Induction With and Without Natural Properties: a New Approach to the New Riddle of Induction

Paul D. Thorn & Gerhard Schurz

Philosophia (2024) 52:935–958
https://doi.org/10.1007/s11406-024-00768-8

Induction With and Without Natural Properties: a New Approach to the New Riddle of Induction

Paul D. Thorn[1] · Gerhard Schurz[1]

Received: 16 October 2023 / Revised: 18 June 2024 / Accepted: 25 July 2024 /
Published online: 22 August 2024

Abstract
Drawing on past work, we introduce a new approach to the New Riddle of Induction, showing that the inductive projection of gruesome properties is unreliable under particular ideal conditions that are sufficient for the reliable inductive projection of non-gruesome properties. As an auxiliary to our account, we introduce rules for resolving conflicts between background information and the conclusions of otherwise reliable inductive inferences. Our approach to the New Riddle of Induction is quite permissive in the range of properties it recognizes as suitable for inductive projection, allowing for the inductive projection of highly gerrymandered non-natural properties. However, as an addendum to our discussion of the New Riddle, we show that natural properties do form a more reliable basis for inductive projection in cases where one's sample is small.

Keywords Problem of induction · New riddle of induction · Nelson Goodman · Natural properties · Projectability

1 Introduction

In 1955, Nelson Goodman formulated so called 'gruesome' predicates in order to present a previously unknown problem concerning the cogency of inductive inference (Goodman, 1955). In his original example, Goodman invites us to imagine a situation where, prior to the present time *t*, we have gathered a large sample of emeralds and observed that they are all green. In this situation, it appears that we have grounds for thinking it likely that all or nearly all of the emeralds that we have yet to observe are

✉ Paul D. Thorn
thorn@phil-fak.uni-duesseldorf.de

Gerhard Schurz
schurz@phil-fak.uni-duesseldorf.de

[1] Institut Für Philosophie, Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany

also green, on the basis of a rudimentary principle of induction that tells us to expect that generalizations that hold for our observed sample will also hold for objects that are not in our sample. To illustrate a problem with this way of thinking (and the rudimentary principle of induction), Goodman introduced the term "grue" that applies to an object if and only if it is green and examined before *t* or blue and not examined before *t*. Within Goodman's example, all of the emeralds in our sample are green and examined before *t*. So application of the rudimentary principle of induction tells us to expect that all or nearly all of the emeralds not in our sample are grue, and thus blue, since these emeralds were not examined prior to *t*. Goodman's example thereby shows that the rudimentary principle of induction readily leads to inconsistent conclusions (i.e., that emeralds not in our sample are both green and blue). The Goodman problem, also known as the New Riddle of Induction, is the problem of identifying which of the generalizations that hold for our sample can reasonably be expected to hold for objects that are not in our sample.

Throughout the article, our concern will be with inductive inference in a narrow sense that is understood as including any inference that precedes from a principal premise stating that the frequency of a property ϕ within one's sample is some value r (or within a range of values R) to the conclusion that the frequency of ϕ among a larger domain is probably approximately r (or within R), with the possibility of auxiliary premises specifying additional conditions (e.g., that the sample was random) whose function is to establish the reasonableness of the accepting the conclusion on the basis of the principal premise. Following past usage, we will refer to the act of inductive inference (in the narrow sense) as "inductive projection". We will say that a predicate is *projectable* in a given situation if and only if the inductive projection of the predicate is reasonable in the given situation.

Typical approaches to the Goodman problem take the form of criteria aimed at blocking a range of intuitively undesirable inductive inferences, and invariably get things wrong in at least one of two possible ways: (1) by not blocking some inductive inferences that should be blocked, or (2) by blocking some inductive inferences that should not be blocked. Rather than identify general criteria aimed at blocking inductive inferences with gruesome predicates, we make the following observation: Under certain conditions that may be regarded as ideal ones under which to perform induction, induction with gruesome predicates is unreliable. Inasmuch as we reasonably take ourselves to be in conditions that sufficiently approximate ideal ones for performing induction, we may reasonably judge that inductive inferences with gruesome predicates are not reliable, and so (obviously) should not be made. In the end, showing that induction with gruesome predicates is unreliable is not sufficient to address all Goodman-type worries, since Goodmanesque problems may arise due to conflicts between our background information and the conclusions of otherwise reliable inductive inferences. In order to address such cases, we introduce rules for resolving conflicts between background information and the conclusions of otherwise reliable inductive inferences.

Our approach to Goodman's problem is relatively permissive regarding the range of predicates that may be employed in making cogent inductive inferences. As a postscript to our discussion of the New Riddle, we consider whether there is any reason to prefer some species of 'natural' predicate as a basis for induction. We answer in the affirmative, and explain why a certain kind of natural predicate tends to support more reliable inductive inferences in cases where the sample size is small.

2 Discussion of Other Approaches

There are quite a number of alternative approaches to the Goodman problem. Rather than present an exhaustive survey, we consider three past approaches that illustrate the kinds of difficulties that arise when attempting to address Goodman's problem. Despite their limitations, none of the approaches that we consider are without merit, and our approach synthesizes aspects of the three approaches that we consider.

2.1 Qualitative Predicates

This approach was proposed by Carnap (1947) and defended by Swinburne (1968), who was keen to point out that Goodman had been too quick in dismissing the proposal to restrict the set of projectable predicates to ones that are qualitative. Although proposed definitions vary, Swinburne defined a qualitative predicate as one whose applicability can be determined without knowing "its spatial or temporal relation to a particular object, event, time, or place" (Swinburne, 1968, 124). Although green is a qualitative predicate and grue is not, Swinburne's proposal gets things wrong in two critical ways: (1) by blocking some inductive inferences that should not be blocked, and (2) by not blocking some inductive inferences that should be blocked.

While inductive inferences with qualitative predicates may appear to be on firmer ground than inductive inferences with non-qualitative predicates, it's clear that many inductive inferences with non-qualitative predicates are cogent and of practical importance. For example, suppose one wants to make an inductive inference concerning what percentage of a group's members were born in France. It is possible to imagine situations where one will want to have such information, and it is reasonable to make an inductive inference based on a random sample to draw a conclusion about the frequency of a group's members that were born in France.

Swinburne's official proposal was that inductive inferences using qualitative predicates take priority over those using non-qualitative ones, in case of conflict, which appears to allow for cases of cogent inductive inferences with non-qualitative predicates. However, in the absence of further principles, Swinburne's official proposal collapses into the doctrine that only inductive inferences with qualitative predicates are cogent, since for any inductive inference formulated using a non-qualitative predicate there is an inference formulated using a gruesome predicate that yields a conflicting conclusion. Since Swinburne's approach offers no means for prioritizing any of the inductive inferences with non-qualitative predicates over others, Swinburne's approach winds up blocking all inductive inferences with non-qualitative predicates, and thereby some inductive inferences that should not be blocked.

Swinburne's approach also fails to block some inductive inferences that should be blocked. To see why this is, notice that in many circumstances it is possible to formulate non-projectable qualitative predicates. We describe a schema for generating such predicates as follows. First, observe that qualitative predicates are closed under disjunction: if x is ϕ is a qualitative predicate and x is ψ is a qualitative predicate, then x is $\phi \vee x$ is ψ is also a qualitative predicate. Next notice that it will frequently be possible to form qualitative predicates of the following form: x is $\chi_1 \vee \ldots \vee x$ is $\chi_n \vee x$ is γ, where one's sample has n elements and for all i, x is χ_i is the most specific qualitative description we are able to provide of the i^{th} member of the sample, and x is γ is an *arbitrary* qualitative predicate.[1] Such predicates will permit the formulation of an inductive inference to the conclusion that all (or almost all) of the objects not in our sample are γ (where γ is arbitrary), so long as the available qualitative descriptions for the sample members are so specific that none (or few) of the objects outside the sample satisfy any of the descriptions. This will not always be the case, but it often will, as for example when one is reasoning about human beings (since no two are precise qualitative duplicates) and similarly for moderate-sized physical objects (since no two objects will have the exact same mass, etc.).

2.2 Counterfactual Independence

The basic idea behind this approach is that a property ϕ is projectable for a given sample only if each ϕ in one's sample would still have been ϕ if it had not been in one's sample. The proposed projectability condition is known as a counterfactual independence condition, since it requires that an object's being ϕ is independent of its being in the sample (i.e., the objects in one's sample that are ϕ would still have been ϕ in the counterfactual situation in which they were not in the sample). The approach was first proposed by Jackson (1975), but more sophisticated variants of the approach have gradually been formulated, in order to address problems raised by critics (Godfrey-Smith, 2003, 2011; Jackson & Pargetter, 1980; Warren, 2023).

We are sympathetic to the idea that the satisfaction of such a condition is necessary for the projectability of a predicate, but it is clear that no such condition will be sufficient to block all intuitively undesirable inductive inferences. To see that this is so, it is sufficient to notice that it is generally possible to formulate a non-projectable predicate where an object's satisfaction of the predicate is counterfactually independent of its membership (or not) in our sample. The highly gerrymandered qualitative predicates that we introduced in the previous subsection will serve this purpose. Moreover, since we need not limit ourselves to qualitative predicates in the formulation of the predicates that are now under consideration, it looks like it will always be possible to formulate a non-projectable predicate that slips by any proposed counterfactual independence condition.[2]

2.3 Derivative Defeat

In a recent account, Freitag introduced a notion of 'derivative defeat' as a means to explaining why the projection of grueness (but not greenness) is defeated in

[1] Similar examples originate with Scheffler (1963).

[2] Warren's (2023) approach to the New Riddle combines a counterfactual independence condition with a 'derivative defeat' condition of the sort we consider in the next subsection, and is supposed to avoid the present problem thanks to the derivative defeat condition. We explain later why Warren's derivative defeat condition gives the wrong results in some cases.

Goodman's example (Freitag, 2016). According to Freitag, "the projection of an (inductively confirmed) hypothesis is derivatively defeated if and only if the pertinent inductive evidence doxastically depends on the inductive evidence for the projection of a discriminating predicate", where a discriminating predicate is one where the agent knows that the predicate applies to all members of the sample, and to no objects outside of the sample (Freitag, 2016, 8). Expressed in more colloquial terms, Freitag's proposal is roughly that induction with a predicate ϕ is defeated if the evidence that the objects in our sample are ϕ depends on the evidence that the objects in our sample are ψ, where we know that an object is ψ if and only if it is a member of our sample. Freitag's notion of derivative defeat applies straightforwardly to Goodman's example, since, within the example, (1) the proposition that *all emeralds in the sample are grue* doxastically depends on the proposition that *all emeralds in the sample were examined before t*, and (2) the predicate *examined before t* is discriminating (i.e., is known to apply to all and only members of our sample).

Freitag's approach seems to correctly identify a problem with inductive inferences with predicates such as grue, since the trouble apparently derives from trying to project a disjunction of a discriminating predicate (green and observed before t) and another arbitrary predicate. In fact, the problem identified by Freitag is closely related to a general problem that was identified by Pollock, which involves inference to an arbitrary proposition T from the disjunction of T with another proposition R, in a case where one has both a defeasible reason to believe R and a defeasible reason to believe not R (Pollock, 1995, 116). Pollock called instances of this kind of reasoning "self-defeating", because some steps in such chains of reasoning conflict with others. In the end, we think that a principle such as Freitag's is needed in addressing Goodman-type cases that arise due to conflicts between our background information and otherwise reliable inductive inferences. Nevertheless, Freitag's approach readily yields the unwelcome results when applied to variations of Goodman's example.

Consider an agent, Tamara, who interacts with her environment using a sensor that detects whether or not objects are grue.[3] Tamara knows what green and blue objects look like, but at present the only 'color' discriminations she is able to make are between grue and non-grue objects. Suppose that Tamara knows that an object is grue just in case it is green and examined before *t* or blue and not examined before *t*, but she does not know the date, and so she does not know whether *t* is in the past or in the future. Under these conditions, Tamara draws a sample of emeralds and 'sees' that they are all grue. After she has examined the emeralds in her sample, Tamara receives the information that the time is *t* and so correctly infers that all of the emeralds in the sample were examined before *t*. Tamara then correctly deduces that all of the emeralds in her sample are green. Tamara finally considers whether to project grueness or greenness to the emeralds that are not in her sample.

Within the preceding example, Tamara's belief that the emeralds in her sample are green is dependent on the evidence that they were examined before *t*. So her proposed projection of greenness to the emeralds that are not in her sample is subject to

[3] We assume that the sensor integrates information about time and the color of objects in order to indicate with 100% reliability which objects are grue and which are not.

derivative defeat, according to Freitag's criterion. On the other hand, Tamara's judgment that the emeralds in her sample are grue is not doxastically dependent on the evidence that they were examined before *t* (or any other discriminating predicate). So her proposed projection of grueness to the emeralds that are not in her sample is not subject to derivative defeat, according to Freitag's criterion.

Despite the recommendations of Freitag's approach, it appears that Tamara should project greenness, and not grueness, to the emeralds that are not in her sample. Indeed, although the means by which Tamara came to know that the objects in her sample are green is indirect, her grounds for projecting greenness is no less cogent than the grounds of an agent who came to know that the emeralds in her sample are green in the normal way.

Despite her unusual situation, Tamara's concepts and innate perceptual capacities are similar to our own. If we consider an agent whose concepts and innate perceptual capacities are quite different than ours, it is evident that Freitag's approach is unsatisfactory as a general solution to the Goodman problem. Indeed, consider an agent whose innate perceptual capacities present objects as grue and bleen, and for whom grue and bleen are basic conceptual categories, where greenness is only derivatively defined as grue and observed before t or not bleen and not observed before t.[4] For such an agent, Freitag's account tells us that, upon observing a sample of grue emeralds prior to t, the projection of greenness rather than grueness is subject to derivative defeat. So, in the end, Freitag's account leads us to a form of relativism about which properties it is correct to project (cf. Dorst, 2018).[5]

3 Ideal Conditions for Induction

It is difficult to specify the problematic properties that give rise to the Goodman problem via a precise definition that blocks all of the inductive inferences that should be blocked, without also blocking some inductive inferences that should not be blocked. However, it is possible to show that certain sorts of properties (namely, green-like ones) support reliable inductive inferences under ideal conditions, while other properties (namely, grue-like ones) do not. This is the cornerstone of the approach taken here, namely, to investigate which properties support reliable inductive inference under the sorts of conditions that are ideal ones in which to perform induction. Roughly stated, the ideal conditions that we consider are ones in which (i) sampling is *random*, and (ii) the properties of objects do not vary according to which objects are in the sample, save those differences implied by membership or non-membership in the sample. We refer to condition (ii) as "domain stability under sampling", or "domain stability", for short.

[4] An object is bleen if and only if it is blue and examined before t or green and not examined before t.

[5] The derivative defeat condition proposed by Warren (2023) yields the same conclusions as Freitag's in this example and in the case of Tamara, and so also leads to what we regard as an undesirable form of relativism.

Where the size of our sample is n, we use the expression σ_n to denote the set of n-membered subsets of the relevant domain D. Condition (i) requires that each element of σ_n has the same probability of being sampled, i.e., our sampling procedure is an instance of a random experiment of drawing an n-membered sample from D, and each element of σ_n has the same probability of being the sample, namely, $1/|\sigma_n|$.

In specifying ideal conditions under which to perform induction, domain stability is needed in addition to random sampling, since for any predicate, ϕ, whose satisfaction by objects is contingent, random sampling is insufficient to support reliable inductive inferences concerning ϕ, due to the possibility that which objects are ϕ depends on which objects are in the sample. To see why this is so, first notice that, for some properties, which objects have the property always depends on which objects are in the sample. For example, which objects have the property 'being a member of the sample' always depends on which objects end up in the sample, and of course induction using the predicate 'x is a member of the sample' is obviously unreliable. But similar issues can arise for almost any property. For example, the inductive projection of greenness will be unreliable in situations where being green is fully dependent upon being a member of the sample, under conditions where any object that is not subject to the causal influence of sampling is non-green.

In order to provide a more precise statement of what domain stability requires, we use the terms $c_1, \ldots, c_k$ as rigid designators for the elements of D, and the terms $s, s_1, \ldots, s_m$ as rigid designators for the elements of σ_n. We use the expression S as a definite description denoting, in each possible world, the set of objects that are in *the* sample in that world, and we use terms of the form w^{si} to refer to the counterfactual variant of the world w in which $S = s_i$. Finally, in order to express which facts must remain stable regardless of which sample is selected, we use a notion of intrinsicness that was explicated by Marshall and Weatherson (2018):

> **Intrinsicness:** F is intrinsic if and only if, necessarily, for all x, if x is F, then x is F is a matter of how x and its parts are and how they are related to each other, as opposed to how x and its parts are related to other things and how other things are.[6]

Given these preliminaries, we define domain stability as follows:

> **Domain stability** holds for S in a world w if and only if there is a situation (or an incomplete possible world) w^{-s} such that
>
> (1) w^{-s} agrees with each w^{si} regarding the intrinsic properties of all objects, and

[6] The present definition of intrinsicness is one of many notions of intrinsicness that Marshall and Weatherson identify and explicate. The present notion is most suitable for our purposes, but we do not here advocate a view concerning the definition's primacy as an explication of intrinsicness.

(2) for all w^{si}, the set of statements that are true in w^{si} is the deductive closure of { α | α is true w^{-s}} ∪ { α | α ≡ (c_j is in S) and $c_j \in s_i$} ∪ { α | α ≡ ¬(c_j is in S) and $c_j \notin s_i$}.[7]

In understanding this definition, it may be helpful to think of the situation w^{-s} as if it were a complete possible world, with the worlds w^{s1}, …, w^{sm} being variants of w^{-s} that differ only through the addition of exogenous labels specifying which objects are in S and which are not. In addition to demanding conformity among all of the w^{si} concerning the intrinsic properties of objects (in condition (1)), domain stability demands conformity concerning all further properties and relations that can be fixed independently of fixing which objects are in S (in condition (2)).

In the case of actual agents drawing samples in order to make inductive inferences about the actual world, domain stability will probably never be satisfied, since any procedure that actual agents use for choosing a sample will always bring about some changes in the intrinsic properties of some objects. Despite this practical limitation, domain stability, as stated, is suggestive of the ideal towards which we would like to proceed when we take a sample as a basis for making an inductive inference (especially if we are open to the inductive projection of virtually any property). Namely, we want to avoid circumstances where our sampling procedure induces changes in those properties about which we are open to making an inductive inference.

Domain stability is obviously a sort of counterfactual independence condition with a focus on certain properties (namely, the intrinsic ones). Our approach thereby synthesizes elements of Jackson's approach with Swinburne's proposal to assign priority to inductive inferences concerning qualitative properties (a close relative of intrinsic properties). We acknowledge that what we call "domain stability" is essentially domain stability *with respect to intrinsic properties*, and that one could achieve conclusions similar to the ones that we reach concerning intrinsic properties by substituting another sort of properties for the intrinsic ones in the definition of domain stability.[8] We are open to considering other forms of domain stability, but focus on intrinsic properties, since (1) it is typically reasonable to expect domain stability to hold approximately for intrinsic properties, and (2) many non-intrinsic properties supervene on the intrinsic ones with the result that stability with respect to intrinsic properties ensures stability with respect to many non-intrinsic properties.

In cases where domain stability holds for a sample in a world w, it will be important to notice later that facts about which objects are green in a given world

[7] A situation is here understood to be an incomplete possible world. Unlike possible worlds wherein the set of propositions true in a world is both consistent and maximal, the set of propositions that is true in a situation is consistent but not maximal. We take no further stand on what kind of entity a situation is. We thereby allow for possible situations where an object is grue and yet there is no fact of the matter about whether the object was observed before time *t*, and so no fact of the matter about whether the object is green and no fact of the matter about whether the object is blue. Condition (1) of domain stability ensures that w^{-s} is not a situation of this kind. Indeed, since being green is an intrinsic property, condition (1) demands that if a given object is green in some w^{si}, then it is green in w^{-s} and in each of the other w^{si}s.

[8] One alternative would be to consider domain stability with respect to the set of properties, where is the largest set of properties such that domain stability holds with respect to those properties.

w are fixed by w^{-s}, whereas facts about which objects are grue in w are frequently indeterminate in w^{-s}. The preceding holds since being green is an intrinsic property, as is being blue. So for each object in D, w^{-s} specifies whether the object is green, blue, or neither. On the other hand, w^{-s} leaves the question open concerning which of the green and blue objects in D are grue, since w^{-s} leaves the question open concerning which objects in D are in S.

In the following section, we show that induction with non-Goodman-type predicates is reliable under conditions of random sampling and domain stability, while induction with Goodman-type predicates is not. Our position in addressing the Goodman problem is, then, that we frequently find ourselves in conditions that sufficiently approximate random sampling and domain stability, and so we frequently have grounds for making inductive inferences with non-Goodman-type predicates, while rejecting inductive inferences with Goodman-type predicates. Before proceeding, we consider some preliminary objections to the assumptions we have made so far.

To begin with, it may be objected that our use of the concept of an intrinsic property is problematic, since, despite ongoing efforts, no widely accepted analysis of this concept has been given. In fact, we don't think that the ongoing controversies present a problem for our proposal, because cogent disagreements over the concept intrinsicness are exclusively at the margins (concerning the precise analysis of the concept and concerning the application of the concept in usual cases), in the context of broad agreement about which properties are intrinsic. To this we would add that our interest in the concept of intrinsicness is largely a matter of its suitability in playing a particular 'functional role' within our approach to the Goodman problem. In particular, stability is demanded with respect to 'intrinsic properties' with the consequence (as we will see) that induction with predicates that denote 'intrinsic properties' is reliable. Since our concern with intrinsicness is limited to the fulfilment of this functional role (moving from stability regarding x to reliability with x) we are happy to adopt a flexible stance toward disagreements concerning the precise analysis and nature of intrinsicness.

Assuming categorical objections to our use of the concept of intrinsicness are waived, further worries are certainly possible. Such worries relate to two propositions that are critical to our approach: (1) that we are able to reliably distinguish between intrinsic and extrinsic properties, and (2) that we are often in a situation wherein it is possible to draw a sample in such a way that we may justifiably believe that domain stability and random sampling hold for the sample (at least approximately). It must be acknowledged that both of these propositions are dubitable. But indubitability is surely the wrong standard to adopt in the present context, since our aim is not to refute skepticism or even skepticism with regard to induction (Hume's problem). While the two propositions are obviously subject to skeptical worries, they are supported by common sense and contemporary science, and this is sufficient for our goal, which is to explain why some inductive inferences are reasonable and others are not. This being our principle response to possible objections, let us briefly consider specific worries that may be raised concerning the two aforementioned propositions.

Consider the claim that we are able to reliably distinguish between intrinsic and extrinsic properties. To begin with, it appears that most of the properties that John Locke called “primary qualities” are intrinsic as a matter of conceptual necessity, and so having the proper conception of these properties is sufficient for understanding that they are intrinsic. For example, if one understands what it is for an object to be cube-shaped, then one sees for any cube-shaped object that its being cube-shaped is a matter of how the object and its parts are and how they are related to each other, and not a matter of how the object and its parts are related to other things and how other things are. The situation is somewhat different with those properties that Locke called “secondary qualities”. Take, for example, the property of being green. Our common sense conception of greenness appears to include the idea that an object’s being green is a matter of how the object itself is, and not a matter of how the object is related to other things. However, there is still controversy concerning which property, if any, our common sense conception of greenness denotes. Faced with this problem, we propose to treat a predicate as denoting an intrinsic property if there is a possible relatively felicitous precising definition of the predicate such that the predicate so defined denotes an intrinsic property. Contemporary science tells us that there are possible definitions of this sort for color predicates such as ‘*x* is green’. Schematically such a definition would take the following form: x is green $=_{\text{dfn}}$ x has a physical surface of sort ϕ (where having a physical surface of sort ϕ is intrinsic and it is understood that objects in the actual world reflect and absorb light in characteristic ways as a consequence of having a physical surface of sort ϕ).

Now consider the claim that we are often in a situation wherein it is possible to draw a sample is such a way that we may justifiably believe that domain stability and random sampling hold for the sample (at least approximately). This claim is justified according to common sense and the picture of the world provided by contemporary science. The picture given by science and common sense tells us that we can select objects in a manner the approximates randomness using urns filled with numbered balls, for example. Similarly, science and common sense tell us that which intrinsic properties an object has is typically not dependent on our sampling procedures. It’s clear that the picture of the world given by science and common sense is dependent on some form or forms of non-deductive inference, and so the picture of the world given by science and common sense is potentially subject to skepticism about induction. But our goal here is not to address the problem of induction. So those worries are not a present concern.

In spite of the preceding, it might be objected that an agent with a suitably formulated ‘grue friendly’ conceptual scheme could make parallel arguments for the agent’s favored gruesome predicates, maintaining that the properties denoted by their favored gruesome predicates are actually the ones that are intrinsic. Our main point in addressing this worry is to observe that the notion intrinsicness appealed to in our account is non-relative in the following sense: Whether a property is intrinsic is not relative to any factor that could vary among different agents. In particular, the concept of intrinsicness includes no explicit or implicit indexical element that would allow for the possibility of one agent truly believing and another truly disbelieving that a given property is intrinsic. Rather if two agents differ in their beliefs about whether a given property is intrinsic, then those beliefs are in conflict and one of the agents is mistaken. So if an agent disagrees with our judgment that green is intrinsic, then the disagreement concerns a factual matter. Regarding such factual disagreements, we repeat our previous point: The claim the green is intrinsic is a matter of common sense and established science. The claim is perhaps dubitable. But since our goal here is not to address skepticism or even skepticism about induction, the dubitability of the claim is not a worry for us.

4 Induction Under Ideal Conditions

The justification of induction, in the case of random sampling, is plausibly pursued by appeal to a combinatorial principle about the frequent similarity of sets with their subsets. The combinatorial principle says: For any finite sets D and G, the vast majority of subsets of D of sufficient size have approximately the same frequency of Gs as D, regardless of what the frequency of G in D is. So if one randomly selects a sufficiently large set S (namely, a sufficiently large sample) of a given finite domain D, then it appears reasonable to expect that the frequency of Gs among S will be approximately the same as the frequency of Gs among D. Of course, not all instances of the preceding way of thinking are reasonable. For example, if G denotes *the set of objects in one's sample*, then one's sample will not agree with D on the frequency of G (assuming one's sample is much smaller than D). It is also apparent that domain stability plays a role in distinguishing which properties can be reliably projected in the case of random sampling. Indeed, the frequency of green objects in one's sample need not be a good indicator of the frequency of green objects in the full domain in cases where facts about which sample is selected influence which objects are green.

In order to get a concrete idea about how reliable induction is in the case of random sampling with domain stability, consider a case where we take a random 100 member sample from a 10,000 member domain and project the frequency of a given property G. It can be shown that induction is reliable in this case regardless of the frequency of G in the domain as a whole, which is a critical point, since our concern is typically with the case where we do not know what the frequency of G in the domain is, and we demand assurance that induction concerning G is reliable regardless of the frequency of G in the domain. For this reason, we consider all possible cases regarding the frequency of G in D, and show that the frequency of G in our sample tends to closely approximate the frequency of G in D, in each case. Because it is possible to compute the distribution of the frequency of G among the 100 member subsets of D, given the number of Gs in D, it would be straightforward (though extremely tedious) to compute the probability of various claims concerning the difference between freq(G|S) and freq(G|D) (i.e., the difference between the frequency of G in S and the frequency of G in D), assuming various values of freq(G|D). Rather than present such calculations, we present the results of computer simulations that estimate the mean difference between freq(G|S) and freq(G|D) for random samples, according to the possible values of freq(G|D).

Figure 1 presents the average difference between the frequency of G in a domain D and in a sample S, with the frequency of G in D as the independent variable. The values presented in Fig. 1 were estimated via 1,000,000 randomly generated cases for each possible frequency of G in D. In each case, the set G and then the set S were selected independently and at random according to a uniform probability distribution. Within the simulations presented in Fig. 1, it is implied by the independent selection of G and S that the extension of G does not change according to which set S is selected. This assumption corresponds to domain stability: Domain stability permits us to infer that the extension of many predicates is invariant according to which sample is selected, and the method for selecting G and S ensures that G is one of these predicates.

Figure 1 illustrates that the inductive projection of G in these circumstances, is absolutely reliable in the case were freq(G|D) = 0 or freq(G|D) = 1, which is obvious since freq(G|D) = 0 implies that it is impossible to select a sample that contains any Gs, and similarly freq(G|D) = 1 implies that it is impossible to select a sample that contains any non-Gs. The projection of G is least accurate in the case where freq(G|D) = 0.5, with the difference between freq(G|D) and freq(G|S) averaging just less than 0.04 in that case. Regardless of the value of freq(G|D), the average difference between freq(GD) and freq(G|S) can be made as close to zero as one likes by increasing the size of the sample S, regardless of the size of D.

The data presented in Fig. 1 shows that the inductive projection of a property, G, based on a moderate-sized sample is reliable, regardless of the frequency of G in the population, assuming random sampling and assuming that the extension of G does not change according to which set S is selected. Of course, inductive projection is not reliable for all properties, under the ideal conditions considered here. The paradigmatic example is the property of being a member of the sample, which is obviously unreliable in any circumstance where the domain is much larger than one's sample. Inductive projection is also unreliable for grueness, under the ideal conditions considered here. We make this point, by considering some further data concerning the simulations presented in Fig. 1.

Suppose that an object is G* if and only if it is (G and S) or (not G and not S). Figure 2 shows that the inductive projection of G* is not reliable, for many values of freq(G|D). For example, in the case where the frequency of G in the population is 0 or 1, the difference between freq(G*|D) and freq(G*|S) is the difference between the size of D and the size of S divided by the size of D (i.e., (|D| − |S|)/|D|), which is 0.99 in the cases presented in Fig. 2, and tends to 1 for increasing |D|, if |S| is constant. The inductive projection of G* is reliable in the case where freq(G|D) = 0.5 (or is near 0.5). In this case, freq(G|S) tends to be close to 0.5 (as shown in Fig. 1), and so both freq(G*|S) and freq(G*|D) also tend to be close to 0.5 by implication from freq(G|S) ≈ 0.5 and freq(G|D) ≈ 0.5 and the definition of G*.

The fact that induction with G is generally reliable, and induction with G* is not, derives from the fact that the extension of G, unlike the extension of G*, is independent of which members of the domain came to be members of S. The fact that the extension of G is independent of the selection of S, and the extension of G* is dependent on the selection of S, is a consequence of choices we made in programming our simulations. Of course, we could easily have set things up so that the extension of G* is independent of the selection of S, and the extension of G is not. Indeed, since $G = (G^* \cap S) \cup (\neg G^* \cap \neg S)$, we could have selected G* (rather than G)

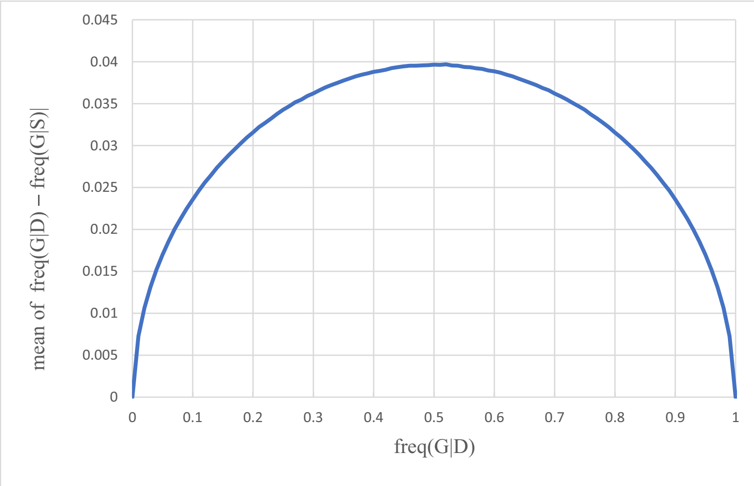

Fig. 1 The average difference in the frequency of G in the domain D (of size 10,000) and the sample S (of size 100, randomly chosen), depending on the frequency of G in D

at random and independently of S. In that case, the inductive projection of G* would generally be reliable, and the inductive projection of G would not.[9] However, such a 'reversal' is impossible for the properties green and grue* (where grue* = (green ∩ S) ∪ (¬green ∩ ¬S)), in the case where domain stability holds, since greenness is an intrinsic property of objects. So, assuming domain stability holds, which objects are green does not change according to which objects are in S.

It might be suggested that a fair comparison of induction with G versus G* would treat freq(G*|D), rather than freq(G|D), as the independent variable in assessing the reliability of induction with G*. However, it would only make sense to treat freq(G*|D) as an independent variable, corresponding to the dependent variable |freq(G*|D) – freq(G*|S)|, in a case where the members of G* are selected independently of the members of S (otherwise the selection of S would affect the value of freq(G*|D)). However, it is not possible to select the elements of G* independently of the elements of S in the case where G is an intrinsic property and domain stability holds.

It may be observed that Goodman's original grue predicate would be better formulated as: G′ = (G and S) or (B and not S), where G ∩ B = ∅. A variant of Fig. 2 for G′ would have an independent variable for freq(B|D), in addition to one for freq(G|D), and a satisfactory condition for the reliability of projecting G′ would be that the mean difference between freq(G′|D) and freq(G′|S) be small regardless of the values of freq(G|D) and freq(B|D). It is clear from the data presented in Fig. 2 that this condition cannot be met, since Fig. 2 implicitly represents the mean difference between freq(G′|D) and freq(G′|S), for the presented values of freq(G|D), in the

[9] In this case, it would be appropriate to treat freq(G*|D) as the independent variable in assessing the general reliability of inductively projecting G and G*.

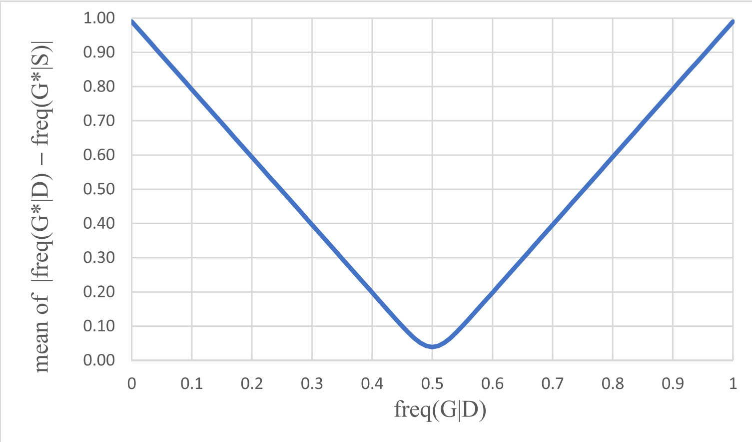

Fig. 2 The average difference in the frequency of G* (grueness) in the domain D and the sample S, depending on the frequency of G (greenness) in D

case where freq(B|D) = freq(¬G|D) (namely, the case where all the non-Gs are Bs). So the data presented in Fig. 2 shows that induction with G′ is not generally reliable.

There are, of course, many other Goodman-type predicates, and the results of our simulations don't directly apply to all of them. Nevertheless, the simulation results presented here are representative of our approach, which is in the first instance to rule out undesirable inductive inferences by showing that they are not reliable, under conditions of random sampling and domain stability. The simplest recipe for generating problematic predicates is to take the disjunction of any predicate '*x* is ϕ' and the predicate '*x* is a member of the sample':

x is ϕ or *x* is a member of the sample

With the help of such compound predicates, we can formulate inductive inferences for any conclusion whatsoever concerning the characteristics of the objects that are not in our sample. It is straightforward to show that inductive inferences with such predicates are typically not reliable, under conditions of random sampling and domain stability. Indeed, it is clear that induction with the predicate '*x* is ϕ or *x* is a member of the sample' is only reliable to the degree that |ϕ| approximates the size of the domain. If ϕ is empty, then regardless of which objects end up in our sample, induction with the predicate '*x* is ϕ or *x* is a member of the sample' will tell us that all objects not in our sample are ϕ, although none are.

The more direct point of the simulations presented in this section is to illustrate that induction with predicates denoting intrinsic properties is reliable under conditions of random sampling and domain stability. More precisely, we claim that the following reliability principle holds for intrinsic properties:

For all predicates ϕ, if ϕ denotes an intrinsic property, then it is highly probable that the frequency of ϕ in one's sample will be approximately the same as the frequency of ϕ in the domain, given all and only the conditions that obtained prior to the determination of which objects are in the sample, assuming random sampling, domain stability, and that one's sample is sufficiently large.

The principal ground for this claim is the already stated fact that, under the conditions of random sampling and domain stability, induction with a relevant ϕ is reliable no matter what the frequency of ϕ is in the domain. Our conclusion, then, is that given the extent to which random sampling and domain stability fix the effects of sampling, no further facts concerning the situation prior to sampling can undermine the claim that it is highly probable that the frequency of ϕ in our sample will be approximately the same as the frequency of ϕ in the domain.

In what follows, when we say that induction with a predicate ϕ is reliable, we mean that it is highly probable that the frequency of ϕ in one's sample will be approximately the same as the frequency of ϕ in the domain, given all and only the conditions that obtained prior to the determination of which objects are in the sample. While domain stability and random sampling are sufficient for the reliability of induction with any predicate that denotes an intrinsic property, denoting an intrinsic property is not a necessary condition for the reliability of induction with ϕ. To see that intrinsicness is not necessary, notice that the reliability of the inductive projection of G, as illustrated in Fig. 1, is not a consequence of G's intrinsicness. G is simply a set of objects that is selected at random and independently of the set of objects in the respective sample S. In fact, the results presented in Fig. 1 do not depend on the assumption that the elements of G were selected at random. Indeed, we could have selected the elements of G by any means that we like (e.g., with the goal of maximizing heterogeneity) and would still obtain the same results, so long as the elements of S were selected independently of the elements of G.

5 Managing Conflicts with Background Knowledge

In the preceding section, we argued that the intrinsicness of a property, ϕ, is a *sufficient* condition for the reliability of induction with the corresponding predicate '*x* is ϕ', under conditions of random sampling and domain stability. However, reliable inferential processes can still lead to conclusions that are inconsistent with one's background information. For example, suppose one takes a large random sample of objects from a domain, in a case where domain stability holds and one knows that 80% of objects in a given domain are green. In this case, one can reasonably expect that about 80% of the objects in one's sample will be green, but it is obviously possible (however unlikely) that one draws a large sample in which only 20% of the objects are green. In such a case, inductive inference to the conclusion that 20% of the objects in the domain are green is obviously unreasonable. Similar situations may also arise in cases where we do not know the frequency of a given intrinsic property ϕ in a domain, but we are *justified* in assigning prior probabilities to the possible frequencies of ϕ in the domain, and we know the likelihood of the observed frequency of ϕ in our sample conditional on the different possible frequencies of ϕ in the domain. In such cases, we are in a position to use Bayesian conditonalization to form justified conclusions about the posterior probabilities of different hypotheses concerning the frequency of ϕ in the domain. In case of conflict, we maintain that justified conclusions reached by Bayesian conditionalization override conclusions reached by inductive inference in the narrow sense (and this result is built into the projectability condition **[PRO]** that we introduce below, which deems any predicate that supports induction to a conclusion that conflicts with propositions we are justified in accepting to be non-projectable).

By exploiting inconsistencies with background information, it is also possible to formulate problematic predicates that generate problems analogous to the Goodman problem. For example, in the situation described in the preceding paragraph (where we know that 80% of objects in a given domain are green but only 20% of the objects in our sample are green), induction with predicates of the form '*x* is non-green or ϕ' is obviously unreasonable, where ϕ denotes an arbitrary intrinsic property. Similarly, in any case where we are in a position to identify an intrinsic property that characterizes all and only members of our sample, it is straightforward to define problematic predicates that denote intrinsic properties. This will be the case, for example, if every object in the domain has a unique mass, and we are able to measure the mass of each object in our sample. In that case, if $m_1, \ldots, m_n$ are the masses of the objects in our sample, M is the property of having mass m_1 or…or mass m_n, and ϕ is an intrinsic property, then $\phi \vee M$ is an intrinsic property. Although induction with '*x* is ϕ or *x* is M' is correctly regarded as reliable prior to the determination of which objects are in the sample (assuming random sampling and domain stability), induction with the predicate is obviously unreasonable in any case where M characterizes membership in our sample. Moreover, it's clear that in any case where we are in a position to formulate such predicates (because we are in a position to identify an intrinsic property that characterizes sample membership) we will be able to formulate inductive inferences for any conclusion whatsoever concerning the properties of objects that are not in our sample.

In order to address the problem of conflicts between background information and the conclusions of inductive inference with reliable predicates, we offer an approach that satisfies 4 desiderata:

1. **Reliability**: A predicate ϕ is projectable in a given situation only if given all and only the conditions that obtained prior to the determination of which objects are in the sample it is probable that the frequency of ϕ in the sample is approximately the same as the frequency of ϕ in the domain.
2. **Consistency**: The set of predicates Φ that is projectable in a given situation is such that the set of conclusions supported by induction from the elements of Φ is consistent with our background information and with other propositions we are justified in accepting.
3. **Non-arbitrary Resolution:** In the case where the conclusions of one or more inductive inferences are inconsistent with each other or with our background information and other propositions we are justified in accepting, the inconsistency is resolved in favor of one conclusion over another only if there are cogent grounds favoring one of the conclusions over the other.

4. **Avoid Unnecessary Exclusion**: Outside of inductive inferences that are deemed unreasonable as a consequence of avoiding inconsistency in a non-arbitrary manner (conditions 2. and 3.), it is reasonable to accept the conclusions of reliable inductive inferences.

In order to satisfy the 4 desiderata, we propose two preliminary projectability conditions and a final definition of projectability that insures that induction with projectable predicates does not lead to inconsistency. Before proceeding, observe that condition 4 (Avoid Unnecessary Exclusion) is critical, since absent this condition, we could easily satisfy the other three conditions by advocating skepticism about induction, i.e., that the set of predicates that is projectable in any given situation is, invariably, empty, thereby guaranteeing reliability, consistency, and non-arbitrary resolution.

We call the additional projectability conditions "self-defeat" and "avoid the projection of overly broad predicates". In stating the two conditions, we use the term "ρ" to denote the set of predicates that are reliable in the given situation (evaluated prior to sampling). Self-defeat is then expressed as follows:

> **Self-defeat [SD]**: A predicate ϕ in ρ is projectable in a given situation only if there is no object c in our sample for which we are justified in believing that c is ϕ, such that for all collections of inferences, Σ, that we have made that are sufficient for justifying our belief that c is ϕ, there exists a proposition, P, appearing in Σ, and a predicate ψ in ρ, such that P implies that c is ψ, and we are justified in believing that the frequency of ψ in the domain is not approximately identical to the frequency of ψ in the sample.

[SD] is similar to Freitag's principle of derivative defeat, and its purpose is to exclude induction with a ϕ predicate in cases where our justification for attributing ϕ to some elements of our sample is derivative of our justification for attributing ψ to those objects, and we are justified in believing that the frequency of ψ in the domain is not approximately identical to the frequency of ψ in the sample (but notice that **[SD]** also applies in the case where $\phi = \psi$). In understating **[SD]**, we assume that we are justified in believing that the frequency of ψ in the domain is not approximately identical to the frequency of ψ in the sample if and only if we are justified in believing that $\text{freq}(\psi|D - S) \in U$ and $\text{freq}(\psi|S) \in V$, for some U and V, such that U and V are subsets of [0, 1] and $U \cap V = \varnothing$ (where D is the domain and S is the sample).

In one respect, **[SD]** is less restrictive than Freitag's principle, because it only applies when a relevant ψ *is in* ρ, which is why **[SD]** does not deem '*x* is green' to be non-projectable in the example of Tamara presented in Sect. 2. Beyond this, **[SD]** is far more demanding the Freitag's principle, and it is straightforward to find examples where **[SD]** deems a predicate to be non-projectable but Freitag's principle does not. For example, suppose we draw an *n* member sample (under conditions of random sampling and domain stability). Suppose we also know that each element of the domain has a unique mass, and that exactly one element of the domain is yellow. Let m_1 through m_n specify *n* distinct possible masses, let M be the property of having mass m_1 or …or mass m_n, and let M_y be the property of being yellow or having mass

m_1 or …or mass m_n. Now suppose that for each object c in our sample we are able to determine that c is M_y, but not whether c is M or whether c is yellow. Now consider the predicate '*x* is M_y or ϕ', where ϕ is arbitrary. Because '*x* is M_y' is not a 'discriminating predicate', Freitag's principle does not tell us that the predicate '*x* is M_y or ϕ' is non-projectable. However, **[SD]** implies that '*x* is M_y or ϕ' is non-projectable in the case described, since in the situation described we are justified in believing that $freq(M_y|D\text{-}S) \leq 1/|D-S|$ and $freq(M_y|S) = 1$.

Our second proposed principle is expressed as follows:

> **Avoid Induction with Overly Broad Predicates [AOB]:** A predicate ϕ in ρ is projectable in a given situation only if there is no predicate ψ in ρ, such that necessarily all ψ are ϕ but not necessarily all ϕ are ψ, and for all objects c in our sample for which we are justified in believing that c is ϕ and for all collections of inferences, Σ, that we have made that are sufficient for justifying our belief that c is ϕ, there is a proposition P appearing in Σ, such that P implies that c is ψ.

[AOB] applies in order to exclude inductive inferences with predicates that are unnecessarily broad. In cases where the principle applies due to some predicate ψ being more specific than another one ϕ, the principle prevents inductive inference concerning ϕ-ness. There are two interesting subcases: In the case where induction with ψ is non-projectable, it is desirable to prevent induction concerning ϕ-ness, in addition to induction concerning ψ-ness, which is what **[AOB]** requires. On the other hand, in the case where ψ is projectable, the principle allows conclusions concerning ϕ-ness to proceed by the deduction from conclusions reached by induction concerning ψ-ness. We will explain later why **[AOB]** is needed in addition to **[SD]**.

While **[SD]** blocks all instances of induction where the proposed inference is based on the projection of a predicate whose frequency among the sample and domain is known to differ, **[SD]** does not block the possibility of inductive inferences with reliable predicates leading to inconsistent conclusions. For example, suppose we take a large random sample of objects, in a case where domain stability holds and we know that 50% of the objects in a domain are red. Now suppose (unlikely as it may be) that 50% of the objects in our sample are blue and 50% are green. Although neither '*x* is blue' nor '*x* is green' is deemed non-projectable by **[SD]** or **[AOB]**, the conclusion that 50% of the objects in the domain are blue and 50% are green is inconsistent with the information that 50% of the objects in the domain are red. A further condition is needed, in order to block inconsistency in this kind of situation.

Recall that ρ denotes the set of predicates that are reliable in a given situation. Now let ρ′ be the set of predicates that result from removing all of the elements of ρ that are deemed non-projectable by **[SD]** or **[AOB]**. We say that a set of predicates P is *inductively consistent* if and only if the set of conclusions given by induction from the elements of P is consistent with our background information in conjunction with all other propositions we are justified in accepting. In that case, define the set of predicates that are projectable in a given circumstance as the intersection of the set of subsets of ρ′ that are inductively consistent and have no inductively consistent proper supersets. Expressed more precisely, we have:

> **Projectability [PRO]:** x is ϕ is projectable in given circumstances if and only if x is $\phi \in \cap\{P: P \subseteq \rho' \ \&\ P$ is inductively consistent & for all $P^+: P \subset P^+ \subseteq \rho' \Rightarrow P^+$ is not inductively consistent$\}$.

Having provided our formal proposal, we finally return to explain the role of **[AOB]**. The purpose of **[AOB]** is to block consideration of predicates whose inclusion in ρ′ would create conflicts with other predicates that should be regarded as projectable. For example, consider the previously described situation where we know that 50% of the objects in the domain are red, and 50% of the objects in our sample are blue and 50% are green (where sampling was random and domain stability holds). In this case, the predicates '*x* is blue' and '*x* is green' are deemed non-projectable, because no subset of ρ′ that contains both predicates is inductively consistent. Now suppose that in addition to their color, we observed that 100% of the objects in our sample are cubes. This gives us a reason for concluding that that the frequency of cubes in the domain is probably very high. But consider the predicates '*x* is blue or not a cube' and '*x* is green or not a cube'. The purpose of **[AOB]** is to exclude such predicates from ρ′: If these predicates were in ρ′, then **[PRO]** would incorrectly classify the predicate '*x* is a cube' as non-projectable, since '*x* is a cube' would not be a member of any inductively consistent subset of ρ′ that includes both '*x* is blue or not a cube' and '*x* is green or not a cube'.

Before considering some additional examples, let us assess how well **[PRO]** satisfies the four desiderata that we proposed earlier. The first desideratum concerns reliability. The set of predicates ρ′ with which our proposed projectability condition is expressed only includes reliable predicates. So clearly the proposal satisfies this condition. Our second desideratum is that the set of projectable predicates be inductively consistent. Our proposal clearly satisfies this desideratum, since our proposed condition says that the projectable predicates are the ones whose conclusions are in the intersection of a set of consistent subsets of a set. Our proposal also apparently satisfies the third condition (Non-arbitrary Resolution). Indeed, the set of projectable predicates is simply the result of taking the reliable predicates, ρ, removing some apparently problematic predicates (according to **[SD]**) and some redundant and potentially problematic predicates (according to **[AOB]**) in order to form ρ′, and then removing any further predicates that contribute to inconsistency (according to **[PRO]**). Our final desideratum tells us that outside of inductive inferences that are deemed non-projectable as a consequence of avoiding inconsistency in a non-arbitrary manner, it is reasonable to accept the conclusions of reliable inductive inferences. It is not 100% clear that **[PRO]** satisfies this desideratum. In fact, we regard **[PRO]** provisional, and acknowledge the possibility that **[SD]** and **[AOB]** are too weak to exclude a sufficient number of predicates from ρ′, in some cases. With regard to the provisionality of **[PRO]**, notice that if further non-trivial conditions are added to **[SD]** and **[AOB]** in defining ρ′, then the only effect will be to increase the number of predicates that are deemed projectable, in some situations. This is due to brute manner by which **[PRO]** eliminates inconsistency: In the case of a conflict among predicates, all are ejected. So the only possible effect of an additional principle that removes a predicate ψ from ρ′, is that a predicate that would have been deemed non-projectable due to conflict with ψ may yet be projectable according to

[PRO]. We are satisfied with the provisionality of **[PRO]**, since it is clear that any reasonable projectability conditions that could be added to our account would only increase the number of predicates that are deemed projectable in a very small number of highly unusual cases.

Variant expressions of the Goodman problem have been formulated (in, e.g., Skyrms, 2000; Johannesson, 2023) that consider the problem of predicting which function best captures the relationship between two (or more) quantities. Inasmuch as such examples can be reformulated as the problem of projecting the frequency with which some property holds among a sample of objects, the approach presented here applies to the examples. For example, suppose we gather a random sample of squirrels, under conditions of domain stability. Suppose that among our sample we notice that the length of the body of each squirrel is twice the length of its tail. The predicate '*x* is a squirrel with a body that is twice as long as its tail' obviously denotes an intrinsic property. So projection of the predicate is reliable, assuming random sampling and domain stability, and we expect that this predicate would typically be projectable. On the other hand, relevant grue-like predicates are non-projectable. For example, consider the predicate '*x* is a squirrel with a body that is twice as long as its tail or *x* is a member of our sample'. It is demonstrable that induction with this predicate is not reliable under the given conditions. But now suppose we know that every squirrel has a unique mass and m_1, …, m_n are the masses of the squirrels in our sample. In that case, projection of the predicate '*x* is a squirrel with a body that is twice as long as its tail or *x* has mass m_1 or … or m_n' denotes an intrinsic property. Although induction with this predicate is reliable, it is not projectable in the situation described (according to **[SD]**).

In an example presented by Johannesson (2023), we are presented with a device that computes an unknown function for input values ranging from 0 to 1,000,000 and output values of 0 and 1. We draw a random sample of 1,000 input–output pairs and observe that 100% of the pairs conform to some function *f*. As Johannesson observes, there is inevitably another function *f** that agrees with *f* for all of the input–output pairs in our sample, but disagrees with *f* for all of the input–output pairs that our not in our sample. To make the example interesting, assume that domain stability holds and that the input–output pairs in the relevant domain are physically realized as intrinsic properties of associated objects. In that case, there are 1,000,001 objects in the relevant domain and each is uniquely identified by an associated input value. So the situation described is one in which we are able to formulate a predicate denoting an intrinsic property that characterizes membership in the sample. It is also apparent that neither of the predicates '*x* agrees with *f*' nor '*x* agrees with *f**' is projectable according to our account. Indeed, let M be the set of 1,000 input values for objects in our sample. In that case, **[SD]** tells us that the predicate '*x* agrees with *f*' is non-projectable, where $\phi =$ '*x* agrees with *f*' and $\psi =$ '*x* has an input value in M' (and similarly for *f**). In the present example, it is possible to foster the intuition that it is possible to draw a reasonable conclusion about which function the device computes, if we add the detail that the output values conform to a familiar pattern (e.g., for all objects in our sample, the device returns 1 for even-numbered inputs, and 0 for odd-numbered ones). However, this intuition is only warranted if we import background information into the example, in the form of justified expectations concerning which functions the device is likely to compute. In a variant of the example where we are justified in assigning prior probabilities to different propositions concerning which function the device computes, it might be possible to use Bayesian conditionalization to assign justified posterior probabilities to various hypotheses concerning which function the device computes, after seeing how the device behaves in a random sample of cases. Our account delivers the correct result in such cases: In cases we are justified in accepting posterior probabilities on the basis of Bayesian conditionalization, our account (and **[PRO]** in particular) deems all predicates that support inductive inference to a conclusion that conflicts with those posteriors to be non-projectable.

6 Induction with Natural Properties

Our approach to Goodman's problem is obviously very permissive regarding the range of predicates that may be employed in making cogent inductive inferences. Indeed, a common way to conceive of natural properties is as properties shared by objects that bear some 'real' similarity to each other, and are more sparse than sets (Bird & Tobin, 2022; Quine, 1970). But as the simulation results presented in Fig. 1 show, inductive projection is reliable for randomly selected sets of objects, where the members of the set bear no similarity beyond shared membership in the randomly selected set. And, as we already mentioned, the results presented in Fig. 1 do not depend on the assumption that the elements of G were selected at random: We could have selected them by any means that we like, so long as the elements of S were selected independently of the elements of G. So, in the end, we do not regard the naturalness of a property as a necessary condition for its inductive projection. Despite this conclusion, simulation results presented in this section show that the inductive projection of natural properties (construed in a particular way) is more reliable in cases where the sample size is small.

The previous discussion focused on bivalent properties, where for each property and individual, the individual simply has the property or it does not. In addition to bivalent properties, we now consider characteristics that admit of *magnitude* (such as height, weight, loudness, brightness, etc.), and represent the basic features of objects by points in a *quality space*, which is a sort of conceptual space (cf. Gärdenfors, 1990). Given a preferred quality space, we suppose that the points representing the qualities of objects are distributed in a non-uniform way, with the points forming clusters which are aptly described as representing natural kinds. In this sort of case, one might expect that predicates that are formulated to match up to the natural kinds will be better as bases for induction than predicates that do not. Within the simulation studies that we performed, this was the result that we observed in the case where the size of the sample is small.

We here present the results of a simulations study involving a four dimensional quality space, namely, $CS = \{ (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : 0 \leq x_i \leq 1 \}$.[10] Our study consisted

[10] The actual numeric values in our simulations were double precision floating-point numbers.

of many iterations of the following procedure: (1) generate a domain of objects, each object being assigned to a point in CS, (2) select a random sample, and (3) compare the sample and domain frequencies according to two competing systems of categories that differ in their naturalness. Each individual domain consisted of 1,000,000 objects organized into natural kinds. A domain of objects was generated by first selecting eight prototypes (for the natural kinds) from among the points in CS by a uniform distribution over [0,1] for each quality dimension. Individual objects were then assigned to the prototypes (i.e., the natural kinds), with probability 1/8 of being assigned to each prototype. After being assigned to a prototype, the characteristics of objects were assigned. For each object, the value of each quality dimension was selected by a normal distribution, where the mean was the value in that dimension of the prototype to which the object was assigned, and the standard deviation was 0.3, with values less than 0 rounded up to 0, and values greater than 1 rounded down to 1.

Given a quality space populated by objects in the manner just described, we compared the reliability of induction relative to two alternative systems of categories. The first category system is non-natural. The categories for this system simply consist in the elements of the partition of CS that results from 'slicing' CS perpendicularly to the first quality dimension into 8 cuboids of identical size. So the first category consist of the region $\{ (x_1, x_2, x_3, x_4) \in \mathbb{R}^4: 0 \le x_1 \le 1/8 \ \& \ 0 \le x_2 \le 1 \ \& \ 0 \le x_3 \le 1 \ \& \ 0 \le x_4 \le 1\}$, and the second category consist of the region $\{ (x_1, x_2, x_3, x_4) \in \mathbb{R}^4: 1/8 < x_1 \le 2/8 \ \& \ 0 \le x_2 \le 1 \ \& \ 0 \le x_3 \le 1 \ \& \ 0 \le x_4 \le 1\}$, etc. The second category system is more natural, and consists in the partition of the domain that maximizes the similarity between the sampled objects that are assigned to the same category. More precisely, k-means clustering, where $k = 8$, was applied to the objects in the sample. This results in a partition of the sample that minimizes the sum of the squared distances between objects and the mean values of the cells to which they are assigned. The mean values for the cells of the selected partition are called centroids. A partition of the full quality space is then given by assigning each point to the nearest centroid.

The independent variable for our simulations was sample size. The dependent variable, for each system of categories, was the mean difference between the frequency of objects falling within a category in the sample and the frequency of objects falling within the category in the domain as a whole (averaged across the 8 categories for each system). We refer to the mean distance between the frequency of objects falling within a category in the sample and in the domain as the "mean error" for inductive inference with the corresponding system of categories. Figure 3 shows the mean error rates for the 2 systems of categories for different size samples. The data presented in Fig. 3 derives from 100,000 randomly generated domains for each of the 6 possible sample sizes, with one sample per domain.

The data presented in Fig. 3 suggests that induction with natural categories is more accurate than with non-natural categories in the case where objects are clustered around prototypes and one's sample size is small, with the advantage conferred by natural categories dissipating as one's sample size becomes large. While the difference in the reliability of induction with the two sorts of categories is small, it is not inconsequential, and we would expect the difference to be of great consequence in situations where it is necessary to make inductive inferences based on small

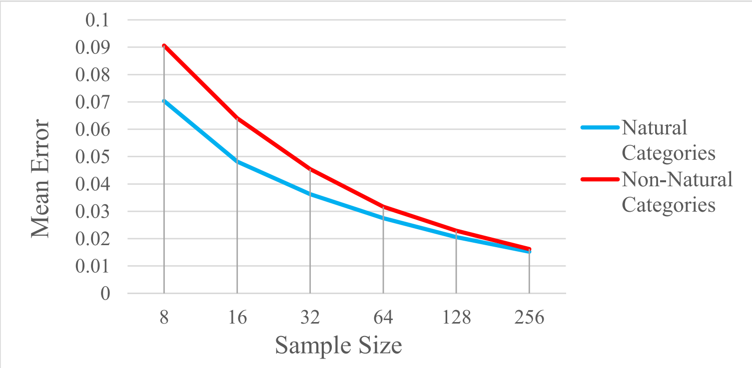

Fig. 3 Mean error rates for natural and non-natural categories as a function of sample size

samples, and where the accuracy of those inferences makes a difference to success and/or survival. Indeed, under plausible evolutionary conditions, agents using non-natural categories will die out, being replaced by agents who use natural ones.

7 Conclusion

Our main goal here was to present an approach to Goodman's New Riddle of Induction that avoids the problems that have beset previous approaches. Rather than propose criteria aimed at blocking a range of intuitively undesirable inductive inferences, we identified ideal conditions under which to perform induction. We then observed that induction with non-gruesome predicates is reliable under the ideal conditions, while induction with gruesome predicates is not. Identifying the predicates that permit reliable inductive inferences is not sufficient to address the possibility of conflicts with background information. To address this issue, we proposed a mechanism for resolving inconsistencies. An ideal solution to the Goodman problem would precisely identify which instances of induction are reasonable and which are not. Our approach comes much closer to meeting this goal than previous approaches. A further advantage of our approach is in the tight connection it draws between projectability and reliability: The projectability criteria that we propose have the important consequence that induction with predicate ϕ is projectable in a given situation only if induction with ϕ is reliable in that situation.

Our approach to the Goodman problem ultimately leads to a permissive view about which properties are suitable for inductive projection. As an addendum to this conclusion, we show, in the penultimate section, that natural properties are more reliable as a basis for induction in cases where the sample is small. Although the effect is small, it is a difference that would obviously confer a selective advantage upon agents who prefer the inductive projection of natural properties.

Acknowledgements Work on this paper was supported by the DFG-Grants TH 2580/1–1, SCHU 1566/17−1, and 432308570. For valuable comments concerning the content of this article, we are indebted to Igor Douven, Ludwig Fahrbach, Peter Gärdenfors, James Hampton, Siegfried Jaag, Markus Schrenk, and two anonymous reviewers for Philosophia.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data presented in the Figures 1, 2, and 3 was generated by computer simulations. Code for the simulations is available at the following webpage: https://www.philosophie.hhu.de/professuren/seniorprofessur-theoretische-philosophie-gerhard-schurz/gerhardschurz/publikationen/computerprogramme.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

Bird, A., & Tobin, E. (2022). Natural kinds. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/natural-kinds/. Accessed 1 Oct. 2023.

Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research, 8*, 133–147.

Dorst, C. (2018). Bet accepted: A reply to Freitag. *Philosophical Quarterly, 68*(270), 175–183.

Freitag, W. (2016). The disjunctive riddle and the Grue-Paradox. *Dialectica, 70*(2), 185–200.

Gärdenfors, P. (1990). Induction, conceptual spaces and AI. *Philosophy of Science, 57*(1), 78–95.

Godfrey-Smith, P. (2011). Induction, samples, and kinds. In Campbell, O'Rourke, & Slater (Eds.), *Carving nature at its joints: Topics in contemporary philosophy: Volume 8* (pp. 33–52). The MIT Press.

Godfrey-Smith, P. (2003). Goodman's problem and scientific methodology. *Journal of Philosophy, 100*, 573–590.

Goodman, N. (1955). *Fact, fiction, and forecast*. Harvard University Press.

Jackson, F. (1975). Grue. *Journal of Philosophy, 72*, 113–131.

Jackson, F., & Pargetter, R. (1980). Confirmation and the nomological. *Canadian Journal of Philosophy, 10*, 415–428.

Johannesson, E. (2023). The statistical riddle of induction. *Australasian Journal of Philosophy, 101*(2), 313–326.

Marshall, D., & Weatherson, B. (2018). Intrinsic vs. extrinsic properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/intrinsic-extrinsic/. Accessed 1 Oct. 2023.

Pollock, J. (1995). *Cognitive carpentry*. MIT press.

Quine, W. V. (1970). Natural kinds. In Rescher (Ed.), *Essays in honor of Carl G. Hempel* (pp. 5–23). D. Reidel Publishing Company.

Scheffler, I. (1963). *Anatomy of inquiry*. Knopf.

Skyrms, B. (2000). *Choice and chance: An introduction to inductive logic*. Wadsworth.

Swinburne, R. G. (1968). Grue. *Analysis, 28*(4), 123–128.

Warren, J. (2023). The independence solution to grue. *Philosophical Studies, 180*, 1305–1326.