

Appeared as Chapter 27 in *EPSA15 Selected Papers: The 5th Conference of the Philosophy of Science Association*, M. Massimi, Romeijn, J-W., and Schurz, G. (eds.), Springer, 2017, pp. 335-345.

Propensities, Probabilities, and Experimental Statistics ¹

5,169 words

Mauricio Suárez

Department of Logic and Philosophy of Science,
Complutense University of Madrid, 28040 Madrid, Spain

Email: msuarez@filos.ucm.es

Abstract: I defend a three-fold form of pluralism about chance, involving a tripartite distinction between propensities, probabilities, and frequencies. The argument has a negative and a positive part. Negatively, I argue against the identity thesis that informs current propensity theories, which already suggests the need for a tripartite distinction. Positively, I argue that that a tripartite distinction is implicit in much statistical practice. Finally, I apply a well-known framework in the modelling literature in order to characterize these three separate concepts functionally in terms of their roles in modelling practice.

Keywords: chance, propensities, probability, statistics, modelling

¹ I thank audiences at the BPS 2015 conference in Manchester and the EPSA15 conference in Dusseldorf for their comments and reactions. Thanks also to the other members of the symposium panel at EPSA15: Luke Glynn, Aidan Lyon, and Philip Dawid, as well as two anonymous referees. Research towards this paper was funded by a Marie Curie personal grant from the European Commission (FP7-PEOPLE-2012-IEF: Project number 329430), and research project FFI2014-57064-P from the Spanish Government (Ministry of Economics and Competitiveness).

1. Pluralism about Objective Probability

Rudolf Carnap (1945, 1950) was one of the first analytical philosophers of science to openly defend and promote the view that there is not just one kind of probability but a variety of kinds; and correspondingly not just one “probability” concept, but a plurality of concepts. Carnap’s pluralism was modest: having rejected one concept, he settled for the next number up, namely two concepts, so minimizing the variety as much as possible. He characteristically referred to these two concepts by means of indexes, as probability₁ and probability₂. Probability₁ is applicable to the confirmation of theories by empirical evidence, and more particularly to the confirmation of theoretical sentences by so-called protocol sentences. Thus we say that a particular theory is more or less probable in the light of evidence; and that it is more or less probable than some competitor in the light of such evidence; and we may even have reason to assert that its degree of confirmation, or probability, is 0.9 or some other such value in the real unit interval. The first type of probability is thus not a mind or language independent feature of the world. It is rather a feature of our theories or linguistic descriptions of the world. In other words the term “probability₁” belongs in what Carnap called the formal mode of speech (Carnap, 1935/37).

The second kind of probability, or “Probability₂”, is by contrast a mind or language independent objective feature of the world. It depends on the way the world is constituted and what the facts are, regardless of our language, cognitive or mental states, beliefs, attitudes or abilities. In other words, “probability₂” is a term that belongs in the material mode of speech, and appears in ordinary descriptions of the objective probabilities or chances of particular events. Science merely extends this ordinary use of language in order to describe particular phenomena as stochastic by means of statistical or probabilistic models. The statements regarding “probabilities” that appear in scientific models – in physics and elsewhere – are therefore all *prima facie* “probability₂” statements.

Carnap went on to associate these statements to statistical frequencies in the empiricist tradition of Von Mises (1928) or Reichenbach (1935). We nowadays

think that no mere statistical interpretation of, say, the quantum state vector, or the probabilities that it entails, can be made to work. However, for Carnap the fortune of a frequency interpretation of objective probability is a matter of secondary importance – and he was in fact acutely critical of some key aspects in Reichenbach’s empiricist account. Carnap’s main concern was not to defend frequencies, but genuine objective probabilities. And while his particular contrast between logical and frequency concepts of probability did not perhaps succeed well, the overall two-fold pluralism did. Thus twenty years on, we find Ian Hacking (1975) drawing a similar two-fold distinction between subjective and objective aspects of probability. More recent work in the philosophy of probability (e.g. Gillies, 2000) if anything entrenches this kind of pluralism, as a positive state of things to be celebrated.

2. Reductive Analyses of Chance

The philosophy of objective chance has throughout much of its history pursued a reductionist agenda. Some philosophers have attempted to reduce objective probabilities or chances to frequencies or ratios in (virtual or real) sequences of experimental outcomes; others have attempted to reduce them to propensities, understood as the probabilistic dispositions of chancy set ups or arrangements.² Such reductive exercises are at least prima facie contrary to appearances. Consider a few statements of paradigmatic objective chances as expressed in the material mode of speech:

- 1) A coin’s propensity to land heads with a certain probability when tossed as displayed in a long sequence of tosses.

- 2) Smoking’s propensity to cause lung cancer with a certain probability, as demonstrated by control population statistics.

² Lewisian analyses of chance in the spirit of Hume may be regarded as a variety of frequency accounts for the purposes of this paper.

- 3) The propensity of a radioactive atom to decay with a certain probability exhibited in experiments run on the material.

There are obvious differences between the cases. The first statement describes an ordinary or everyday chance; the second one involves a chance to cause a particular effect; and the final statement refers to a putatively fundamental and therefore irreducible chance in atomic physics.³ Nevertheless all these statements appears to involve three distinct properties: the “propensities” of the chancy object; the “certain probabilities” that such propensities give rise to; and the (finite, actual) frequencies of the corresponding outcomes observed in an experimental trial which display such probabilities.

In other words, the “appearances”, as I shall call them, involve three distinct properties. Yet, reductive analyses of chance (frequency and propensity interpretations of probability) aim to reduce them all to just one, or at best two. On the frequency interpretation propensities are redundant and can be discarded altogether; and probabilities can be fully analysed in terms of either long run actual frequencies, or hypothetical limiting frequencies. Carnap refers to such an identification of probabilities with frequencies as the “identity conception” (Carnap, 1945, p. 527). And while there is debate amongst different frequency schools, in particular regarding the status and nature of the limiting hypothetical frequencies, they are all agreed on the essential facts about reduction. On any of these schools there are only really frequencies; every reference to any other apparent concept in the statements above is in fact redundant.

On the other hand the propensity interpretation of probability defended by Karl Popper (Popper, 1959) notoriously embraced a similar (but incompatible) identification of probabilities with propensities, which I have elsewhere referred

³ One may in turn wonder whether all bona fide chances ultimately reduce to physical chances. The answer turns on the thorny question of whether the “special” sciences, and indeed ordinary cognition of macroscopic objects and phenomena, ultimately reduce to physics. I very much doubt such reduction is possible or desirable, but my claims in this paper are independent and require neither reductionism to physical chances, nor its denial.

to as the “identity thesis” (Suárez, 2013). On this view, there are of course finite frequencies in actual experimental runs of any experiment, but they need have no limiting properties. Probabilities are at any rate not to be identified with either the actual or the hypothetical limiting frequencies. They are instead propensities. So, on this account there are only really frequencies and propensities; any apparent reference in the statements above to “probability” as a distinct kind or property is ultimately redundant.

Each of these reductions has had formidable champions throughout the history of the subject; in fact barely any philosopher of probability has failed to attempt one or another version of this reduction of chance. Yet, there are by now very strong arguments against both kinds of reduction, which suggest that the prospects of a reduction of probability are dim. I shall here only review arguments to the effect that probability cannot be reduced to propensity. But the arguments by Alan Hajek and others against frequency interpretations of probability are at least as convincing.⁴ All three concepts (propensity, probability, frequency) seem to be required for a satisfactory understanding of objective chance.

My main claim in this essay is that what Carnap called probability₂ is not in fact a monolithic notion. It too is plural, and composed of an array of three different concepts holding interestingly complex relations to each other. In addition I do of course accept subjective probabilities or credences, and perhaps also distinct logical or epistemological probabilities (confirmatory probabilities). In other words, I very much share Carnap’s pragmatic pluralism, but whereas Carnap tried to minimize the pluralism by restricting it to two kinds of probability, I find good reasons nowadays to want to maximize the pluralism in order to achieve a full understanding of objective chance. There are both negative and positive reasons for maximal pluralism. The negative reasons have all to do with

⁴ Some of Hájek’s arguments (1997) rely on the well-known reference class problems. I am not so interested in them here because they leave open any claim regarding a reduction to propensities, and I am arguing for a full tripartite distinction.

the failures of reductive programmes (section 3). The positive reasons are connected with the presuppositions of scientific practice (section 4)

3. Against the Identity Thesis

Let me briefly review the argument from the philosophy of probability against the identity thesis between propensities and probabilities.⁵ The identity thesis has two parts, or halves, which we may refer to as the propensity-to-probability half and the probability-to-propensity half. The former asserts that all propensities are, or can be represented as, probabilities. The latter states that all probabilities are propensities, or can be interpreted as such. Together they make the full claim that probabilities and propensities are extensionally identical.

Both parts of the identity thesis are in fact false, as is shown by different forms of what is known as Humphreys' paradox. The falsity of the probability-to-propensity half is a trivial consequence of the asymmetries of propensities. This is best understood by considering a causal propensity such as smoking's propensity to cause lung cancer (my example 2 above). Suppose we estimate for a particular population the incidence of lung cancer amongst smokers at 1%, which we may write as $P(C/S) = 0.01$. And suppose that we also have estimates for the prior probabilities of smoking and lung cancer across the population at, say, 20% and 0,5% respectively ($P(S) = 0.2$ and $P(C) = 0.005$). We may then easily estimate the inverse probability by means of Bayes' theorem:

$$P(S/C) = \frac{P(C/S)P(S)}{P(C)} = \frac{0.01 \times 0.2}{0.005} = 0.4.$$

Now, the first half of the identity thesis (the probability-to-propensity half) holds that probabilities may be interpreted as propensities. If so, $P(C/S)$ may be understood as the propensity of smoking to cause cancer. But it follows from our

⁵ The full argument may be found in Suárez (2013, 2014) of which this section is an elaboration and summary.

derivation that $P(S/C)$ is then also well-defined at 40%, so it must also receive a propensity interpretation, which seems just impossible: There is simply no propensity of lung cancer to cause smoking.

While this simple type of argument is well known the consequences for the identity thesis are not always fully appreciated. Propensities are asymmetric in a way that probabilities are not. The asymmetry is revealed most strikingly in the case of causal propensities, but is more generally a feature of all propensities whether or not causal. Thus a coin's propensity to land heads, and a radioactive atom's propensity to decay are also asymmetric in a way that generates a similar problem for their Bayes inverse probabilities. We may hope that the other half of the identity thesis (the propensity-to-probability half) holds nonetheless, and that it is still the case that all propensities are probabilities. While this half on its own fails short of a full reductive analysis for probabilities, it is still an account of propensities as probabilities.

However, Paul Humphreys' original argument defeats this half of the identity thesis as well. Humphreys (1985) considered a thought experiment where, regardless of the outcome, the propensities of the system described are not and cannot be represented as probabilities. He considered a source emitting one photon at a time t_1 , reaching a half silver mirror at time t_2 , and being transmitted at time t_3 . He then plausibly stipulated that the following three claims hold regarding the propensities of the photon in the thought experiment:

- i) Any photon that reaches the half silver mirror has some finite (non-zero) propensity to be transmitted.
- ii) Any photon that is emitted has some propensity greater than zero but not one to reach the mirror.
- iii) Any photon that is emitted and fails to reach the mirror has propensity zero (i.e. it has no propensity) to be transmitted.

These claims may all be regarded as uncontroversial regarding the thought experiment at hand. They all certainly seem very plausible. Humphreys rendered

these claims in a conditional probability formulation that is however far from innocuous or obvious, as follows:

i) $P_{t_3}(T_{t_3}/I_{t_2} \& B_{t_1}) = p > 0.$

ii) $1 > P_{t_1}(I_{t_2}/B_{t_1}) = q > 0.$

iii) $P_{t_1}(T_{t_3}/\neg I_{t_2} \& B_{t_1}) = 0.$

Each of these formal conditions is meant to capture fully each of the corresponding physical claims regarding the propensities at work in the thought experiment. This assumes that there is always a unique representation for propensities in terms of conditional probabilities. Yet, these three formal conditions are inconsistent with the Kolmogorov axioms, and in particular with the fourth axiom for conditional probability (also known as the ratio analysis of conditional probability).

Now, there are a number of caveats to Humphreys' proof, which I cannot discuss here in full, but deserve a brief mention. Firstly, the proof assumes a principle of conditional independence whereby propensities do not act backwards in time: $P_{t_1}(I_{t_2}/T_{t_3} \& B_{t_1}) = P_{t_1}(I_{t_2}/\neg T_{t_3} \& B_{t_1}) = P_{t_1}(I_{t_2}/B_{t_1})$. The principle is indeed questionable in general, since there is no reason why propensities should be any more forward-looking than causes. In other words, the asymmetry of propensities is not the asymmetry of time, but is rather a *sui generis* asymmetry, which may or not coincide with temporal asymmetry. Yet, in the thought experiment at hand, the application of conditional independence is legitimate – in other words the propensities that obtain in the thought experiment are all as a matter of fact forward looking. So the proof does not hang on this assumption being generally valid.

The second caveat is that relinquishing the ratio analysis of conditional probability (i.e. giving up on the Kolmogorov calculus, or at least on the implicated

fourth axiom: $P(A/B) = \frac{P(B/A)}{P(B)}$), does not actually deliver us from contradiction,

and cannot in fact get us out of trouble. The ratio analysis is indeed a presupposition of the Kolmogorov calculus, but no other calculi that we have so far developed is in a better position to overcome Humphreys' paradox.⁶

To sum up, Humphreys' proof is rightly widely understood to show that the representation of propensities in terms of conditional Kolmogorov probabilities is flawed: Propensities are not in general probabilities.

4. Chance Assumptions in Statistical Modelling

The second and main argument for pluralism does not rely on formal issues in the axiomatization of probability. It is rather related to the practice of statistical modelling. I shall argue that the "appearances" (i.e. the tripartite distinction between propensities, probabilities, and frequencies) are tacitly presupposed in much of this modelling practice. And philosophers of science as it happens are well equipped to understand the tripartite distinction as part of any modelling practice – since it follows from an influential account of modelling in general.

A statistical model is often presented as a pair structure, consisting of a sample or outcome space S and a set of probability distributions, or distribution functions P_i , defined over this sample: $\langle S, P_i \rangle$. The domain of each of the probability functions is a subset or power set of the elements in the sample and the range of each probability function is of course the unit real number interval.⁷ However, the simple definition has buried within it a fair amount of tacit structure that is rarely made explicit. In particular the selection of the sample or domain of a statistical model is not a trivial matter and involves considerable judgement. In his influential (2002) paper, Peter McCullagh shows how any statistical model of a stochastic phenomenon involves two strictly distinct domains: the domain of the

⁶ For a very nice treatment of this issue in connection with Renyi's axiom system, see Lyon (2013).

⁷ The literature on statistical modelling is large. My understanding is informed mainly by Cox (2006), Freedman (2009), and Krzanowski (1998), in addition to the McCullagh paper discussed in the text.

phenomenon in question, and the domain of the probability functions contained in the model. The idea is that the phenomenon is first described as a set of parameters Θ , in what may be called a prepared description. A statistical model is then a function that maps each parameter point in Θ onto one of the probability functions $\wp(S)$ defined over the sample space. In other words a statistical model is functionally a map: $P : \Theta \rightarrow \wp(S)$ which assigns to every point $\theta_i \in \Theta$ in the parameter set that represents the phenomenon a corresponding probability function $\wp_{\theta_i}(S)$ defined over the sample space. As McCullagh (2002, p. 1225) notes: “it is important to distinguish between the model as a function $P : \Theta \rightarrow \wp(S)$, and the associated set of distributions $\wp_{\theta_i}(S) \subset \wp(S)$ ”.

A statistical model comprises both the parameter set and the set of probability functions over the sample space. So, implicitly, a statistical model is defined over two distinct domains: Θ and S . The former domain appears merely as a subscript to the probability distribution function. It is the latter domain, the sample or outcome space, that is the proper sigma field over which the probabilities are defined. It follows then that the probabilities in a statistical model are not defined over the parameter set that represents the phenomenon in question. The relationship between the sample space and the parameter set is rather indirect; and the most important question for any modeller is precisely how to ‘convert’ the parameter space meaningfully into the sample or outcome space. There is no trivial algorithmic procedure: It is rather a highly contextual matter of judgement, relative to the particular problem at hand. It effectively transforms a question regarding the phenomenon and its causes into a question regarding the probabilities in the model. The model must of course be consistent with known data relative to the phenomenon, but it is hard to see how it would not involve idealization of one sort or another. This is after all one more instance of ‘modelling’ the phenomena in a streamlined description. To quote from the distinguished statistician David Cox (2006, p. 197):

“Formalization of the research question as being concerned with aspects of a specified kind of probability model is clearly of critical

importance. It translates a subject-matter question into a formal statistical question and that translation must be reasonably faithful and, as far as feasible, the consistency of the model with the data must be checked. How this translation from subject-matter problem to statistical model is done is often the most critical part of an analysis. Furthermore, all formal representations of the process of analysis and its justification are at best idealized models of an often complex chain of argument.”

The most important constraint in statistical modelling is this: The derivation of the sample space from the parameter space must be responsive to the features of the phenomenon. The function that takes from the parameter set Θ to the probability distribution $\wp_{\theta_i}(S)$ is not arbitrary, but depends sensitively upon the nature of the phenomenon in question. In particular, if the phenomenon is dynamical the function must respond to its dynamical laws. And if the laws are stochastic, then it must respond to the objective chances that appear in those laws.⁸ So the function that yields the sample or outcome space of any statistical model of an indeterministic phenomenon already has built into it a dependence upon some antecedent chances, which appear in the phenomenon as described. The model is then of course tested against experiments run upon the system and yielding finite frequencies. It then follows that a tripartite distinction between propensities, probabilities, and frequencies is already implicitly assumed in the practice of statistical modelling.⁹

⁸ Consider as a rudimentary example two fair coins, each independently obeying a binomial distribution. Suppose that the coins are then physically connected in accordance to a dynamical law that implies correlations amongst them (you can imagine some kind of invisible thread connecting both tail sides). They are thereafter always tossed simultaneously and more likely to fall on the same side. The sample space in the statistical model for this phenomenon must then include both outcome events (“head” and “tails”) for each of the coins, as well as all the joint events (“heads & heads” “heads & tails”, etc). And the probability distribution function defined in this formal model must be consistent with these underlying dynamical facts.

⁹ A referee helpfully points out that Spanos (2006) defends a similar distinction between structural theory models, statistical models, and observational data, with similar consequences regarding the role of “chance set-ups”.

5. Saving the Statistical Phenomena

The statistical modelling of a phenomenon thus presupposes a three-fold distinction between the propensities responsible for the phenomenon, the probabilities that appear in the statistical model of the phenomenon, and the (actual, finite) frequencies in the experimental sequences that test the model. This tripartite distinction in statistical modelling practice is a natural consequence of the application to stochastic phenomena of a corresponding tripartite distinction in modelling more generally: Bogen and Woodward's (1988) distinction between theory, phenomena, and data.¹⁰ On their account, which is backed up by a number of detailed cases studies, the main role of a theory is to explain a phenomenon; and while theories can be consistent with data, they are not in the business of explaining or predicting data. Correlatively, an experiment typically yields observable data (that is: finite, actual records of particular observations or measurement outcomes); but the phenomenon itself cannot be so observed. Rather our knowledge of a phenomenon is the result of a number of low-level inferences that establish a particular model for it. The inferred phenomenon is then described in the model and with luck explained by a theory. It should be clear that the account presupposes that there are independent functional roles for each of the three components: theories, phenomena, and data.

The best-known illustration of the tripartite model that Bogen and Woodward provide us with is the old example of the melting temperature of lead.

¹¹ I do not need to review their discussion in great detail for my purposes here. It is

¹⁰ I do not thereby endorse here any of their epistemological claims beyond the tripartite distinction. My account of statistical modelling, for instance, is also consistent – at least for the purposes of the present essay – with the widely accepted claim that models are autonomous relative to both theory and data. See the essays in Morrison and Morgan (1999) for an articulation and defence.

¹¹ They derive the example from Ernst Nagel's (1961) discussion. One of Bogen and Woodward's main claims is that the logical positivist accounts of explanation and confirmation suffers from oversimplification of the empirical content of science. The logical positivist emphasis on "observable phenomena" is, according to Bogen and Woodward, an oxymoron. As explained in the text, phenomena are on their account never observable, but always the result of some low level generalizing inferences.

enough to emphasise that the fact that lead melts at 327.5 degrees Celsius is not a piece of observable data on their account, but may only be inferred from a very complex array of data by some sophisticated data analysis. The data points that are in fact observed correspond to single recordings of measurements taken on particular samples of lead under very particular conditions – and hence subject to huge variation in experimental and / or systematic error. The variation is so large in fact that there may exist no data point corresponding to the precise melting temperature of lead. And, contrary to what the logical positivists thought, no theory in solid-state physics may be able to explain a single data point.

The theory of phase transitions for metals indeed explains the different melting temperatures of the different metals in terms of their intermolecular forces, and it predicts the critical energy required to overcome the molecular bonds typical of a crystalline solid. But on the tripartite account, it does not need to predict or explain each or any particular measurement record. But then the theory is not intended to ever do that. It is rather meant to account for, and explain, the one true claim about the phenomenon, i.e. the inferred fact that lead melts precisely at 327,5 degrees Celsius.

In other words, on this account of modelling practice, there is a clear-cut functional distinction between theories, models of phenomena, and observable data. The theories are in the business of explaining and predicting phenomena, not data. The phenomena are described by means of models that are in turn inferred from the data by complex statistical analyses. The observable data are used to confirm these models of phenomena but they cannot directly be employed in the confirmation (or refutation) of theory. No finite number of contradicting data points can refute a theory, especially when the data taken together vindicate a phenomenon that is compatible with the theory. In this tripartite account the connections between data and theory are always mediated by (a model of) phenomena.

Statistical modelling is one type of scientific modelling practice. So it stands to reason that it should exhibit the same functional distinctions that are operative

in modelling in general. And indeed it does. Statisticians draw the relevant distinctions, particularly when they reflect upon their practice. In particular, as shown in the previous section, they distinguish carefully the formal “statistical model” pair $\langle S, P_i \rangle$ from the phenomenon itself, however parametrized under some theoretical description. With some ingenuity a theory may be found that describes the dynamical properties behind the phenomenon, including its propensities. If so, an explanation is thus provided for the phenomenon as described in a formal statistical model – by means of a set of probability distribution functions defined upon an appropriate sample space. These models are in turn tested against the data – namely the frequency ratios revealed in long but finite sequences of experimental outcomes.¹²

Radioactive decay rates are a good illustration. An atom’s chance or probability of decay is a propensity of the material, as described by atomic theory. The theory invokes such powers as part of the explanation of the typical rate of decay (half life) of the material. It does not thereby explain any particular atom’s event of decay (or otherwise). This is an indeterministic single event that cannot be so explained. And no particular event of decay (or non-decay) can on its own provide any confirmation or refutation for or against atomic theory. The only way data can possibly impinge on theory is indirectly via the probability distribution functions in the statistical model that accounts for the phenomena of radioactive decay.¹³

¹² In our rudimentary two-coin system example, the theory that describes the dynamics of the system (including the hidden mechanism, such as the connecting thread) is not meant to account for, or explain, any particular two-coin outcome. It is only meant to explain the probability distribution that appears in the formal statistical model for the phenomenon. Similarly no particular outcome may refute this theory other than by compromising the distribution function in the model – for which much more than just one observation will certainly be needed.

¹³ How would a subjectivist try to account for these distinctions? One way that occurs to me is via Skyrms’ (1977) notion of propensity as resilient subjective probability. Roughly, a probability is resilient if it is immune to (or invariant under) further conditionalization by admissible evidence. While this type of stability strikes me as a good way to mark out one difference between what I call chances and frequencies (the former being stable in a way the latter are not), I do not see how it can possibly account for the robust form of explanatory power that I here ascribe to propensities in relation to chances.

1. Conclusion

I have argued for maximal pluralism about chance, by providing negative and positive arguments for a tripartite distinction between propensities, probabilities, and frequencies. Humphreys' paradox provides grounds for the distinction, since it makes it very implausible that chances may be reduced or analysed away in any fewer terms. In addition, I have positively displayed elements in the practice of statistical modelling recommending the same distinction. Finally, I have argued that the tripartite distinction makes full sense within an influential current account of modelling practice.

REFERENCES

Bogen, Jim and Woodward, James. 1988. Saving the Phenomena. *The Philosophical Review*, XCVII (3): 303-352.

Carnap, Rudolf. 1935/1937. *The Logical Syntax of Language*. Kegan Paul.

Carnap, Rudolf. 1945. Two Concepts of Probability. *Philosophy and Phenomenological Research*, 5 (4), pp. 513-532.

Carnap, Rudolf. 1950/1966. *Logical Foundations of Probability* (2nd ed. 1966). Chicago: Chicago University Press.

Cox, David. 2006. *Principles of Statistical Inference*. Cambridge: Cambridge University Press.

Freedman, David. 2009. *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.

Gillies, Donald. 2000. *Philosophical Theories of Probability*. Routledge.

Hacking, Ian. 1975. *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hajek, Alan. 1997. Mises Redux – Redux: Fifteen Arguments Against Finite Frequentism. *Erkenntnis* 45: pp. 209-227.

Humphreys, Paul. 1985. Why Propensities Cannot Be Probabilities. *The Philosophical Review* 94: 557-70.

Krzanowski, Wojtek. 1998. *An Introduction to Statistical Modelling*. Chichester: Wiley and Sons.

Lyon, Aidan. 2013. From Kolmogorov to Popper, to Rényi: There's no Escaping Humphreys' Paradox (When Generalized). In *Chance and Temporal Asymmetry*, ed. Alastair Wilson, 112-125. Oxford: Oxford University Press.

McCullagh, Peter. 2002. What is a Statistical Model? *The Annals of Statistics* 30 (5): 1225-1310.

Morrison, Margaret and Mary Morgan. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.

Nagel, Ernst. 1961. *The Structure of Science*. New York: Harcourt, Brace and World.

Reichenbach, Hans. 1935/1949. *The Theory of Probability: An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Los Angeles: California University Press.

Spanos, Aris. 2006. Where Do Statistical Models Come From? Revisiting the Problem of Specification. *IMS Lecture Notes- Monograph Series 2nd Lehmann Symposium – Optimality*, vol. 49, pp. 98-119.

Skyrms, Brian. 1977. Resiliency, Propensity and Causal Necessity. *The Journal of Philosophy* 74 (11): 704-711.

Suárez, Mauricio. 2013. Propensities and Pragmatism. *The Journal of Philosophy* 110 (2): 61-92.

Suárez, Mauricio. 2014. A Critique of Empiricist Propensity Theories. *European Journal for Philosophy of Science* 4 (2): 215-231.

Von Mises, Richard. 1928 /1957. *Probability, Statistics and Truth* (2nd ed. 1957). New York: Dover.