Episteme (2021), **18**, 640–653
doi:10.1017/epi.2019.41

ARTICLE

Direct Inference, Reichenbach's Principle, and the Sleeping Beauty Problem

Terry Horgan*

University of Arizona, Tucson, USA
*Email: thorgan@email.arizona.edu

(Received 25 June 2019; revised 22 September 2019; accepted 28 September 2019; first published online 11 November 2019)

Abstract
A group of philosophers led by the late John Pollock has applied a method of reasoning about probability, known as direct inference and governed by a constraint known as Reichenbach's principle, to argue in support of 'thirdism' concerning the Sleeping Beauty Problem. A subsequent debate has ensued about whether their argument constitutes a legitimate application of direct inference. Here I defend the argument against two extant objections charging illegitimacy. One objection can be overcome via a natural and plausible definition, given here, of the binary relation 'logically stronger than' between two properties that can obtain even when the respective properties differ from one another in 'arity'; given this definition, the Pollock group's argument conforms to Reichenbach's principle. Another objection prompts a certain refinement of Reichenbach's principle that is independently well-motivated. My defense of the Pollock group's argument has epistemological import beyond the Sleeping Beauty problem, because it both widens and sharpens the applicability of direct inference as a method for inferring single-case epistemic probabilities on the basis of general information of a probabilistic or statistical nature.

Keywords: Sleeping Beauty; probability; direct inference; Reichenbach's principle

A familiar version of the much-discussed Sleeping Beauty problem goes as follows. Beauty is a subject in a sleep-lab experiment that will run from Sunday evening until Wednesday morning. On Sunday she learns the following, reliably. The researchers will put her into a dreamless sleep on Sunday night, and will awaken her in the lab on Monday. Later on Monday while she is still in the lab, they will put her back into a dreamless sleep, and then will toss a fair coin. If the Monday evening coin toss lands heads, then they will keep her in a dreamless sleep until Wednesday morning, when she will awaken by herself and know the experiment is over. But if the coin toss lands tails, then they will erase her Monday memories while she is sleeping and awaken her a second time on Tuesday morning; later on Tuesday while she is still in the lab, they will put her back into a dreamless sleep until Wednesday morning, when she will awaken by herself and know the experiment is over. When she finds herself having been awakened by the experimenters, with no memory of a prior awakening and now not knowing whether it is currently Monday or currently Tuesday, what probability should she assign to the proposition that the coin toss comes up heads?

A large group of philosophers, led by the late John Pollock and writing collectively under the name 'The Oscar Seminar' (Seminar 2008),[1] have applied the conceptual machinery of 'objective probability' theory – specifically, what is called 'direct inference' in that theory – to argue in support of 'thirdism'. In subsequent literature spawned by this paper, Joel Pust (2011) argues that the Oscar Seminar's argument is unsound according to the standards of objective probability theory itself; Paul Thorn (2011) defends the argument against Pust's objection; and Kaila Draper (2017) first argues that Thorn's response to Pust's objection is unsuccessful and then offers a further argument of her own for the claim that the Oscar Seminar's argument is unsound according to the standards of objective probability theory itself.

Here I will engage this philosophical dispute, with two principal goals. First, I will defend the Oscar Seminar's argument against Pust's objection, in a way that implements the guiding idea behind Thorn's reply to Pust but also avoids two shortcomings in Thorn's own implementation. (One of these shortcomings was noted by Draper, and the other one is first noted in the present paper.) Second, I will defend the Oscar Seminar's argument against Draper's objection. Both defenses will involve proposing certain elaborations of the theory of direct inference that are independently well-motivated, rather than being ad hoc or question-begging.

In addition, I will argue that the specific way in which the Oscar Seminar's argument deploys direct inference is largely neutral about disputed questions about the nature of probability – and hence that this argument can be endorsed by virtually anyone, regardless of their views about the nature of probability itself. Opponents of thirdism therefore cannot repudiate the argument just by rejecting all versions of "frequentism" about probability – despite the fact that frequentism typically has been endorsed, in some version or other, by proponents of objective probability theory such as Fisher (1922: 311), Popper (1956), Kyburg (1974), Bacchus (1990), Halpern (1990), and Pollock (1990).

The potential import of my discussion extends well beyond the Sleeping Beauty problem. Direct inference is a frequently applicable technique for drawing conclusions about single-case epistemic probabilities on the basis of general information of a probabilistic or statistical nature. If I am right in defending the Oscar Seminar's argument against Pust's and Draper's charge that the argument is a mis-application of direct inference, then direct inference turns out to be considerably more widely applicable than it otherwise would be. And if I am right that this inferential technique is largely neutral about the foundations of probability, then direct inference can be safely employed without a commitment to objective probability theory.

1. Objective probability theory, direct inference, and the Oscar Seminar's argument for thirdism

One version of the theory of objective probability, advocated by Pollock, invokes a distinction between 'definite' probabilities [indicated with small caps 'PROB' as in 'PROB (P/Q)' or 'PROB(P)'] which attach to propositions, and 'indefinite' probabilities [indicated with lower case 'prob' as in 'prob(F*x*/G*x*)'] which attach to properties (including n-adic relations) or propositional functions. What is called 'a theory of direct inference' within objective probability theory becomes, within this version of the theory, an

[1]Adam Arico, Nathan Ballantyne, Matt Bedke, Jacob Caton, Ian Evans, Don Fallis, Brian Fiala, Martin Frike, David Glick, Peter Gross, Terry Horgan, Jenann Ismael, John Pollock, Daniel Sanderman, Paul Thorn, Orlin Vakerelov.

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

account of the conditions under which one is justified in inferring definite probabilities from indefinite probabilities. In 'prob(F*x*/G*x*)' G is called 'the reference property' and F 'the consequent property'. Following the pioneering work of Reichenbach (1949: 374), a core principle is that one should identify the probability that a given particular, *a*, is F (PROB(F*a*)), with prob(F*x*/G*x*), where G is the logically strongest evidentially relevant property such that one knows prob(F*x*/G*x*) and one knows G*a*.

The Oscar Seminar's argument invokes direct inference in defense of thirdism, as follows. Where 'a Sleeping Beauty scenario' is an instance of the Sleeping Beauty problem, the Oscar Seminar's members use 'B(*t*,*s*)' to mean '*s* is a Sleeping Beauty scenario and *t* is a time during *s*' and 'Toss(*x*,*s*)' to mean '*x* is a (the) coin toss involved in *s*'. Where *x* is a coin toss, they use 'H*x*' to mean '*x* lands heads'. And they use 'W(*t*,*s*)' to mean 'Sleeping Beauty is awakened in the scenario *s* during an interval Δ since *t*, without remembering being awakened at any prior time during *s*'. According to the Oscar Seminar, the following indefinite probability claims are both true:

(1) prob(H*x*/B(*t*,*s*) & Toss(*x*,*s*)) = 1/2
(2) prob(H*x*/W(*t*,*s*) & B(*t*,*s*) & Toss(*x*,*s*)) = 1/3

where '*x*', '*t*', and '*s*' are free variables within the formulas to which the operator 'prob (/)' applies, and are bound within (1) and (2) by 'prob' itself.

Let σ be a particular Sleeping Beauty scenario and let τ be the coin toss in σ. On Sunday in σ, according to the Oscar Seminar, Beauty knows B(now, σ) and Toss(τ,σ) and so can conclude by direct inference from (1) that PROB(Hτ) = 1/2. But upon being awakened during the experiment, Beauty comes to know W(now,σ) & B(now, σ) and Toss(τ,σ). Since (2) involves 'a more specific reference property than (1)', the Oscar Seminar claims that Beauty should, by Reichenbach's principle, conclude by direct inference deploying (2) that PROB(Hτ) = 1/3 (2008: 152).

2. The Pust/Thorn/Draper dialectic regarding the Oscar Seminar's argument

Joel Pust (2011) and Kaila Draper (2017) both maintain that the Oscar Seminar's argument for the thirdist answer to the Sleeping Beauty problem is unsound. Neither author contests statement (2), and neither author launches a general objection to direct inference as a technique for drawing conclusions about single-case probabilities. Rather, they both argue that the Oscar Seminar's argument violates the constraints on direct inference that are imposed by the conceptual machinery of objective probability theory itself. Pust's reasoning appeals to the following indefinite probability claim:

(3) prob(H*x*/Toss(*x*,*s*)) = 1/2

Pust argues as follows:

> Upon awakening, Beauty knows Toss(τ,σ) and so she is, given (3), prima facie justified in concluding that PROB(Hτ) = 1/2 … In (2) (and (1)) the reference property is a property of time-Sleeping Beauty scenario-coin toss *triples*, while in (3) the reference property is a property of Sleeping Beauty scenario-coin toss *pairs* … [B]ecause (2) and (3) concern property possession by n-tuples of different n, neither trumps the other as a basis for direct inference. Instead, as (2) and (3) prima facie justify direct inferences to contradictory claims, such inferences defeat each other and neither conclusion is all-things-considered justified. Therefore, Beauty's situation is one in which she cannot engage in an all-things-considered

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

> justified direct inference to PROB(Hτ) = 1/3 or to PROB(Hτ) = 1/2. (Pust 2011: 292)

In response to Pust, Paul Thorn (2011) contends that 'Sleeping Beauty should disregard Pust's direct inference [viz., the prima facie justified direct inference based on (3)], and accept the direct inference to the 1/3-conclusion' (Thorn 2011: 662). The heart of Thorn's reasoning is a proposed account of comparative logical strength for two reference properties that is applicable even when one of the reference properties has different 'arity' than the other. Thorn writes:

> In general, where $n \geq m$, I will say that an n-place reference property R′ is logically stronger than an m-place reference property R if and only if it is a logical truth that $\forall x_1, \ldots, x_n$: $R'(x_1, \ldots, x_n) \supset R(x_1, \ldots, x_m)$. In those cases where a candidate direct inference is in fact defeated because the reference property of another direct inference is logically stronger, I will say (following Pollock) that the candidate direct inference is subject to *undercutting defeat* …
>
> The intuition behind Reichenbach's principle is simply that we should prefer direct inferences based on reference properties that incorporate more of the things we know concerning the objects about which we wish to make probability judgments. This intuition *does* support the conclusion that direct inference based on (2) should be preferred to direct inference based on (3). (Thorn 2011: 663, 665)

In reply to Thorn, Kaila Draper (2017) argues that Thorn's defense of the Oscar Seminar's argument for thirdism fails because Thorn's proposed definition of 'logically stronger reference property' is unacceptable. Draper writes:

> That definition … has odd consequences. In the first place, it seems to allow for the possibility of a pair of properties each being logically stronger than the other. For example, the property of being a three-angled closed polygon and the property of being a three-sided closed polygon each appear to be logically stronger than the other on Thorn's definition …
>
> Thorn's definition also allows a property $R'(x_1, \ldots, x_n)$ to be logically stronger than a second property $R(x_1, \ldots, x_m)$ even though it is logically impossible for anything that satisfies $R'(x_1, \ldots, x_n)$ to also satisfy $R(x_1, \ldots, x_m)$. (Draper 2017: 34–5)

Draper goes on to offer a new argument of her own, different from Pust's, aimed at establishing that the direct inference in the Oscar Seminar's argument for thirdism is unsound. I will address this further argument in §5 below.

3. Draper vs. Thorn on comparative logical strength between reference properties of different arity

I have several comments about Thorn's proposed definition of 'logically stronger reference property' and about Draper's objections to it.

First, the natural way to try characterizing comparative logical strength for properties of different arity would be in two steps: first, offer a definition of *logical entailment* between properties, in a way that allows this relation to hold between properties of differing arity. Second, define greater logical strength as follows: property R′ is logically stronger than property R just in case R′ logically entails R but R does not logically entail R′.

Second, it is also very natural to construe Thorn's own proposed definition of greater logical strength as actually constituting a proposed definition of logical entailment between an n-ary property R′ and an m-ary property R (where $n \geq m$) – and to construe

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

Thorn as having mistakenly conflated the task of defining logical entailment for properties with the distinct task of defining the 'logically stronger than' relation between properties.

Third, this means that if Thorn's proposed definition does indeed adequately characterize logical entailment between properties, then Draper's objection is easily overcome – by saying that property R′ is logically stronger than property R just in case R′ entails R but not conversely.

However, fourth, Thorn's proposal (as thus reconstrued) does not constitute an adequate definition of logical entailment between properties. The problem is that the proffered definition has the following (probably unintended) consequence:

> For $m < n$: an n-place reference property R′ logically entails an m-place reference property R only if: for any items $i_1,\ldots, i_n$,
> **if** $< i_1,\ldots, i_n>$ satisfies R′
> **then** R is satisfied by $< i_1,\ldots i_m>$, *the initial m-element sub-sequence* of $< i_1,\ldots, i_n>$.

This putatively necessary condition on the logical entailment relation (and on the 'logically stronger than' relation) is far too strong, in two respects. First, the specific items from $\{i_1,\ldots, i_n\}$ that are logically guaranteed to satisfy R, given that $< i_1,\ldots, i_n>$ satisfies R′, need not be the first m items in the n-tuple $< i_1,\ldots, i_n>$; rather, they could be *any m* items from the set $\{i_1,\ldots, i_n\}$. Second, the order in which the R-satisfying items from $\{i_1,\ldots, i_n\}$are logically guaranteed to satisfy R, given that $< i_1,\ldots, i_n>$ satisfies R′, need not be the same as the order in which those items occur, left to right, in the n-tuple $< i_1,\ldots, i_n>$; rather, it could be any successive ordering whatever of those items.[2]

Fifth, the problem just noted appears, prima facie, to be a mere technical problem. There ought to be a natural and appropriate way to define logical entailment between properties, as a relation that can hold even when the related properties differ in arity from one another. (Technical problems have technical solutions!) If such a definition can be given, then it will turn out that greater logical strength, construed as one-way logical entailment, can indeed obtain between properties of differing arity. This would vindicate the underlying idea behind Thorn's reply to Pust.

[2]Essentially the same problem arises for Thorn's proposed definition, in Thorn (in press), of the notion of one reference class being *more specific* than another. For example, in a pertinent hypothetical long run of Sleeping Beauty scenarios, the reference class of time/scenario/toss ordered triples satisfying the tertiary formula

> t is a time in scenario s & Beauty is awake in s at t & x is a coin-toss in s

should count as more specific than the reference class of toss/scenario ordered pairs satisfying the binary predicate

> x is a coin-toss in scenario s.

But Thorn's definition does not yield this result, because the left-to-right first occurrences of variables are different in the two formulas. The problem can be resolved by adopting a different definition of the *more specific than* relation between reference classes, similar in form to the definition I offer in the next section of the *logically stronger than* relation between properties.

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

4. Defining logical entailment, and greater logical strength, between properties

I will now offer a definition of logical entailment, as a relation that can hold between two properties of differing arity. Greater logical strength will then be easily definable, as one-way logical entailment. The two definitions together will elucidate the content of the intuitive thought that the greater logical strength of property R′, vis-à-vis property R, is a matter of R′ being 'informationally richer' than R – the idea behind Reichenbach's principle. If satisfactory, the proposed definition of logical entailment for properties will render the reference property in claim (2) logically stronger than the reference property in claim (3), thereby defusing Pust's objection to the Oscar Seminar's argument for thirdism. And the definition will be a contribution to the theory of direct inference, as regards the proper construal of Reichenbach's principle.

Two criteria of adequacy should be met by the sought-for definition of logical entailment for properties. First, the definition should avoid the problems, described in §3, encountered by Thorn's own proposal. His is inadequate, either as a definition of greater logical strength between properties or as a definition of logical entailment between properties, because it is not sufficiently permissive about the various ways that the elements of an n-tuple $< i_1,\ldots, i_n>$ might be related to the elements of an m-tuple $< j_1,\ldots, j_m>$ in instances where it is logically true that if $< i_1,\ldots, i_n>$ satisfies property R′ then $< j_1,\ldots, j_m>$ satisfies property R.

Second, the definition should specify some *uniform* connection that must obtain between (i) an n-tuple of items $< i_1,\ldots, i_n>$ that satisfies the n-ary property R, and (ii) the specific corresponding m-tuple of items, all from the set $\{i_1,\ldots, i_n\}$, that must satisfy the m-ary property R′. The idea here is that if R′ really *logically* entails R, then the pertinent, definitive, connection between an n-tuple that satisfies R′ and the corresponding m-tuple that must satisfy R should be the same regardless of the specific items that satisfy R′ and R, rather than depending upon which particular items are involved. The connection should be a matter of logic alone.

I begin with two preliminary definitions. First is the notion of a *position-pairing function*, which pairs element-positions in n-tuples with element-positions in m-tuples. (There are n respective positions in an n-tuple, and m respective positions in an m-tuple.)

For natural numbers n and m, an ***n*-to-*m* position-pairing function** is a 1–1 function f that meets the following conditions:

(1) if $n > m$, then (i) the domain of f is a set containing m members of the set of all n-tuple positions, (ii) the range of f is the set of all m-tuple positions, and (iii) for each distinct m-tuple position q there is a distinct n-tuple position p such that $f(p) = q$, and

(2) if $n \leq m$, then (i) the domain of f is the set of all n-tuple positions, (ii) the range of f is a set containing n members of the set of all m-tuple positions, and (iii) for each distinct n-tuple position p there a distinct m-tuple position q such that $f(p) = q$.

Second is the notion of *conformity* between an m-tuple and an n-tuple, relative to an n-to-m position-pairing function.

For natural numbers n and m, n-tuple $< i_1,\ldots, i_n>$, m-tuple $< j_1,\ldots, j_m>$, and n-to-m position-pairing function f, the m-tuple $< j_1,\ldots, j_m>$ ***f*-conforms** to the n-tuple $< i_1,\ldots, i_n>$ just in case the following condition obtains:

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

for every pair $< p,q >$ such that (i) p is an i-tuple position, (ii) q is a j-tuple position, and (iii) $q = f(p)$,
the element i_r of $< i_1,\ldots, i_n>$ that occupies position p in $< i_1,\ldots, i_n>$ is identical to the element j_s of $< j_1,\ldots, j_m>$ that occupies position q in $< j_1,\ldots, j_m>$.

Using these preliminary definitions, I now propose the following definition of *logical entailment*, as a relation between an n-ary property R′ and an m-ary property R.

For natural numbers n and m, n-ary property R′ **logically entails** m-ary property R just in case
there exists an n-to-m pairing function f such that
for any i-tuple $< i_1,\ldots, i_n>$, and any j-tuple $< j_1,\ldots, j_m>$ that f-conforms to $< i_1,\ldots, i_n>$,
it is logically true that
if $< i_1,\ldots, i_n>$ satisfies R′ **then** $< j_1,\ldots, j_m>$ satisfies R.

This definition meets the first criterion of adequacy, the permissiveness criterion, because it allows any m of the elements in an R′-satisfying n-tuple $< i_1,\ldots, i_n>$, in any order, to constitute the m successive elements in the corresponding R-satisfying m-tuple $< j_1, \ldots, j_m>$. And the definition meets the second criterion of adequacy, the uniqueness criterion, by requiring the pairing function f to select certain specifically-positioned elements of $< i_1,\ldots, i_n>$ as the successive elements of $< j_1, \ldots, j_m>$ solely on the basis of the n-tuple *positions* of the selected items from $< i_1,\ldots, i_n>$, independently of what the items themselves are.

Of course, comparative logical strength for properties is now easily defined this way: n-ary property R′ is **logically stronger** than property R just in case (i) R′ logically entails R and (ii) R does not logically entail R′.

Return now to the guiding intuition behind the idea that an n-ary property R′ can be logically stronger than an m-ary property R even when $n > m$. That intuition can be formulated this way: R′ is logically stronger than R only if the claim R′($i_1,\ldots, i_n$), about the items $i_1,\ldots, i_n$, says (or entails) everything *about* ***some*** *of those items* (about m specific ones of them) that is said by predicating R *of those items*; and in addition, the claim R′ ($i_1,\ldots, i_n$) also says *more besides*. My proposed definition of 'logically stronger than', I submit, does full justice to this intuition.[3]

The upshot, as regards the Oscar Seminar's argument for the thirdist answer to the Sleeping Beauty problem, is that the reference property in claim (2) above is indeed logically stronger than the reference property in claim (3). Thorn's response to Pust's objection to the argument is thereby vindicated in spirit – even though Draper is correct in claiming that Thorn's own proposed definition of 'logically stronger than' is not acceptable, and even though it also is not an acceptable definition of logical entailment between properties.

The wider upshot is the licensing of preference for direct inferences based on logically stronger reference properties, regardless of arity – in alignment with existing

[3]My definition of logical entailment for properties leaves open the possibility that an n-ary property R′ can logically entail an m-ary property R even if R has higher adicity than R′; this, in turn, leaves open the possibility that R′ can be logically stronger than R when n < m. Are there such cases? Indeed there are, viz., ones in which the more 'besides' that R attributes to the extra elements of $<j_1,\ldots, j_m>$ is all tautological. For instance, if the monadic predicates 'F' and 'G' express distinct monadic properties, then the monadic property expressed by the expression 'Fx & Gx' is logically stronger than the binary property expressed by the expression 'Fx & y = y', and is logically stronger than the tertiary property expressed by the expression 'Fx & y = y & z = z', etc. (This assumes a broad notion of property, and a fine-grained conception of property-individuation; but the same is true of Reichenbach's principle itself.)

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

treatments of direct inference that also purport to have this feature (Pollock 1990; Kyburg and Teng 2001: 216; Thorn 2011, in press).[4] This greatly enhances the scope and power of direct inference as a method for drawing conclusions about single-case epistemic probabilities.

5. Draper's new objection to the Oscar Seminar's argument

Draper raises an objection of her own to the Oscar Seminar's argument for thirdism about the Sleeping Beauty problem – an objection that does not involve properties of different arity. If we let 'M*t*' abbreviate '*t* is a time on Monday', she says, then it is clear that

(4) prob(H*x*/W(*t*,*s*) & B(*t*,*s*) & Toss(*x*,*s*) & M*t*) = 1/2.

It also is clear that the reference property in (4) is logically stronger than the reference property in (2) – even in the ordinary sense, since both reference properties have the same, tertiary, arity. She now argues as follows:

> Consider … the time that I will refer to as 'this time on Monday (of the scenario)', which we can abbreviate as 'ttm'. It is important to recognize that I am using the expression 'this time' as it is used when we say things like, 'At this time on Monday I will be safely at home'. Thus, 'this time on Monday' means 'Monday at whatever time of day it is now'. On Monday morning, then, Beauty knows that ttm is a time that might be now and might be exactly 24 hours ago. Beauty also knows that (4) and that W(ttm,σ) & B (ttm,σ) & Toss(τ,σ) & M(ttm). Thus, by direct inference, Beauty arrives at the prima facie conclusion that PROB(Hτ) = 1/2 … [I]t seems clear that … the direct inference that relies on (4) rebuts and thereby defeats the direct inference that relies on (2). Therefore … the Oscar Seminar's argument fails. (Draper 2017: 35–6)

At first blush, this argument might seem very difficult to fault. After all, the reference property in (4) is indeed logically stronger than the reference property in (3), even in the noncontroversial sense of 'logically stronger' involving two properties with the same arity. Also, since Beauty knows both (4) itself and that W(ttm,σ) & B(ttm,σ) & Toss(τ,σ) & M(ttm), it would seem that she can indeed arrive by direct inference at the prima facie conclusion that PROB(Hτ) = 1/2. How, then, could one possibly fault Draper's contention that this prima facie direct inference rebuts and thereby defeats the prima facie direct inference that relies on (2)?

I maintain that one can, and one should, fault this contention. To see the problem, consider some of the pertinent statements that Beauty knows to obtain. She knows both (2) and (4). She also knows the following two statements concerning the scenario σ (the one she is now in) and the coin toss τ in this scenario.

(5) W(now,σ) & B(now,σ) and Toss(τ,σ)
(6) W(ttm,σ) & B(ttm,σ) & Toss(τ,σ) & M(ttm)

Now, the reference property in (4) is clearly logically stronger than the reference property in (3). But how do (5) and (6) stack up against one another in terms of comparative logical strength? Well, the statement 'M(ttm)' is trivially true, since it says 'This time on Monday is a time on Monday'. (It is *logically guaranteed* to be true, given that the

[4]Concerning Pollock, Thorn (2011: 665) says this: 'Although Pollock does not discuss the point, his principle *DI* (1990: 190) entails a preference for logically stronger reference properties of higher arity, via applications of his principle *IND* (1990: 46).'

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

indexical expression 'this time' is logically guaranteed to refer to the current time and the statement 'B(ttm,σ)' entails that there is unique referent of 'Monday' within the scenario.) Thus, (5) is logically equivalent to

(5*) W(now,σ) & B(now,σ) and Toss(τ,σ) & M(ttm)

But (5*) is logically stronger than (6), because (5*) entails (6) but not conversely. (This is because (6) is consistent with the statement '¬W(now,σ) & B(now,σ) and Toss(τ,σ)', which would have been true had today been Tuesday and had the coin toss come up Heads – in which case Sleeping Beauty would have slept dreamlessly all day today rather than being awakened today by the experimenters.) Thus, since (5) and (5*) are logically equivalent, (5) is logically stronger than (6).[5]

So something odd and curious is at work in Draper's argument: although the reference property in (4) is indeed logically stronger than the reference property in (2), nevertheless the pertinent *instantiation* of the reference property in (4) – the instantiation that is deployed in Draper's prima facie direct inference relying on (4) – is logically **weaker** than the pertinent instantiation of the reference property in (2) that is deployed in the Oscar Seminar's direct inference relying on (2).[6,7,8]

[5]I maintain that because (i) the statement '¬W(now,σ) & B(now,σ) and Toss(τ,σ)' expresses an (essentially indexical) possibility that is consistent with the combination of Beauty's Sunday information about the sleep experiment plus her current, essentially indexical, knowledge that today is either or Monday or Tuesday, whereas (ii) statement (5) precludes this possibility, Beauty's knowledge of (5) thereby constitutes *evidence* she possesses that is pertinent to her single-case epistemic probability for 'Hτ'. This evidence is what makes it the case that PROB(Hτ) = 1/3 for her, despite the fact she knows the objective chance of Hτ to be 1/2.

In Horgan (2004, 2008) I defend thirdism, in a way that emphasizes the evidential import of Beauty's essentially indexical information that she was awakened *today* by the experimenters, by appeal to a mode of single-case probabilistic reasoning that I dub 'synchronic Bayesian updating'. My argument requires Beauty to assign what I call a 'preliminary probability' of 1/4 to the statement 'The coin toss comes up heads and today is Tuesday and I'm in a dreamless sleep all day today' – which, as I acknowledge, is intuitively somewhat peculiar.

A nice feature of the Oscar Seminar's argument is that it avoids this peculiarity, while still helping make intuitively clear why Beauty's (essentially indexical) knowledge that she was awakened *today* by the experimenters constitutes evidence in virtue of which her epistemic probability for Hτ deviates from its known chance of 1/2.

[6]I emphasize that an instantiation of a property, as I am using the term 'instantiation' here and below, is a *statement*. Thus, the statements 'Today is a day on Monday' and 'This day on Monday is a day on Monday' are distinct instantiations of the property *being a day on Monday* – even if the singular expressions 'today' and 'this day on Monday' happen to refer to the same day (viz., Monday during the experiment). Property-instantiations, in the operative sense, are subject to fine-grained criteria of individuation.

[7]Does fine-grained individuation come into play, in direct inference, with respect to the operative reference-property and consequent property *themselves*? This answer is Yes, I would maintain – although this point is not as pertinent to the present discussion of Draper as is the fine-grained individuation of the statements that I am calling reference-property *instantiations*. (The crucial thing here is that statements deploying distinct, co-referring, *singular* expressions count as distinct property-instantiations, even when they deploy the same *predicative* formula to express the given reference-property.) An available prima facie direct inference involves a *known* (or justifiably believed) indefinite-probability statement, deploying a specific reference-**predicate** and a specific consequent-**predicate** (in a broad sense of 'predicate' that applies to logically complex formulas containing free variables). Although some other predicate might pick out the same **property** as one or the other of these two predicates, this fact might not be part of one's available evidence.

[8]In light of footnote 7, I suggest that Reichenbach's principle itself really should be formulated in terms of reference predicates and consequent predicates, rather than in terms of reference properties and consequent properties – and likewise for my proposed definitions of logical entailment, and of greater logical strength, in §4 above. But in the body of this paper I stick to talk of reference and consequent **properties**,

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

What moral should be drawn from this curious inversion in comparative logical strength? The answer, I suggest, is this: the Oscar Seminar's direct inference from (2) and (5) to the conclusion $PROB(H\tau) = 1/3$ defeats, but is not itself defeated by, Draper's prima facie direct inference from (4) and (6) to the conclusion $PROB(H\tau) = 1/2$. This is so because the reference-property *instantiation* invoked in the Oscar Seminar's direct inference – viz., statement (5) – is logically stronger than the reference-property instantiation invoked in Draper's prima facie direct inference – viz., statement (6).

The immediate upshot is that Draper's objection to the Oscar Seminar's argument for thirdism concerning the Sleeping Beauty problem is unsuccessful, because she is mistaken to claim that 'the direct inference that relies on (4) rebuts and thereby defeats the direct inference that relies on (2)'. On the contrary, the rebut-and-defeat relation obtains only in the other direction: the inference that relies on (2) rebuts and thereby defeats the direct inference that relies on (4), but not conversely.

The more general upshot concerns the proper formulation and proper application of Reichenbach's principle. The intuitive idea behind the principle is that a proper direct inference will be one that deploys one's *strongest pertinent evidence*. It turns out, however, that this intuitive idea is not perfectly captured by the idea that a proper direct inference to a conclusion $PROB(Fa_1,\ldots, a_n) = r$ will deploy the logically strongest evidentially relevant reference property G such that a known, or justifiably believed, instantiation of G, plus a known, or justifiably believed, indefinite-probability claim whose reference property is G, together provide a *prima facie* direct inference to the conclusion $PROB(Fa_1, \ldots, a_n) = r$. Trouble can arise in cases where the instantiation of G is a conjunctive statement in which at least one conjunct is *degenerate* (as I will put it) – i.e., this conjunct is an atomic statement whose singular-term constituent(s) is/are so constructed, vis-à-vis its predicate-constituent, as to render the statement trivially true.

This is exactly what happens in Draper's prima facie direct inference deploying statements (4) and (6). Statement (6), the pertinent instantiation of the reference-property in (4), contains a degenerate conjunct – viz., the conjunct 'M(ttm)', which symbolizes 'This time on Monday is a time on Monday'. This statement is trivially true, because the indexical expression 'ttm' is constructed so as to trivially satisfy the predicate 'M'. And that is why statement (6) is logically weaker than statement (5).

The following claim is very plausible: if property G is logically stronger than property G, but a given instantiation I_G of property G is logically weaker than a given instantiation I^*_H of property H, then this is because there is at least one degenerate conjunct in I_G. And the following further claim also is very plausible: if a statement $PROB(Fa_1,\ldots, a_n) = r$ can be legitimately derived from a given body of evidence by direct inference deploying the reference property G, then it can be derived by a G-deploying direct inference in which the pertinent instantiation of G is a statement containing no degenerate conjuncts.

In light of these two highly plausible claims, a simple and natural way to amend Reichenbach's principle is to rule out as improper any putative direct inference employing a reference-property instantiation containing a degenerate conjunct. The revised principle is this: a proper direct inference to a conclusion $PROB(Fa_1,\ldots, a_n) = r$ should deploy the logically strongest evidentially relevant reference property G such that a known, or justifiably believed, G-instantiation **containing no degenerate conjunct**, plus a known, or justifiably

since this is standard in the pertinent philosophical literature. (This is safe enough vis-à-vis ordinary applications of direct inference, since normally the premises of a pertinent direct inference will be statements that either are known or at least are justifiably believed, given the reference predicate and the consequent predicate that occur in these premises.)

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

believed, indefinite-probability claim whose reference property is G, together provide a *prima facie* direct inference to the conclusion PROB($Fa_1,\ldots, a_n$) = r. (Reichenbach's principle also can be amended in a way that does not appeal to the notion of a degenerate conjunct, although the alternative revised principle will be more cumbersome.[9])

6. The Oscar Seminar's argument and the nature of probability

Could opponents of thirdism grant my replies to Pust and Draper, but then resist the Oscar Seminar's argument by repudiating certain assumptions about the nature of probability to which the argument is committed? The Oscar Seminar's paper begins with the following remarks:

> The literature on the Sleeping Beauty problem has been dominated by Bayesians. Even those authors who are not Bayesians have addressed the problem without using much of the rich machinery available to objective probability theorists. We show that the objective probability theorist has a *very* simple argument for thirdism. (Seminar 2008: 149)

Shortly thereafter the paper says, "Most objective approaches to probability tie probabilities to relative frequencies in some way" (p. 150). These passages might encourage the thought that one way to resist the Oscar Seminar's argument for thirdism would be to repudiate all philosophical views that construe probabilities as relative frequencies of some kind.

It is true enough that the sort of position typically espoused by advocates of objective probability theory themselves asserts that (i) objective probability (so-called "chance") is relative frequency of some kind, (ii) indefinite probability is chance as thus construed, and (iii) definite probability is a derivative attribute that is determined by known, or justifiably believed, indefinite probability. But one need not embrace any of this just by virtue of embracing the Oscar Seminar's argument for thirdism. Instead one could construe the pertinent conceptual machinery of objective probability theory in some other way.

Here is one alternative construal, for example. (I myself find it very attractive.) So-called indefinite probability, as expressed by 'prob', is not really a distinct kind of probability from the kind expressed by 'PROB'. Rather, a statement 'prob($Fx \mid Gx$) = r' is just a notational variant of the universally quantified statement '$\forall x$(PROB ($Fx \mid Gx$) = r)'. Moreover, epistemic probability, as expressed by 'PROB', is *quantitative*

[9]Suppose that D is an available prima facie direct inference with conclusion PROB($Fa_1,\ldots, a_n$) = r, and that D* is a distinct available prima facie direct inference with conclusion PROB($Fa_1,\ldots, a_n$) = s, where $r \neq s$; let D* be called an *available competitor* to D. Then the suitably modified version of Reichenbach's principle now can be formulated this way:

> If D is an available prima facie direct inference deploying reference property G, together with G-instantiation I(G), to generate conclusion PROB($Fa_1,\ldots, a_n$) = r, then D is a **legitimate** direct inference only if the following condition obtains: for any available prima facie direct inference D* that (i) is an available competitor to D, (ii) deploys a reference property H distinct from G, and (iii) deploys a specific H-instantiation I*(H), the statement I(G) is logically stronger than the statement I*(H).

A legitimate direct inference to a conclusion PROB($Fa_1,\ldots, a_n$) = r should not only deploy the strongest evidentially relevant reference property G such that one knows, or justifiably believes, prob($Fx_1,\ldots, x_n/Gx_1,\ldots, x_n$) and one knows, or justifiably believes, $Ga_1,\ldots, a_n$. In addition, the G-instantiation $Ga_1,\ldots, a_n$ should itself be logically stronger than any known, or justifiably believed, instantiation of any competing reference-property H.

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

degree of evidential support, relative to one's total available evidence (cf. Horgan 2017*a*, 2017*b*: section 6). Thus, 'prob(Fx | Gx) = r' is to be understood as asserting that for any arbitrary item x, considered as such (and apart from any additional knowledge one might happen to have about this x specifically), one's degree of evidential support for x's being F, given that x is G, is equal to r. Direct inference is a transition, from an evidentially sanctioned premise '∀x(PROB(Fx | Gx) = r)' plus an evidentially sanctioned premise 'Gτ', to a conclusion 'PROB(Fτ) = r'. (Multiple variables can be involved, instantiated by multiple singular terms.) Quantifiers in statements of the form '∀x (PROB(Fx | Gx) = r)' range over *possible instances* of the reference property G; hence such a statement, when evidentially sanctioned itself, thereby determines a *rationally expectable hypothetical long run* of G-instances – a long run in which the relative frequency of F-instances is r. But relative frequency across such a hypothetical long run is a derivative matter, rather than being either fundamental or a kind probability itself.

Other construals of the pertinent conceptual machinery of objective probability theory seem available too. In particular, there is no obvious reason why one could not adopt a "Bayesian" construal. This would be similar to the construal described in the preceding paragraph, but would identify epistemic probability with quantitative degree of *partial belief* rather than with quantitative degree of evidential support.

The upshot is that those who wish to repudiate the Oscar Seminar's argument for thirdism bear a substantial burden of proof. Since the conceptual machinery used in the argument looks to be construable in conformity with most any familiar way of interpreting the kind of probability that is at issue in the Sleeping Beauty problem, opponents of the Oscar Seminar's argument cannot repudiate it just by rejecting frequentism. Instead they must confront the argument on its own terms.

7. Conclusion

I have argued that Pust and Draper are both mistaken in claiming that the Oscar Seminar's argument is unsound according to the standards of objective probability theory itself. Pust's argument is mistaken because the theory of direct inference can, and should, embrace the contention that reference properties of differing arity can bear relations of comparative logical strength to one another. Draper's argument is mistaken because it deploys a reference-property instantiation containing a degenerate conjunct.

I also have urged that the Oscar Seminar's argument is effectively neutral about disputed issues concerning the nature of the kind(s) of probability involved in the Sleeping Beauty problem. Opponents of the argument therefore cannot repudiate it simply by rejecting frequentist views of probability that traditionally have been championed by advocates of objective probability theory.

References

Bacchus F. (1990). *Representing and Reasoning with Probabilistic Knowledge*. Cambridge, MA: MIT Press.

Draper K. (2017). 'Even for Objectivists, Sleeping Beauty isn't so Simple.' *Analysis* **77**(1), 29–37.

Draper K. (in press). 'Direct Inference and the Sleeping Beauty Problem.' *Synthese*.

Fisher R.A. (1922). 'On the Mathematical Foundations of Theoretical Statistics.' *Philosophical Transactions of the Royal Society* A **222**, 309–68.

Follesdal D. (1967). 'Knowledge, Identity, and Existence.' *Theoria* **33**(1), 1–27.

Halpern J. (1990). 'An Analysis of First-Order Logics of Probability.' *Artificial Intelligence* **46**, 311–50.

Horgan T. (2004). 'Sleeping Beauty Awakened: New Odds at the Dawn of the New Day.' *Analysis* **64**(1), 10–21. Reprinted in *Essays on Paradoxes*, pp. 209–19. New York, NY: Oxford University Press.

Horgan T. (2008). 'Synchronic Bayesian Updating the Sleeping Beauty Problem: Reply to Pust.' *Synthese* **160**(2), 155–9. Reprinted in *Essays on Paradoxes*, pp. 220–5. New York, NY: Oxford University Press.

https://doi.org/10.1017/epi.2019.41 Published online by Cambridge University Press

Horgan T. (2017*a*). 'Epistemic Probability.' In *Essays on Paradoxes*, pp. 281–318. New York, NY: Oxford University Press.
Horgan T. (2017*b*). 'Troubles for Bayesian Formal Epistemology.' *Res Philosophica* **94**(2), 233–55.
Kyburg H.E. (1974). *The Logical Foundations of Statistical Inference*. Dordrecht: Reidel.
Kyburg H.E. and Teng C. (2001). *Uncertain Inference*. Cambridge: Cambridge University Press.
Pollock J. (1990). *Nomic Probability and the Foundations of Induction*. New York, NY: Oxford University Press.
Popper I. (1956). 'The Propensity Interpretation of Probability.' *British Journal for the Philosophy of Science* **10**(37), 25–42.
Pust J. (2011). 'Sleeping Beauty and Direct Inference.' *Analysis* **71**(2), 290–3.
Reichenbach H. (1949). *A Theory of Probability*. Berkeley, CA: University of California Press.
Seminar O. (2008). 'An Objectivist Argument for Thirdism.' *Analysis* **68**(2), 149–55.
Thorn P. (2011). 'Undercutting Defeat via Reference Properties of Different Arity: A Reply to Pust.' *Analysis* **71**(4), 662–7.
Thorn P. (2012). 'Two Problems of Direct Inference.' *Erkenntnis* **76**(3), 299–318.
Thorn P. (in press). 'A Formal Solution to Reichenbach's Reference Class Problem.' *Dialectica*.

Appendix

Upon completing a version of this paper under the title 'Direct Inference and the Sleeping Beauty Problem,' I came upon a paper with that same title by (of all people!) Kaila Draper – viz., Draper (in press). Drawing on recent work by (of all people!) Paul Thorn – viz., Thorn (2012) – Draper describes her project this way:

> This article is an attempt to use the insights of objective probability theory to solve the Sleeping Beauty problem. The approach is to develop a partial theory of direct inference and apply that partial theory to the problem. One of the crucial components of the partial theory is that expected indefinite probabilities provide a reliable basis for direct inference. The article relies heavily on recent work by Paul D. Thorn to defend that thesis. The article's primary conclusion is that Beauty (the perfectly rational agent in the Sleeping Beauty story) can by way of a justifiable direct inference reach the conclusion that the epistemic probability that the relevant coin toss lands heads is 1/3. (Draper in press: 1)

Her argument for the thirdist conclusion is different from the Oscar Seminar's argument, and is much more complicated. (Although direct inference plays a role, the overall argument has numerous premises and numerous inferential steps.[10])

Concerning the Oscar Seminar's argument and the dialectic that I have addressed in the present paper, Draper says this:

> Writing collectively as the Oscar Seminar (2008), John Pollock, Paul D. Thorn, and several of their colleagues have also defended 1/3 on objectivist grounds. The Oscar Seminar's argument has been challenged, however, by Joel Pust (2011) and by me (2017). In addition to raising an objection to that argument, Pust advances the positive thesis that objectivists are committed to rejecting both and 1/2 and 1/3. I take this thesis seriously partly because I believe that some prominent theories of direct inference do yield the result that neither 1/2 nor 1/3 is correct. Ultimately however, I reject that thesis and any theory of direct inference that would, if accurate, provide a basis for it. (Draper in press: 2)

If I understand her correctly, she still rejects the Oscar Seminar's argument, because she continues to agree with the following claims in the passage from Pust (2011) that I quoted in §2 above, involving statements (2) and (3):

[10]Another important respect in which Draper's argument is more complicated is that hers involves indefinite-probability claims in which the indexical word 'today' occurs, used in an *essentially* indexical way. (One might well wonder whether this will make serious trouble for her argument; I have my suspicions.) The pertinent indefinite-probability claim that figures in the Oscar Seminar's argument, on the other hand, is statement (2) above, which contains no indexicals.

> [B]ecause (2) and (3) concern property possession by n-tuples of different n, neither trumps the other as a basis for direct inference. Instead, as (2) and (3) prima facie justify direct inferences to contradictory claims, such inferences defeat each other. (Pust 2011: 292)

But because of her own new argument for thirdism – which includes applications of direct inference to certain claims about expected indefinite probabilities – she now rejects the following conclusion in the above-quoted passage from Pust:

> Therefore, Beauty's situation is one in which she cannot engage in an all-things-considered justified direct inference to PROB(Hτ) = 1/3 or to PROB(Hτ) = 1/2. (Pust 2011: 292)

She maintains, instead, that even though the two competing prima facie direct inferences based respectively on (2) and (3) do indeed defeat each other, nevertheless her own new direct-inference argument renders the thirdist conclusion both defeasibly justified and undefeated.

I do not attempt here to assess Draper's new argument. If it turns out to be sound, so much the better for thirdism. But regardless whether it is sound or not, the fact remains that the Oscar Seminar's own argument is immune from Pust's objection. This is because, as shown in §4 above, greater logical strength is definable for reference properties of different arity – and given the appropriate definition, the reference property cited in statement (2) is logically stronger than the reference property cited in statement (3). The fact also remains that the Oscar Seminar's argument is immune to Draper's own earlier objection, once Reichenbach's principle is suitably modified in the manner described in §5 above.[11,12]

Terry Horgan is Professor of Philosophy at the University of Arizona. He has published in philosophy of mind, metaphysics, metaethics, and epistemology. He also writes on philosophical paradoxes.

[11]In her new paper, Draper herself offers a proposed rebuttal of her earlier objection to the Oscar Seminar's argument. This response deploys reasoning similar to that which figures in her new direct-inference argument for thirdism; see note 11 in the paper, and the paragraph to which that note is appended. My own critique remains applicable, however, regardless of how her proposed critique fares. (Perhaps her earlier objection to the Oscar Seminar's argument was mistaken in two different ways.)

[12]For helpful discussion of issues addressed here, I thank Don Fallis, Reina Hayaki, Kay Mathiesen, Joel Pust, and Paul Thorn.

Cite this article: Horgan T (2021). Direct Inference, Reichenbach's Principle, and the Sleeping Beauty Problem. *Episteme* **18**, 640–653. https://doi.org/10.1017/epi.2019.41