assessed. I then present a specific account of causal probability with the following features: (i) First, the link between causal probability and a particular account of induction and causation is established, namely eliminative induction and the related difference-making account of causation in the tradition of Bacon, Herschel, and Mill. (ii) Second, it is shown how a causal approach is useful beyond applications of the method of arbitrary functions and is able to deal with various shades of both ontic and epistemic probabilities. (iii) Furthermore, I clarify the notion of causal symmetry as a central element of an objective version of the principle of indifference and relate probabilistic independence to causal irrelevance. According to the proposed account, probability distributions are interpreted in terms of causal symmetries in the circumstances rather than relative frequencies.

[1] pietsch@cvl-a.tum.de

1. Introduction

Research on probabilistic causality has been a thriving enterprise since about the 1980s addressing the mainly methodological question how causality can be inferred from statistical data. By contrast, this article is about causal probability, i.e. the conceptual question how probability can be integrated into a general framework of induction and causation.

In recent discussions on the foundations of probability, a novel class of objective interpretations has been proposed that is distinct from the more familiar propensity and frequency accounts (Strevens 2006, 2011; Rosenthal 2010, 2012; Abrams 2012). The interpretations essentially stand in the tradition of an approach by 19th-century methodologist Johannes von Kries and of related work on the method of arbitrary functions. For reasons that will soon become clear, I subsume these and similar approaches under the notion of causal probability. Two common features are particularly important: (i) First, causal interpretations replace or supplement the principle of insufficient reason by an objective version of the principle of indifference[2] that refers to physical or causal symmetries. This distinguishes causal interpretations both from frequentist approaches, which exclusively refer to relative frequencies as fundamental evidence for probabilities, and from logical accounts, which base probabilities on ignorance via the principle of insufficient reason, i.e. a purely epistemic reading of the principle of indifference. As we will see, the objective variant of the principle of indifference is not troubled by the central objections brought forth against the principle of insufficient reason, in particular the ambiguities in its application called Bertrand's paradox. (ii) Second, causal interpretations employ a notion of probability in terms of the ratio between favorable *conditions* and all conditions. This is another subtle but crucial difference to frequency interpretations which define probability in terms of the ratio between the number of *events* leading to a certain outcome and the total number of events. As will be shown in

[2] A note on terminology: In the following the term 'principle of indifference' will be used to refer to both an epistemic version, called 'principle of insufficient reason', and an objective version, called 'principle of causal symmetry'.

Section 3, rendering probability relative to the conditions determining an ensemble or collective provides for a simple solution to a specific version of the reference class problem.

Note that propensity interpretations also frame probability in terms of circumstances or conditions and they sometimes make the link to causation. In fact, the proposed account owes in many ways to various versions of the propensity interpretation, but it also differs in important respects. First of all, propensity accounts rely on a distinct ontological category, namely propensities in the sense of tendencies or dispositions. The relation with causality is not always clarified, but if it is, as in Karl Popper's later work, then propensities are often considered to be more general than causation. By contrast, the causal interpretation presented in Sections 3 to 6 tries to situate probability within a framework of causal reasoning. While propensity accounts focus conceptually on dispositions or tendencies and rather casually remark upon the parallel with causation, the interpretation proposed here starts with a detailed and specific account of causation and then examines how probability fits into the picture. Furthermore, a number of concepts are central to the causal approach that are not usually evoked in the exposition of propensity interpretations, in particular the notion of causal symmetry leading to an objective version of the principle of indifference (Section 4) and the causal construal of probabilistic independence based on judgments of causal irrelevance (Section 5).

In Section 2, I discuss various proponents of a causal[3] approach to probability from the 19th century as well as more recent developments in the tradition of the method of arbitrary functions. The latter are mainly due to Michael Strevens, Jacob Rosenthal, and Marshall Abrams, and are henceforth abbreviated as SRA-approach. I briefly indicate how causal probability resolves several objections against other interpretations of probability, e.g. the problem of distinguishing between accidental and necessary relations in the frequentist approach, or problems regarding the principle of indifference in the logical approach. I then point out some shortcomings of the SRA-approach. Besides some technical difficulties, it makes no connection with a general framework of induction and causation. Also, it cannot handle indeterminism and epistemic probabilities. Later in the article, I suggest how causal probability can deal with these issues.

Starting from Section 3, I will develop a specific account of causal probability, according to which probabilities are understood as degrees or grades of causal determination of a phenomenon by a given set of circumstances or conditions, which can be considered both in direction from causes to effects and vice versa to avoid Humphreys' paradox. First, two fundamental inductive frameworks are outlined, enumerative and eliminative induction. For each, I show how probability can be integrated. Enumerative induction leads to a naïve frequency view of probability, which suffers from the familiar problems, in particular that it cannot distinguish between law-like and accidental frequencies. Eliminative induction resolves this issue by carefully keeping track of all conditions under which a phenomenon happens. The corresponding account of probability, which carefully distinguishes different types of conditions, is termed causal probability. What I will call the *collective conditions*

[3] Some do not explicitly use the term causality, but instead refer to nomic dependencies. Many of the central ideas and concepts nevertheless remain the same.

determine the possibility space of a probabilistic phenomenon, i.e. all possible outcomes. The outcomes are categorized and the classes are labeled, where the labels are called *attributes*[4]. The *range conditions* (together with the collective conditions) then determine exactly which of the attributes occurs, at least in deterministic contexts. While the collective conditions remain constant for a probabilistic phenomenon, the range conditions will vary. A *measure* over the input space, spanned by the range conditions, is also fixed by the collective conditions, more exactly by symmetries in the causal set-up. One could say that symmetrism replaces frequentism. Via the mathematical theorem, called the law of large numbers, the measure denotes the limiting relative frequency with which the different input states are instantiated. *Causal probability then is calculated as the fraction of outcome states, weighted with the measure, that pertain to a certain attribute.* Rendering probability relative to collective conditions and measure resolves the mentioned technical problems of the SRA-approach while adding an irreducible epistemic element.

Section 4 introduces the notion of a causal symmetry which allows inferring probabilities without taking recourse to relative frequencies of input states or of outcome events. A causal symmetry basically consists in a possible relabeling of the outcome space that does not affect the causal structure responsible for the probability distribution. The concept leads to an objective version of the principle of indifference, which I term *principle of causal symmetry*. In the simplest case, two attributes that exhibit a causal symmetry are assigned equal probability. Furthermore, I argue that the epistemic principle of insufficient reason yields the same results as the principle of causal symmetry, whenever the resulting probabilities are predictive, i.e. essentially whenever these probabilities correspond to the actual limiting frequencies. If the relevant causal symmetries are not epistemically accessible, as is often the case, relative frequencies can be consulted as a weaker type of evidence for predictive probabilities.

In Section 5, the notion of probabilistic independence is explicated at some length establishing its relationship with causal irrelevance as determined by eliminative induction. Independence guarantees randomness in the sequence of input states and consequently of attributes. Since many theorems in probability theory like the law of large numbers presuppose independence of trials, a causal construal of independence is another crucial ingredient of the causal interpretation of probability. It broadly corresponds to the notion of randomness in frequentism and exchangeability in the subjectivist approach to probability. Furthermore, I outline how a probability measure can be established and interpreted based on arguments of symmetry and irrelevance without having to take recourse to relative frequencies. To sum up, the definition of probability in Section 3b, the principle of causal symmetry, and the causal rendering of probabilistic independence should be considered as a coherent package of the account of causal probability proposed in this essay.

Finally, various ontic and epistemic aspects in probability statements are identified in Section 6, and it is shown how the framework of causal probability can cover a wide range of

[4] The terms ‘attribute’ (translated from the German ‘Merkmal’) and ‘range’ (German ‘Spielraum’) are used in reverence to von Mises and von Kries, respectively, on whose ideas the present essay draws substantially.

applications from indeterministic phenomena to probabilities from causal symptoms to the probabilities of hypotheses.

2. Predecessors and contemporary debate

2a. Historical proponents: Cournot, Mill, von Kries

The two main ingredients of a causal interpretation as sketched in the introduction and elaborated later on in the article can be found with a variety of writers until the end of the 19^{th} century. As already indicated, the viewpoint is rather rare in the 20^{th} century presumably due to a widespread hostility towards inductive or causal approaches in science.

The distinction between an epistemic principle of insufficient reason and an objective principle of causal symmetry may be foreshadowed already in Laplace's classic 'Philosophical Essay on Probabilities': "The theory of chance consists in reducing all events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought." (Laplace 1902, 6-7; see also Strevens, Ch. 3.2) Of course, Laplace has in mind what was later called the classical definition of probability, i.e. the ratio of favorable to all possible cases. But everything hinges on the exact interpretation of equal possibility and how it is determined. Curiously, Laplace alludes to both epistemic and objective aspects, though these are not clearly held apart in his writing. In the quote given above, equal undecidedness implies an epistemic reading of equal possibility. But a later discussion of a loaded die evokes objective connotations in that Laplace distinguishes between judgments with respect to the knowledge of the observer and the presumably objective bias manifest in the coin. Laplace adds that the determination of respective possibilities is "one of the most delicate points of the theory of chances" (p. 11).

Other authors have been more explicit in drawing the distinction between epistemic and objective versions of the principle of indifference. One of the clearest expositions is due to Antoine-Augustin Cournot, who in the following quote delineates a principle of insufficient reason, which cannot establish objective probabilities: "If, in an imperfect state of our knowledge, we have no reason to believe that one combination is realized rather than another, even though in reality these combinations are events that may have unequal mathematical [i.e. objective] probabilities or possibilities, and if we understand by the probability of an event the ratio of the number of combinations that are favorable to the event to the total number of combinations that we put on the same line, this probability could still serve, in lack of a better option, to fix the conditions of a bet […]; but this probability would not anymore express the ratio that really and objectively exists between things; it would take on a purely subjective character and could vary from one individual to the other depending on the extent of her knowledge."[5] (1843, 438, my translation)

[5] "Si, dans l'état d'imperfection de nos connaissances, nous n'avons aucune raison de supposer qu'une combinaison arrive plutôt qu'une autre, quoiqu'en réalité ces combinaisons soient autant d'événements qui peuvent avoir des probabilités mathématiques ou des possibilités inégales, et si nous entendons par probabilité

Cournot also sketches the role of frequencies with respect to objective probabilities leading to the following colloquial statement of the law of large numbers: "If one considers a large number of trials of the same chance process, the ratio of the number of trials where the same event happens to the total number, becomes perceptibly equal to the ratio of the number of chances favorable to the event to the total number of chances, or what one calls the *mathematical probability* of an event."[6] (437, my translation) According to Cournot, the chances are measured in terms of the possibilities that certain conditions occur together to produce a particular type of event. Obviously, he employs a notion of probability distinct from relative frequencies referring to the ratio of favorable to all conditions or circumstances.

Thus, Cournot's account shows both ingredients of causal probability that were identified in the introduction: the distinction between an epistemic and an objective version of the principle of indifference and a definition of probability that refers to the number of favorable conditions, not instances.

The basic idea of an objective causal interpretation distinct from a frequentist approach is present with several other authors in the 19th century, for example in the writings of John Stuart Mill: "The probability of events as calculated from their mere frequency in past experience affords a less secure basis for practical guidance than their probability as deduced from an equally accurate knowledge of the frequency of occurrence of their causes." (1886, 355) Mill also recognizes the distinction between an epistemic and an objective reading of the principle of indifference. For example, he criticizes the alleged purely epistemic reading by Laplace: "To be able […] to pronounce two events equally probable, it is not enough that we should know that one or the other must happen, and should have no grounds for conjecturing which. Experience must have shown that the two are of equally frequent occurrence." (351) Mill sketches several options how the latter could happen, e.g. for the case of a coin toss: "We may know [that two events are of equal occurrence] if we please by actual experiment; or by the daily experience which life affords of events of the same general character; or deductively, from the effect of mechanical laws on a symmetrical body acted upon by forces varying indefinitely in quantity and direction." (351) Here, Mill introduces the important distinction between evidence in terms of frequencies and in terms of causal symmetries to establish objective equipossibility (cf. Section 4d). On this basis, he roughly formulates the notion of causal probability referring not to the frequency of events, but to causal conditions: "We can make a step beyond [the frequentist estimation of probabilities] when we can ascend to the causes on which the occurrence of [an event] A or its non-occurrence will depend, and form an estimate of the comparative frequency of the causes favourable and of those unfavourable to the occurrence." (355)

---

d'un événement le rapport entre le nombre des combinaisons qui lui sont favorables, et le nombre total des combinaisons mises par nous sur la même ligne, cette probabilité pourra encore servir, faute de mieux, à fixer les conditions d'un pari, d'un marché aléatoire quelconque; mais elle cessera d'exprimer un rapport subsistant réellement et objectivement entre les choses; elle prendra un caractère purement subjectif, et sera susceptible de varier d'un individu à un autre, selon la mesure de ses connaissances."

[6] "Lorsque l'on considère un grand nombre d'épreuves du même hasard, le rapport entre le nombre des cas où le même événement s'est produit, et le nombre total des épreuves, devient sensiblement égal au rapport entre le nombre des chances favorables à l'événement et le nombre total des chances, ou à ce qu'on nomme probabilité mathématique de l'événement." (437)

Curiously, Mill eventually retreats from this position that he so clearly formulated in the first edition of his ‘Logic’, adding the following comment in later editions: “I have since become convinced that the theory of chances, as conceived by Laplace and by mathematicians generally, has not the fundamental fallacy which I had ascribed to it [essentially referring to the epistemic reading of the principle of indifference].” (351) Mill claims that probability is fundamentally an epistemic notion and that probabilistic statements have no objective meaning anyways, because in a deterministic world any future event is fully determined by preceding conditions.

It remains somewhat unclear where Mill is heading with these remarks. Does he just want to rehabilitate the epistemic reading of the principle of indifference or does he want to deny the distinction between epistemic and objective readings altogether? From the viewpoint of this essay, Mill is correct that in a deterministic world, there is an epistemic element to any probabilistic statement, but he apparently fails to recognize that a fairly[7] objective meaning of probability nevertheless remains feasible: if one always relates probability *to a causally determined collective* (as elaborated in Section 3b). In any case, it is quite remarkable to observe how even an ingenious thinker like Mill struggles with the concept of probability.

Finally, the approach of Johannes von Kries should be mentioned (as summarized in his 1886, vii-viii).[8] His account was highlyinfluential on 20th-century philosophy, both on discussions within the Vienna Circle (Waismann, Wittgenstein) and on recent proposals regarding a novel class of objective probabilities (Strevens, Rosenthal, Abrams). Central to von Kries’ notion of probability is the *spielraum*[9] concept denoting the range of initial conditions that lead to a certain result. In principle, probability is determined by the ratio of the measure of the spielraum leading to a specific outcome to the measure of the entire spielraum. Based on this idea, von Kries formulates three conditions for numerical probabilities: (i) the different possibilities must correspond to *comparable* (‘vergleichbar’) spielräume[10]. In particular, it should be feasible to establish the equality in terms of measure of the various spielräume leading to different outcomes. (ii) Furthermore, the spielräume should be *original* (‘ursprünglich’), i.e. the equality of the spielräume must not cease to be the decisive criterion for our expectations when tracing the further history of the conditions making up the spielräume. (iii) Third, von Kries requires that the spielräume be *indifferent* (‘indifferent’), i.e. only the size of the spielräume and no other logical conditions should be relevant for the probability. According to von Kries, the most important criterion in this respect is that a small change in conditions may already lead to a different outcome. The various outcomes are supposed to alternate rapidly when continuously changing the conditions.

It is mainly this last criterion that establishes the parallel with the *method of arbitrary functions*, a term coined by Henri Poincaré (1912, p. 148). The French mathematician is

[7] depending on whether the various epistemic elements discussed in Section 6 are present. Certainly, causation has to be interpreted objectively as well.
[8] For a recent discussion consult the edited volume by Rosenthal & Seck (2016). In an interesting contribution, Helmut Pulte (2016) elaborates von Kries’ conceptions of natural laws and nomological knowledge as a conceptual background for his approach to probability and examines to what extent these are rooted in the historical context of his time.
[9] Spielraum translates to ‘range of possibilities’.
[10] I am using the German plural Spielräume.

usually seen as the originator of this tradition, although many ideas are already present in the mentioned work by von Kries (1886; later proponents are von Smoluchowski 1918, Hopf 1936; for a philosophical-historical overview, see von Plato 1983). In general, proponents of the method of arbitrary functions aim to establish objective probability for deterministic phenomena. Building on physical instability, they argue that any sufficiently regular distribution over the initial conditions leads to roughly the same ratio of occurrences in macroscopic outcomes. Primary applications are games of chance like roulette, which already Poincaré discussed in much detail, or the throwing of dice and coins.

Von Kries' account can broadly be classified as causal probability because the two criteria outlined in the introduction are present in his theory as well. First, his treatise on probability contains one of the most insightful assessments of the principle of insufficient reason in the history of probability (1886, Ch. 2). Second, he defines probability not in terms of frequencies of events but in terms of the ratio between different spielräume, i.e. conditions.

The outlined accounts are meant to be exemplary, a deeper look into 19th-century discussions on probability would presumably reveal that similar causal viewpoints were widespread.[11] In the first half of the 20th century, the ideas of von Kries were picked up and developed into an objective interpretation of probability by Friedrich Waismann (1930/1931), who claims in turn to have been influenced by Wittgenstein.[12] These accounts are somewhat similar to independent suggestions elaborated in recent years by Michael Strevens, Jacob Rosenthal, and Marshall Abrams, to which we will turn now.

2b. Contemporary debate: Abrams, Rosenthal, Strevens

Apparently, the history of causal interpretations of probability before the 20th century is quite rich and it seems plausible that the demise of this perspective more or less parallels the rise of causal skepticism in the beginning of the 20th century. At the same time, the distinction between frequentist evidence for objective probabilities and evidence in terms of causal symmetries largely disappears from the debate leading to a purely frequentist view of objective probabilities. Furthermore, only the epistemic version of the principle of indifference remains as a centerpiece of the logical interpretation, while the objective reading is largely abandoned. A notable exception in the latter respect are the writings of John Maynard Keynes who clearly recognizes a difference between ascribing equal probabilities on the basis of no evidence as opposed to evidence in terms of frequencies or relevant circumstances. He believes that the distinction is gradual and introduces the notion of *weight of argument* to account for it (1921, Ch. VI). But the idea has not caught on in 20th-century literature on probability.[13]

In recent years, one can observe a revival of objective interpretations that go beyond the frequency account by making explicit reference to initial conditions as well as system

[11] In fact, already Jacob Bernoulli in his *Ars conjectandi* interpreted equipossibility in a causal manner: "All cases are equally possible, that is to say, each can come about as easily as any other" (1713, 219; cited in Hacking 1971, 344).

[12] For a historical overview, see Heidelberger (2001).

[13] As Keynes himself stated, he was influenced by von Kries in framing the notion of weight of argument (cp. Fioretti 1998).

dynamics and thus bear resemblance to the historical accounts depicted in the previous section. This type of objective interpretations, which has been more or less independently developed by Marshall Abrams, Jacob Rosenthal, and Michael Strevens, substantially relies on ideas from the method of arbitrary functions.[14]

The best-known account in this modern tradition is Michael Strevens' *microconstant probability* (2011; see also 1998, 2006, 2013). In part, his approach is inspired by Maxwell's derivation of the molecular velocity distribution in an ideal gas which was carried out without empirical data about those velocities, i.e. without frequency data. Strevens elaborates in much detail the distinction between an objective and an epistemic reading of the principle of indifference (2013, Ch. 3). In his recent book 'Tychomancy', he lays out the most important principles for applying the objective version, which he terms *equidynamics*, by analyzing exemplary processes such as stirring or shaking (Ch. 5-8).

In one recent article, Strevens defines microconstant probability as an objective physical probability for deterministic systems along the lines of the method of arbitrary functions: "The event of a system S's producing an outcome of type e has a microconstant probability equal to p if (1.) the dynamics of S is microconstant with respect to e, and has strike ratio p, (2.) the actual initial conditions of nearly all long series of trials on systems of the same type as S make up macroperiodically distributed sets, and (3.) the macroperiodicity of the initial conditions is robust." (2011, 359)

Apparently, the crucial notions are *microconstancy* and *macroperiodicity*. The former refers to the premise that "within any small but not too small neighborhood, the proportion of initial conditions producing a given outcome is [approximately] the same" (2013, 11). This proportion is called *strike ratio* and it essentially determines the probability modulo substantial problems concerning the limiting process to infinitesimal neighborhoods and thus to exact probability values. Macroperiodicity denotes a certain smoothness in the probability distribution over initial conditions, such that neighboring initial conditions leading to different results should occur with approximately the same frequency in long series of trials.[15] This uniformity together with microconstancy leads to stable strike ratios and thus probabilities that are largely independent of the exact probability distribution over initial conditions. Finally, robustness in Strevens' third premise refers to counterfactual robustness, i.e. that counterfactual and predictive statements about frequencies are sufficiently reliable. Typical applications for microconstant probability are games of chance like roulette or playing dice, but Strevens believes that the notion also covers scientific applications from statistical physics[16] to the theory of evolution. Obviously, Strevens' approach features both characteristics of causal probability mentioned in the introduction.

[14] One should also mention the work of Richard Johns, who proposed a causal account of chance: "the chance of an event is the degree to which it is determined by its cause" (2002, 4). Moreover, propensity accounts are related to the causal approach, as already pointed out in the introduction and discussed further in Section 3c.
[15] In 'Tychomancy', Strevens replaces the term by the "in essence identical" (2013, 58) notion of *microequiprobability* that the probability density is approximately uniform over any small contiguous interval or region in the initial state space (2013, 246).
[16] For a related discussion, cf. Myrvold 2011.

A further prominent account in the tradition of the method of arbitrary functions is Marshall Abrams' *far-flung frequency (FFF) mechanistic probability* (2012). His approach, although independently developed, bears close resemblance to the accounts of both Strevens and Rosenthal. In particular, he relies on the same concepts of microconstancy and macroperiodicity as coined by Strevens. Abrams introduces the idea of a *causal map device*, which maps the input space to the outcome space, and partitions the outcome space into basic outcomes. A bubble is defined as a region in the input space containing points leading to all possible outcomes. A partition of the entire input space into bubbles he calls a bubble partition. Probability then is determined in the following manner: "There is a bubble partition of the [causal map] device's input space, such that many 'far flung' large natural collections of inputs together determine an input measure which makes most of the collections macroperiodic (and such that moderately significant changes in the spatiotemporal range across which natural collections are defined don't significantly affect outcome probabilities)." (Sec. 6) For lack of space, I won't go into details what exactly Abrams understands by "'far flung' large natural collections of inputs", but essentially they fulfill two conditions: they are microconstant and they reflect actual input patterns in the world (Sec. 4.2). Abrams emphasizes that he intends an objective probability interpretation that can account for a wide range of applications in games of chance, statistical mechanics and perhaps also the social and biological sciences (Sec. 6).

Finally, Rosenthal presents a very clear and thoroughly argued account of an objective probability interpretation largely construed around the notion of arbitrary functions, which he terms *natural range conception* in reminiscence of von Kries' spielraum-concept. He formulates two equivalent versions, one in terms of an integral over those initial states that lead to a desired outcome and the other referring to the ratio of ranges in the initial state space. I will focus on the second explication, which Rosenthal frames as follows: "Let $\boldsymbol{E}$ be a random experiment and A a possible outcome of it. Let $\boldsymbol{S}$ be the initial-state space attached to $\boldsymbol{E}$, and $\boldsymbol{S}_A$ be the set of those initial states leading to A. We assume that $\boldsymbol{S}$ and $\boldsymbol{S}_A$ are measurable subsets of the n-dimensional real vector space $\mathbf{R}^n$ (for some n). Let μ be the standard (Lebesgue-)measure. If there is a number p such that for each not-too-small n-dimensional (equilateral) interval $\boldsymbol{I}$ in $\boldsymbol{S}$, we have

$$\frac{\mu(\mathbf{I} \cap \mathbf{S}_A)}{\mu(\mathbf{I})} \approx \text{ p}$$

then there is an objective probability of A upon a trial of $\boldsymbol{E}$, and its value is p." (2012, 224)

Thus, Rosenthal explicitly frames his account as an objective probability interpretation for deterministic systems (2010, Sec. 5.3). In summary, the idea is that the probability of an outcome is proportional to that fraction of the initial-state space leading to the outcome, as determined by the Lebesgue measure. Since Rosenthal aims to develop an account for deterministic chance, i.e. he wants to eliminate epistemic aspects as far as possible, he has to require that in the initial-state space the conditions leading to the different outcomes are everywhere equally distributed, at least when looking with sufficient coarse-graining. This implies that any sufficiently smooth density function over the initial-state space will lead to approximately the same probability, which establishes the connection to the approach of arbitrary functions and the close relatedness with Strevens' microconstant probability relying on the notions of microconstancy and macroperiodicity. Of the three accounts discussed in this section, Rosenthal's definition remains closest to the original ideas of von Kries' spielraum conception by referring explicitly to a specific measure over the initial space. Without this element, the method of arbitrary functions could also be understood in terms of a frequentist approach with respect to the occurrence of initial states.

Rosenthal discusses a central objection against his own approach which comes in two slightly differing versions (2012, Sec. 4; 2010, Sec. 5.5). First, an eccentric distribution over initial states might be realized in nature leading to observed frequencies deviating substantially from p. Rosenthal suggests that at least in some such cases a nomological factor has been overlooked that determines the eccentric distribution. According to the second variant of the objection, there usually exist various ways in which the initial-state space could be reformulated such that it loses the characteristics required for Rosenthal's definition of probability. In particular, the Lebesgue measure might cease to be an appropriate choice to account for observed frequencies. Thus, one has to motivate why a certain formulation of initial conditions suitable for the natural-range conception is superior to others that are not suitable. Rosenthal essentially acknowledges that these are open problems for his approach.

Note that they are equally troublesome for Strevens' and Abrams' account since the concepts of microconstancy and macroperiodicity already presuppose a choice of measure. As a solution, Strevens suggests to always use standard variables, measured in standard ways. Because these tend to be macroperiodically distributed, microconstancy with respect to standard variables is meaningful. While Strevens' account is quite sophisticated in this respect (2006, Sec. 2.5; 2013, Ch. 12), I believe that the rejoinder eventually fails due to the blurriness and context-dependence of the notion of standard variable. After all, most phenomena can be accounted for in a large number of ways and it is just not plausible that all formulations will always yield microconstancy and macroperiodicity to the same extent.

A related problem concerns the various imprecisions and approximations figuring in the definition of probability of all three accounts. For example, Rosenthal's definition refers to "not-too-small" intervals and that the ratio of ranges only approximately determines the probability "$\approx$ p". In fact, the strike ratio will in general slightly fluctuate between different regions of the initial-state space. Thus, all theorems concerning microconstancy and macroperiodicity also hold only approximately. Especially, when aiming at a purely objective interpretation, these features are troublesome.[17] In Section 3b, I suggest how the outlined technical problems can be avoided by rendering probability measure-dependent.

Due to the close similarity of the accounts developed by Strevens, Rosenthal, and Abrams, I will in the following refer to them as the SRA-approach to objective probability.

[17] In personal communication, Michael Strevens has suggested as a response to consider microconstant probability as objective, but slightly indeterminate.

2c. Some praise

The causal approach referring to initial or boundary conditions can resolve a number of problems for traditional accounts of probability. These issues are discussed extensively by the authors mentioned in the previous section, so I will not delve into details. Let me just briefly comment on a few points.

According to Strevens, the "fundamental flaw" of the frequency account is that it cannot distinguish between meaningful and arbitrary frequencies and thus cannot reliably ground counterfactual statements and predictions in probabilities (2011, Sec. 2). The issue largely parallels the standard problem of induction. In reply, the causal account offers as a criterion that frequencies are only meaningful, when they result from a collective determined by causal conditions. Of course, this solution can only get off the ground given a defensible notion of causation, a topic that will be addressed in Section 3.

A further major advantage in comparison with frequency theories is that causal interpretations can establish probability independently of observed frequencies, for example by referring to symmetries or by rendering probabilistic phenomena largely independent of the probability distribution over initial states. Among other things, this allows for a non-circular reading of the law of large numbers if probabilities are not themselves defined in terms of limiting frequencies (e.g. Abrams 2012, Sec. 1.1; Rosenthal 2010, Sec. 5.2).

By relying on some version of the principle of indifference, causal probabilities bear resemblance to logical interpretations of probability. However, the principle of insufficient reason referring to ignorance, as it is used in the logical approach, is notoriously flawed by challenging objections—in particular Bertrand's paradox, which highlights ambiguities in the application of this principle (Bertrand 1889; van Fraassen 1990, Ch. 12). The causal approach resolves these ambiguities by introducing an objective variant of the principle of indifference, later referred to as principle of causal symmetry in the specific account of causal probability to be developed in Sections 3 to 6 (cp. esp. Section 4c).

2d. Critical reflections

While clearly being a major step in the right direction, the recent attempts to develop an objective account of probability in the tradition of the method of arbitrary functions suffer from a number of shortcomings. There are the technical objections already pointed out towards the end of Section 2b. In addition, there are three more general issues, which I will delineate in the following.

First, the SRA-approach tries to establish that probabilities are largely independent of the measure over the input space. But this does not eliminate the need to interpret the measure in a way that does not refer to relative frequencies which would lead us back to essentially a frequency interpretation. To solve this problem, I argue in Section 5b that the measure can be interpreted in terms of symmetries in the circumstances determining a probabilistic phenomenon.

Second, the objective accounts of the SRA-approach mostly fail to clarify the relation to epistemic probabilities and therefore implicitly subscribe to an in my view misguided sharp distinction between ontic and epistemic probabilities. Instead, I will pin down in Section 6 several epistemic features that can but need not be present in the assignment of probabilities. I will sketch how the various shades of epistemic and ontic probabilities can all be accounted for in terms of a single causal interpretation. Thus, the range of application is widely extended beyond cases in which the method of arbitrary functions can be employed.

Third, the mentioned accounts all rely on physical or causal laws determining the dynamics of the considered phenomena largely without explaining the origin of these laws. In the worst case, they need to be established inductively leading us back to the problem of distinguishing between meaningful and arbitrary relations, which the SRA-approach aimed to resolve in the first place. Thus, a major task for any approach to probability is to clarify how it fits into a more general framework of induction and causation. This will be attempted in the following.

3. Induction, causation, and probability

In the previous section, a shortcoming of the SRA-approach was identified that the probabilities rely on physical knowledge in terms of dynamics and laws of motion but fail to make a connection with a specific account of induction and a corresponding notion of causation. In the following, I try to ameliorate the situation by comparing two distinct accounts of induction, namely enumerative and eliminative, and by examining how in each case a notion of probability could be integrated. Enumerative induction leads to a naïve frequency account of probability that must be rejected in particular for failing to draw a distinction between accidental and lawlike regularities. By contrast, eliminative induction offers a solution to this problem in terms of a difference-making account of causation, while of course some amount of uncertainty remains for any inductive inference. Trying to implement probability in eliminative induction will lead to an account of causal probability that resembles those presented in Sections 2a and 2b. From now on, the terms 'causal interpretation of probability' and 'causal probability' more narrowly refer to the specific account to be developed in the remainder of the essay. According to the proposed viewpoint, probabilities are understood as degrees or grades of causal determination by a given set of circumstances or conditions. It should be added that such determination may be considered both in the direction from causes to effects and from effects to causes (for the latter cp. in particular Section 6d).[18,19]

[18] One referee has suggested that the proposed notion of probability should be interpreted as an abstract framework into which every causal interpretation of probability has to fit. In principle, I am happy with such a pluralistic reading in terms of a class of interpretations rather than a single one. In particular, there certainly is some room for allowing different understandings of causality.

[19] Marshall Abrams, in an interesting recent paper (2015), formulates a notion of causal probability that is strictly speaking neither an interpretation nor a class of interpretations. Rather, he considers the causal nature an additional feature of some interpretations of probability including long run propensities and his own 'mechanistic probability' (2012). Broadly speaking, Abrams terms probabilities causal when a change in properties of a chance set-up affects the relative frequencies of outcomes. For a more exact definition, he employs Woodward's interventionist framework. While Abram's approach is certainly related, the proposal in

3a. Enumerative induction and the frequency theory

Enumerative induction is the rather naïve view that general laws can be deduced from the observation of mere regularities: If in all observations, one finds two events, objects, properties, etc. A and B always conjoined then there supposedly exists a causal connection between A and B. This basic idea is shared by all naïve regularity conceptions of natural laws and causation.

The generalization to statistical laws is straight-forward although some technical complications arise due to possible fluctuations in the observed frequencies. Basically, if in a sequence of events of type A one finds a more or less constant ratio p for another type of event B, then one can conclude to a statistical law connecting A and B with probability p. For example, if a coin lands heads in approximately one half of all trials, then the probability of this event probably is somewhere close to one half. Serious problems arise because the true value of the probability is usually identified with the limiting frequency in an infinite number of trials. The naïve frequency view thus grants epistemic access only to observed frequencies but not to the underlying probabilities themselves. Therefore, it exhibits considerable difficulties dealing with cases, where the frequencies by pure coincidence deviate from the actual probabilities.

However, at this point we can neglect the problems arising in this regard since the naïve frequency view falls prey to a much more fundamental flaw, the same as the naïve regularity conception of laws and causation: it cannot distinguish between accidental and lawlike statistical relationships, i.e. between those that can ground predictions and successful manipulations and those that cannot (cp. Strevens 2011, Sec. 2; as already discussed in Section 2c). For example, the naïve frequency view cannot handle the following situation of an exchanged coin. Consider a sequence of throws, during which the coin is exchanged at some point with another one looking very much alike. Presumably, the naïve frequentist would have to derive predictions about future events from the whole sequence. He cannot make the crucial distinction between the case, where both coins are structurally similar, and the case, where the coins are structurally distinct, e.g. one fair the other loaded. As we will see shortly, such distinctions can be systematically established only within the context of eliminative induction. In other words, the naïve frequency view leads to an essentially unresolvable reference class problem since it lacks clear rules how to determine structural similarity.

In comparison, the causal interpretation elaborated in this essay accepts that any single event can be attributed to different collectives, which in general imply different probabilities for the event. In other words, there is an ambiguity in the choice of reference class, which however is not fatal to the causal interpretation, since causal probability is defined with respect to a collective. This dissolves what Alan Hájek has termed the metaphysical reference class

---

the present paper ascribes a much more central role to causality extending to a number of fundamental concepts in probability theory like probabilistic independence or the principle of indifference. Causal probability therefore should be understood as a specific interpretation of probability in its own right (or at least a class of interpretations determined by different accounts of causation) that is conceptually incompatible with other interpretations. For a discussion of those differences, see in particular Section 3c.

problem (2007). Note that an epistemic agent acting on the basis of probabilities should use the collective that is as specific as possible in terms of causally relevant conditions under the additional constraint that the agent has epistemic access to some evidence for the corresponding probabilities in terms of symmetries or relative frequencies. By contrast, the fatal reference class problem for the naïve frequentist is that he may construct an ensemble of seemingly similar events, which are however structurally dissimilar, and therefore the resulting frequencies are not predictive. This problem is avoided in the causal approach because the collective conditions are by definition causally relevant for the considered phenomenon and must remain constant during all trials, while the range conditions are supposed to vary randomly.

3b. Eliminative induction and the causal conception of probability

Eliminative induction is distinguished from enumerative induction in that it examines not the mere repetition of phenomena but rather phenomena under varying circumstances or conditions. Eliminative methods determine the causal relevance or irrelevance of conditions for a certain phenomenon. The main methods are the method of difference and the strict method of agreement. The first establishes causal relevance of a condition C to a phenomenon P from the observation of two instances which are alike in all conditions that are causally relevant to P except for C. If in one instance both C and P are present and in the other both C and P are absent, then C is causally relevant to P. The strict method of agreement establishes causal irrelevance in much the same manner, except that the change in C has no influence on P.[20] According to this view of eliminative induction, causal (ir-)relevance is a three-place notion: Condition C is causally (ir-)relevant to P with respect to a background B consisting of further conditions that remain constant if causally relevant to P or that are allowed to vary if causally irrelevant. For further details, see Pietsch (2014).

The outlined approach to induction has a counterpart in an account of causation that broadly stands in the counterfactual tradition and that was elsewhere termed difference-making account.[21] It is distinguished from conventional counterfactual approaches, in particular that of David Lewis, by the following characteristics: a notion of causal irrelevance is introduced; all causal relationships are rendered background-dependent; and counterfactual propositions are not evaluated in terms of possible worlds but on the basis of refined versions of the method of difference and the strict method of agreement and therefore by referring to instances in the actual world.

The main ingredients of this difference-making account are (i) counterfactual definitions of the fundamental notions of causal relevance and causal irrelevance: ‘in a context B, in which a condition C and a phenomenon P occur, C is causally relevant (irrelevant) to P, iff the following counterfactual holds: if C had not occurred, P would also not have occurred (if C had not occurred, P would still have occurred)’; (ii) obviously, these definitions implement background- or context-dependence, an idea roughly taken from John Mackie’s work: in principle, a background or context is defined by conditions that must remain constant and

[20] Note that as a complication, judgments of causal irrelevance depend on measurement accuracy.
[21] In the following I can only provide a very brief sketch of the account. A basic outline can be found in Pietsch (2015, Sec. 4.1), a detailed defense in Pietsch (2016).

others that are allowed to vary; (iii) finally, an account of counterfactuals is employed that takes its inspiration directly from the method of difference: '"If C were not the case, P would not be the case" is true with respect to an instance in which both C and P occur in a context B, if first, at least one instance is realized in which neither C nor P occurs in the same context B and second, if B guarantees homogeneity.' The latter is the case, iff only conditions that are causally irrelevant to P can change, except for C itself and conditions that are causally relevant to P in virtue of C being causally relevant to P, i.e. in particular conditions that lie on a causal chain through C to P. Let me emphasize again that the definitions of causal relevance and irrelevance correspond directly to the method of difference and the strict method of agreement, respectively.

How does probability fit into this picture of induction and causation? Note first that both principal methods of eliminative induction and the corresponding definitions of causal relevance and irrelevance presuppose determinism, i.e. that P is fully determined by causal conditions (Pietsch 2014, Sec. 3f). Consequently, we will in the following delineate an essentially epistemic probability conception for deterministic phenomena, while indeterministic probabilities can be integrated later on, as discussed in Sections 6a and 6b.

Let me begin with a simple example to outline the basic idea of the proposed causal interpretation of probability. Consider a wheel of fortune with four different areas of equal size, which are labeled, say, as green, blue, red, and yellow. Let a blindfolded person determine the moment, when to stop the wheel. Apparently, certain conditions remain constant in different instances or trials of this set-up, for example the mentioned distribution of labels on the wheel, maybe also the velocity with which the wheel is turning etc. These constant conditions in probabilistic phenomena shall be called collective conditions. A number of other conditions may change from trial to trial, in particular the moment and the position at which the wheel starts turning and the moment when the blindfolded person stops the wheel. Let these conditions be called range conditions. Obviously, collective and range conditions taken together causally fix the specific event that will happen.

The range conditions span an outcome space and each point in this space is labeled in terms of the resulting attribute, which in the discussed example is the color at which the wheel stops. In the next step, one is interested in the distribution of attributes in the outcome space. According to the proposed interpretation, this issue is tackled using symmetry arguments, e.g. by examining the dynamics of the probabilistic phenomenon. And indeed it turns out that the considered causal structure is invariant under permutation of the different colors, which implies that all colors appear in the outcome space to the same extent, i.e. they all have equal measure. Note that at this point the measure does not yet have to be a probability measure, e.g. it can follow from a regular dynamics, but it has to be characteristic of the extent in which the various attributes are realized, in order to eventually establish the connection with frequencies. In the example, we know that the color sequence of the rotating wheel of fortune follows a perfectly regular pattern with each color appearing an equal amount of time.

Thus, an additional argument is needed to establish the measure as a probability measure. In particular, it has to be shown that different trials are independent of each other. Generally, this can be guaranteed based on causal irrelevance. Roughly, since the person stopping the wheel is blindfolded and the wheel is turning at considerable speed, we know on the basis of fairly simple scientific laws that the rotational state of the wheel is causally unrelated with the moment when the blindfolded person stops the wheel. Given that the measure designates how often an attribute is realized and that in addition independence of trials can be established by means of arguments from causal irrelevance, the ratios of the various attributes in the outcome space weighted by the measure can be interpreted as probabilities. Also, since the main premises for the law of large numbers are fulfilled, a link to relative frequencies can be established.

Let me now introduce the formal framework. In developing a probability concept for eliminative induction, the focus must lie on the variation of conditions, which constitutes the crucial change in perspective compared with enumerative induction which focuses on the number of instances (cp. Federica Russo's variational epistemology for causation, e.g. Illari & Russo 2014, Ch. 16). In particular, a careful distinction between various types of circumstances or conditions needs to be introduced.

We are interested in the impact of a number of potentially relevant conditions $C_1, \ldots, C_M$ on a statistical[22] phenomenon P with respect to a background B. Since P is statistical, it must be linked to a space O of possible outcome states, which may be continuous and many-dimensional, but will for the sake of simplicity from now on be assumed as discrete and one-dimensional. No additional conceptual difficulties arise in the former case. The outcome space is divided into mutually exclusive regions covering the whole space. These regions are labeled and the labels are called *attributes* $M_1, \ldots, M_N$.[23] Note that the labels are introduced in addition to the parameters spanning the outcome space for reasons that will become clear later on when the notion of causal symmetry is defined.

Let me now introduce various types of conditions, in particular the distinction between *collective*[24] *conditions* and *range*[25] *conditions*. Both types are causally relevant (in the sense of difference-making) to P. When examining a particular probabilistic phenomenon, the collective conditions must remain constant, while the range conditions are allowed to vary. The collective conditions fix the occurrence of the class P but do not determine which of the attributes $M_1, \ldots, M_N$ will actually happen, i.e. these conditions determine the probability space regarding the various manifestations of the phenomenon P. Note that the collective conditions include all causally relevant conditions in the background or context B. Collective and range conditions together causally fix the exact outcome state and thereby also which event M_X of the $M_1, \ldots, M_N$ will actually happen in a specific instance or trial. Furthermore, a measure W needs to be introduced denoting the probability with which certain combinations of range conditions appear and thus the probability of the corresponding outcome states. In principle, this measure is determined by the collective conditions as further discussed in Section 4. It is normalized over the whole outcome space and should, via the various laws of

[22] ‚Statistical phenomenon' here is not identical with ‚probabilistic phenomenon' as defined below, but is more broadly understood as a phenomenon that is not fully determined by the considered circumstances or conditions.
[23] In reverence to von Mises who used the German term 'merkmal' that translates to feature, attribute, characteristic.
[24] Again, we rely on the terminology of von Mises.
[25] The terminology here is of course in reverence to von Kries.

large numbers, correspond to the limiting frequencies with which the outcome states of a specific probabilistic phenomenon will be instantiated. The exact interpretation of this measure constitutes a crucial challenge for the causal account mainly for two reasons: (i) there is an immediate threat of conceptual circularity if the measure is itself explicated in probabilistic terms; (ii) in particular, if the measure is interpreted in terms of relative frequencies we are thrown back on a frequentist interpretation of probability. The suggested solution in the framework of the causal approach is to establish the measure quantitatively on the basis of causal symmetries in the collective conditions and to identify it as a probability measure that corresponds to limiting frequencies by arguments based on causal irrelevance. In other words, the measure is interpreted in terms of causal symmetries and causal irrelevance (cf. Section 5b).[26]

Sometimes, when it is possible to clearly specify the process determining the measure, it may make sense to distinguish between two types of collective conditions: *set-up conditions* determining the possible combinations of range conditions; and *measure conditions*, which fix the measure over the space spanned by the range conditions and thus the probabilities of the outcomes.[27] Note that in the exceptional case of indeterministic phenomena, there are no range conditions that vary. The measure therefore becomes dispensable, and the probabilities directly result from the system's indeterministic dynamics.[28]

This leads to the notions of a *probabilistic phenomenon* and of *causal probability* (definition 1):

> *A probabilistic phenomenon P is determined by collective conditions C that remain constant; range conditions R that are allowed to vary and that span an outcome space O; as well as a probability measure W over the outcome space. The causal probability of a specific attribute M_X, combining a set of possible outcomes of the phenomenon P, is given by the fraction of outcome states pertaining to attribute M_X, weighted*[29] *with the measure W.*[30]

As already said, the measure is in principle determined by causal symmetries in the collective conditions (cf. Section 4) and the nature as a probability measure (i.e. a measure that corresponds to the actual limiting frequencies) must be established in terms of causal irrelevance (Section 5). In summary, probabilities according to the proposed view denote *degrees of causal determination of the attributes by the collective conditions*.

In some situations, it may be useful to add more structure to the probabilistic phenomenon by introducing an input space of possible input states $S_1, \ldots, S_Q$, which is now spanned by the range conditions, as well as a causal mapping $S \stackrel{c}{\to} O$. Again, we assumed a discrete, one-

---

[26] Like the frequentist interpretation this constitutes an operationalist approach, only on the basis of symmetries rather than relative frequencies.

[27] There is often a normative component to the measure and thus also to the collective conditions, since it is partly a matter of choice which events to include in a collective and which not.

[28] We will return to this topic in Section 6a.

[29] Henceforth, I will speak of the 'weighted fraction of outcome states'.

[30] I claim that this is the notion of probability that many of the classical thinkers mentioned in Section 2a had in mind. Strevens (2006) makes a similar suggestion, but sees it as a special kind of probability, namely 'complex probability', in contrast with 'simple probabilities' that appear in or depend on fundamental laws of nature.

dimensional input space for the sake of simplicity, a generalization would add no further difficulties. Most importantly, this extra structure of the probabilistic phenomenon allows to better show the connection with the method of arbitrary functions and the SRA framework (definition 2):

> *A probabilistic phenomenon P is determined by collective conditions C, range conditions R spanning the input space S, a probability measure W over the input space and a causal mapping* $S \overset{c}{\to} O$ *of the input space on the outcome space O. The causal probability of a specific attribute* M_X*, combining a set of possible outcomes of the phenomenon P, is given by the fraction of input states leading to attribute* M_X*, weighted with the measure W.*

Obviously, this second definition is a special case of the first.

According to both definitions, probability is always relative to collective conditions—which is very much in the spirit of von Mises' famous statement "first the collective—then the probability" (1981, 18).[31] Sometimes, when the range conditions and the measure over those conditions are not explicitly known, one may express a probabilistic phenomenon in terms of the attribute space, but a constant collective is nevertheless always required. Note finally that the basic axioms of probability will be satisfied since the definitions are based on fractions referring to a normalized measure.

Let me briefly elaborate on the issue, why the presented approach merits to be called 'causal'. Most importantly, the collective conditions *causally* determine the probabilistic phenomenon P, and collective[32] and range conditions taken together *causally* determine a specific manifestation of P. While sophisticated frequency accounts like von Mises' approach also require a collective, they do not consider it fixed strictly by causal conditions, but presumably other types of conditions may also appear, e.g. these accounts lack the important distinction between causal variables and proxy variables as discussed in Section 6c. Note that this constitutes the decisive step with respect to frequency accounts to solve the problem of properly distinguishing between arbitrary and meaningful frequencies.

In the following sections, I will introduce several further concepts that are central to the causal interpretation. The notion of causal symmetry, referring to invariance of causal structure with respect to attribute permutations, and the related principle of causal symmetry, as explicated in Section 4, allow establishing the measure to an extent that the probability distribution of the attributes can be fixed without relying on relative frequencies as evidence. In Section 5, a causal construal of the notion of independence will be provided ensuring that sequences of outcome states will be random. Without independence (or related concepts like exchangeability), one could hardly speak of a probabilistic phenomenon, since many theorems of probability theory like the various laws of large numbers justifying the convergence of relative frequencies to the actual probabilities rely on independence of subsequent events.

[31] "we shall not speak of probability until a collective has been defined" (ibid.)

[32] more exactly, the set-up conditions, if these can be distinguished from the measure conditions

The notions of causal symmetry and the causal construal of independence further underline the causal nature of the proposed account of probability. The definition of probability given above, the principle of causal symmetry, and a causal construal of the notion of independence should be seen as one package making up causal probability. The connection with eliminative induction can be understood in terms of a coarse-grained formulation. Instead of examining particular instances, where specific R and O are realized, statistical phenomena P as a whole can be considered—determined by certain collective conditions and an attribute distribution, e.g. an ideal gas in a box or a long sequence of throws with a die. The causal relations how changes in collective conditions affect the respective statistical phenomena within this macro-perspective can again be established by the method of difference and the strict method of agreement. Predictions and counterfactual statements can thus be derived.

Let me illustrate the proposed notion of probability with another simple example regarding the throw of a coin (P). The attributes partitioning the outcome space are heads-up (M_1) or tails-up (M_2). The collective conditions are the causal conditions of the set-up, e.g. concerning the type of coin, the allowed types of throwing, the types of surface on which the coin lands, etc. These conditions are held fix in all instances of the phenomenon. The range conditions are also causally relevant to the outcome but randomly vary from throw to throw: including the exact initial state of the coin before the throw, the initial speed, direction, and torque of the throw, etc. Assuming determinism, the attribute is fixed by the range conditions. Finally, the measure W denotes the probability, with which the various range conditions occur. In principle, W is fixed by causal symmetries in the collective conditions. In particular, the dynamics of the throw as well as the process determining the initial state of the coin might both be invariant with respect to exchanging the labels on the coin. It should be added that it generally suffices that the instructions how to throw the coin determine the measure over range conditions to an extent that the attribute distribution is fairly stable. In other words, the measure is seldom fixed to full extent. This is the lesson learned from the method of arbitrary functions. Note finally that the range conditions can usually be formulated in different ways for a probabilistic phenomenon, which requires a complementary adjustment of the measure.

As long as the collective for the throws remains the same, including that the initial states vary sufficiently, long-run frequencies will almost always closely approximate the actual probabilities according to the mathematical theorem called the law of large numbers. This solves the problem of the exchanged coin of Section 3a. As long as both coins are structurally similar, e.g. fair, the collective conditions stay the same when the coin is exchanged, and therefore predictions based on combined frequencies can be expected to hold. If one coin is fair and the other loaded, then the instances do not form a collective, because a causally relevant condition has changed and therefore predictions based on relative frequencies will in general fail to hold (though there may be ways of formulating a combined collective, see Section 6b).

Another classic application of probability concerns population statistics, e.g. the question whether a certain person will die at a given age. Regarding this type of problem Mill has claimed that probability lacks an objective meaning since for every individual death is supposedly a matter of deterministic fact (cf. Section 2a). With respect to single-case probabilities in deterministic settings, this assessment is certainly correct. However, there is a fairly objective meaning to probability if relating it to a specific collective as required by the definition of causal probability given above (regarding a discussion of various epistemic elements in causal probabilities, cf. Section 6).

To determine the probability whether someone will die at a specific age we thus first have to fix a collective specifying causally relevant circumstances, for example the gender of a person, certain habits, e.g. whether he/she smokes, is active in sports, or has pre-existing diseases. The collective conditions leave open the two possibilities of interest that the person dies at a given age or not. Probabilities result from the range conditions and a measure over the space spanned by the range conditions, although these need not—and often cannot—be made explicit. While admittedly it is impossible to list all the relevant causal conditions for phenomena with a complex causal structure like the death of a person, in principle the construction of a collective according to the definition above is possible assuming determinism. And the fact that insurance companies manage to arrive at fairly stable probability distributions suggests that they have some epistemic access to appropriate collectives.

In combination, collective and range conditions causally determine whether a person will die or not. Of course, the exact boundary between collective and range conditions is usually quite arbitrary. In the case of population statistics, the collective is mostly determined by choosing a certain group of the total population, for example white male living in New York State. Since epistemic access to causal symmetries is implausible for phenomena of such complexity, the required information about range conditions and measure is derived from past frequency data—under the assumption that this data is representative of the group and that the collective conditions will approximately stay the same for the time period that is to be predicted. Note again that the collective should generally be chosen in such a way that it includes all conditions that are known to be causally relevant in a considered instance, if one wants to act on the basis of the resulting probabilities. For example when someone is known to have prostate cancer, this information should be included in the collective conditions concerning an imminent death, if, of course, there is also sufficient frequency data available to determine the corresponding probabilities.

3c. A brief comparison with other accounts

In the introduction, I had already pointed out the main differences between the causal approach and the logical as well as the frequentist accounts. With respect to the former, the causal approach relies on an ontic and not on an epistemic version of the principle of indifference. With respect to the latter, the causal approach defines probability in terms of the ratio of favorable boundary or initial conditions and not in terms of relative frequencies of events.[33]

---

[33] One should also mention best-systems interpretations originating with Lewis (1994); Hoefer (2007) is a more recent development in this tradition. These interpretations pursue a different aim compared with causal interpretations bytrying to situate probability within a framework of Lewisian metaphysics. Let me further

The account proposed in Section 3b is conceptually closest to the SRA-approach and to the propensity theory. It is therefore worthwhile to briefly address the most important differences in each case. Without any loss of generality, I will rely on definition 2 in the following discussion. With respect to the SRA-approach based on the method of arbitrary functions, a crucial difference is that causal probability[34] is always relative to the collective conditions and thereby also to the measure over the input space while the SRA-approach tries to establish that probabilities are independent of the choice of measure. Rendering probability relative to the measure resolves in a simple manner the central objection against the natural-range conception that was described towards the end of Section 2b. Concerning the first situation, i.e. the problem of eccentric distributions over initial states, the causal perspective is the following. If the collective conditions determine an eccentric distribution, the measure must reflect this distribution. By contrast, if an eccentric sequence of initial states occurs by coincidence given a non-eccentric measure, then the eccentric sequence must be attributed to chance.

The second situation, Rosenthal worries about, is that reformulations of the initial conditions lead to a change in probabilities. Indeed according to his natural range conception, which relies on the Lebesgue measure over the initial-state space, reformulations could easily imply probabilities in contradiction with observed frequencies. Rosenthal suggests excluding such “unphysical” descriptions, but it remains completely unclear how to construe a suitable notion of unphysicality. Rather, the various debates on conventionality in physics have shown that supposedly unphysical descriptions are often feasible and empirically adequate. Furthermore, opinions about physicality habitually change over the course of history. This difficulty is also resolved in a simple manner by the account of causal probability. Essentially, any change in the formulation of the range conditions has to be compensated by a complementary change in measure in order to stay consistent with the collective conditions and the observed frequencies. Obviously, this option is not available to Rosenthal since he insists on using the Lebesgue measure as probability measure. Note again that the same difficulties which Rosenthal makes explicit are hidden in the conditions of microconstancy and macroperiodicity in Strevens’ and Abrams’ account which presuppose a measure. Strevens’ response in terms of standard variables was already described in Section 2b and is largely equivalent to Rosenthal’s proposal.

Furthermore, there is no need for approximations or imprecisions in the causal account in contrast with Rosenthal’s definition of probability or the related definitions of microconstancy and macroperiodicity in Strevens’ and Abrams’ accounts (cf. the end of Section 2b). Rather, the probability according to the causal interpretation corresponds *exactly* to the weighted fraction of outcome states. Again, this move is possible since the causal account renders probability relative to the measure, but also because the causal construal of independence ensures randomness in the sequence of initial conditions and thus convergence of relative frequencies to the causal probabilities by the law of large numbers.

---

briefly point to an interesting recent attempt to combine a spielraum approach with a best-systems interpretation by Claus Beisbart (2016).

[34] Remember that the terms ‘causal interpretation’ and ‘causal probability’ now refer exclusively to the account developed in Section 3b.

The price to pay is that probability becomes relative to the essentially epistemic choice of a collective (cf. Section 6), which thwarts the project of a purely objective probability interpretation in deterministic settings. On the other hand, I don't see why accepting some epistemic aspects in probability is problematic except if one adheres to an overly realist view of science. And again, this very step enables the causal interpretation to cover a wide range of applications from indeterministic probabilities to probabilities of hypotheses as described in Section 6—compared with the rather narrow range of applications of the SRA-approach requiring microconstancy and macroperiodicity.

Of course, phenomena accessible to the method of arbitrary functions can be treated within the causal approach as well. In such cases, the collective conditions[35] and the measure need to be fixed only to the extent that the probability distribution is approximately stable. As an example, consider the throw of a die. The probability distribution does not depend much on the exact instructions for the collective, e.g. concerning the original position of the die, the way it is thrown etc. Generally speaking, the exact choice of collective conditions and measure is largely irrelevant, if the dynamics of the system is sufficiently complex—a topic that is discussed today mainly in the domain of ergodic theory.

On a deeper level, the introduction of measure-dependence in the causal approach calls for new concepts that are not central to the SRA-approach. First, the measure over input states must be determinable independently of relative frequencies in the causal approach—otherwise we would be thrown back on frequentism. To this purpose, the principle of causal symmetry is introduced in the next Section 4. Second, when the condition of microconstancy is dropped, it cannot be assumed anymore that the occurrence of attributes will be sufficiently random due to slight variations in initial conditions. Therefore, in the causal interpretation randomness has to be established by other means leading to the causal construal of independence proposed in Section 5. By referring to causal symmetries in the collective conditions and to causal irrelevance establishing probabilistic independence, the causal interpretation resolves one of the fundamental problems of the SRA-approach, namely how to interpret the measure over input space (cp. Section 5b).

The causal approach also owes considerably to various versions of the propensity interpretation. Most importantly, they share the broad (and important) idea that probabilities arise from circumstances or conditions. However, a direct comparison is rendered somewhat difficult by the enormous spectrum of propensity accounts in the literature (a good recent overview can be found in Berkovitz 2015). In fact, the various accounts differ so substantially that some scholars subsume under the notion of propensity any objective approach that is not a frequency interpretation (Gillies 2000a, 114). The most fundamental distinction is between long-run and single-case propensity theories: "A long-run propensity theory is one in which propensities are associated with repeatable conditions, and are regarded as propensities to produce, in a long series of repetitions of these conditions, frequencies which are approximately equal to the probabilities. A single-case propensity theory is one in which propensities are regarded as propensities to produce a particular result on a specific occasion." (Gillies 2000a, 126) One crucial problem of single-case propensity interpretations, especially

[35] in particular the measure conditions, if these can be separated from the set-up conditions

those in which probabilities depend on the whole state of the universe at a given time (the later Popper, David Miller), is to establish a connection between propensities and relative frequencies. On the other hand, long-run propensity interpretations run the risk of collapsing into a frequency interpretation, since relative frequencies are invoked on a fundamental conceptual level. By contrast, the proposed account of causal probability does not rely on relative frequencies but on symmetries to establish probability distributions on a fundamental level, while at the same time making a connection with relative frequencies via the law of large numbers, if probabilistic independence can be shown by arguments of causal irrelevance.

There are a number of crucial differences between propensity theories and the proposed account. The first point concerns ontology. If the proponents of propensities were thinking of causal determination, why not call it causation? Why use a rather obscure term like propensity? Popper and other proponents of propensity accounts seem to have felt the need to introduce a novel ontological category to account for probabilistic phenomena. In later years, Popper considered causation to be a special case of propensities, namely when the propensity equals one. As another example, Donald Gillies claims that non-causal correlations for example between a low barometer reading and subsequent rainfall also constitute propensities (2000b, 829-830). Other proponents of a propensity interpretation like D.H. Mellor reject the view that chance is a sort of "weak or intermittently successful causal link" maintaining that "causal talk is not really illuminating in statistical contexts" (Mellor cited in Berkovitz 2015, 657). To resolve this confusing disagreement concerning the relationship between causation and probability, the approach proposed in this essay tries to situate probability within a specific framework of causation. While in general propensity accounts focus conceptually on dispositions or tendencies and rather casually remark upon the parallel with causation, the interpretation proposed here starts with a detailed and specific concept of causation as difference making and examines how probability fits into the picture.

On a more methodological level, causal probability is relative not only to the collective conditions but—unlike propensities—also to the measure over the space spanned by the range conditions. Relatedly, propensity approaches are often silent on the question how exactly the circumstances determine the probabilities. They typically lack the notion of causal symmetry, the ontic version of the principle of indifference, and the causal construal of probabilistic independence. With respect to the last issue, the randomness of subsequent events is often considered as implicit in the notion of tendency in propensity accounts.

Finally, the fact that propensities are framed in a language of tendencies or dispositions appears to explicitly exclude the formulation of inverse probabilities, i.e. evidential probabilities or the probabilities of hypotheses (for an elaboration of this criticism, cp. Humphreys 1985).[36] How causal probabilities as proposed in this essay can be inversed is briefly indicated in Section 6d. Furthermore, while influential propensity theorists like Popper have argued that inductive concepts like confirmation are not explicable in terms of probabilities at all, the causal interpretation explicitly establishes the link with an inductive

[36] Various responses from propensity theorists to this so-called Humphreys' paradox can be found in Berkovitz (2015, Sec. 5). Several scholars like Mauricio Suarez conclude that propensities cannot be probabilities (2013).

framework. Part of the project of a causal interpretation is to show how the basic idea that probabilities arise from circumstances can be extended to epistemic probabilities like the probabilities of hypotheses (cp. Sec. 6).

4. Causal symmetries and the principle of causal symmetry

4a. Causal symmetries

I will now argue that given full knowledge of the causal setup, the measure over different combinations of range conditions can always be determined by means of symmetry considerations without taking recourse to relative frequencies. More exactly, the symmetries must fix the measure only to the extent that a stable probability distribution results. Note also, that in this section the measure does not yet have to be considered a probability measure, e.g. it can result from perfectly regular dynamics. How the random nature of the attribute sequence can be established in addition will then be discussed in the next section. That symmetries and invariances play a crucial role in the determination of probabilities is of course quite obvious, just think of games of chance or Maxwell's derivation of the velocity distribution in an ideal gas. Of course, for many phenomena the underlying symmetries may not be fully known, which then requires resorting to relative frequencies as a weaker kind of evidence. Referring to the examples of the previous section, population statistics constitutes a typical case of a frequentist approach to the measure, while the die is a good example for a symmetry approach.

But how exactly the notion of symmetry must be framed in a probabilistic context is not entirely clear from the relevant literature. Let me therefore define as the most basic, if not yet fully general notion of a *causal symmetry*:

> *A causal symmetry with respect to a probabilistic phenomenon exists if the probability distribution, as determined by the weighted fractions of outcome states, is invariant under a permutation*[37] *of the attribute space—corresponding to a mere relabeling of the outcome space while the collective conditions determining the causal structure of the probabilistic phenomenon remain unchanged.*

In other words, a causal symmetry consists in a possible relabeling of the attribute space that leaves the relevant causal structure unchanged. The idea that invariance under reformulations can fix a probability distribution has long been used with respect to epistemic symmetries in belief states, reaching back at least to the work of Bolzano (1837/1972, § 161; see also e.g. Jaynes 2003, Ch. 12; Norton 2007). Above, the same kind of reasoning was employed with respect to objective causal symmetries.

Only causal symmetries—in contrast to symmetries in belief states—imply the truth of counterfactual statements, such as: If trials of a probabilistic phenomenon were carried out with a different labeling, the probability distribution would remain the same, i.e. any event M_X according to the old labeling would have the same probability as the event M_X according

[37] A generalization to continuous attribute spaces is straightforward.

to the new labeling. With respect to the account of eliminative induction sketched in Section 3b, counterfactual invariance is established by showing the irrelevance of a change in circumstances, in this case of the relabeling of the outcome space, for the causal structure of the probabilistic phenomenon as determined by the collective conditions.

The definition of a causal symmetry directly implies a *principle of causal symmetry* as an objective variant of the principle of indifference:

> *In the case of a causal symmetry regarding the exchange of two attributes, these attributes have equal probability.*[38]

Admittedly, the principle verges on tautology, given the previous definition of a causal symmetry. However, the crucial point is that causal symmetries can often be established non-probabilistically, e.g. on the basis of the laws of classical mechanics as in the paradigmatic cases of throwing dice and coins or of a roulette wheel.

As a simple example, consider the fair throw of a fair die. The attribute space consists in the numbers 1 to 6, located on the different sides of the die. Now, a well-established physical symmetry exists that the numbers on the sides can be permuted in arbitrary ways without affecting the probability distribution, given typical processes of choosing initial conditions and of throwing the die. This symmetry can be justified by referring to well-known laws of classical mechanics, e.g. concerning the mixing of trajectories in certain dynamical systems. Given the principle of causal symmetry, it follows immediately that all attributes must have the same probability 1/6. It is straightforward to apply this type of reasoning to more complex geometrical structures, e.g. a triangular prism with three congruent rectangular sides and two congruent equilateral triangles. Clearly, one can deduce from the corresponding symmetry transformations of the attribute space —without having to refer to relative frequencies—that the triangles and the rectangles all have the same probabilities respectively, while not much can be said about the relative probability between rectangles and triangles, except of course that they must add up to one.

The notion of causal symmetry can be extended to more complex transformations of the attribute space including attributes with different probabilities. Such transformations consist in a permutation of the attributes while taking into account the weighted fractions of outcome states with the respective attributes. Let $\{M\} = \{M_1, M_2, \dots, M_n\}$ be the attribute space, with $P(M_i)$ denoting the probabilities given by the weighted fractions of outcome states with attributes M_i. Furthermore, let $\{M'\} = \{M'_1, M'_2, \dots, M'_n\} = T(\{M\}) = \{M_{T(1)}, M_{T(2)}, \dots, M_{T(n)}\}$ be the relabeled attribute space, where T() denotes a permutation of the original attribute space {M}. Let $P'(M'_i)$ denote the probability of attribute M'_i. Under these circumstances, we can define:

> *A generalized causal symmetry with respect to a probabilistic phenomenon exists, if for the probability distribution of the permuted attribute space {M'} we have:*

[38] Note that any permutation can be reconstructed from a sequence of exchanges of attributes. In the case of a continuous attribute distribution and invariance under a certain transformation, the principle of causal symmetry states that an attribute has the same probability as the attribute that it is mapped on.

$P'(M'_i) = P'(M_{T(i)}) = P(M_{T(i)}) * w(M_i \rightarrow M_{T(i)}) = P(M_i)$, where $w(M_i \rightarrow M_j)$ denotes the ratio of weighted fraction of outcome states with attribute M_i to weighted fraction of outcome states with attribute M_j.

To avoid circularity, the relative weights $w(M_i \rightarrow M_j)$ should again be established non-probabilistically, e.g. by means of the laws of mechanics or by causal irrelevance arguments. A corresponding principle of indifference results:

In case of a generalized causal symmetry, we have: $P(M_{T(i)}) * w(M_i \rightarrow M_{T(i)}) = P(M_i)$.

Obviously, the simpler version of a causal symmetry formulated at the beginning of this section results if w =1. Again a generalization to continuous attribute distributions and their invariance under certain transformations is straight-forward.

Consider as an example of a generalized symmetry a die that is labelled '1' on one side and '6' on all other five sides. The attribute space is $\{M\} = \{1{,}6\}$ with $\{P\} = \{\frac{1}{6}, \frac{5}{6}\}$. If the attributes are exchanged $\{M'\} = \{6{,}1\}$ we can calculate as expected $P'(6) = P(6) * w(1 \rightarrow 6) = \frac{5}{6} * \frac{1}{5} = P(1)$ and $P'(1) = P(1) * w(6 \rightarrow 1) = \frac{1}{6} * 5 = P(6)$. Of course, the tricky part is to non-probabilistically establish the causal symmetry and to non-probabilistically determine the relative weights of the attributes w(). In the described case of a die, this is rather simple, since the mechanical symmetry with respect to the six sides is fairly obvious, but certainly most applications will be more complex than that.

Instead of transforming the attribute space one could also introduce a complementary mapping of the space spanned by the range conditions, which leads to a further rendering of the notion of causal symmetry, for example:

A causal symmetry with respect to a probabilistic phenomenon exists if there is a mapping of the space spanned by the range conditions onto a different space, which is still consistent with the collective conditions, leading to a permutation of the attribute space. The attributes that are thereby mapped onto each other have the same probability.[39]

Consider for example the throw of a fair coin with a certain set of input states and a measure. Now, by physical reasoning we know: (i) if for every input state the coin is rotated by exactly 180°, then the attributes after the throw will be exchanged: heads ⇔ tails; (ii) this mapping of the input space is measure-preserving, since for every throw in the original input space there is a corresponding one with equal weight in the mapped input space. Of course, the mapped input space is still consistent with the collective conditions for the fair throw of a fair coin. Finally, let me stress again that causal symmetries are not epistemic judgments in lack of knowledge, but statements concerning the irrelevance of attribute transformations—or,

[39] It is again straight-forward to extend this idea to more complex causal symmetries, where attributes have different weights.

equivalently, transformations of the space spanned by the range conditions—for the causal structure of a phenomenon and in particular for the probability distribution.

4b. Further examples

Let us look at more examples of causal symmetries to show that the notion can be applied widely. An interesting case in point is Maxwell's derivation of the equilibrium distribution for molecular velocities in an ideal gas from symmetry considerations. Here, the attributes are labels corresponding to different velocities $\boldsymbol{v} = (v_x, v_y, v_z)$ and positions in space $\boldsymbol{s} = (s_x, s_y, s_z)$. Various symmetry assumptions enter in the derivation (Maxwell 1860; cp. Strevens 2013, Ch. 1): (i) homogeneity in space, i.e. there is a causal symmetry with respect to all measure-preserving transformations (relabeling) of the considered spatial volume. It follows that the probability distribution is independent of the spatial coordinates within the considered container (and zero outside the container); (ii) isotropy, i.e. there is a causal symmetry with respect to all rotations (and reflections at the origin) of the velocity space. This symmetry implies that all velocities with the same absolute value $\sqrt{|v_x^2 + v_y^2 + v_z^2|}$ have the same probability;[40] (iii) independence of the one-dimensional velocity distributions along the three Cartesian axes: $P(\boldsymbol{v}) = f_x(v_x)f_y(v_y)f_z(v_z) = f(v_x)f(v_y)f(v_z)$. Strictly speaking, only the second equality relies on causal symmetry, the first on probabilistic independence.[41] As elaborated in Section 5b, probabilistic independence can be established by showing the irrelevance of one attribute distribution for the other. For the sake of simplicity, let us assume just two dimensions x and y. A condition for irrelevance is that the probability $f_y(v_y)$ for any v_y has no influence on the probability $f_x(v_x)$ for any v_x. This holds, since in equilibrium the number of collisions with v_y for one of the particles before the collision and v_x for one of the particles after the collision should be equal to the number of collisions with v_x for one of the particles before the collision and v_y for one of the particles after the collision. Due to this relation, which follows from the constancy of the distribution in equilibrium and from symmetry considerations, changing $f_x(v_x)$ has no influence on $f_y(v_y)$ and vice versa. That the probability distribution is the same f(.) for all coordinates again follows from isotropy. Somewhat surprisingly, these relatively weak conditions (i)-(iii) already hint at the correct probability distribution.

Another causal symmetry is evoked in a later derivation of the equilibrium velocity distribution by Maxwell (1867, 63). In equilibrium one should have the following equality for the probability distributions before and after collisions between two particles: $P(\boldsymbol{v_1})P(\boldsymbol{v_2}) = P(\boldsymbol{v_1}')P(\boldsymbol{v_2}')$ under the assumption that momentum and kinetic energy is conserved, e.g.

[40] Maxwell argues: "the directions of the coordinates are perfectly arbitrary, and therefore [the probability] must depend on the distance from the origin alone" (Maxwell 1860, 153). This reasoning is criticized by Strevens (2013, 14) on the grounds that Maxwell's remark supposedly holds for any probability distribution over velocities, which would be an absurd consequence. However, if one understands 'arbitrary' in the sense that the choice of coordinates is irrelevant for the probability distribution, then Maxwell's reasoning is basically correct, evoking a causal symmetry as we had defined it in the previous section.

[41] As pointed out by Strevens (2013, 14), Maxwell's own reasoning in this regard is not entirely convincing, although Maxwell does appeal to independence: "the existence of velocity x does not in any way affect the velocities y or z, since these are all at right angles to each other and independent" (1860, 153).

$v_1^2 + v_2^2 = v_1'^2 + v_2'^2$ and $\boldsymbol{v_1} + \boldsymbol{v_2} = \boldsymbol{v'_1} + \boldsymbol{v'_2}$ if all particle masses are the same. Here, primed quantities refer to the velocities after the collision and unprimed before the collision. Again, the relation is not justified by frequency data but by physical reasoning. In fact, it essentially follows from the definition of equilibrium, i.e. the requirement that collisions between particles shall not change the probability distribution: "When the number of pairs of molecules which change their velocities from $[\boldsymbol{v_1}, \boldsymbol{v_2}]$ to $[\boldsymbol{v'_1}, \boldsymbol{v'_2}]$ is equal to the number which change from $[\boldsymbol{v'_1}, \boldsymbol{v'_2}]$ to $[\boldsymbol{v_1}, \boldsymbol{v_2}]$, then the final distribution of velocity will be obtained, which will not be altered by subsequent exchanges."[42] (Maxwell 1867, 63) The equality $P(\boldsymbol{v_1})P(\boldsymbol{v_2}) = P(\boldsymbol{v_1}')P(\boldsymbol{v_2}')$ can be interpreted as a generalized causal symmetry with respect to transformations of the attribute space $\boldsymbol{v_1} \leftrightarrow \boldsymbol{v'_1}$. It yields direct access to the relative measure $w(\boldsymbol{v_1} \rightarrow \boldsymbol{v'_1}) = \frac{P(\boldsymbol{v_2}')}{P(\boldsymbol{v_2})}$. Since supposedly the Maxwell distribution is the only plausible function satisfying the equality, the argument allows establishing this distribution merely by appeal to physical symmetries.

A further notable example of causal symmetries concerns the ubiquitous binomial distribution for the calculation of k successes in n trials of an event with probability p: $P_{n,p}(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$. For the sake of simplicity let us focus on the special case p = 1/2. A physical process that generates the corresponding distribution is the Galton board. The essential mechanical symmetry of the Galton board is that at each pin there is no difference between a ball going right or left. Therefore, there is a causal symmetry for each pin i that the probability distribution will not change if one exchanges the labels left l and right r. It follows from the principle of causal symmetry for all i: *P(l|i) = P(r|i) = 1/2*. Based on this insight, the distribution of balls at each level n of the Galton board can be calculated in a purely combinatorial manner by tracing the possible trajectories of the balls through the board. The resulting recursive formula denotes a rather complex causal symmetry that allows to completely determine the binomial distribution at each level $P_n(k) = \frac{1}{2}[P_{n-1}(k-1) + P_{n-1}(k)]$ with $P_0(0) = 1$, $P_n(-1) = P_n(n+1) = 0$. Let me stress again that in deriving the probability distribution for the Galton board we need not make reference to any frequency data whatsoever.

Note that the mentioned complex symmetry does not immediately fit into the framework described in the previous Section 4a, since the recursive formula relates distributions for different levels n. But it is straight-forward to reformulate it in a way that it fits with the form of generalized causal symmetries $P\left(M_{T(i)}\right) * w\left(M_i \rightarrow M_{T(i)}\right) = P(M_i)$. Special cases follow directly from further mechanical symmetries of the physical set-up, e.g. $P_n(k) = P_n(n-k)$. A generalization to $p \neq \frac{1}{2}$ is also straight-forward if one can establish a causal symmetry of the form $P(l|i)p = P(r|i)(1-p)$.

To conclude, let me stress again that the reasoning in these examples does not rely on an epistemic principle of indifference but rather on an objective principle of causal symmetry. Causal symmetries do not refer to lack of knowledge, but follow from the invariance of the

[42] The same relation was used above when arguing for mutual independence of the one-dimensional velocity distributions.

causal structure determining the probability distribution under certain transformations of the attribute space.

4c. The principle of causal symmetry

In Section 4a, I defined the notion of causal symmetry and based on it a principle of causal symmetry as an objective version of the principle of indifference. In its simplest form the principle of causal symmetry states that given a causal symmetry one should ascribe equal probabilities to the corresponding attributes.

How does the epistemic version of the principle of indifference fit into the picture, i.e. the principle of insufficient reason that we should ascribe equal probability when our knowledge about a process does not favor one or the other outcome? Note that there seem to be clear-cut examples, where this epistemic version is employed, for example in Laplace's treatment of the loaded coin: In lack of evidence regarding the way in which the coin is loaded, so the reasoning goes, we should ascribe equal probability to both sides (cp. Section 2a).

Several authors like Cournot or Strevens suggest grounding the distinction between epistemic and ontic probabilities on whether they have been established by an epistemic or an objective version of the principle of indifference, respectively. By contrast, I will now argue that apparent applications of the principle of insufficient reason yield the same results as the principle of causal symmetry whenever the resulting probabilities are predictive.[43] The key idea lies in constructing an adequate collective so that the principle of causal symmetry can be applied. Here, predictiveness requires two things, (i) that the causal structure in terms of collective conditions is sufficiently specified to warrant an unambiguous ascription of causal probabilities according to the definitions given in Section 3b and (ii) that these collective conditions are compatible with the actual conditions realized in the considered event(s). By means of the law of large numbers, predictiveness then implies certain limiting frequencies to be realized in the world for the specified collective.

As an example, assume that we know to which extent a coin is loaded, say p=2/3, but do not know in which direction. As mentioned, it seems a straight-forward application of the principle of insufficient reason, when one ascribes probability 1/2 to both heads and tails before the first throw. However, we can also construe an adequate collective to subsume the reasoning under the principle of causal symmetry. The collective conditions should include the premise that the coin is loaded, while the measure ascribes equal weight to both possibilities p(heads)=1/3 and p(heads)=2/3. The set-up corresponds to a probabilistic phenomenon, where we are given two coins that are loaded in opposite ways, randomly pick

[43] One may be tempted to speak of a reduction of the principle of insufficient reason to the principle of causal symmetry whenever the probabilities are predictive, but such a formulation can be misleading. The starting point of the present section is the question how the principle of insufficient reason could be supplemented or changed such that the notorious Bertrand type ambiguities disappear. A clear criterion how much causal structure is necessary for this task is given in terms of causal symmetries. Of course, this strategy is somewhat opposed to the original idea and spirit of the principle of insufficient reason, namely to assign probabilities on the basis of ignorance, no matter how little we know about a phenomenon. Such a universal principle of insufficient reason is not sensible according to the proposed approach. By contrast, if one insists on the latter, which is of course possible (Shackel 2007 is an example, referring himself to van Fraassen), as a consequence one will always be stuck with Bertrand type ambiguities.

one of them, and throw it. With respect to this collective and measure, a probability 1/2 for both heads and tails results.

When *we know what we don't know* in terms of causal influences on the probability distribution, i.e. when the lack of knowledge can be expressed in terms of causal conditions, one can always proceed in this manner, i.e. construct a collective that accounts for the lack of knowledge and determine the corresponding probability distribution. Of course, lack of knowledge can come in different degrees. For example, it might be the case that we are only given a probability distribution for the extent to which the coin is loaded. But again, this knowledge already determines the measure and thus an appropriate collective.

Apparently, there are two types of situations, (i) when the collective refers only to conditions that are known to be realized in the considered event(s) and (ii) when for some conditions it is unknown whether they are realized and thus they have to be postulated (cp. also Section 6b). As an example, the two coins that are loaded in different directions could both really exist, e.g. lie on a table before us. Or, there could be just a single coin of which we do not know in what direction it is loaded and the origin of which is unclear. In the latter case, the process of randomly choosing between two coins has to be postulated to avoid contradictions, since otherwise a collective and measure cannot be assigned. Certainly, the resulting probabilities are only predictive, if the postulated collective conditions are compatible with the partly unknown actual conditions of the considered event. One might be tempted to ground the distinction between the epistemic principle of insufficient reason and the ontic principle of causal symmetry on this difference between an actual and a postulated collective. But note that conventionally the principle of insufficient reason does not require constructing a causal collective. Also, the mentioned distinction is certainly not sharp but rather blurry, since clearly it is somewhat contextual whether one considers a collective actual or postulated. In any case, the distinction cannot serve to establish a substantial difference between epistemic and ontic probabilities.

Are there applications of the principle of insufficient reason that cannot be accounted for in terms of the principle of causal symmetry? These must be instances where the collective is not sufficiently specified to warrant the ascription of probabilities. In other words, *we do not know what we don't know* in terms of causal influences on the probability distribution. But if collective and measure are underdetermined then we are immediately confronted with Bertrand-type paradoxes. Consider the notorious example concerning the probabilities of different colors, e.g. red, blue, and green. Do red and non-red have the same probability according to the principle of insufficient reason? That cannot be since it would be incompatible with the analogous case that blue and non-blue have the same probability. According to the perspective of this essay, such contradictions arise because the causal context is not specified in terms of collective conditions, range conditions and measure insofar as they are relevant to the probability distribution of attributes. Without the causal context, the principle of indifference leads to contradictions and thus cannot be meaningfully applied.

Thus, Bertrand-type paradoxes are resolved by rendering probabilities relative to a collective, i.e. essentially by the requirement that the causal set-up is sufficiently specified. Consider another classic example dating back to Joseph Bertrand himself (1889, 4-5): What is the probability that the length of a random chord in a circle is shorter than the side of an equilateral triangle inscribed in the same circle? Bertrand points out that there are various incompatible answers depending on which measure one chooses, e.g. equal measure for the distance of the middle of the chord to the center of the circle, equal measure for the angle between chord and the corresponding tangent to the circle, or equal measure for the surface element into which the middle of the chord falls. Again, the ambiguity is resolved by sufficiently specifying the causal process that determines the location of the chord, e.g. the way a stick is dropped on a circle drawn on the floor.

When the causal context is sufficiently specified in terms of collective conditions, then the corresponding probabilities are automatically predictive about the respective probabilistic phenomenon. Also, under such circumstances, every supposed application of the epistemic principle of insufficient reason can be reconstructed as an application of the principle of causal symmetry.[44] By contrast, probabilities resulting from applications of the principle of insufficient reason that cannot be rendered in terms of the principle of causal symmetry are in general not predictive because the causal structure is not sufficiently specified to allow an unambiguous ascription of probabilities.

Note finally that the principle of causal symmetry is not affected by another standard objection against the principle of insufficient reason that it supposedly derives something from nothing, namely probabilities from ignorance. Rather, the principle of causal symmetry presupposes considerable knowledge in terms of causal circumstances in order to establish probabilities that are predictive for a specific probabilistic phenomenon. Henceforth, we suggest excluding from the theory of probability all cases where the relevant context in terms of collective conditions is not specified and therefore predictiveness cannot be guaranteed.

5. Causal irrelevance and probabilistic independence

5a. Independence

As indicated in Section 4, symmetry arguments primarily establish equal measure for the realization of different attributes. However, in order to definitely identify this measure as a probability measure, the independence of trials has to be shown in addition. In the following, I will argue that the causal approach can also throw some light on the notion of independence—an issue that has been called "one of the most important problems in the philosophy of the natural sciences"[45] by Kolmogorov. In a recent paper, Strevens essentially concurs and adds that the "matter has, however, received relatively little attention in the literature" (forthcoming, 3). The notion of independence is a major issue in the controversy between

[44] In this respect, the viewpoint of this essay resembles the position of North (2010), who also denies that there exist distinct objective and epistemic versions of the principle of indifference.

[45] "one of the most important problems in the philosophy of the natural sciences is—in addition to the well-known one regarding the essence of the concept of probability itself—to make precise the premises which would make it possible to regard any given real events as independent. This question, however, is beyond the scope of this book." (Kolmogorov 1956, 9)

subjectivist and objectivist readings of probability. For example, Bruno de Finetti, as a main proponent of subjectivism, aimed to eliminate the essentially objectivist concept of independence altogether and to replace it with exchangeability. In the following, a causal construal of independence will be sketched linking it to causal irrelevance.

For further discussion, it is helpful to distinguish two notions of independence, (i) the independence of consecutive trials of the same probabilistic phenomenon and (ii) independence of random variables associated with different probabilistic phenomena. Roughly speaking, independence of two variables A and B means that (a) one outcome does not affect the other P(A|B)=P(A) or, equivalently from a mathematical point of view, that (b) the corresponding probabilities factorize P(A,B)=P(A)P(B).[46] Independence is often defined in terms of such factorization, for example by Kolmogorov (1956, §5). But certainly this does not solve the difficult methodological question how to determine independence in the world. Why, for example, are two consecutive draws from an urn generally considered independent in case of replacement and otherwise not?

Let us take up a widespread intuition and relate independence to irrelevance. In Section 3b, I argued for a link between eliminative induction and the notion of causal probability. Now, eliminative induction as introduced there also provides a framework for determining causal irrelevance in the sense of difference-making with respect to background conditions. Regarding the first notion of independence (i), consider two trials with the same collective conditions and the same measure. A sufficient criterion for probabilistic independence is:

> *Two trials are probabilistically independent, if the range conditions in one trial are causally irrelevant[47] for the collective conditions in the other trial and thereby for the probability distribution of range conditions in the other trial.[48]*

In other words, arguments based on causal irrelevance shall establish that whatever range conditions are realized in one trial, the probability distribution in the other trial will be the same—which corresponds to the usual framing of independence.

As outlined in the beginning of Section 3b, causal irrelevance can be understood in counterfactual terms: if the range conditions had been different in one trial, the collective conditions and in particular the process determining the range conditions in the next trial would not have changed. In many situations, we have fairly reliable intuitions about such counterfactual statements which are usually evaluated based on the absence of plausible causal influences, as e.g. in the case of a blind-folded person drawing from an urn with replacement or stopping a wheel of fortune several times in a row.

One might object to the above definition that causal irrelevance is not sufficient since there could still be correlations between the range conditions of the first trial and the collective conditions of the second trial that do not result from a direct causal relationship. In particular,

[46] Note that this covers also the first notion of independence (i), if one interprets the consecutive trials as different probabilistic phenomena.

[47] i.e. irrelevant with respect to a causal background constituted by the collective conditions of the first trial.

[48] This definition assumes the absence of a definitional connection between the range conditions of the first trial and the collective conditions of the second trial.

there might be a common cause that influences the range conditions in both trials. However, the notion of causal irrelevance from Section 3b excludes such cases.

To recall, in a context B, in which a condition C and a phenomenon P occur, C was defined as causally irrelevant to P, iff the following counterfactual holds: if C had not occurred, P would still have occurred. Now, in cases with a common cause for C and P, the mentioned counterfactual generally does not have a determined truth value. After all, there are situations in which P would not have occurred if C had not occurred, namely exactly those, in which C and P are due to a common cause. In those situations, the absence of the common cause implies the absence of both C and P. Thus, whenever a common cause exists, there is *no* causal irrelevance and consequently *no* independence of trials.

Note that this line of reasoning is itself not obvious, but depends intricately on the specific understanding of counterfactuals that is employed. David Lewis, for example, would disagree with the above assessment on the basis of his possible-worlds approach to counterfactuals. For reasons that are beyond the scope of this essay, Lewis in his analysis excludes so-called backtracking counterfactuals of the type that if the effect had not happened then the cause would not have happened either. Thus, in the case of a common cause for C and P but in the absence of a direct causal connection between C and P, Lewis would generally claim that if C had not happened, P would still have happened implying causal irrelevance between C and P.

Therefore, a different analysis of counterfactuals is required that was very briefly delineated in the beginning of Section 3b. According to this approach which takes inspiration from the method of difference, backtracking counterfactuals are true if the context fulfills the requirement of homogeneity as also defined in Section 3b. Given homogeneity, the absence of an effect *must* result from the absence of the considered cause.

Thus, the proposed definition for probabilistic independence of trials excludes correlations due to direct causal relevance but also due to common causes. Now what about other kinds of correlations? A further important type does not result from causal dependencies, but rather from definitional relationships. After all, if there is a definitional connection between C and P, of course, there could be correlations as well. But in the case of such relationships, a completely analogous treatment in terms of a counterfactual analysis is possible. After all, the mentioned counterfactual would not be true either, only that a different kind of necessity is involved compared with the case of causal irrelevance.[49]

Last not least, there may be correlations that are neither due to causal nor due to definitional connections between C and P. However, in such cases, it is plausible to assume that the correlations are purely accidental, i.e. that they are merely fluctuations in the observed frequencies that may of course always occur in probabilistic phenomena, even in the case of probabilistic independence. Thus, the proposed account of causal probability again manages to draw the correct distinction between correlations that are meaningful and those that are not.

[49] This treatment can easily be extended to cover still further types of necessity.

Now, the independence of random variables (ii) concerns different probabilistic phenomena that can have different collective conditions. Each random variable is associated with a specific probabilistic phenomenon. A sufficient criterion for independence is:

> *Two random variables are probabilistically independent, if the range conditions in one probabilistic phenomenon are causally irrelevant*[50] *for the collective conditions in the other probabilistic phenomenon, in particular for the probability distribution of range conditions in the other probabilistic phenomenon.*

This criterion broadly stands in the tradition of definition (a) for independence, but it also differs in important respects. Most importantly, it makes reference not to the attribute distribution but to the usually more fine-grained distribution of range conditions. Thus, the evaluation of the criterion is more intuitive since it makes explicit reference to the processes that are causally responsible for the probability distributions of attributes. As an example the throw of a coin and the probability of rain tomorrow are independent, because there is no causal connection between the corresponding processes determining the range conditions in each case. On the other hand, the probability of smoking and the probability of getting lung cancer are in general not independent in an individual, because there is a plausible causal influence from the range conditions of smoking to those of getting lung cancer.

Note again that with respect to the conventional definition of independence the criteria given above are only sufficient but not necessary. As an example, consider two consecutive draws of a ball with replacement. The first ball is drawn arbitrarily from one of two urns B and W both of which have the same ratio of black and white balls. The second draw depends on the result of the first draw. If the ball is black, the next one is drawn from urn B, otherwise from urn W. Now, even though there is some causal relevance of the range conditions in the first draw for the collective conditions of the second draw, the draws are still independent in the conventional sense: for the attribute distribution black/white in the second draw the attribute of the first draw does not matter. The trick is of course that while there is causal dependence, this has no influence on the probability distribution in the second draw.

Thus, one could conceptually distinguish probabilistic independence as framed above in terms of irrelevance of the range conditions from the conventional concept of probabilistic independence referring to the irrelevance of attributes. Of course, the former implies the latter—simply because the attributes are defined on the outcome space spanned by the range conditions. A sufficient and necessary criterion for independence in the conventional sense is:

> *Two trials are probabilistically independent iff the attributes in one trial are causally irrelevant*[51] *for the probability distribution of attributes in the other trial*.

Thus, there may be causal relevance for the collective conditions in the other trial, as long as the resulting collective conditions imply the same probability distribution as in the first trial. For example, in cases, where the method of arbitrary functions can be applied, there may be

[50] i.e. irrelevant with respect to a causal background constituted by the collective conditions of the first phenomenon.
[51] i.e. irrelevant with respect to a causal background constituted by the collective conditions of the first trial.

causal relevance between subsequent initial conditions on a macroscopic scale, which however will be irrelevant for the probability distribution due to microconstancy. Equally:

> *Two random variables are probabilistically independent in the conventional sense iff the attributes in one probabilistic phenomenon are causally irrelevant[52] for the probability distribution of attributes in the other probabilistic phenomenon*.

Essentially, this is only the familiar requirement P(A|B)=P(A), while specifying that the criterion is to be understood in terms of causal irrelevance according to eliminative induction. An example was discussed in Section 4b concerning the mutual independence of velocity distributions along different coordinate axes in an ideal gas at equilibrium.

Note finally that the notions of independence and randomness are closely related. Most importantly: *If subsequent trials are independent, then the sequence of outcomes will be random.* Certainly, this perspective on randomness within the causal approach differs considerably from traditional explications, where randomness has mostly been defined with respect to certain mathematical or formal properties in the sequence of attributes. Von Mises' notion of irregularity, essentially that all subsequences chosen without reference to the attributes must exhibit the same attribute distribution as the sequence itself, and Kolmogorov's work on algorithmic complexity are just two examples in this respect.

In summary, we have suggested how probabilistic independence could be derived from causal irrelevance of probabilistic phenomena as determined by eliminative induction. Of course, these few sketchy ideas cannot fully account for the enormous complexity of the notion.

5b. Interpreting the measure

In one article, Rosenthal describes as the "main problem of the range approach" (2010, 81) that it inherits the circularity of the classical approach to probability in that the measure determining the weights of certain combinations of range conditions itself requires justification in terms of probabilities, i.e. probabilities of initial conditions. For authors like Rosenthal, who argue on the basis of the method of arbitrary functions, the solution is to establish that for certain phenomena, most choices of measure lead to roughly the same probabilities. However, as pointed out towards the end of Section 2b, a number of problems result from this approach. Most importantly, the equivalence of different measures holds only approximately and there are even some measures for which the probability distribution is far off from the correct result. These problems were resolved in the causal approach by rendering probability relative to collective conditions and thereby to the measure over the state space spanned by the range conditions (cf. Section 3b).

For the causal approach the challenge remains to give an interpretation of the measure without having to refer to other concepts of probability, in particular to relative frequencies, which essentially would throw us back on a frequentist account of probability. However, we now have the necessary conceptual tools to tackle this problem. Essentially, the probability measure over range conditions can be construed in terms of causal symmetries in the

[52] i.e. irrelevant with respect to a causal background constituted by the collective conditions of the first phenomenon.

collective conditions[53] and in terms of independence of different trials resulting from causal irrelevance. By means of symmetry arguments, the measure can be quantitatively determined. It is often a measure in time resulting from the system dynamics, which notably may be deterministic and even quite regular. Arguments from causal irrelevance then allow establishing the independence of range conditions in different trials and thereby interpreting the measure as a probability measure. Given the existence of a measure and independence of trials, which constitute the main premises for the various laws of large numbers, the link to limiting relative frequencies can be made via those laws. Note that the second-order probabilities occurring in these laws are not problematic for the causal approach. Rather, they can be interpreted in a straightforward manner in terms of different copies of the same causal set-up as determined by the collective conditions.

Independence of subsequent trials is usually guaranteed by allowing two processes that are causally irrelevant for each other to interfere within the same probabilistic phenomenon, where, as argued before, causal irrelevance excludes direct causal relevance as well as common causes. All probabilistic phenomena (except those with indeterministic dynamics) appear to have such an element that seems required to ensure the random nature of the attribute sequence. Causal irrelevance allows establishing counterfactual statements of the following type: if the range conditions realized in one of the mentioned processes would have been different, the distribution of range conditions of the other process would still have been the same.

A good example is the wheel of fortune as already discussed in Section 3b. The dynamics of the wheel, which is perfectly regular, establishes the measure for the different outcome states of the wheel, in particular equal measure in time for all four colors. Also, the initial conditions determining the rotation of the wheel are causally irrelevant for the moment when the wheel is stopped. To ensure this the person stopping the wheel is blind-folded and all information concerning speed and state of the wheel is withheld from her. Again, the causal irrelevance can be informally tested by evaluating the counterfactual, whether the person would have picked another moment when to stop the wheel had the initial conditions of the wheel been different.

Applications of the method of arbitrary functions can be explained in the same manner. Take the roulette wheel as an example. The assumption of microconstancy requires that slight changes in initial conditions may already lead to a change in attribute. Again, the process determining the measure over initial conditions has to be causally irrelevant for the resulting probability distribution of attributes. Obviously, this is the case when for example we look at what is happening in casinos around the world. The rotation of the roulette wheel is too fast and the dynamics of the ball on the roulette wheel too irregular that the croupier could influence the result by letting the ball enter in a certain way. The advantage of phenomena

[53] One referee has objected that it is not always obvious whether the collective conditions really determine the measure and in particular whether there are the mentioned symmetries in the collective conditions. However, for any probabilistic phenomenon the collective conditions just amount to the instructions which events to include as trials of the phenomenon. If these trials have a stable causal structure then there will automatically be a corresponding probability distribution, which must necessarily correspond to causal symmetries in the collective conditions as should be obvious from the definitions in Section 4a.

falling under the method of arbitrary functions is that the resulting probability distributions are robust, i.e. there is a broad range of processes how to choose the initial conditions of ball and wheel that fulfills the requirement of irrelevance for the probability distribution. Thus, the present analysis does not identify as decisive element in those examples microconstancy or macroperiodicity, but rather the interference of two causally unrelated processes, the rotation of the wheel and the entering of the ball. Again, microconstancy and macroperiodicity just guarantee, that this result is fairly stable across a wide variety of processes. But they are not decisive for the probabilistic nature of the phenomenon.

By contrast, consider the following example originally due to Richard von Mises which for him explicitly does not constitute a probabilistic phenomenon. Let there be a sequence of posts along a road, a large always following a small and vice versa. Certainly, collective conditions can be formulated, e.g. regarding someone driving along the road and writing down the sequence of posts. Also, a symmetry exists with respect to large and small posts. However, at this stage one is not dealing with a probabilistic phenomenon since it lacks the feature of randomness. But again, the latter could be implemented by adding a further causally unrelated process, e.g. a process that puts a person on the road at an arbitrary location to then determine the size of the nearest post.

Thus, the short answer to the problem of circularity is that the measure over different combinations of range conditions designates a probability measure but that it can be construed conceptually in terms of causal symmetries in the collective conditions which quantitatively determine the measure and in terms of causal irrelevance implying the independence of subsequent realizations of range conditions or at least attributes.

6. Ontic and epistemic probabilities

6a. Single-case probabilities and indeterminism

Indeterministic phenomena can easily be integrated into the suggested framework of causal probability. For a fully indeterministic phenomenon, there are no hidden variables, i.e. no range conditions that determine outcome and attribute. More exactly, with respect to definition 2 of Section 3b, there is only one input state determined by the collective conditions and the measure over input space thus becomes trivial. With respect to the terminology introduced in Section 3b, there are no measure conditions and the collective conditions consist only of set-up conditions, which by means of the indeterministic dynamics $S \xrightarrow{c} O$ fix a measure over the outcome space and thus the probability distribution for the attributes. This distribution is given by definition 1 of Section 3b referring to a probability measure over the outcome space instead of the input space.

The orthodox interpretation of quantum mechanics provides a prime example. Via the Schrödinger equation, the collective conditions determine the wave function and thereby the probability distribution upon measurement for certain attributes like position or momentum. The orthodox interpretation explicitly excludes range conditions which would correspond to hidden variables rendering the phenomenon deterministic.

These remarks can also help to clarify the role for single-case probabilities according to the perspective of this essay. In principle, there are no probabilities without collective. However, fully indeterministic events could be viewed as single-case probabilities, since for these a natural choice of collective conditions exists, namely those that maximally determine the probability distribution. Thus, the collective is to some extent already implied by the description of a single event. Note further that according to the causal approach of this essay one can speak of the probability of an event, even though the corresponding probabilistic phenomenon may have occurred only once. As long as one has epistemic access to the measure over outcome space, the phenomenon need not even be repeatable. This distinguishes the causal approach from the naïve frequency view which obviously has to rely on a sufficient number of instantiations.

6b. Epistemic and ontic probabilities

The discussion of indeterminism in the previous section directly leads to one of the basic themes in the debate on interpreting probability, namely the distinction between epistemic and ontic probabilities. As emphasized before, unlike the SRA-approach, the causal framework delineated in this article is meant to extend to cases of indeterminism and also to epistemic probabilities such as probabilities of hypotheses. In fact, causal probability is intended to cover all applications of the probability axioms in which probability is predictive, i.e. in which the main premises for the law of large numbers hold, in particular existence of a measure and independence of trials.

The definitions from Section 3b allow identifying different types of probabilities along the ontic-epistemic spectrum. (i) Purely ontic probabilities are those for which a specific collective is singled out by the statistical event. The typical example concerns indeterminism as discussed in Section 6a, e.g. the decay of a radioactive atom according to the orthodox interpretation of quantum mechanics. In the case of indeterminism, collective conditions exist that maximally determine an event in question with a probability unequal to one, in contrast to deterministic settings where, obviously, the conditions that maximally determine an event yield probability one.

(ii) When the event does not single out the collective conditions (as in the case of indeterminism just discussed), there will automatically be an epistemic element in the choice of these conditions. Most importantly, there remains some leeway, which causal circumstances to consider as collective conditions and which as range conditions, usually implying a change in probabilities. Notably, different probability measures may result from different choices of collective conditions. These epistemic aspects are not problematic for the causal approach since it always relates probability to a specific collective. In this sense, we could still speak of objective probabilities. Note that the mentioned epistemic aspects in principle also exist for the deterministic probabilities established by the method of arbitrary functions, if somewhat less pronounced.

(iii) A further epistemic element concerns the distinction between a situation, where the collective conditions are known to be realized in one or more instances in the world, and situations, where the known conditions of a specific event do not suffice to unambiguously assign a probability and thus additional conditions have to be imagined or postulated in order to construct an appropriate collective. With respect to the example of the two coins that was already discussed, does one actually choose between two coins that are loaded in different ways—or is there only one coin and is the ensemble of two coins just imagined as a subjective range of alternatives? These two situations roughly correspond to the distinction between an objective and an epistemic reading of the principle of indifference, as introduced in Section 4c. In the case that some conditions have to be imagined or postulated, we must resort to statements like: 'if such and such collective conditions are compatible with the considered instance(s), which we do not know for sure, then the resulting probability distribution is predictive with respect to this collective.'

As noted before, the distinction is not sharp and depends considerably on context. But of course, the fewer conditions are known about a phenomenon, the more flexibility exists how to construct the collective—corresponding to a more pronounced subjective element in the assignment of probabilities.

In the following, I will discuss two further variants of epistemic probabilities, first concerning predictions that rely on symptoms instead of the actual causes and second probabilities of hypotheses.

6c. Probabilities from causal symptoms

Sometimes, the space spanned by the range conditions is parametrized not in terms of causes of the probabilistic phenomenon, but rather in terms of symptoms or proxy variables that are somehow causally related. Without loss of generality, this problem is best discussed in terms of definition 2 of Section 3b. A typical example concerns the correlation between barometer and weather. One can quite reliably predict the weather by referring to a barometer reading, but of course the barometer reading is not a cause of the weather. Rather, air pressure is a common cause that influences both barometer and weather. Since air pressure is not easily accessible epistemically, one might be tempted to postulate a probabilistic phenomenon that has as input space the barometer reading and as outcome space a certain parametrization of the weather. While in practice such probabilities predicting from symptoms or proxies of common causes are widespread, let us briefly examine if they are consistent with the viewpoint of causal probability.

Formally, we have an outcome space O, a space spanned by the parametrization of the symptoms I, and an unknown input space S that causally determines the outcome space. In the example above, O would be the weather, I would be the barometer reading, and S would be spanned by some microparameters determining the weather, including air pressure. Two situations need to be distinguished: (i) the symptoms I are fully determined by S; (ii) there are other causes of I that are not in S.

In the first case, probabilities from symptoms easily fit into the framework of causal probability in the following manner. For the sake of simplicity, assume that to any S can be attributed an I. The symptoms I can then be considered as labels of the input space and thus as a reparametrization of the input space, which allows to establish a probability distribution over the attributes based on the symptoms. Note that the mapping I → O will in general not be fully deterministic, i.e. the same value of I can lead to different values of O.

By contrast, such a probability distribution does not exist in the second case, because there are other unrelated causes for I. For example, someone may mechanically interfere with the barometer reading or the spring in the barometer may break. If such external causes are possible, then a probability distribution for the attributes based on symptoms cannot be given. The situation can only be resolved, if one includes in the parametrization of the input space S all possible external causes of I and if one knows the probability measure over those causes. In that case, we can again interpret the symptoms I as a reparametrization of the extended input space and a meaningful probability distribution results for the attributes.

In summary, probabilities from symptoms are only meaningful if they can in principle be reduced to causal probabilities as defined in Section 3b.

6d. Probabilities of causal hypotheses

Thus far, we have treated probabilities of events or types of events as determined by their causal circumstances. But the inductive framework of Section 3b can also cover inverse probabilities, i.e. probabilities of hypotheses regarding possible causes generating the given evidence. The reason is that the eliminative logic underlying causal probability works in both directions—from given causes to possible effects and from given effects to hypotheses about causes. This resolves Humphreys' paradox for the proposed account in a way that corresponds quite closely to a suggestion by Donald Gillies (2000a, 131-133).

Consider again a probabilistic phenomenon determined by certain collective conditions, an input space, a measure W over the input space and a causal mapping from input space to outcome space. When determining the probability of hypotheses, a labelling of the input states must be introduced, which allots these to the different hypotheses $H_1, \ldots, H_N$ (i.e. each hypothesis is about a certain cause being active in some of the input states to bring about a certain outcome). This labelling must be mutually exclusive and must cover the whole input space. If, for the sake of simplicity, it is assumed that the causal mapping is bijective[54], a corresponding labelling of the outcome space results. The causal mapping also determines a measure W_O over the outcome space from the measure over the input space. Relevant evidence leading to an adjustment of the probabilities of the various hypotheses can concern the input space and the outcome space. We can now define:

> *The probability of a causal hypothesis H_X, combining a set of input states of the probabilistic phenomenon P, is given by the fraction of input states weighted with measure W carrying the label H_X or, equivalently, by the fraction of outcome states weighted with measure W_O carrying the label H_X.*[55]

[54] Generalizations are straightforward, e.g. to indeterministic mappings or when it is only surjective. In the latter case, the probabilities have to be calculated in the input space, of course.

[55] Note that the probabilities of hypotheses can be interpreted in terms of probabilities of events, when it is possible to look up which of the hypotheses is actually realized in the world. For example, in the Monty Hall problem discussed below, the corresponding event would consist in opening all doors to verify where the car is.

Let us look at the Monty Hall problem as a simple example for probabilities of causal hypotheses generating a given evidence. In a quiz show, a candidate is presented with three doors A, B, C, behind one of which is a car, behind the two others there are goats. The candidate chooses one of the doors, e.g. A. At the beginning, the evidence conveyed by the quizmaster does not favor any of the hypotheses H_A, H_B, H_C that the car is behind the respective door. In other words, there is a causal symmetry in the set-up of the game with respect to permutations of the doors A, B, C. Consequently, the labels are equally distributed in both weighted input and weighted outcome space, resulting in equal probability for all three hypotheses. Here, the input space is determined by different instances in which the game is originally set up, while the output space is determined by corresponding instances how the game is ended by the candidate.

Now, the quizmaster opens a door, e.g. C, of which he knows that there is a goat behind it and which is not the one chosen by the player. Thereby, new information E is conveyed—which can be accounted for in terms of an additional collective condition. In light of this new condition, the input states which are incompatible with E have to be erased. In particular, all input states associated with hypothesis H_C have to be eliminated, because the truth of H_C is incompatible with the evidence. Furthermore, half of the input states of hypothesis H_A have to be eliminated, namely those, in which the quizmaster would have opened door B. By contrast, none of the input states of H_B are deleted because all of them already imply that the quizmaster opens door C. This leads to the familiar result that in light of the new evidence we have $P(H_A)=1/3$ and $P(H_B)=2/3$.

Obviously, this result can also be calculated via Bayes' Theorem: $P(H_X|E) = \frac{P(E|H_X)P(H_X)}{\sum_{i=1}^{N} P(E|H_i)P(H_i)}$. The quantities on the right side refer to the old collective, $P(H_X|E)$ on the left side is equivalent to the probability $P(H_X)$ relative to the new collective incorporating evidence E. In summary, the change in collective conditions due to novel evidence corresponds to a process of Bayesian updating.

Another example concerns the loaded coin as already discussed in previous sections—except that this time we are not interested in the event of throwing the coin, but in the probability of the two hypotheses H_1 and H_2 that the coin is loaded P(heads)=2/3 or P(heads)=1/3, respectively. Before the coin is thrown for the first time, the evidence does not favor any of the hypotheses and therefore both hypotheses have equal probability 1/2 with respect to a suitably constructed collective. After the first throw, the situation ceases to be symmetric since there is now evidence in which way the coin might be loaded. Again, this evidence can be integrated in the collective conditions leading to a change in measure and thus a new probability distribution over the causal hypotheses. For example, if the result is 'head', then all those input states have to be eliminated that would have led to 'tail' in the first throw, i.e. 1/3 of the input states belonging to H_1 and 2/3 of the input states belonging to H_2. The new probabilities are consequently $P(H_1)=2/3$ and $P(H_2)=1/3$, which is exactly the result given by Bayes' Theorem. From the causal perspective, Bayesian updating can be interpreted as describing how in light of new evidence, which leads to additional constraints in the collective conditions, the measure over the hypothesis space has to be adapted.

Also in the case of probabilities of hypotheses, the ascription of probabilities is predictive only if one specifies collective and measure, i.e. in particular if one knows the complete set of (mutually exclusive) causal hypotheses and if one knows or assumes a measure over these hypotheses that is determined by the collective conditions. Of course, one also needs to know with which probabilities the different hypotheses lead to various pieces of evidence, i.e. essentially the causal mapping of the input to the outcome space. These requirements delineate a fairly restricted range of application for probabilities of hypotheses—excluding for example several 'standard' applications of subjective Bayesianism like the probabilities of abstract scientific theories or hypotheses. Since the range of alternatives is not known in these cases, it seems implausible to construct a collective and relatedly the measure remains undetermined. If one requires probabilities to be predictive, the range of hypotheses to which probabilities should be ascribed is thus rather restricted.[56]

We are therefore in the position to assess the plausibility of the various Bayesian programs from the perspective of causal probability. Sometimes, the hypothesis space and the measure are objectively determined by the causal set-up. Consider for example the following experiment with three urns, each containing both black and white balls but in different ratios, e.g. 1:2, 1:1, 2:1, corresponding to three hypotheses. Now, one of these urns is randomly chosen and then balls are drawn with replacement. Given a certain sequence of draws as evidence, e.g. w-w-b, a probability for each of the three hypotheses can be calculated, whether it holds for the chosen urn. In this specific situation, an objective Bayesian approach is feasible because all relevant elements are determined by the physical set-up: the hypothesis space, the initial probability measure over the hypothesis space, and the probability of evidence given a certain hypothesis is true.

In other circumstances, we might not be so lucky. We may for example be confronted with limited information about a single urn, e.g. that the colors of the balls are only black and white and that there are no more than five balls in the urn. In this case, the hypothesis space is determined by the set-up but there is flexibility in the choice of measure since the actual process with which the urn was prepared is unknown. In analogy to the discussion in point iii) of Section 6b, the Bayesian can now construct in her mind a collective to which the urn is attributed, e.g. an ensemble in which every ratio of balls has equal prior probability. With respect to such a collective, the posterior probabilities of the various hypotheses can then be calculated taking into account additional evidence. However, the Bayesian might just as well have chosen a different measure over the hypothesis space and would have come up with a different result for the posterior probabilities. There is no contradiction, since strictly speaking the probabilities only hold relative to the respective collective and if the collective conditions are compatible with the partly unknown conditions of the considered instances. In cases, where the measure is underdetermined by given knowledge and somewhat arbitrarily construed with respect to an imagined collective, we may plausibly speak of subjective Bayesianism.

[56] An argument in this direction was already given by Popper, who claimed in a reductio ad absurdum that given an infinite number of alternatives, the probabilities of scientific theories would always be zero. See also Pietsch (2014) for a different argument against ascribing probabilities to scientific theories or abstract scientific hypotheses.

Of course, much more should be said how Bayesianism is to be integrated into the framework of causal probability. But the brief discussion above already suggests how the notion of causal probability allows determining the limits of a Bayesian approach.

7. Conclusion

We have proposed a specific account of causal probability that ties in with recent work on objective probabilities in the tradition of the method of arbitrary functions and with earlier accounts mainly from the 19th century, for example by Cournot, Mill, or von Kries. The causal probability of the present essay broadly fits with eliminative induction and the corresponding difference-making account of causation. Probability is interpreted as degree of causal determination of a phenomenon by a given set of conditions. The proposed notion of probability is the following: *The causal probability of a specific attribute M_X of a probabilistic phenomenon P is given by the fraction of outcome states pertaining to attribute M_X, weighted with the probability measure W.*

As a further constraint, we required that one should speak of probabilities only when the respective weighted ratios are predictive, i.e. when the causal structure in terms of collective conditions is sufficiently specified such that probabilities can be unambiguously determined and if the causal structure corresponds to an actual structure in the world. This delineates the range of application for probabilities both of events and of hypotheses. It also allows for a refined version of the principle of indifference, which was termed principle of causal symmetry. Note again that the principle of causal symmetry does not fall prey to Bertrand-type ambiguities exactly because it requires that the causal context is sufficiently specified. Regarding the difficult notion of probabilistic independence a suggestion was sketched how to connect it to causal irrelevance. On this basis, randomness in the attribute sequence generated by a probabilistic phenomenon can be established. The mentioned definition of probability, the notion of causal symmetry, and the causal construal of probabilistic independence should be considered as a coherent conceptual package making up causal probability. In a way, causal probability constitutes an extension of the essentially deterministic framework of eliminative induction and the corresponding difference-making account of causation to statistical and indeterministic contexts.

Acknowledgments

I am much grateful to Jacob Rosenthal and Michael Strevens for helpful comments on an earlier version of this essay as well as to two referees for Erkenntnis who provided very valuable feedback over several rounds of corrections. Last, but not least I want to thank the students of my history and philosophy of probability seminar for many insightful discussions.

References

Abrams, Marshall. 2012. “Mechanistic Probability.” *Synthese* 187(2):343-375.

Abrams, Marshall. 2015. “Probability and Manipulation: Evolution and Simulation in Applied Population Genetics.” *Erkenntnis* 80(3 Suppl.):519-549.

Bacon, Francis. 1620/1994. *Novum Organum*. Chicago, Il: Open Court.

Baumgartner, Michael & Gerd Graßhoff. 2004. *Kausalität und kausales Schließen*. Norderstedt: Books on Demand.

Beisbart, Claus. 2016. A Humean Guide to Spielraum Probabilities. *Journal for General Philosophy of Science* 47(1):189-216.

Berkovitz, Joseph. 2015. “The Propensity Interpretation of Probability: A Re-evaluation.” *Erkenntnis* 80:629-711.

Bernoulli, Jacob. 1713. *Ars conjectandi*. Basel: Thurneysen.

Bertrand, Joseph. 1889. *Calcul des probabilitiés*. Paris: Gauthier-Villars.

Bolzano, Bernard. 1837/1972. *Theory of Science.* Berkeley, CA: University of California Press.

Cournot, Antoine-Augustin. 1843. *Exposition de la théorie des chances et des probabilities*. Paris: Hachette.

Fioretti, Guido. 1998. “John Maynard Keynes and Johannes von Kries.” *History of Economic Ideas* VI/1998/3: 51-80.

Gillies, Donald. 2000a. *Philosophical Theories of Probability*. London: Routledge.

Gillies, Donald. 2000b. “Varieties of Propensity.” *The British Journal for the Philosophy of Science* 51:807-835.

Glynn, Luke. 2010. “Deterministic Chance.” *The British Journal for the Philosophy of Science* 61:51-80.

Hacking, Ian. 1971. “Equipossibility Theories of Probability.” *The British Journal for the Philosophy of Science*, 22(4):339-355.

Hájek, Alan. 2007. “The Reference Class Problem is Your Problem Too.” *Synthese* 156:563-585.

Heidelberger, Michael. 2001. “Origins of the Logical Theory of Probability: von Kries, Wittgenstein, Waismann.” *International Studies in the Philosophy of Science* 15:177-188.

Herschel, John F. W. 1851. *Preliminary Discourse on the Study of Natural Philosophy*. London: Longman, Brown, Green, and Longmans.

Hoefer, Carl. 2007. “The Third Way on Objective Probability: A Skeptic’s Guide to Objective Chance”. *Mind* 116:549–596.

Hopf, Eberhard. 1936. „Über die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie.“ *Jahresbericht der Deutschen Mathematikervereinigung* 46:179-195.

Humphreys, Paul. 1985. “Why Propensities Cannot be Probabilities.” *The Philosophical Review* 94:557-570.

Illara, Phyllis & Federica Russo. 2014. *Causality. Philosophical Theory Meets Scientific Practice.* Oxford: Oxford University Press.

Jaynes, Edwin T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Johns, Richard. 2002. *A Theory of Physical Probability*. Toronto: University of Toronto Press.

Kamlah, Andreas (1983) "Probability as a Quasi-theoretical Concept: J. v. Kries' Sophisticated Account After a Century." *Erkenntnis* 17:135-169.

Keynes, John M. 1921. *A Treatise on Probability*. London: Macmillan.

Kolmogorov, Andrey N. 1956. *Foundations of the Theory of Probability.* New York: Chelsea Publishing Company.

Laplace, Pierre Simon de. 1814. *Essai Philosophique sur les Probabilités*. Paris. Translated into English as *A Philosophical Essay on Probabilities*. New York: Wiley, 1902.

Lewis, David. 1994. "Humean Supervenience Debugged". *Mind* 103:473-90.

Maxwell, James C. 1860. "Illustrations of the Dynamical Theory of Gases." Reprinted in Steven G. Brush (ed.) *Kinetic Theory. Selected Readings in Physics. Vol. 2* (148-171). Oxford: Pergamon Press.

Maxwell, James C. 1867. "On the Dynamical Theory of Gases." *Philosophical Transactions of the Royal Society of London* 157:49-88.

Mill, John S. 1886. *System of Logic*. London: Longmans, Green & Co.

Myrvold, Wayne. 2011. "Deterministic Laws and Epistemic Chances." In Yemima Ben-Menahem & Meir Hemmo (eds.). *Probability in Physics*. New York: Springer.

North, Jill. 2010. "An Empirical Approach to Symmetry and Probability." *Studies in History and Philosophy of Modern Physics* 41:27-40.

Norton, John. 2008. "Ignorance and Indifference." *Philosophy of Science* 75:45-68.

Pearl, Judea. 2000. *Causality. Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Pietsch, Wolfgang. 2013. "The Limits of Probabilism." In Vassilios Karakostas & Dennis Dieks (eds.). *EPSA11 Perspectives and Foundational Problems in Philosophy of Science* (55-65). Dordrecht: Springer.

Pietsch, Wolfgang. 2014. "The Nature of Causal Evidence Based on Eliminative Induction." In P. Illari & F. Russo (eds.), *Topoi*. Doi:10.1007/s11245-013-9190-y

Pietsch, Wolfgang. 2015. "The Causal Nature of Modeling with Big Data." *Philosophy & Technology*. Doi: 10.1007/s13347-015-0202-2.

Pietsch, Wolfgang. 2016. "A Difference-Making Account of Causation." http://philsci-archive.pitt.edu/11913/

Poincaré, Henri. 1912. *Calcul des probabilités*. 2nd edition. Paris: Gauthier-Villars.

Poincaré, Henri. 1914. *Science and Method*. London: Thomas Nelson and Sons.

Pulte, Helmut. 2016. "Johannes von Kries's Objective Probability as a Semi-classical Concept. Prehistory, Preconditions and Problems of a Progressive Idea." *Journal for General Philosophy of Science* 47(1):109-129

Rosenthal, Jacob. 2010. "The Natural-Range Conception of Probability." In Gerhard Ernst & Andreas Hüttemann (eds.). *Time, Chance and Reduction. Philosophical Aspects of Statistical Mechanics* (71-91). Cambridge: Cambridge University Press.

Rosenthal, Jacob. 2012. "Probabilities as Ratios of Ranges in Initial-State Spaces." *Journal of Logic, Language, and Information* 21:217-236.

Rosenthal, Jacob & Carsten Seck (eds.). 2016. *Kries and Objective Probability. Special Issue of Journal for General Philosophy of Science.* Berlin: Springer.

Shackel, Nicholas. 2007. “Bertrand’s Paradox and the Principle of Indifference.” *Philosophy of Science* 74(2): 150-175.
Spirtes, Peter, Clark Glymour & Richard Scheines. 2000. *Causation, Prediction and Search*. Cambridge, MA: M.I.T. Press.
Strevens, Michael. 1998. “Inferring Probabilities from Symmetries.” *Noûs* 32:231-246.
Strevens, Michael. 2006. *Bigger Than Chaos*. Cambridge, MA: Harvard University Press.
Strevens, Michael. 2011. “Probability out of Determinism.” In Claus Beisbart & Stephan Hartmann (eds.). *Probabilities in physics*. Oxford: Oxford University Press.
Strevens, Michael. 2013. *Tychomancy*. Cambridge, MA: Harvard University Press.
Strevens, Michael. Forthcoming. „Stochastic Independence and Causal Connection.“ *Erkenntnis*. Preprint: http://www.strevens.org/research/prob/Interdependence.pdf
Suárez, Mauricio. 2011. „Propensities and Pragmatism.“ *Journal of Philosophy* 110(2):61-102.
Waismann, Friedrich. 1930/1931. „Logische Analyse des Wahrscheinlichkeitsbegriffs.“ *Erkenntnis* 1:228-248.
Williamson, John. 2005. *Bayesian Nets and Causality: Philosophical and Computational Foundations.* Oxford: Oxford University Press.
Van Fraassen, Bas C. 1990. *Laws and Symmetry*. Oxford: Oxford University Press.
von Kries, Johannes A. 1886. *Die Principien der Wahrscheinlichkeitsrechnung*. Tübingen: Mohr Siebeck.
von Mises, Richard. 1981. *Probability, Statistics and Truth.* New York, NY: Dover.
von Plato, Jan. 1983. “The Method of Arbitrary Functions.” *The British Journal for the Philosophy of Science* 34:37-47.
von Smoluchowski, Marian. 1918. “Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik.“ *Die Naturwissenschaften* 6:253-263.