On the Preference for More Specific Reference Classes*

Paul D. Thorn

Abstract

In attempting to form rational personal probabilities by direct inference, it is usually assumed that one should prefer frequency information concerning more specific reference classes. While the preceding assumption is intuitively plausible, little energy has been expended in explaining why it should be accepted. In the present article, I address this omission by showing that, among the principled policies that may be used in setting one's personal probabilities, the policy of making direct inferences with a preference for frequency information for more specific reference classes yields personal probabilities whose accuracy is optimal, according to all proper scoring rules, in situations where all of the relevant frequency information is *point-valued*. Assuming that frequency information for narrower reference classes is preferred, when the relevant frequency statements are point-valued, a dilemma arises when choosing whether to make a direct inference based upon (i) relatively *precise-valued* frequency information for a broad reference class, R, or upon (ii) relatively *imprecise-valued* frequency information for a more specific reference class, R′ (R′ ⊂ R). I address such cases, by showing that it is often possible to make a precise-valued frequency judgment regarding R′ based on precise-valued frequency information for R, using standard principles of direct inference. Having made such a frequency judgment, the dilemma of choosing between (i) and (ii) is removed, and one may proceed by using the precise-valued frequency estimate for the more specific reference class as a premise for direct inference.

Keywords: Direct inference, Statistical syllogism, Specificity, Scoring rules, The reference class problem, Imprecise probabilities, The principle of indifference

[0]The final publication is available at Springer via http://link.springer.com/article/10.1007/s11229-016-1035-y.

1 Introduction

Direct inference typically proceeds from two premises of the following sort: The first (minor) premise states that a given object, c, is an element of a reference class R. The second (major) premise states that the frequency with which members of R are members of a respective target class, T, is r. The conclusion of the direct inference is then that the probability that c is a member of T is r. In order to abbreviate the description of such inferences, I use the notation "PROB" to refer to a probability function that takes propositions as arguments, and is understood as designating the (potentially imprecise) personal probabilities (or degrees of belief) that are rational for a respective agent, given the evidence that the agent has. So the injunction to infer a given personal probability statement, PROB(α) = r, is tantamount to the injunction to infer that the personal probability r is rational for the proposition α, given one's evidence. I use the notation "freq" (for "frequency") to refer to a function that takes a pair of sets as an argument, and returns the relative frequency of the first set among the second. So "freq(T|R) = 0.5" expresses that the relative frequency of Rs (elements of R) that are Ts (elements of T) is 0.5. Given this notation, typical instances of direct inference satisfy the following schema:

From c $\in$ R and freq(T|R) = r infer that PROB(c $\in$ T) = r.

Instances of the preceding schema are, of course, defeasible. A particular condition under which instances of the schema are (usually taken to be) defeated is the central concern of the present article. In particular, it is typically held that, in cases where two instances of the preceding schema yield *conflicting* conclusions regarding the probability of some proposition c $\in$ T, one should form one's conclusion regarding the value PROB(c $\in$ T) by the direct inference that employs the narrower reference class, provided the reference class for one of the two direct inferences is narrower than the other (according to the proper subset relation).[1]

The doctrine that one should prefer frequency information for more specific reference classes in conducting direct inference is intuitively plausible. The prior intuitive plausibility of the doctrine probably explains why its advocates haven't taken much care to argue for it, including Venn (1866),

[1]In fact, further qualifications are required in order to exclude *degenerate* direct inferences (cf. Pollock 1990, Kyburg & Teng 2001, Thorn 2012). The problem of excluding degenerate direct inferences does not arise within the simple sorts of population model considered in Section 2. In Section 3.5, I will say a little bit about the problem of excluding degenerate direct inferences.

Reichenbach (1949), Kyburg (1974), Pollock (1990), Bacchus (1990), Kyburg and Teng (2001), and Thorn (2012).[2] My primary aim in the present article is to address this omission. From the outset, I acknowledge that many may find the prior intuitive plausibility of the doctrine that one should favor more specific reference classes to be greater than some of the assumptions that I will make in defending the doctrine. But my goal is not, merely, to preach to the converted, i.e., those that find the doctrine highly plausible solely on the basis of prior intuitions. Rather, my goal is to provide independent reasons for the policy of preferring more specific reference classes. To this end, I show that the policy of using direct inference with the most specific applicable reference classes yields personal probabilities whose accuracy is optimal, according to all proper scoring rules. The optimality results presented here are similar to the accuracy based considerations adduced by Joyce (1998) in favor of probabilism. The proposed defense of direct inference, with a preference for more specific reference classes, is also similar to the defense of updating by conditionalization by appeal to expected accuracy maximization, as found in (Greaves & Wallace 2006) and (Leitgeb & Pettigrew 2010b), and generalized in (Easwaran 2013).

The following section of the paper presents the basic optimality results for the policy of preferring frequency information for more specific reference classes in conducting direct inference. While the results are suggestive, they properly apply only to situations where an agent has access to point-valued frequencies (for the relevant target class) for the most specific relevant reference classes. Section 3 introduces several measures that are aimed at mitigating the limitations of the results presented in Section 2. First, analogues of the results of Section 2 are introduced that show that the policy of

[2] Pollock (1990, 86) asserts that the preference for narrower reference classes is a 'kind of' total evidence requirement. This may be. However, there is no straightforward way to direct the force of arguments in support of Carnap's Principle of Total Evidence in order to support a preference for narrower reference classes. Indeed, Carnap's Principle of Total Evidence (1962, 211) prescribes that one's posterior probability for a proposition α be identical to one's prior probability for α conditional on one's complete body of evidence. So in a case where one's complete body of evidence consists of $\text{freq}(\text{T}|\text{R}) = 0.6$, $\text{freq}(\text{T}|\text{R}')$ $= 0.9$, $\text{R}' \subseteq \text{R}$, and $\text{c} \in \text{R}'$, Carnap's principle prescribes that one's posterior probability for $\text{c} \in \text{T}$ be identical to one's prior probability for $\text{c} \in \text{T}$ conditional on $\text{freq}(\text{T}|\text{R}) = 0.6$ $\wedge\ \text{freq}(\text{T}|\text{R}') = 0.9 \wedge \text{R}' \subseteq \text{R} \wedge \text{c} \in \text{R}'$. However, since one's prior probability for $\text{c} \in \text{T}$ conditional on $\text{freq}(\text{T}|\text{R}) = 0.6 \wedge \text{freq}(\text{T}|\text{R}') = 0.9 \wedge \text{R}' \subseteq \text{R} \wedge \text{c} \in \text{R}'$ need not be 0.9, the preference for narrower reference classes does not follow from Carnap's principle. Perhaps rational personal probabilities are structured in such a way as to generate a preference for narrower reference classes, when updating by conditionalization (cf. Thorn 2014). Whether this is the case is something that would need to be argued for, independently of the Principle of Total Evidence.

making direct inferences (with a preference for narrower reference classes) based on *expected* frequencies maximizes *expected* accuracy. Similar to the results of Section 2, the results concerning direct inference based on expected frequencies apply only to situations where an agent has access to point-valued expected frequencies for the most specific relevant reference classes. To mitigate this limitation, two methods of inferring precise-valued expected frequencies are introduced. In a wide range of cases, the methods permit one to infer a precise-valued expected frequency for a reference class, R′, based on precise-valued frequency information for a set R that is a superset of R′. Both methods proceed by locating R′, itself, in an appropriate reference class (a set of subsets of R), and then drawing a series of conclusions (by direct inference) about the probability that the frequency of T among R′ takes various values. These conclusions are then used to infer the expected frequency of T among R′. As I will explain, in Section 3, the two proposed methods are of independent interest (beyond mitigating the limitations of the described optimality results), since they are applicable to addressing the general problem of choosing between direct inferences based on (i) precise-valued frequency information for broad reference classes, versus (ii) imprecise-valued frequency information for more specific reference classes.

2 The Optimality of Preferring More Specific Reference Classes

In order to demonstrate the virtues of reasoning by direct inference using the most specific applicable reference classes, I propose that we use a simple 'test environment', in order to evaluate various 'policies' for forming personal probabilities. For this purpose, I introduce the notion of a population model M, which is a triple ⟨U, T, Π⟩, consisting of a domain of objects U, a subset T of U (where "T" stands for "target class"), and a partition $\Pi = \{\pi_1, ..., \pi_n\}$ of U, where Π corresponds to the set of maximally specific descriptions within which we are able to assign elements of U.[3]

[3] In the circumstance of making a judgment about the probability that an object c is a member of T, it is always possible to introduce a description of maximal specificity, which denotes the unit set consisting of the very object about which one is reasoning. It is reasonable to ignore such descriptions, in cases where we have no substantive information concerning the value of freq(T|{c}). So, in the present section, I take the reasonable course of ignoring such descriptions. But see Section 3.3, which considers the proper treatment of reference classes for which one has no prior frequency information, and provides a more adequate treatment of the present issue.

The task of respective policies will be to recommend accurate probability judgments concerning which members of U are in T. In making their recommendations, our policies may avail themselves of information about the relative frequency of membership in T among objects falling within various categories. While Π specifies the most specific categories that are distinguishable, our policies may consider the relative frequency of membership in T among the broad class of categories consisting of the algebra of subsets of U that is formed by unions of elements of Π. I call this algebra "F", which is defined: $\mathrm{F} = \{\mathrm{f} : \mathrm{f} = \cup\mathrm{A} \wedge \mathrm{A} \subseteq \Pi\}$.[4]

I begin by considering the case where it is known which objects are elements of which elements of Π, and our policies have access to the relative frequency of T among each and every element of Π, where $\mathrm{freq}(\mathrm{T}|\pi) = |\{\mathrm{x} : \mathrm{x} \in \pi \wedge \mathrm{x} \in \mathrm{T}\}|/|\{\mathrm{x} : \mathrm{x} \in \pi\}|$, for all π in Π. For each object, x, in U, the task of a policy is to recommend a degree of belief in the proposition that x is in T. In other words, the task is to recommend a (credence) function from U into [0, 1], which represents degrees of belief regarding the truth value of $\mathrm{x} \in \mathrm{T}$, for each x in U. My intention here is to demonstrate the optimality of the following policy, δ, which corresponds to using direct inference with the most specific applicable reference classes:

Relative to a respective population model M, let $\delta(\mathrm{x} \in \mathrm{T}) = \mathrm{freq}(\mathrm{T}|\pi)$, for all x, where π is the element of Π containing x.

The policy corresponding to δ is not optimal in comparison to all possible policies, with respect to all possible population models (for example, in comparison to the 'oracular' policy, ν, that precisely tracks the truth value of all relevant propositions, i.e., $\nu(\mathrm{x} \in \mathrm{T}) = 1$, if $\mathrm{x} \in \mathrm{T}$, and $\nu(\mathrm{x} \in \mathrm{T}) = 0$, otherwise). However, δ is optimal (given a restriction on admissible accuracy measures) in comparison to (the policies represented by) the following credence functions, whose value assignments are 'principled':

Definition. A credence function, χ, is *principled* in M if and only if

$$\forall\pi\in\Pi\text{: } \forall\mathrm{x},\mathrm{y}\in\mathrm{U}\text{: if } \mathrm{x} \in \pi \text{ and } \mathrm{y} \in \pi\text{, then } \chi(\mathrm{x} \in \mathrm{T}) = \chi(\mathrm{y} \in \mathrm{T}).$$ [5]

The preceding definition tells us that a credence function is principled just in case, for each pair of objects, the same credence is assigned to both

[4] Note that the present specification of categories represents a generalization of the case where F is the set of all subsets of U, which corresponds to the case where of $\Pi = \{\{\mathrm{x}\} : \mathrm{x} \in \mathrm{U}\}$.

[5] Easwaran (2013, 124) appeals to a similar condition in showing that updating by conditionalization maximizes expected accuracy, including the case of probability functions that are defined over infinite sets of possible worlds.

elements of the pair, regarding membership in T, if the two objects have exactly the same properties, among the set of properties that one is able to distinguish. Notice that δ is principled. On the other hand, the restriction of our concern to principled credence functions excludes oracles, along with other policies that succeed by assigning different probabilities to objects that are indistinguishable, from the point of view of the policy.[6]

The optimality of δ is dependent on how we measure accuracy. I here adopt the common parlance, and refer to accuracy measures as "scoring rules". Formally, I here treat a scoring rule, S, as a function from pairs consisting of the credence assigned to a proposition, $\chi(\alpha)$, and the proposition's truth value, as represented by a standard truth-valuation function, ν. So "$\mathrm{S}(\chi(\alpha), \nu(\alpha))$" would return the score for the credence function, χ, regarding the proposition, α, given α's truth value, $\nu(\alpha)$. Since we only consider propositions concerning whether given elements of U are in T (according to a given population model M), the application of scoring rules, in the present article, takes the following form: $\mathrm{S}(\chi(\mathrm{x} \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))$ (where everything is implicitly relativized to M).

As it turns out, the optimality of the policy represented by the credence function δ holds for a broad class of highly esteemed scoring rules, namely the set of all proper scoring rules. A scoring rule is proper just in case the *expected* score earned according to the measure is *maximized* by reporting one's actual personal probabilities, i.e.:[7]

Definition. S is a *proper scoring rule* if and only if

$$\forall \mathrm{r,s}\text{: } \mathrm{S}(\mathrm{s}, 1)\times \mathrm{s} + \mathrm{S}(\mathrm{s}, 0)\times(1-\mathrm{s}) \geq \mathrm{S}(\mathrm{r}, 1)\times \mathrm{s} + \mathrm{S}(\mathrm{r}, 0)\times(1-\mathrm{s}).$$

This is not the place to summarize all of the arguments that have been made in favor of particular proper scoring rules (especially quadratic scoring rules), or proper scoring rules, generally (but see Brier (1950), de Finetti (1974), Joyce (1998), Selten (1998), Greaves & Wallace (2006), Leitgeb & Pettigrew (2010a), and Levinstein (2012)). Nevertheless, one consideration that counts in favor of such accuracy measures is worth mentioning, as illustrated by the following situation: Imagine circumstances where one is asked to report one's personal probability for some proposition, α, upon the understanding that one will receive a payoff, measured in units of utility,

[6]Note that the oracular policy, ν, will be principled in some population models, such as in population models where $\Pi = \{\{\mathrm{x}\} : \mathrm{x} \in \mathrm{U}\}$. In all such cases, $\delta(\mathrm{x} \in \mathrm{T}) = \nu(\mathrm{x} \in \mathrm{T})$, for all x in U.

[7]For the sake of uniformity, negatively oriented scoring rules (such as Brier scoring) are treated as *loss functions*, where the scores corresponding to such loss functions are determined by multiplying the loss earned according to such a rule by -1.

according to the accuracy of the report, as determined by some scoring rule, S. If S is proper, then one will expect to do best by accurately reporting one's personal probability for α. On the other hand, if S is improper, then one will expect to do better by reporting a value that differs from one's personal probability, in at least some cases. Assume, in the present case, that (1) one's personal probability for α is r, (2) S is improper, and (3) one expects to do better by reporting that one's personal probability for α is s (s $\neq$ r). Now since the 'impropriety' of improper scoring rules applies irrespective of whether one's personal probability is rational, assume that one's degree of belief regarding α is rational, given one's evidence. In that case we have a curious situation which speaks against treating S (an arbitrary improper scoring rule) as an accuracy measure, namely: PROB(α) = r is rational, but the expected accuracy of PROB(α) = s is greater than the expected accuracy of PROB(α) = r.

Many notable proper scoring rules (e.g., quadratic and logarithmic scoring rules) are sometimes called "strictly proper". A scoring rule is *strictly* proper just in case reporting one's actual personal probabilities is *unique* in maximizing one's expected score, according to the measure, i.e.:

Definition. S is a *strictly proper scoring rule* if and only if
$\forall$r,s: if r $\neq$ s, then $S(s, 1)\times s + S(s, 0)\times(1-s) > S(r, 1)\times s + S(r, 0)\times(1-s)$.

Since all strictly proper scoring rules are also proper scoring rules, any result that holds for proper scoring rules holds for strictly proper scoring rules.

The first optimality result regarding δ is as follows (with a proof appearing in the appendix):

Theorem 1. $\forall$M,χ: if χ is principled in M, then $\forall$S:

(1) if S is a *proper* scoring rule, then $\forall\pi\in\Pi$:
$\Sigma_{x\in\pi}\, S(\delta(x \in T), \nu(x \in T)) \geq \Sigma_{x\in\pi}\, S(\chi(x \in T), \nu(x \in T))$, and

(2) if S is a *strictly proper* scoring rule and $\chi \neq \delta$, then $\exists\pi\in\Pi$:
$\Sigma_{x\in\pi}\, S(\delta(x \in T), \nu(x \in T)) > \Sigma_{x\in\pi}\, S(\chi(x \in T), \nu(x \in T))$.

Theorem 1 derives from the following fact: In situations where one must assign the same credence, r, to propositions of the form x $\in$ T, for each x in some set R, one is guaranteed to maximize accuracy when r = freq(T|R), so long as accuracy is measured by a proper scoring rule. The assignment r = freq(T|R) is unique in maximizing accuracy, if accuracy is measured by a strictly proper scoring rule.

Theorem 1 establishes the optimality of reasoning by direct inference and preferring more specific reference classes, within a restricted range of cases. Accepting such limitations in applicability, it may still be objected that Theorem 1 only establishes that δ maximizes the *sum* of the scores earned for a set of personal probabilities, and that some χ might be preferable to δ, in virtue of making judgments that are more accurate in the cases which 'count for more', by making judgments that are more accurate regarding propositions $x \in T$, concerning objects x that one is more likely to encounter, for example. Although we can imagine the possibility of such a χ, Theorem 1 shows (within the sort of population models to which it applies and assuming the suitability of proper scoring rules) that there can be no sensible reason for deviating from credences that conform to δ. Indeed, if χ is principled, then whatever indicators χ employs as a basis for discerning which elements of the population 'count for more' are already reflected within Π – recall that Π corresponds to the set of maximally specific descriptions within which we are able to assign the elements of U. But Theorem 1 asserts that χ's aggregate score cannot exceed δ's regarding any element of Π. So if one limits oneself to principled strategies, then in any situation where one considered deviating from δ, regarding some category of propositions (those ones regarded as counting for more, for example) one could apply Theorem 1, and see that one would score at least as well, with respect to those propositions, by adopting credences that conform to δ.

The applicability of Theorem 1 is limited to cases where it is possible to locate each object within a reference class corresponding to a maximally specific description (as represented by an element of Π). We can generalize Theorem 1 to apply to cases where it is not possible to locate each object within such a reference class, by considering cases where the elements of U are presented under the guise of descriptions that do not necessarily correspond to membership in an element of Π, but rather merely to an element of F (i.e., an element of $\{f : f = \cup A \wedge A \subseteq \Pi\}$). In this case, I assume that our policies have access to the relative frequency of T among each and every element of F, where $\text{freq}(T|f) = |\{x : x \in f \wedge x \in T\}|/|\{x : x \in f\}|$, for all f in F.

In order to present the proposed generalization of Theorem 1, let population models be defined as before. Now consider objects under descriptions, represented as pairs $\langle x, f\rangle$, where $\langle x, f\rangle$ functions as a name for x, and f is the most specific description that x is known to satisfy under the name $\langle x, f\rangle$. It is assumed that all such descriptions are accurate, and that an agent may be acquainted with the same object under different names without realizing that the names refer to the same object (so that $\chi(\langle x, f_i\rangle \in T) \neq \chi(\langle x, f_j\rangle \in$

T) may hold of a coherent credence function). Let U^F be the set of names with respect to a population model M, i.e., $U^F = \{\langle x, f\rangle : x \in f \wedge f \in F\}$. A credence function, χ, regarding U^F is then defined as a function from U^F into [0, 1], which (intuitively) represents degrees of belief regarding whether the bearers of respective names, of the form '$\langle x, f\rangle$', are elements of T. By extension, I permit terms and expressions such as: $\chi(\langle x, f\rangle \in T)$, and $\chi(\langle x, f\rangle \in T) = s$, etc. The following generalizes the notion of principledness, in order to apply in the present context:

Definition. A credence function, χ, is *principled* in M if and only if
$\forall f\in F$: $\forall x,y\in U$: if $x \in f$ and $y \in f$, then $\chi(\langle x, f\rangle \in T) = \chi(\langle y, f\rangle \in T)$.

The extension of δ in the present context is as follows: Relative to a respective population model M, let $\delta(\langle x, f\rangle \in T) = \text{freq}(T|f)$, if x is in f. The optimality of the policy represented by δ is expressed by the following theorem:[8]

Theorem 2. $\forall M,\chi$: if χ is principled in M, then $\forall S$:
(1) if S is *proper*, then $\forall f\in F$:
$\Sigma_{x\in f}\, S(\delta(\langle x, f\rangle \in T), \nu(x \in T)) \geq \Sigma_{x\in f}\, S(\chi(\langle x, f\rangle \in T), \nu(x \in T))$, and
(2) if S is *strictly proper* and $\chi \neq \delta$, then $\exists f\in F$:
$\Sigma_{x\in f}\, S(\delta(\langle x, f\rangle \in T), \nu(x \in T)) > \Sigma_{x\in f}\, S(\chi(\langle x, f\rangle \in T), \nu(x \in T))$.

In addition to recommending a preference for direct inferences based on frequency information for narrower reference classes, Theorems 1 and 2 are applicable to explaining why one should prefer direct inferences that employ 'standard' reference classes (where the object about which one would like to make a judgment is an element of the reference class) over direct inferences where the reference class is a partition of the standard reference class. This preference is relevant to correctly arbitrating between competing direct inferences. For example, suppose that the members of a certain group are distributed among the categories *small*, *medium*, and *large*, and one would like to form a judgment about the likelihood that a particular member of the group, called "c", is large. Suppose one's information is limited, as follows: One knows that c is a member of the group (but not whether c is small, medium, or large), and that the ratio of small to medium to large members of the group is 1:8:1. In that case, it is (apparently) correct to conclude that the probability that c is large is 0.1. Where the sets of

[8]The proof of Theorem 2 is identical to that of Theorem 1, where we replace instances of π and Π, with f and F, and instances of x, x_i, and U, with $\langle x, f\rangle$, $\langle x_i, f\rangle$, and U^F, excluding instances of x and x_i in the scope of ν.

small, medium, and large members of the group are S, M, and L, the direct inference that yields this conclusion may be expressed as follows: From c $\in$ S$\cup$M$\cup$L and freq(L|S$\cup$M$\cup$L) = 0.1 infer PROB(c $\in$ L) = 0.1. On the other hand, someone might reason that one of three classifications is applicable to c (i.e., S, M, or L), and that the applicability of each classification is equally likely. Where f(c) denotes the element of {S, M, L} of which c is a member, we may attempt to underwrite the proposed conclusion by appeal to the following (highly suspect) direct inference: From f(c) $\in$ {S, M, L} and freq({L}|{S, M, L}) = 1/3 infer PROB(f(c) $\in$ {L}) = 1/3 (which entails that PROB(c $\in$ L) = 1/3). Although the latter direct inference is suspect, its reference class, {S, M, L}, is a partition rather than a superset of S$\cup$M$\cup$L. Nevertheless, accuracy considerations of the sort encapsulated by Theorems 1 and 2 may be applied in explaining the preference for the former over the latter direct inference. In particular, the policy of adopting a credence of 0.1 for propositions of the form x $\in$ L, for each x in S$\cup$M$\cup$L, yields more accurate degrees of belief than adopting a credence of 1/3 for propositions of the form x $\in$ L, for each x in S$\cup$M$\cup$L. The point illustrated by the present example is completely general: The *aggregate* accuracy of credences formed by direct inference from point-valued frequency information for a given reference class is guaranteed to be at least as great as the *aggregate* accuracy of credences formed (in the described manner) by a partition of that reference class (assuming proper scoring).[9]

It is not my intention to exaggerate the weight of Theorems 1 and 2 in providing a justification for the policy of forming one's personal probabilities via direct inference with a preference for more specific reference classes. Many have argued for the correctness of one or another proper scoring rule (especially quadratic scoring rules), and the results expressed by Theorems 1 and 2 hold for all such rules. That said, if one is open to linear scoring, which has some prima facie plausibility, one may be unmoved by Theorems 1 and 2. Moreover, Theorems 1 and 2 (while suggestive) only demonstrate the optimality of direct inference with a preference for more specific reference classes, in comparison to principled policies, in cases where an agent makes an inference about every object in the relevant domain (in the case of Theorem 1), or about every object under every accurate description (in the case of Theorem 2).[10] Finally, the optimality results expressed by The-

[9]Note that I have not argued here that all direct inferences based on reference classes that are partitions are degenerate. I only maintain that there is a preference for direct inferences that employ 'standard' reference classes versus their partitions.

[10]The results also imply the *expected* optimality of δ, in the case where an agent makes inferences about uniformly randomly selected elements of the domain.

orems 1 and 2 are limited (while suggestive), inasmuch as the theorems are only properly applicable in cases where one has access to the (point-valued) frequency of T, for every element of Π and F, respectively. In the following section, I make *some* progress in addressing the preceding limitation of the present results, by presenting analogous results that apply when one is able to make a judgment about the *expected* frequency of T for every relevant reference class. I then go on to show that we are often in a position to make point-valued expected frequency judgments, in cases where we are not warranted in accepting a respective point-valued frequency statement.

3 In the Absence of Precise-Valued Frequency Information

In the preceding section, I offered reasons for forming one's personal probabilities via direct inference with a preference for more specific reference classes, in situations where one has access to point-valued frequencies for all of the relevant reference classes. Assuming that frequency information for a narrow reference class is preferred over frequency information for a broad reference class, when the relevant frequency statements are point-valued, a further dilemma arises when choosing whether to make a direct inference based upon (i) *point-valued* frequency information for a broad reference class, R, or upon (ii) *non-point-valued* frequency information for a narrower reference class, R′ (R′ ⊂ R). More generally, there is a dilemma concerning the choice of direct inferences based upon (i) relatively *precise-valued* frequency information for broad reference classes (i.e., cases where the value of a respective frequency is known to reside within a relatively narrow interval), or upon (ii) relatively *imprecise-valued* frequency information for narrower reference classes (i.e., cases where the narrowest interval within which the value of a respective frequency is known to reside is relatively broad). The present dilemma is not inconsequential. When applied without qualification, the doctrine that one should always prefer direct inference based on more specific reference classes yields the result that one should always prefer direct inference based on the single element reference class consisting of the very object about which one would like to draw a conclusion. But if one should always prefer such single element reference classes, then it appears that direct inference can never be used to draw an informative conclusion (cf. Kyburg 1974).

I will here present a new approach to the problem of choosing between direct inferences based on (i) relatively precise-valued frequency information for broad reference classes, versus (ii) relatively imprecise-valued frequency information for more specific reference classes.[11] I address the problem by showing that it is usually possible to use direct inference to make a (relatively) precise-valued frequency estimate for a more specific reference class based on (relatively) precise-valued frequency information for a respective broad reference class. Having made such a frequency estimate, the dilemma of choosing between (i) and (ii) is removed, and one may proceed by using the precise-valued frequency estimate for the more specific reference class as a premise for direct inference. The proposed approach turns on some observations concerning the combinatorial properties of sets, and on the thesis that the 'proper' statistical statements that may serve as major premises for direct inference are actually statements of *expected* frequency.

I will proceed, in Section 3.1, by articulating and defending the claim that the proper major premises for direct inference are statements of expected frequency. In Sections 3.2 and 3.3, I propose two methods that often permit one to infer a (relatively) precise-valued expected frequency for a more specific reference class, R', based on (relatively) precise-valued frequency information for a broad reference class, R ($R' \subset R$). In Section 3.4, I describe the respective roles of the two methods, and how they work together. Finally, in Section 3.5, I argue that it is rational to adopt the expectations generated by the proposed methods.

3.1 Expected Frequencies as the Basis for Direct Inference

The thesis that it is statements of *expected* frequency that are the proper statistical premises for direct inference is found in (Bacchus 1990), and defended at some length in (Thorn 2012). I here rehearse some of the considerations adduced in (Thorn 2012) which are sufficient to lend plausibility to the thesis.

I begin by noting that expected frequencies are simply probability weighted averages of frequencies, relativized to a respective probability function. So the expected frequency of T among R, written "E[freq(T|R)]", relativized to a probability function, PROB, is defined as follows:

$$E[\mathrm{freq}(T|R)] = \Sigma_r\ r \times \mathrm{PROB}(\mathrm{freq}(T|R) = r).$$

Notice that the use of known frequencies as premises for direct inference is a special case of the use of expected frequencies, since if $\mathrm{PROB}(\mathrm{freq}(T|R) = r) = 1$, then $E[\mathrm{freq}(T|R)] = r$. More generally, if $\mathrm{PROB}(\mathrm{freq}(T|R) \in$

[11] For a survey of past approaches to the present problem, including those of Bacchus (1990), Pollock (1990), and Kyburg & Teng (2001), see Thorn (2012).

S) = 1 and U is the smallest interval such that $S \subseteq U$, then E[freq(T|R)] $\in$ U (Thorn 2012). It is also of interest to note that expected frequencies generalize single case probabilities, in the following manner: PROB($c \in T$) = PROB(freq(T|{c}) = 1) = E[freq(T|{c})].

A good reason for regarding statements of expected frequency as the proper statistical premises for direct inference is connected with the assumption one makes when one performs a direct inference. When making a direct inference, one assumes that the object about which one is reasoning, c, is as likely to be a member of the respective target class, T, as a uniformly randomly selected element of the proposed reference class, R. Specifying the precise conditions under which it is reasonable to make this assumption is an open problem.[12] Regardless, if one does assume that c is as likely to be in T as a random element of R (selected according to a uniform distribution), then one is obliged to conclude that the probability that c is in T is equal to the frequency of elements of T among R, in cases where one is aware of the value of this frequency. Similarly, one is obliged to conclude that the probability that c is in T is equal to the expected frequency of T among R, since the probability that a random element of R is an element of T is identical to the expected frequency of T among R (provided one makes the reasonable assumption that independence obtains between what value freq(T|R) takes and which element of R is selected) (cf. Thorn 2012).

Given the preceding, I now assume that properly formulated direct inferences satisfy the following schemata:

From $c \in R$ and E[freq(T|R)] = r infer that PROB($c \in T$) = r.

From $c \in R$ and E[freq(T|R)] $\in$ S infer that PROB($c \in T$) $\in$ S.

Since PROB(freq(T|R) = r) = 1 implies E[freq(T|R)] = r, it is also acceptable to formulate direct inferences using the schema introduced in Section 1.

It may be observed, at this point, that the results concerning the optimality of direct inference based on more specific reference classes (Theorems 1 and 2) are inapplicable in the case where we base our direct inferences on expected frequencies. Nevertheless, it is straightforward to generalize the two theorems, in order to demonstrate that direct inference using expected frequencies for the most specific relevant reference classes has the highest *expected* accuracy among the set of principled policies. These generalizations

[12] I will touch on this problem briefly in Section 3.5, in connection with the classical principle of indifference.

of Theorems 1 and 2 provide further reasons in favor of the claim that it is statements of expected frequency that are the proper statistical premises for direct inference.

In order to carry out the generalization of Theorem 1, redefine δ, so that $\delta(\mathrm{x} \in \mathrm{T}) = \mathrm{E}[\mathrm{freq}(\mathrm{T}|\pi)]$, if x is in π. We then have the following result:

Theorem 3. $\forall$M,χ: if χ is principled in M, then $\forall$S:

(1) if S is *proper*, then $\forall\pi\in\Pi$:
$\mathrm{E}[\Sigma_{\mathrm{x}\in\pi}\ \mathrm{S}(\delta(\mathrm{x} \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))] \geq \mathrm{E}[\Sigma_{\mathrm{x}\in\pi}\ \mathrm{S}(\chi(\mathrm{x} \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))]$, and

(2) if S is *strictly proper* and $\chi \neq \delta$, then $\exists\pi\in\Pi$:
$\mathrm{E}[\Sigma_{\mathrm{x}\in\pi}\ \mathrm{S}(\delta(\mathrm{x} \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))] > \mathrm{E}[\Sigma_{\mathrm{x}\in\pi}\ \mathrm{S}(\chi(\mathrm{x} \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))]$.

Theorem 3 is a straightforward consequence of Theorem 1: Theorem 1 tells us that the described inequalities hold regardless of the value of $\mathrm{freq}(\mathrm{T}|\pi)$. Theorem 3 tells us that the inequalities hold for any weighted average of the values of $\mathrm{freq}(\mathrm{T}|\pi)$.

In order to carry out the generalization of Theorem 2, redefine δ, so that $\delta(\langle\mathrm{x}, \mathrm{f}\rangle \in \mathrm{T}) = \mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{f})]$, if x is in f. In that case, we have the following result (which is a straightforward consequence of Theorem 2):

Theorem 4. $\forall$M,χ: if χ is principled in M, then $\forall$S:

(1) if S is *proper*, then $\forall\mathrm{f}\in\mathrm{F}$:
$\mathrm{E}[\Sigma_{\mathrm{x}\in\mathrm{f}}\ \mathrm{S}(\delta(\langle\mathrm{x}, \mathrm{f}\rangle \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))] \geq \mathrm{E}[\Sigma_{\mathrm{x}\in\mathrm{f}}\ \mathrm{S}(\chi(\langle\mathrm{x}, \mathrm{f}\rangle \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))]$, and

(2) if S is *strictly proper* and $\chi \neq \delta$, then $\exists\mathrm{f}\in\mathrm{F}$:
$\mathrm{E}[\Sigma_{\mathrm{x}\in\mathrm{f}}\ \mathrm{S}(\delta(\langle\mathrm{x}, \mathrm{f}\rangle \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))] > \mathrm{E}[\Sigma_{\mathrm{x}\in\mathrm{f}}\ \mathrm{S}(\chi(\langle\mathrm{x}, \mathrm{f}\rangle \in \mathrm{T}), \nu(\mathrm{x} \in \mathrm{T}))]$.

Theorems 3 and 4 demonstrate that the policy of forming one's personal probabilities by direct inference based on expected frequencies for the most specific relevant reference classes yields personal probabilities whose expected accuracy is maximal, among the field of principled competitors. So if one cares about the expected accuracy of one's personal probabilities, then it behooves one to form one's personal probabilities in the described manner. The urgency of the present injunction is, of course, contingent upon the 'correctness' of the probabilities with which one's expectations are defined. In the present case, the relevant probabilities are fixed by one's expected frequencies for the relevant reference classes. So the force of the injunction to form one's personal probabilities by direct inference based on expected frequencies (for the most specific relevant reference classes) is, in some sense, dependent on the correctness of one's expected frequencies.

Beyond considerations of 'normative' applicability, the formal applicability of Theorems 3 and 4 is contingent upon forming point-valued expected frequency judgments for all of the relevant reference classes. This marks an improvement over Theorems 1 and 2, whose applicability was contingent on having access to the actual frequencies for all of the relevant reference classes. Despite this improvement (or extension of applicability), it is still demanding to suppose that one is generally in a position to make point-valued expected frequency judgments for all relevant reference classes, especially if one's concern is to make expected frequency judgments that are rational, as measured by some appropriate normative standard. While I doubt that one is always in a position to make a rational point-valued expected frequency judgment for every reference and target class, my aim in the following subsections is to outline two methods that permit one to make point-valued expected frequency judgments in a wide range of cases.

3.2 Inferring Expected Frequencies for More Specific Reference Classes: Method I

It is intended that the methods proposed in the present and following subsection be applied in reasoning from information concerning the incidence of T among a reference class R to a conclusion concerning the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, where $\mathrm{R}' \subset \mathrm{R}$. In the cases that interest us, our aim is to form a personal probability about the proposition $\mathrm{c} \in \mathrm{T}$, for an object c, where c is a member of a reference class R', R' is a subset of R, and we are not in a position to make a point-valued judgment concerning the value of $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$. By assumption, the relative frequency of T among R is known. For the moment, I also assume that for some numeric value z, we are warranted in accepting $|\mathrm{R}| = \mathrm{z}$. I will explain, below, how to dispense with this assumption.

In applying the proposed method in order to form a judgment about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, we proceed in two steps.

Step 1: Make a series of direct inferences to form conclusions of the form $\mathrm{PROB}(\mathrm{freq}(\mathrm{T}|\mathrm{R}') = \mathrm{v}_i) = \mathrm{p}_i$, for each v_i within the smallest set of numerically expressed values in which $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$ is known to lie.

Step 2: Use the conclusions formed in Step 1 to infer the expectation of $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$, according to the equation: $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] = \Sigma_i\ \mathrm{v}_i \times \mathrm{p}_i$.

Executing the first step of the proposed method is somewhat complicated. Before providing a general description of how to proceed in Step 1, I illustrate the proposed method using the following simple example.

Suppose we are trying to assign a probability to the proposition that a member of Company B, named “Bill”, is an NCO. Suppose we know that 25% of the 100 soldiers in Company B are NCOs. However, suppose we also know that Bill is a member of the command unit of Company B, which has 10 members, and we know that either 20% or 30% of the members of the command unit are NCOs, and we do not know which. (Suppose the percentage of NCOs in command units varies according to whether a respective company is an artillery or infantry company, and we do not know whether Company B is artillery or infantry.) The problem now is to draw a conclusion about the expected frequency of NCOs among the command unit of Company B (which can then be used to infer the probability that Bill is an NCO).

By assumption, the frequency of NCOs among the command unit of Company B takes one of two possible values, namely, 0.2 or 0.3. So we could draw a conclusion about the expected frequency of NCOs among the command unit, if we could draw a conclusion about the probability that the frequency of NCOs among the command unit of Company B is 0.2, and a conclusion about the probability that the frequency of NCOs among the command unit of Company B is 0.3. As a basis for drawing the needed conclusions, notice that the command unit of Company B is an element of the reference class composed of the ten membered subsets of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3. Next notice that we are in a position to compute the values of the following frequencies concerning this reference class: (1) the frequency of sets with a frequency of 0.2 NCOs among the ten membered subsets of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3, and (2) the frequency of sets with a frequency of 0.3 NCOs among the ten membered subsets of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3. Given that there are 100 soldiers in Company B, and 25 are NCOs, the former frequency is identical to the number of ways of selecting 2 of the 25 NCOs along with 8 of the 75 non-NCOs divided by *the sum of* the number of ways of selecting 2 of the 25 NCOs along with 8 of the non-NCOs *and* the number of ways of selecting 3 of the 25 NCOs along with 7 of the non-NCOs. The former frequency is, thus, $(\binom{25}{2} \times \binom{75}{8}) \,/\, (\binom{25}{2} \times \binom{75}{8} + \binom{25}{3} \times \binom{75}{7}) \approx 0.515$. The latter frequency can also be computed via straightforward combinatorial methods. That frequency is $(\binom{25}{3} \times \binom{75}{7}) \,/\, (\binom{25}{2} \times \binom{75}{8} + \binom{25}{3} \times \binom{75}{7}) \approx 0.485$.

Armed with the preceding frequency statements, we are in a position to (use direct inference) to draw conclusions about the two probabilities that are required for computing the expected frequency of NCOs among the command unit of Company B, as follows:

1. The set of soldiers in the command unit of Company B is a ten membered subset of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3. Among the ten membered subsets of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3, about 51.5% have a frequency of 0.2 NCOs. So it is reasonable to infer (by direct inference) that the probability is (about) 51.5% that the frequency of NCOs among the command unit of Company B is 0.2.

2. The set of soldiers in the command unit of Company B is a ten membered subset of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3. Among the ten membered subsets of the set of soldiers in Company B whose frequency of NCOs is either 0.2 or 0.3, about 48.5% have a frequency of 0.3 NCOs. So it is reasonable to infer (by direct inference) that the probability is (about) 48.5% that the frequency of NCOs among the command unit of Company B is 0.3.

The preceding two direct inferences license assignments of probability to the two propositions concerning the possible frequencies of NCOs among the command unit of Company B. Given these conclusions, we can compute the expected frequency of NCOs among the command unit of Company B (Step 2), namely: The expected frequency of NCOs among the command unit of Company B is 0.2 times the probability that the frequency of NCOs among the command unit is 0.2 *plus* 0.3 times the probability that the frequency of NCOs among the command unit is 0.3. In accordance with the two above direct inferences, the present sum is approximately $0.2 \times 0.515 + 0.3 \times 0.485 = 0.2485$. The conclusion that the expected frequency of NCOs among the command unit of Company B is (approximately) 0.2485 can now be used to draw a conclusion regarding the probability that Bill is an NCO, namely: From the fact that Bill is a member of the command unit of Company B and the fact that the expected frequency of NCOs among the command unit is (approximately) 0.2485 infer that the probability that Bill is an NCO is (approximately) 0.2485.

The two step procedure that was used to infer the expected frequency of NCOs among the command unit of Company B is completely general. The complicated part of the procedure is Step 1, which (in the general case) involves drawing a conclusion of the form $\mathrm{PROB}(\mathrm{freq}(\mathrm{T}|\mathrm{R}') = \mathrm{v}_i) = \mathrm{p}_i$, for each v_i within the smallest set of numerically expressed values in which $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$ is known to lie. In the case of the command unit of Company B, the smallest set of values in which $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$ was known to lie was $\{0.2, 0.3\}$. In the general case, the set of possible values of $\mathrm{freq}(\mathrm{T}|\mathrm{R}')$ may be larger. As in the case of Company B, we will have to make a direct inference concerning the value of $\text{PROB}(\text{freq}(\text{T}|\text{R}') = \text{v}_i)$, for each such v_i. In order to comply with considerations of specificity, the reference class for the direct inferences will be $\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) \in \text{V}\}$, where W and V are the *smallest* sets of numerically expressed values in which $|\text{R}'|$ and $\text{freq}(\text{T}|\text{R}')$, respectively, are known to lie. As with the two direct inferences used in the case of the command unit of Company B, the direct inferences, in the general case, take the following form:

From $\text{R}' \in \{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) \in \text{V}\}$ and
$\text{freq}(\{\text{S} : \text{freq}(\text{T}|\text{S}) = \text{v}_i\}|\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) \in \text{V}\}) = \text{p}_i$
infer that $\text{PROB}(\text{R}' \in \{\text{S} : \text{freq}(\text{T}|\text{S}) = \text{v}_i\}) = \text{p}_i$
(i.e., infer that $\text{PROB}(\text{freq}(\text{T}|\text{R}') = \text{v}_i) = \text{p}_i$).

As in the case of Company B, we are in a position to make the described direct inferences, since R' is an element of the respective reference class, and since we are in a position to compute the value of p_i, for each major premise, by appeal to elementary combinatorial principles. The following theorem outlines a means of computing the value of p_i, given v_i, $\text{freq}(\text{T}|\text{R})$, and $|\text{R}|$:

Theorem 5. $\forall \text{T,R,W,V,v}_i$: $\text{freq}(\{\text{S} : \text{freq}(\text{T}|\text{S}) = \text{v}_i\}|\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) \in \text{V}\}) = \Sigma_{\text{w}\in\text{W}} \binom{|\text{R}|\times\text{freq}(\text{T}|\text{R})}{\text{w}\times\text{v}_i} \times \binom{|\text{R}|\times(1-\text{freq}(\text{T}|\text{R}))}{\text{w}\times(1-\text{v}_i)} \Big/ \Sigma_{\text{w}\in\text{W},\text{v}\in\text{V}} \binom{|\text{R}|\times\text{freq}(\text{T}|\text{R})}{\text{w}\times\text{v}} \times \binom{|\text{R}|\times(1-\text{freq}(\text{T}|\text{R}))}{\text{w}\times(1-\text{v})}$.[13]

The described method permits inference to a numeric point-valued conclusion about the value of $\text{E}[\text{freq}(\text{T}|\text{R}')]$, in cases where one is warranted in accepting $|\text{R}| = \text{z}$, for some numeric value z. When the latter condition is not met, we may apply a variation of the proposed method, so long as there is a smallest finite set of (finite) numeric values, Z, such that we are warranted in accepting that $|\text{R}|$ is in Z. In such situations, we reason by cases. Figure 1 illustrates how such reasoning would proceed in the case of the command unit of Company B, in a situation where we did not know the number of soldiers in Company B, and knew only that there was either 100 or 120 soldiers. In that case, we would apply two instances of Method I: one according to the assumption that there is 100 soldiers in Company B, and another according to the assumption that there is 120. We would then

[13]Both of the mentioned values are identical to: $|\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) = \text{v}_i\}| \,/\, |\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| \in \text{W} \wedge \text{freq}(\text{T}|\text{S}) \in \text{V}\}|$.

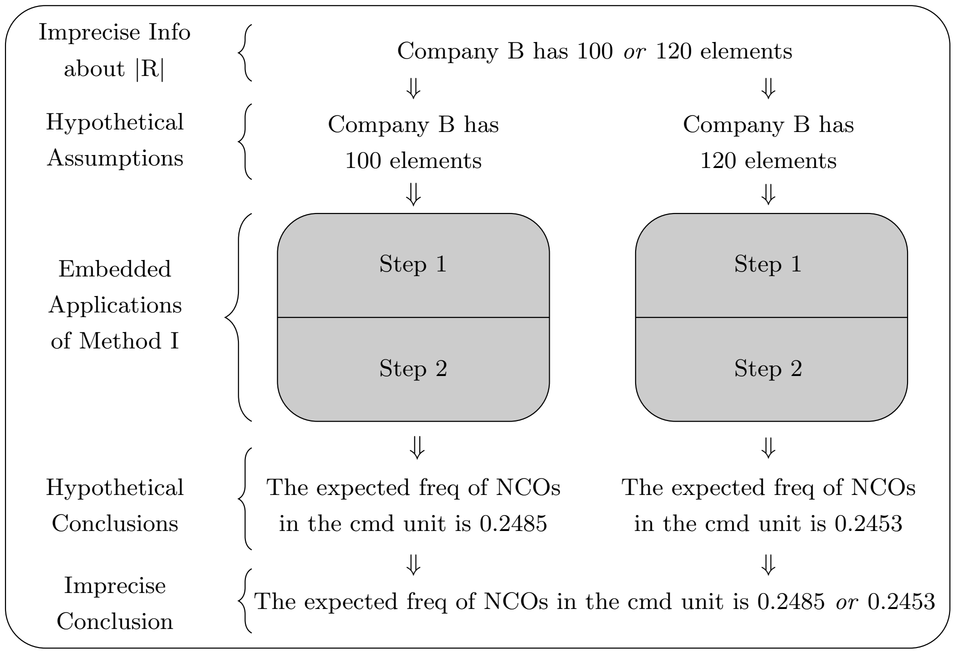

Figure 1: Inference by cases with embedded applications of Method I

accept the disjunction of the two (hypothetical) conclusions. Fortunately, the size of |R| has only a limited bearing upon the value of E[freq(T|R′)] as determined by the basic form of the proposed method. As a consequence, our inability to make a precise-valued judgment about the value of |R| will not prevent us from making a relatively precise judgment about the value of E[freq(T|R′)]. For example, if we hold all other features of the example of Bill of Company B fixed, but suppose that the size of Company B may take any value within the sequence of values ranging from 20 to 2,000,000,000, then we may reason by cases in order to infer that the expected frequency of NCOs among the command unit of Company B lies within the interval [0.231, 0.262].

Before proceeding, note that the just described case-based application of the proposed method is very general. In addition to applying in the case where we are not warranted in making a point-valued judgment about the value of |R|, we may reason in a similar manner when we are not warranted in making a point-valued judgment about the value of freq(T|R). Further, in the absence of point-valued information concerning the value of both |R| and freq(T|R), we may proceed by treating the combinations of possible values of $|R|$ and $freq(T|R)$ as hypothetical assumptions for a case-based application of the proposed method. Evidently, it would also be reasonable to apply a variant of such case-based inference, in situations where we are in a position to assign probabilities to the respective cases. In such situations, our judgment concerning the value of $E[freq(T|R')]$ would be determined by taking a weighted average of the hypothetical conclusions reached within the respective cases, where the respective weights are the probabilities assigned to the corresponding hypothetical assumptions.

Proper (non-case-based) applications of Method I permit inference to a point-valued conclusion concerning the value of a respective expected frequency, $E[freq(T|R')]$, thereby extending the applicability of the optimality results presented in the preceding subsection. Beyond this, case-based applications of Method I are of general interest, since they are applicable to removing the dilemma of choosing between direct inferences based upon precise-valued frequency information for broad reference classes, and direct inferences based upon imprecise-valued frequency information for more specific reference classes.

3.3 Inferring Expected Frequencies for More Specific Reference Classes: Method II

The method proposed in the present subsection represents a streamlined variant of the method of the preceding subsection. The limitation of the streamlined variant is that it is only applicable in cases where one is not warranted in accepting that $freq(T|R') \in v$, for any set of numerically expressed values, v, such that $v \subset \{0/|R'|, 1/|R'|, ..., |R'|/|R'|\}$. In such cases, the method permits one to infer that $E[freq(T|R')] = freq(T|R)$. For example, in a situation where we had *no* information about the frequency of NCOs in the command unit of Company B, we could use the method to infer that the expected frequency of NCOs in the command unit of Company B is identical to the frequency of NCOs in Company B.

As with Method I, Method II follows a two step procedure in inferring the expectation of $freq(T|R')$. In Step 1, we make a series of direct inferences, in order to draw a conclusion of the form $PROB(freq(T|R') = v_i) = p_i$, for each v_i in $\{0/|R'|, 1/|R'|, ..., |R'|/|R'|\}$. In Step 2, we use the conclusions formed in the first step in order to infer the expectation of $freq(T|R')$. The main difference between the two methods is in the reference class used in the first step. In the case of Method II, the reference class simply consists in the set of subsets of R whose size is identical to the size of R'. For example, in a situation where we had no information about the frequency of NCOs in the command unit of Company B, the reference class used in the first step would be: the subsets of the members of Company B, whose size was identical to the size of the command unit of Company B. Using this reference class, we would make one direct inference for each of the eleven possible frequencies of NCOs among the command unit of Company B (i.e., 0.0, 0.1. 0.2, ..., 0.9, 1.0), as follows:

1. The set of soldiers in the command unit of Company B is a subset of the set of soldiers in Company B whose size is identical to the size of the set of soldiers in the command unit of Company B. About 4.8% of the sets among this reference class have a frequency of 0.0 NCOs. So it is reasonable to infer (by direct inference) that the probability is (about) 4.8% that the frequency of NCOs among the command unit of Company B is 0.0.

2. The set of soldiers in the command unit of Company B is a subset of the set of soldiers in Company B whose size is identical to the size of the set of soldiers in the command unit of Company B. About 18.1% of the sets among this reference class have a frequency of 0.1 NCOs. So it is reasonable to infer (by direct inference) that the probability is (about) 18.1% that the frequency of NCOs among the command unit of Company B is 0.1.

Etc.

Written formally, the preceding direct inferences are instances of the following schema:

From $\mathrm{R}' \in \{\mathrm{S} : \mathrm{S} \subseteq \mathrm{R} \wedge |\mathrm{S}| = |\mathrm{R}'|\}$ and
$\mathrm{freq}(\{\mathrm{S} : \mathrm{freq}(\mathrm{T}|\mathrm{S}) = i/|\mathrm{R}'|\}|\{\mathrm{S} : \mathrm{S} \subseteq \mathrm{R} \wedge |\mathrm{S}| = |\mathrm{R}'|\}) = \mathrm{p}_i$
infer that $\mathrm{PROB}(\mathrm{R}' \in \{\mathrm{S} : \mathrm{freq}(\mathrm{T}|\mathrm{S}) = i/|\mathrm{R}'|\}) = \mathrm{p}_i$
(i.e., infer that $\mathrm{PROB}(\mathrm{freq}(\mathrm{T}|\mathrm{R}') = i/|\mathrm{R}'|) = \mathrm{p}_i$).

As I already mentioned, Method II always yields the result that $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] = \mathrm{freq}(\mathrm{T}|\mathrm{R})$. To see why this is so, notice that the content of the conclusions of the direct inferences made in the first step may be re-written as follows:

$\mathrm{PROB}(\mathrm{freq}(\mathrm{T}|\mathrm{R}') = i/|\mathrm{R}'|) =$
$\mathrm{freq}(\{\mathrm{S} : \mathrm{freq}(\mathrm{T}|\mathrm{S}) = i/|\mathrm{R}'|\}|\{\mathrm{S} : \mathrm{S} \subseteq \mathrm{R} \wedge |\mathrm{S}| = |\mathrm{R}'|\})$.

Given the set of such conclusions, the antecedent of the following theorem is satisfied, thereby permitting us to infer that $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] = \mathrm{freq}(\mathrm{T}|\mathrm{R})$:

Theorem 6. $\forall$T,R,R$'$: if $\mathrm{R}' \subseteq \mathrm{R}$ and $\forall i$: $\mathrm{PROB}(\mathrm{freq}(\mathrm{T}|\mathrm{R}') = i/|\mathrm{R}'|) = \mathrm{freq}(\{\mathrm{S} : \mathrm{freq}(\mathrm{T}|\mathrm{S}) = i/|\mathrm{R}'|\}|\{\mathrm{S} : \mathrm{S} \subseteq \mathrm{R} \wedge |\mathrm{S}| = |\mathrm{R}'|\})$, then $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] = \mathrm{freq}(\mathrm{T}|\mathrm{R})$.

Theorem 6 is based on the fact that, for any sets T and R, the average value of the frequency of T among subsets of R of any given size is identical to the frequency of T among R. For example, if (as in the case of Company B) the size of R is 100, and 25 elements of R are elements of T (i.e., $\text{freq}(\text{T}|\text{R}) = 0.25$), then the average number of elements of T among 10 element subsets of R will be 2.5 (and the average frequency of T among such sets will be 0.25). Now notice that such an average is, obviously, identical to the average of the possible frequency values for such subsets, weighted according to the number of subsets having the respective frequency value. So, in the case of Company B, the average of the (eleven) possible values of the frequency of T among the 10 element subsets of R, weighted according to the number of subsets having the respective frequency, is 0.25. Theorem 6 weds the preceding observation to the fact that expectations are (probability) weighted averages, and provides a means of inferring the expected frequency of T among R′, when R′ is a subset of R. In particular, Theorem 6 tells us that if, for each value $i/|\text{R}'|$, we identify the probability that $\text{freq}(\text{T}|\text{R}')$ takes that value with the frequency with which subsets of R (of size $|\text{R}'|$) take that value as their frequency of T, then $\text{E}[\text{freq}(\text{T}|\text{R}')]$ will be identical to $\text{freq}(\text{T}|\text{R})$. The latter identity holds, since the average value of the frequency of T among the subsets of R of size $|\text{R}'|$ (or of any size), weighted according to the number of subsets having the respective frequency, is $\text{freq}(\text{T}|\text{R})$.[14]

In applying Method II to infer the expected frequency of NCOs among the command unit of Company B, it was assumed that we knew the sizes of R and R′, i.e., the size of Company B and the size of the command unit of Company B. A key advantage of Method II is its applicability even when we lack such knowledge. In such cases, we may treat the respective p_i, that appear in the major premises and conclusions of the direct inferences of Step 1, as constants that are introduced according to the following definition:

$$\text{p}_i = \text{freq}(\{\text{S} : \text{freq}(\text{T}|\text{S}) = i/|\text{R}'|\}|\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| = |\text{R}'|\}).$$

In that case, the major premises of the direct inferences made in Step 1 are analytic, and the conclusions of the direct inferences may be re-written (once again) as follows:

$$\text{PROB}(\text{freq}(\text{T}|\text{R}') = i/|\text{R}'|) =$$
$$\text{freq}(\{\text{S} : \text{freq}(\text{T}|\text{S}) = i/|\text{R}'|\}|\{\text{S} : \text{S} \subseteq \text{R} \wedge |\text{S}| = |\text{R}'|\}).$$

[14] A proof of Theorem 6 is given in the appendix.

Given the set of such conclusions, we are in a position to apply Theorem 6, in order to infer that $E[freq(T|R')] = freq(T|R)$.[15]

In cases where we are in a position to see, in advance, that it is possible to use the proposed method, it is practical to reason in accordance with the following defeasible inference schema:

From $R' \subseteq R$ and $freq(T|R) = r$ infer that $E[freq(T|R')] = r$.

The present method is, thus, of theoretical interest, since in the case where $R' = \{c\}$, it recapitulates the inference schema introduced in Section 1, which is equivalent to:

From $\{c\} \subseteq R$ and $freq(T|R) = r$ infer $E[freq(T|\{c\})] = r$.[16]

Now note that the present method for inferring the value of $E[freq(T|R')]$ generalizes to the case where our inferences are based on a judgment concerning the value of $E[freq(T|R)]$, rather than the value of $freq(T|R)$. In this case, it is possible to infer that $E[freq(T|R')] = E[freq(T|R)]$. Our inference to the value of $E[freq(T|R')]$, in such cases, proceeds by direct inferences of the following form:

From $R' \in \{S : S \subseteq R \wedge |S| = |R'|\}$ and
$E[freq(\{S : freq(T|S) = i/|R'|\}|\{S : S \subseteq R \wedge |S|] = |R'|\}) = p_i$ infer that
$PROB(freq(T|R') = i/|R'|) = p_i$.

The value of $E[freq(T|R')]$ then follows by the following theorem:

Theorem 7. $\forall$T,R,R': if $R' \subseteq R$ and $\forall i$: $PROB(freq(T|R') = i/|R'|) = E[freq(\{S : freq(T|S) = i/|R'|\}|\{S : S \subseteq R \wedge |S| = |R'|\})]$, then $E[freq(T|R')] = E[freq(T|R)]$.

Theorem 7 is a straightforward consequence of Theorem 6: Theorem 6 tells us that the identity expressed in its consequent holds regardless of the value of $freq(T|R)$, given an assignment of probabilities according to the frequencies mentioned in its antecedent. Theorem 7 tells us that a corresponding identity holds for any weighted average of the values of $freq(T|R)$,

[15] In cases where the size of $|R'|$ is unknown, let s^+ be the least upper bound that one is warranted in accepting regarding the size of $|R'|$. In Step 1, we then make one direct inference for each value of i in $\{0, ..., s^+\}$.

[16] Recall that $PROB(c \in T) = E[freq(T|\{c\})]$.

given an assignment of probabilities according to a corresponding weighted average of the frequencies mentioned in the antecedent of Theorem 6.

It is noteworthy that Theorem 7, in conjunction with its intended application, illustrate how one may derive (using terms introduced by Pollock (1990)) *non-classical* direct inference from *classical* direct inference, as represented by the following defeasible inference schema:

From $R' \subseteq R$ and $E[freq(T|R)] = r$ infer that $E[freq(T|R')] = r$.

So the present method yields a picture of direct inference that stands in contrast to the accounts of direct inference articulated by Pollock (1990) and Bacchus (1990), which propose to derive classical direct inference from non-classical direct inference. Furthermore, the present method allows us to uphold the doctrine that $E[freq(T|R)] = E[freq(T|R')] = E[freq(T|\{c\})] = r$, for all non-gerrymandered R', such that $\{c\} \subseteq R' \subseteq R$, in cases where it is appropriate to use direct inference to infer that $PROB(c \in T) = r$, based on frequency information for the reference class R.[17]

A worry regarding Method II concerns the peculiar use of constants, in cases where the sizes of R and R′ are unknown. In order to dispel this concern, I will now show that the use of such constants is not essential to reaching the conclusions licensed by the proposed method. Indeed, it is possible to infer those conclusions using case-based inference of the sort introduced in the preceding subsection. I adapt an example of Bradley and Steele (2014) in order to illustrate how such inferences proceed.[18]

According to the example of Bradley and Steele (2014), we know that there are 10 black marbles and 10 white marbles, which are divided between two urns, with each urn containing 10 marbles. Our task in Bradley and Steele's original example is to judge the probability that a given ball drawn from the 'first urn' (an urn selected at random) is white. In order to generate additional uncertainty, I here consider a variant of the example that introduces uncertainty both concerning the total number of marbles, and the number of marbles in the first urn: Rather than know that there are 10 black marbles and 10 white marbles, assume we know that there is either (case 1) 10 black marbles and 10 white marbles, or (case 2) 20 black marbles

[17] The restriction in the applicability of theorem 7 to cases where one is not warranted in accepting that $freq(T|R') \in v$, for any v, such that $v \subset \{0/|R'|, 1/|R'|, ..., |R'|/|R'|\}$ is also suggestive of where past accounts of direct inference (with the possible exception of Thorn 2012) go wrong in the face of Stone's Ace Urn example (Stone 1987, 251).

[18] The example of Bradley and Steele (2014) is meant to serve as a plausible example of credence dilation. If the present treatment of the example is correct, it cannot serve as an example of *rational* credence dilation.

and 20 white marbles. Next, suppose we know that one of the urns contains more balls than the other. In particular, suppose we know that the first urn contains 40% of the balls (subcase A), or that the first urn contains 60% of the balls (subcase B). In this situation, we can infer the expected frequency of white balls among the first urn using case-based reasoning of the sort introduced at the end of subsection 3.3, where hypothetical conclusions are determined by applying Method II, rather than Method I. There are four cases (1A, 1B, 2A, and 2B). Where M is the set of marbles in the two urns, U the set of marbles in the first urn, and W the set of white marbles, our objective is to compute the value of E[freq(W|U)], within each case. We proceed, according to the assumed values of |M| and |U|, as given by the relevant case, by conducting a direct inference of the following form, for each of the possible values of i ($i \in \{0, ..., |U|\}$):

From $U \in \{S : S \subseteq R \wedge |S| = |U|\}$ and
$\text{freq}(\{S : \text{freq}(W|S) = i/|U|\}|\{S : S \subseteq R \wedge |S| = |U|\}) = p_i$
infer that $\text{PROB}(\text{freq}(W|U) = i/|U|) = p_i$.

For each such direct inference, it is possible to directly compute the value of the respective p_i, as follows:

$$\text{freq}(\{S : \text{freq}(W|S) = i/|U|\}|\{S : S \subseteq M \wedge |S| = |U|\}) = \binom{|M| \times \text{freq}(W|M)}{i} \times \binom{|M| \times (1-\text{freq}(W|M))}{|U|-i} \Big/ \Sigma_{k \in \{0,...,|U|\}} \binom{|M| \times \text{freq}(W|M)}{k} \times \binom{|M| \times (1-\text{freq}(W|M))}{|U|-k}.$$

Since instances of the preceding equation contain no numeric constants whose values are unknown within the respective cases, it is possible to directly calculate the value of p_i, for each possible value of i, within each case. Before embarking on such a cumbersome calculation, Theorem 6 informs us of the result of concluding that $\text{PROB}(\text{freq}(W|U) = i/|U|) = \text{freq}(\{S : \text{freq}(W|S) = i/|U|\}|\{S : S \subseteq M \wedge |S| = |U|\})$, for all $i \in \{0, ..., |U|\}$, namely: $E[\text{freq}(W|U)] = \text{freq}(W|M) = 0.5$. Since the preceding identities hold in every case, it is correct to conclude that $E[\text{freq}(W|U)] = 0.5$.

The treatment of the preceding example provides an illustration of how we could have reached the conclusions outlined by Method II by directly computing the values for the major premises that underpin the method, despite being unable to make a precise-valued judgment about the sizes of R and R′.

As with Method I, applications of Method II permit inference to a point-valued conclusion concerning the value of a respective expected frequency, thereby extending the applicability of the optimality results presented in Section 3.1. Similar to Method I, applications of Method II may also be embedded within case-based inferences of the sort described at the end of Section 3.2. Such embeddings permit inference according to the following schemata:

From $\mathrm{R}' \subseteq \mathrm{R}$ and $\mathrm{freq}(\mathrm{T}|\mathrm{R}) \in \mathrm{S}$ infer that $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] \in \mathrm{S}$.

From $\mathrm{R}' \subseteq \mathrm{R}$ and $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R})] \in \mathrm{S}$ infer that $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')] \in \mathrm{S}$.

Such case-based applications of Method II are of general interest, since they are applicable to removing the dilemma of choosing between direct inferences based upon precise-valued frequency information for broad reference classes, and direct inferences based upon imprecise-valued frequency information for more specific reference classes.

3.4 Combining the Two Methods

It is intended that the methods of the preceding subsections have their own 'spheres of influence'. Before outlining the respective spheres (and the manner in which the two methods combine), it is helpful to observe the manner in which a variant of Method I may be used in order to form a judgment about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, given frequency information concerning two nested reference classes R_1 and R_2, where $\mathrm{R}' \subset \mathrm{R}_1 \subset \mathrm{R}_2$. In such cases, one would first make a series of direct inferences (of exactly the sort one makes in Step 1 of typical applications of Method I), in order to form an assignment of probabilities to the possible values of $\mathrm{freq}(\mathrm{T}|\mathrm{R}_1)$ (given one's frequency information for R_2). One would, then, apply a case-based variant of Method I, where each case yields a hypothetical conclusion about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, given a particular assumption about the value of $\mathrm{freq}(\mathrm{T}|\mathrm{R}_1)$. One's final conclusion concerning the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$ is then determined by weighting one's hypothetical conclusions about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, according to the probabilities previously assigned to the assumed values of $\mathrm{freq}(\mathrm{T}|\mathrm{R}_1)$. Similar applications of Method I may also be used in order to form a judgment about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, given frequency information concerning three, or more, nested reference classes.

There are, in fact, many possible variants of Method I that one may apply according to the particularities of one's available information (where reference class selection is driven by considerations of specificity). Indeed, as a step to forming a conclusion about a proposition, $\mathrm{c} \in \mathrm{T}$, by direct inference, Method I should be regarded as an instance of a general strategy for inferring the value of $E[\text{freq}(T|R')]$, where R' is the narrowest (non-gerrymandered) reference class containing c for which we have genuine frequency information (i.e., the narrowest reference class about which one is warranted in accepting $\text{freq}(T|R') \in v$, for some set of numerically expressed values, such that $v \subset \{0/|R'|, 1/|R'|, ..., |R'|/|R'|\}$). Having used Method I (or variants) to infer the value of $E[\text{freq}(T|R')]$, for the narrowest (non-gerrymandered) reference class, R', for which we have genuine frequency information, it is time for Method II to take over. Given the value of $E[\text{freq}(T|R')]$ (or a set of values), Method II may be used to infer the value of $E[\text{freq}(T|R^*)]$ (or a set of values), for all non-gerrymandered R^*, such that $\{c\} \subseteq R^* \subseteq R'$.

3.5 Are the resulting expectations rational?

In proposing the applicability of the methods of the preceding subsections, my guiding assumptions were: (1) that non-degenerate direct inferences provide defeasible reasons for accepting their conclusions, and (2) that one should prefer non-degenerate direct inferences based on more specific reference classes (and, similarly, non-degenerate direct inferences based on 'standard' reference classes versus their partitions), in the case where non-degenerate direct inferences yield conflicting conclusions. Notice that the applicability of (1) and (2) is limited to non-degenerate direct inferences. There are, I think, two sorts of degenerate direct inferences. I briefly discuss each sort, in turn.

One sort of degenerate direct inference involves gerrymandered reference or target classes. I am reasonably confident that it is possible to give formal criteria for identifying such direct inferences, but I will not attempt that task here (but see (Thorn 2012) which purports to provide such criteria). As an alternative, I describe two paradigmatic examples. The first type of example involves a reference class, R, that is formed by taking the union of two sets R_1 and R_2, where (i) R_1 is much larger than R_2, (ii) the frequency of members of the relevant target class, T, among R_1 is known to be very high (or low), (iii) R_2 is known to contain the object, c, about which we wish to form a judgment, and (iv) we are only warranted in making an imprecise-valued judgment about the frequency of members of T among R_2. For example, in the case of Bob of Company B, consider the reference class formed by the union of the set of NCOs in Company B, and the unit set containing Bob. The frequency of NCOs among this set is in $\{25/26, 25/25\}$. As with other reference classes that could be generated in accordance with the preceding recipe, it is clear that the described reference class should not be used in drawing a conclusion about the probability that Bob is an NCO.

Similar problems can be generated by forming target classes by disjunction. For example, suppose we know that more than 90 percent of Mondays are workdays. This implies that more than 90 percent of Mondays are either workdays or Aristotle's birthday. Now suppose that we know that next Monday is a holiday. In that case, it would be inappropriate to use the latter frequency statement to infer that next Monday is very probably a workday or Aristotle's birthday, and thus very probably Aristotle's birthday.

Another sort of degenerate direct inference involves the use of a reference class consisting of a partition of the set of all possible worlds. Such direct inferences correspond to applications of the traditional principle of indifference (cf. White 2009, 169-71). These direct inferences are rightly regarded as problematic since they inherit the defect characteristic of the traditional principle of indifference, namely: It is always possible to re-partition the set of possible worlds in order to form new reference classes, for new direct inferences, that yield conclusions that conflict with the ones licensed by the original partition, and there is no analogue of specificity conditions that can serve as a principled means of arbitrating between such direct inferences. Due to their unprincipled nature, it is sensible to suppose that direct inferences whose reference class consists of a partition of the set of possible worlds (or a partition of a similar infinite set) are *degenerate*. More precisely, I propose that we regard a direct inference as degenerate *if* it is formulated using a reference class that is a partition of an infinite set. Regarding this proposal, it should be acknowledged that it is often possible to 'emulate' a regular direct inference, whose reference class is not a partition of an infinite set (or even a partition of a set), by a series of direct inferences whose reference class consists of a set of mutually exclusive and jointly inclusive propositions (i.e., a partition of the set of all possible worlds).[19] However, there is no reason to regard the emulatability of a regular direct inference by direct inferences whose reference class is a partition of an infinite set as impugning the regular direct inference. Beyond this, it is apparent that the proposed degeneracy condition prohibits those direct inferences that it was intended to prohibit (i.e., those direct inferences that would be sufficient

[19]Consider a regular direct inference of the form: From $c \in R$ and $freq(T|R) = r$ infer that $PROB(c \in T) = r$, where $r = i/|R|$. To achieve such an emulation of this direct inference, proceed as follows: (i) Introduce a set of names $\{c_1, ..., c_i, c_{i+1}, ..., c_{|R|}\}$ for the elements of R, where c_1 through c_i denote elements of T, and c_{i+1} through $c_{|R|}$ do not. (ii) Form the reference class $R_\pi = \{c{=}c_1, ..., c{=}c_{|R|}\}$. (iii) Where V is the set of all true propositions, make direct inferences of the form: From $c{=}c_j \in R_\pi$ and $freq(V|R_\pi) = 1/|R|$ infer that $PROB(c{=}c_j \in V) = 1/|R|$, for each j in $\{1, ..., i\}$. (iv) Given the conclusions of the direct inferences made in step (iii), deduce that $PROB(c \in T) = i/|R|$.

to emulate unprincipled applications of the traditional principle of indifference), while not accosting regular direct inferences. So it appears that the proposed degeneracy condition provides an adequate means of disentangling a sensible fragment of direct inference from unprincipled applications of the traditional principle of indifference.[20]

It is relatively clear that the direct inferences underlying the proposed methods are not degenerate. Neither their reference nor target classes are gerrymandered, nor do their reference classes consist of partitions of an infinite set. This is not to say that the direct inferences underlying the proposed methods are not subject to defeat. Indeed, (2) acknowledges one sort of case where non-degenerate direct inferences are subject to defeat, and there are surely others. For example, I assume, as is usually assumed, that non-degenerate direct inferences to conflicting conclusions are mutually defeating, in cases where specificity considerations are not applicable in yielding a preference for one or the other of the two direct inferences. Nevertheless, (1) generates a presumption in favor of the direct inferences underlying the proposed methods: If there is no reason to think that these direct inferences are defeated, then they are not. One could, of course, object to (1) and (2). I will not take that objection too seriously. I have already provided some reason to accept the two theses, by appeal to optimality results. To the extent that those results are suggestive of the virtues of forming our personal probabilities by direct inference with a preference for narrower reference classes, they provide support for the application of the two proposed methods for inferring expected frequencies, in cases where the direct inferences underlying the methods employ the narrowest relevant reference class. As an alternative, we can fall back to the commonly held 'raw' intuition that something very like (1) and (2) must be correct.

Given (1) and (2), the proposed methods do not generally yield undefeated reasons for accepting respective expected frequencies. Indeed, since the methods are based on direct inferences, and direct inferences based on more specific reference classes are to be preferred, the direct inferences underlying the proposed methods will be defeated in *some* cases. A more significant worry is that the outputs of the proposed methods are always (or nearly always) defeated. This would be the case if there were 'worthy' competing methods that normally generated different expected frequencies. For the reasons that follow, I am optimistic that there are no such methods.

There are plenty of methods/algorithms that one could employ in order

[20] The present issue obviously deserves are more detailed and careful treatment than is given here. For reasons of space, I leave this task to another occasion.

to 'assign' a value to a respective expected frequency. But typical methods for assigning values do not yield reasons for accepting the respective value assignments. In other words, *genuine* competitors for the proposed methods would be backed by inference methods that confer reasons for accepting their conclusions. The set of possible genuine competitor methods can be divided into two classes: (i) ones that depend wholly on instances of direct inference (along with deduction), and (ii) ones that do not. It is plausible to think that there are 'part time' competitors of the latter sort, which apply in some cases. For example, it is plausible that one could form a judgment about the value of $\mathrm{E}[\mathrm{freq}(\mathrm{T}|\mathrm{R}')]$, for respective T and R', based on (some form of) enumerative induction from an observed sample, in some cases. However, one is not normally in a position to form such judgments. Beyond this, it is fair to say that there are no known inference methods, excluding ones grounded in direct inference, that generally (or nearly always) provide reasons for accepting such expected frequencies. The partition dependent methods proposed by objective Bayesians represent a possible exception to the preceding claim. But these methods are generally, and rightly, regarded with skepticism. For this reason, I assume that such methods do not provide reasons for accepting their outputs (at least not in cases where one is in a position to make a competing judgment based on empirical data). In any case, given the problem of partition selection involved in applying such methods, we ought to explore alternatives, such as the methods proposed here.

It is clear that it is possible to formulate methods that are grounded in direct inference that are competitors to the methods proposed here. It *appears* that all such competitor methods fall into one of *three* categories. First, there are competitor methods that depend on direct inferences based on broader reference classes (or on reference classes that consist of partitions of the reference classes upon which the proposed methods are based – recall the discussion of such cases that immediately followed Theorem 2). For this reason, the direct inferences underlying the methods proposed in the present paper should be preferred. Second, it is possible to formulate alternatives that employ gerrymandered reference classes. These direct inferences, unlike the ones proposed here, are degenerate. Third, there are direct inferences whose reference class consists of a partition of the set of possible worlds. Such direct inferences, unlike the ones proposed here, are also degenerate.

The preceding considerations count in favor of the wide spread (though not universal) applicability of the proposed methods. That said, the adduced considerations are not conclusive, since they fall short of demonstrating the non-existence of a genuine competitor that normally generates reasons for accepting different expected frequencies.

4 Conclusion

In the present article, I offered reasons in favor of the policy of forming personal probabilities by direct inference using the most specific applicable reference classes. The main considerations in favor of the policy were presented in Section 2, where it was shown that, among the set of principled policies that could be used in setting one's personal probabilities, the policy of reasoning by direct inference using the most specific applicable reference classes yields personal probabilities whose accuracy is optimal, according to all proper scoring rules, in all situations where all of the applicable frequency information is point-valued. In Section 3, methods were introduced that often permit one to infer point-valued expected frequencies for subsets of sets for which one has point-valued frequency information. These methods go some distance in extending the applicability of the kind of optimality results presented in Section 2. The methods of Section 3 also apply to the dilemma of choosing between direct inference based on relatively precise-valued frequency information for broad reference classes, and direct inference based on relatively imprecise-valued frequency information for more specific reference classes.

Acknowledgements

Work on this paper was supported by DFG Grant SCHU1566/9-1 as part of the priority program "New Frameworks of Rationality" (SPP 1516). For helpful comments on a presentation of this paper, I am thankful for an audience at EPSA 2015. For helpful discussions, I am thankful to Ludwig Fahrbach, Gerhard Schurz, and Ioannis Votsis. Finally, I am especially thankful two anonymous referees for Synthese who provided excellent comments and suggestions concerning an earlier draft of the paper.

5 Appendix

Theorem 1. $\forall$M,χ: if χ is principled in M, then $\forall$S:

(1) if S is a *proper* scoring rule, then $\forall\pi\in\Pi$:
$\Sigma_{x\in\pi}$ S(δ(x $\in$ T), ν(x $\in$ T)) $\geq$ $\Sigma_{x\in\pi}$ S(χ(x $\in$ T), ν(x $\in$ T)), and

(2) if S is a *strictly proper* scoring rule and $\chi \neq \delta$, then $\exists\pi\in\Pi$:

$\Sigma_{x \in \pi}$ S(δ(x $\in$ T), ν(x $\in$ T)) > $\Sigma_{x \in \pi}$ S(χ(x $\in$ T), ν(x $\in$ T)).

Proof. Part (1): Consider an arbitrary π in Π, and an arbitrary x_i in π. We have $\Sigma_{x \in \pi}$ S(δ(x $\in$ T), ν(x $\in$ T)) = $|\pi| \times$[S(δ($x_i \in$ T), 1)$\times\delta$($x_i\in$ T) + S(δ(x_i $\in$ T), 0)$\times$(1$-\delta$($x_i \in$ T)], and $\Sigma_{x \in \pi}$ S(χ(x $\in$ T), ν(x $\in$ T)) = $|\pi|\times$[S(χ(x_i $\in$ T), 1)$\times\delta$($x_i \in$ T) + S(χ($x_i \in$ T), 0)$\times$(1$-\delta$($x_i \in$ T)] (since δ and χ are principled). Since S is proper, we have for all x: S(δ(x $\in$ T), 1)$\times\delta$(x $\in$ T) + S(δ(x $\in$ T), 0)(1$-\delta$(x $\in$ T)) $\geq$ S(χ(x $\in$ T), 1)$\times\delta$(x $\in$ T) + S(χ(x $\in$ T), 0)(1$-\delta$(x $\in$ T)).□ Part (2): For some π, we have δ(x $\in$ T) $\neq$ χ(x $\in$ T), for all x in π (since δ and χ are principled and $\delta \neq \chi$). Consider such a π. For such a π, S(δ(x $\in$ T), 1)$\times\delta$(x $\in$ T) + S(δ(x $\in$ T), 0)$\times$(1$-\delta$(x $\in$ T)) > S(χ(x $\in$ T), 1)$\times\delta$(x $\in$ T) + S(χ(x $\in$ T), 0)$\times$(1$-\delta$(x $\in$ T)), for all x in π, since S is strictly proper.□

Theorem 6. $\forall$T,R,R$'$: if R$' \subseteq$ R and $\forall i$: PROB(freq(T|R$'$) = i/|R$'$|) = freq({S : freq(T|S) = i/|R$'$|}|{S : S $\subseteq$ R $\wedge$ |S| = |R$'$|}), then E[freq(T|R$'$)] = freq(T|R).

Proof. Let T, R, and R$'$ be arbitrary sets such that R$' \subseteq$ R. Note that, for all i, freq({S : freq(T|S) = i/|R$'$|}|{S : S $\subseteq$ R $\wedge$ |S| = |R$'$|}) = $\binom{g}{i}\times\binom{|R|-g}{|R'|-i}$ / $\binom{|R|}{|R'|}$, where g = freq(T|R)$\times$|R|. So, for all i, PROB(freq(T|R$'$) = i/|R$'$|) = $\binom{g}{i}\times\binom{|R|-g}{|R'|-i}$ / $\binom{|R|}{|R'|}$. So E[freq(T|R$'$)] = $\Sigma_{i\in\{0,\ldots,|R'|\}}$ i/|R$'$|$\times$PROB(freq(T|R$'$) = i/|R$'$|) = $\Sigma_{i\in\{0,\ldots,|R'|\}}$ i/|R$'$| $\times$ $\binom{g}{i}\times\binom{|R|-g}{|R'|-i}$ / $\binom{|R|}{|R'|}$ = 1/|R$'$| $\times$ 1/$\binom{|R|}{|R'|}$ $\times$ $\Sigma_{i\in\{0,\ldots,|R'|\}}$ $\binom{g}{i}\times\binom{|R|-g}{|R'|-i}\times\binom{i}{1}$ = 1/|R$'$| $\times$ 1/$\binom{|R|}{|R'|}$ $\times$ $\binom{g}{1}\times\binom{g+|R|-g-1}{|R'|-1}$ [by Vandermonde's Identity (cf. Gould 2010, 6.17)] = 1/|R$'$| $\times$ (|R$'$|!$\times$(|R|$-$|R$'$|)!)/|R|! $\times$ g $\times$ (|R|$-$1)!/(|R$'-$1|!$\times$(|R|$-$|R$'$|)!) = g/|R| = freq(T|R).□

References

F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge.* MIT Press, Cambridge, Massachusetts, 1990.

S. Bradley and K. Steele. Uncertainty, Learning, and the "Problem" of Dilation. *Erkenntnis*, 79(6):1287–1303, 2014.

G. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

R. Carnap. *Logical Foundations of Probability.* University of Chicago Press, Chicago, 1962.

B. de Finetti. *Theory of Probability (vol. 1).* John Wiley, New York, 1974.

K. Easwaran. Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection. *Philosophy of Science*, 80 (1):119–142, 2013.

H. Gould. *Combinatorial Identities, Volume 4: Table I: Intermediate Techniques for Summing Finite Series*. 2010.

H. Greaves and D. Wallace. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 114:607–632, 2006.

J. Joyce. A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65 (4):575–603, 1998.

H. Kyburg. *The Logical Foundations of Statistical Inference.* Reidel Publishing Company, Dordrecht, 1974.

H. Kyburg and C. Teng. *Uncertain Inference.* Cambridge University Press, Cambridge, 2001.

H. Leitgeb and R. Pettigrew. An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science*, 77 (2):201–235, 2010a.

H. Leitgeb and R. Pettigrew. An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. . *Philosophy of Science*, 77 (2):236–272, 2010b.

B. Levinstein. Leitgeb and Pettigrew on Accuracy and Updating. *Philosophy of Science*, 79 (3):413–424, 2012.

J. Pollock. *Nomic Probability and the Foundations of Induction.* Oxford University Press, Oxford, 1990.

H. Reichenbach. *A Theory of Probability.* Berkeley University Press, Berkeley, 1949.

R. Selten. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1:43–62, 1998.

M. Stone. Kyburg, Levi, and Petersen. *Philosophy of Science*, 54 (2):244–255, 1987.

P. Thorn. Two Problems of Direct Inference. *Erkenntnis*, 76:299–318, 2012.

P. Thorn. Defeasible Conditionalization. *Journal of Philosophical Logic*, 43: 283–302, 2014.

J. Venn. *The Logic of Chance.* Chelsea Publishing Company, New York, 1866.