Counterfactuals and the ‘Grue-Speaker’

Alfred Schramm
University of Graz
Anonymized version for review
(corrections June 2018)

Freitag (2015) and Schramm (2014) have proposed different, although converging, solutions of Goodman’s New Riddle of Induction. Answering their proposals, Dorst (2016 and 2018) has used the fictitious character of a ‘grue-speaker’ as his principal device for criticizing counterfactual-based treatments of the Riddle. In this paper, I argue that Dorst’s arguments fail: On the observation of no other than green emeralds, the ‘grue-speaker’ cannot use the symmetry between the ‘green’- and ‘grue’-languages for claiming ‘grue’-instead of ‘green’-evidence, and the counterfactuals involved (explicitly by Schramm and implicitly by Freitag) remain unaffected by Dorst’s proposal for how to evaluate them.

Keywords: counterfactual conditionals, evidence, Goodman’s paradox, grue, induction, New Riddle of Induction

1 *Introduction*

Wolfgang Freitag’s (2015) interesting (and entertaining!) paper on Goodman’s New Riddle of Induction was recently discussed by Christopher Dorst (2018) who objects that Freitag’s proposal must fail for relying on faulty counterfactual claims. Like already once before in his (2016) criticism of Schramm (2014), Dorst uses the fictitious character of a ‘grue-speaker’ as his principal device for criticizing counterfactual-based treatments of the Riddle. In conclusion, Dorst states that all counterfactual-based proposals

> ‘… share the same general problem: they beg the question against the grue-speaker by assuming counterfactuals that he would reject. The fact that appeals to counterfactuals are question-begging has been emphasized by authors such as Roskies (2008) and Dorst (2016), but the continued appearance of proposals that rely on such appeals indicates that the message is not being taken up’ (2018: 182).

This complaint calls for having a closer look at the matter. I shall argue that both, Freitag’s and Schramm’s solutions (despite their marked differences), remain unaffected by Dorst’s criticism.

This paper proceeds as follows. In *Section 2*, I recall as much pre-history of our subject as needed for the ensuing discussion. *Section 3* explains Schramm's proposal of 2014 and discusses Dorst's purported counterexamples. In *Section 4*, I explain Freitag's proposal, in particular that it is not counterfactual-based. *Section 5* deals with Dorst's principal objections against both Schramm's and Freitag's solutions: the grue-speaker's purportedly symmetric answer and how to evaluate counterfactual conditionals.

2 Some pre-history

Selecting from the rich pre-history of our subject, I restrict this section to points that are relevant for later concerns.

2.1 Goodman: Projectibilily as a prerequisite for 'genuine' confirmation

Already in his (1946) Goodman had offered two examples in support of his opinion that 'positive instances' accounts of the confirmatory relation between evidence and hypotheses are to no avail as long as the deeper problem of projectibility remains unresolved. Referring to various contemporary accounts of confirmation (in particular Carnap's and Hempel's), Goodman concluded that

> 'What we have … is an ingenious and valuable logico-mathematical apparatus that we may apply to the sphere of projectible or confirmable predicates whenever we discover what a projectible or confirmable predicate is.' (Goodman 1946: 385)

This view, that projectibility poses a problem to be solved as a prerequisite for any successful application of formal confirmatory relations, became the central tenet of his 'New Riddle of Induction' (Goodman 1983: 72-81). By his famous 'grue'-example presented there, he intended to prove by argument the need for distinguishing 'projectible' from 'non-projectible' predicates (or classes or hypotheses).[1] For later use, I split up the argument schematically into separate steps, all of them taken from one page (*ibid.*: 74).[2]

1 '[T]he new riddle of induction … is more broadly the problem of distinguishing between projectible and non-projectible hypotheses' (Goodman 1983: 83), and 'The problem of what statements are confirmable merely becomes the *equivalent problem* of what predicates are projectible from known to unknown cases' (*ibid.*: 26; my emphasis).

2 I standardize the differing notations in the field by adapting that of Schramm (2014). *Predicates*: '*E*' for emerald; '*O*' for observed (examined, sampled, etc) by time *T*; '*G*' for green; '*B*' for blue; '*GR*' for grue; '*BL*' for bleen. *Individual constants*: $a, b, \ldots, a_1, \ldots, a_n, a_{n+1}$. *Individual variables*: $x, y, \ldots$. *Evidence statements*

SO: On the basis of the very *Same Observations*, represented by evidence statements $e_1, \ldots, e_n$, summarized by $\mathcal{E}$, and by

ADP: choosing *Appropriately Defined Predicates*, we arrive at

PES: *Parallel Evidence Statements*[3] $e_1', \ldots, e_n'$, summarized as $\mathcal{E}'$, such that

CIP: $\mathcal{E}$ respectively $\mathcal{E}'$ equally well *Confirm Incompatible Predictions* (hypotheses) h respectively h'. But obviously

GC: only h is *Genuinely Confirmed* while h' is not.

Consequently: In order to separate the *genuine* ('valid') confirmations from the ('invalid') others, we must differentiate between 'projectible' and 'non-projectible' predicates (or classes or hypotheses).

Here is the 'grue'-example, schematized in this sense:

After having *observed* no other than green emeralds $a_1, \ldots, a_n$ (*SO*), the according *evidence statements* e_1: $Ea_1 \wedge Oa_1 \wedge Ga_1, \ldots, e_n$: $Ea_n \wedge Oa_n \wedge Ga_n$ are summarized as

$\mathcal{E}$: All emeralds examined by now are green: $\forall x((Ex \wedge Ox) \rightarrow Gx)$.

In terms of the *appropriately defined predicate* 'grue'[4] (*ADP*), the *parallel evidence statements* e_1': $Ea_1 \wedge Oa_1 \wedge GRa_1, \ldots, e_n'$: $Ea_n \wedge Oa_n \wedge GRa_n$ (*PES*) are summarized as

$\mathcal{E}'$: All emeralds examined by now are grue: $\forall x((Ex \wedge Ox) \rightarrow GRx)$.

Then $\mathcal{E}$ confirms the prediction (hypothesis)

h: The next, as yet unexamined, emerald a_{n+1} is green: $(Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow Ga_{n+1}$,

and $\mathcal{E}'$ *confirms the incompatible prediction* (*CIP*)

h': The next, as yet unexamined, emerald a_{n+1} is grue (hence blue): $(Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow GRa_{n+1}$ [$\equiv (Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow Ba_{n+1}$]

But obviously only h is *genuinely confirmed* (*GC*) while h' is not.

In a slightly different order, Goodman puts all this as follows:

(observational): $e_1, \ldots, e_n$. *Observational evidence* (= summary of all observational evidence statements believed at the respective present): $\mathcal{E}_1, \mathcal{E}_2, \ldots$; $\mathcal{E}_1', \mathcal{E}_2', \ldots$. *Hypotheses* (both, predictive singular and general): $h_1, \ldots, h_n, h_1', \ldots, h_n'$. *Background hypotheses*: $\ell_1, \ldots, \ell_n$. David Lewis's possible worlds $h, i, j, \ldots, @$ are denoted by 'w_h', 'w_i', 'w_j', ..., '$w_@$'. Quantifiers and logical connectives are standard.

3 To my knowledge, Goodman never explained what 'parallel' should mean in this connection, in particular, how two 'parallel' statements might equally achieve *evidential* (i.e., *observational*) status.

4 One of Goodman's various definitions is: '... the predicate "grue" ... applies to all things examined before t just in case they are green but to other things just in case they are blue' (1983: 74). We may take Jackson's 'grue$_3$' instead: 'x is grue at t iff (x is examined by T and x is green at t) or (x is not examined by T and x is blue at t)' (Jackson 1975: 118); here 'T' refers to some time in the future such that $t<T$. The time-indexes will be dropped where not necessary.

> '… although we are well aware which of the two *incompatible predictions* is *genuinely confirmed*, they are *equally well confirmed* according to our present definition [of confirmation by positive instances]. Moreover, it is clear that if we simply choose an *appropriate predicate*, then on the basis of these *same observations* we shall have equal confirmation, by our definition, for any prediction whatever about other emeralds - or indeed about anything else' (*ibid.*; my emphases).

Goodman concluded from this that the problem of projection needs prior and separate treatment:

> 'This difficulty cannot be set aside as an annoying detail to be taken care of in due course. It has to be met *before* our definition [of confirmation] will work at all' (*ibid.*: 75; my emphasis).

So, Goodman devised his own theory of projection for providing '... an accurate and general way of saying which hypotheses are confirmed by, or which projections are validly made from, any given evidence' (*ibid.*: 84). Although much labour has been invested in this by Goodman and some of his followers, there have also been some promising alternative proposals for resolving the Riddle. The present paper is concerned with alternatives that employ counterfactual conditionals. Dorst (2018) takes all such proposals as belonging (basically) to one kind. We shall see that the proposals of Freitag (2015) and Schramm (2014), converging in their solutions despite their marked differences, remain unaffected by Dorst's criticism if we assess them on their own respective merits.

2.2 Jackson: A counterfactual condition on inductive support

The first notable counterfactual-based proposal was by Jackson (1975). He claimed, contrary to Goodman, that there is no need for a separate theory of projection because 'all (consistent) predicates are projectible and … there is no paradox resulting from "grue" and like predicates' (*ibid.*: 114). While he remained in agreement with Goodman (and the traditional view since) by leaving *SO*, *ADP*, and *PES* undisputed, he located the root of the problem in *CIP,* the *relation* of evidence and hypothesis, or the *projection* from the former to the latter. Jackson discusses confirmation by *straight rule* '*SR*', which accords by and large with Goodman's 'confirmation by positive instances'. His idea is that any application of the *SR* for confirming some hypothesis h from given evidence $\mathcal{E}$ is subject to a *counterfactual condition*. How this should work is explained intuitively, amongst others, by the following example:

> 'Every lobster I have observed has been red. This supports that the next lobster I observe will be red …. But every lobster I have observed has been cooked, and I know that it is the cooking that makes them red - that is, that the lobsters I have observed *would not have been* red if they *had not been* cooked. Hence I do not regard myself as having good evidence that the next uncooked lobster I observe will be red' (*ibid.*: 123; my emphases).

This idea of separating valid from invalid *SR*-inferences is caught in what Jackson calls the *counterfactual condition* [*CC*]. As this condition is imposed on inductive support, let us call it *Counterfactual Condition on Support* [*CC-S*]:

> '… certain *F*s which are *H* being *G does not support* other *F*s which are not *H* being *G* if it is known that the *F*s in the evidence class *would not have been G* if they *had not been H*' (*ibid.*; my emphases).

We may note already here that the employment of the *CC-S* affords (*in addition* to the evidence that certain *F*s of the evidence class which are *H* are *G*) some *further* knowledge about these very same *F*s, namely, that, had they not been *H*, they would not have been *G*, as caught in the conditional. If this condition applies, then the inference ('projection' from the *Fs* of the evidence to other *Fs*) is not admissible.

Let us see in terms of the schema introduced above how this additional knowledge[5] and *CC-S* interact in green/grue-cases.

Like above, after having *observed* no other than green emeralds $a_1, \ldots, a_n$ (*SO*), the according *evidence statements* e_1: $Ea_1 \wedge Oa_1 \wedge Ga_1, \ldots, e_n$: $Ea_n \wedge Oa_n \wedge Ga_n$ are summarized as

> $\mathcal{E}$: All emeralds examined by now are green: $\forall x((Ex \wedge Ox) \rightarrow Gx)$.

In terms of the *appropriately defined predicate* 'grue' (*ADP*), the *parallel evidence statements* e_1': $Ea_1 \wedge Oa_1 \wedge GRa_1, \ldots, e_n'$: $Ea_n \wedge Oa_n \wedge GRa_n$ (*PES*) are summarized as

> $\mathcal{E}'$: All emeralds examined by now are grue: $\forall x((Ex \wedge Ox) \rightarrow GRx)$.

So far (from *SO* to *PES*), Jackson *accords* with Goodman, but the next step is blocked by his *CC-S* plus additional knowledge: 'the counterfactual condition is that the emeralds $a_1, \ldots, a_n$

5 Jackson speaks also of 'additional information'. We might as well take 'additional belief'.

would still have been green even if they had not been examined; and, *in the world as we know it, this condition is satisfied*' (*ibid*.: 124; my emphases), therefore the hypothesis

h: The next, as yet unexamined, emerald a_{n+1} is green: $(Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow Ga_{n+1}$

is *confirmed* by $\mathcal{E}$, while the analogous confirmation of the hypothesis

h´: The next, as yet unexamined, emerald is grue (hence blue):

$(Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow GRa_{n+1}$ [$\equiv (Ea_{n+1} \wedge \neg Oa_{n+1}) \rightarrow Ba_{n+1}$]

is *not confirmed* by $\mathcal{E}'$, because the counterfactual condition blocks the application of *SR* in view of the already mentioned additional knowledge that the emeralds $a_1, \ldots, a_n$ would still have been green (therefore not blue) if they had not been examined.

In Jackson's own words:

> 'If we use the *SR* with the evidence that the emeralds $a_1, \ldots, a_n$ are green and examined, and grue and examined, … [then,] if we bring in the fact that a_{n+1} is unexamined, we no longer are dealing with a case of certain *F*s being *G* supporting other *F*s being *G*, but of certain *F*s which are *H* being *G* supporting certain other *F*s which are not *H* being *G*; … hence, [we] must take note of the counterfactual condition. But if we take note of this condition, we do not get an inconsistency because - although $a_1, \ldots, a_n$ would still have been green if they had not been examined - they would not have been grue if they had not been examined' (*ibid*.: 124).

Summarized: Jackson claimed that abiding by his *CC-S* avoids unwelcome projections of incompatible predictions from 'parallel evidence' produced by reformulating the evidence grueishly. But the judgment *how* the *CC-S* applies affords additional knowledge that is not included in the evidence itself.[6] This provides a pivotal point for Roskies' criticism of Jackson's proposal.

[6] For instance, he argues that if '… an emerald that is green and examined, and so, grue … had not been examined, it would still have been green, *because examining emeralds* (*and indeed examining most things*) *doesn't alter their color* …' (*ibid*.: 125, my emphasis). Obviously, such generalizations cannot be observational knowledge (evidence) and are in need of inductive justification themselves.

2.3 *Roskies' criticism of Jackson: Counterfactual robustness*

Roskies (2008) contains everything that can fairly be said in criticism of Jackson's proposal.[7] So, I merely repeat here what we need for later considerations.

Roskies agrees with Jackson on his denial of non-projectible 'properties' (as she says) and she also presupposes (without explicit discussion) the correctness of the steps from *SO* to *PES*. But she questions the legitimacy of bringing in the already mentioned additional knowledge (or beliefs) needed for employing the *CC-S*. And, most importantly, she spots there a problem that she claims to revive the Riddle: not all predicates, resp. properties, appear to be *counterfactually robust*: 'Jackson smuggles into the application of his Counterfactual Condition knowledge that we are not entitled to, namely, that greenness of the observed emeralds is counterfactually robust whereas grueness is not' (*ibid.*: 222). As we are dealing with a problem of *justification* (like all problems of induction are), the question arises how to justify appeals to the background hypotheses required for employing the *CC-S* correctly. There looms a circularity, Roskies observes, because '… [t]he background knowledge Jackson appeals to is … not the knowledge required in applying the *SR*, but rather [it is] the knowledge required in applying the [*CC-S*], that is the problem' (*ibid.*: 224). The counterfactual condition cannot be *evaluated* on the exclusive basis of having observed that the emeralds are green because observation alone does not deliver knowledge of their '*counterfactual properties*'.

Thus, there arises a new riddle, the *riddle of robustness*. Taking account of the '… symmetry of normal and grueish colors noted by Goodman … [and that] our choice of language should not make it the case, in one instance, that the next emerald we see will look green, and in the other, blue' (*ibid.*: 227), Roskies arrives at the conclusion that

> '… to apply Jackson's counterfactual condition, and thereby prevent ourselves from faulty inductions, we already need to have resolved a riddle, though not the classical form of Goodman's riddle. Rather, it is the *epistemological problem* that arises in justifying, in a given context, which of the infinitely many predicates consistent with our evidence are the counterfactually robust ones, and thus the appropriate ones to use as an inductive base. If green is counterfactually robust, then grue is not, and vice versa' (*ibid.*: 229; my emphasis).

---

7 The basics of her argument can also be extended for criticizing the revised proposal of Jackson and Pargetter (1980), so I need not discuss it here.

This completes the pre-history of our discussion of counterfactual-based proposals for solving the Riddle.

3 Schramm: A counterfactual condition on the evidence

Of the views discussed in this paper, Schramm (2014) presents the most radical departure from the schema introduced above: he disputes head-on the steps leading from *SO* to *PES*. An *admissible* (because 'significant', as he calls it[8]) evidence statement *e*, gained from the *observation* of an individual fact like some emerald *a* being *green* (*SO*) can never be turned into a *likewise admissible* (significant) 'parallel' evidence statement *e´* (*PES*) by merely introducing appropriately defined predicates like *grue* (*ADP*), except *e* and *e´ express the same proposition*. By this, Schramm denies the effect created by grueish predicates as intended and also reaffirmed by Goodman in a later publication: '… an examined thing is *determined* to be green if and only if it is *determined* to be grue, so that we have *exactly equal* and parallel evidence' (Goodman 1972: 359; my emphases).

In contrast, Schramm's view on evidence is that a thing determined by examination to be objectively green *cannot* be determined by examination to be objectively grue as well because an object *a* has property *P objectively, iff a has P whether or not it is examined*. This is the point where Schramm's counterfactual condition comes in. He imposes it *on the evidence*, naming it an *objectifying condition of an evidential proposition* (so I refer to it by '*OC-E*'):

> 'In order to express the *full* content of what we accept as evidence (and, by this, make the claim of objectivity explicit), we must use a *counterfactual condition*: for any one of our examined green emeralds Ea_i we must accept (*in addition* to accepting that it is examined and green) that it *would* also be green if, *counterfactually*, *it had not been examined*, i.e., $\neg Oa_i \Box\!\!\rightarrow Ga_i$. …. To *deny* this counterfactual of an examined and (found to be) green emerald would come to claiming that it were merely green *just in case* it were examined, contrary to our evidential claim of it being *objectively* green.' (Schramm 2014: 576)

[8] Schramm's terminology is somewhat unfortunate, such that his point is not easy to grasp. Instead of distinguishing *evidence* statements (or *evidential* propositions) *e* and *e´* by their 'significance' as distinguishing mark of their 'admissibility', he might as well have called right away only the significant ones 'evidential' and denied the other ones that status: a non-significant proposition cannot serve as evidence and, thus, provides no base for inductions.

Note that Schramm imposes by this a necessary condition on singular external-world beliefs that is suited to serve as a barrier against (a kind of) solipsism: an experientially gained belief that *Pa* (for instance by observation, examination, etc., such that $Oa \wedge Pa$) is significant and, thus, admissible as evidence, only if the according counterfactual $\neg Oa \Box\!\!\rightarrow Pa$ is accepted (believed) as well. In short: evidential significance of *Pa* is already gained by observation *together* with the belief that *Pa*. This indicates no causal modality (that being *P* was causally dependent on being either *O* or $\neg O$) but a conceptual or metaphysical modality: being *objectively P* means being *P whether or not being observed*. And *believing* that objectively Pa_i means *believing* that Pa_i whether or not Oa_i. Consequently:

> '… although we *base* our acceptance that Pa_i on experiential procedures like observation, the acceptance of it *as an objective* truth affords also the acceptance of the according counterfactual' (*ibid*: 578).[9]

By this, Schramm cuts off Goodman's very first step towards *PES*: after having *observed* no other than green emeralds $a_1, \ldots, a_n$ (*SO*), and for resulting in 'significant' evidence, the according *evidence statements* e_1: $Ea_1 \wedge Oa_1 \wedge Ga_1, \ldots, e_n$: $Ea_n \wedge Oa_n \wedge Ga_n$ must be summarized *together with* their objectifying conditions: $\neg Oa_1 \Box\!\!\rightarrow Ga_1, \ldots, \neg Oa_n \Box\!\!\rightarrow Ga_n$ as

$\mathcal{E}_1$: All emeralds examined by now are green and would have been green if they had not been examined: $\forall x((Ex \wedge Ox) \rightarrow (Gx \wedge (\neg Ox \Box\!\!\rightarrow Gx)))$; (cp. *ibid*: 577).

The *proposition* expressed by $\mathcal{E}_1$ can, of course, be equivalently put in grueish terms; however, based on the same definitions as in *ADP*, this reads no longer 'parallel' in Goodman's sense but, instead,

$\mathcal{E}_1'$: All emeralds examined by now are grue and would have been bleen if they had not been examined: $\forall x((Ex \wedge Ox) \rightarrow (GRx \wedge (\neg Ox \Box\!\!\rightarrow BLx)))$; (cp. *ibid*: 582).

$\mathcal{E}_1$ and $\mathcal{E}_1'$ are merely different expressions of that evidence that was gained from observing the emeralds and 'determining' them to be (objectively) green. And this evidence provides no base for predicting that the next, as yet unobserved, emerald Ea_{n+1} would be grue. On the other hand, the Goodmanian 'parallel' *statement* that the observed emeralds 'are' also grue is no parallel *evidence statement* for lack of evidential

---

9 This is reminiscent of a remark by Jackson (1975: 126) although he relates it, differently from Schramm, with the issue of support: 'Examining, observing, sampling, and so on, are how we - human beings - come to know that certain *A*s are *B*; but our knowing this is separate from these *A*s being *B* supporting certain other *A*s being *B*. *What* we come to know does the supporting (if any), not our *coming to know* it.' (The latter emphasis is mine.)

significance: the emeralds, having been observed and determined to be green, would not have been grue (hence blue) if they had not been observed, but would still have been green. In short: significant green-evidence is not merely incompatible with purported grue-evidence, it is *counterevidence* to the grue-hypothesis.

It will be important to keep in mind the difference between Schramm's *OC-E* and Jackson's *CC-S*. Jackson concedes equal evidential status to all examined emeralds being green *and* being grue as well; in his view, it is the *support* lent by the purported grue-evidence to the grue-hypothesis that is then eliminated by his *CC-S* (plus additional information or background knowledge, as Roskies puts it). In contrast, Schramm's *OC-E* precludes the 'parallel' grue-statements from gaining evidential status right from the start because he links the counterfactual condition *directly* with the green-*observations*. Without accepting the according counterfactual conditional we cannot gain *evidential* belief from observation.

For our later discussion, we should also keep in mind that on Schramm's view evidential significance, and thus admissibility as evidence, need not by itself already result in the positive confirmation of an eventual hypothesis *h*. Confirmation (or disconfirmation) will, as a rule, involve eventual further admissible evidence or additional background hypotheses. This is why Schramm represents support *neutrally* '… for *any* qualitative, or … quantitative functions *that are appropriately defined*' (*ibid.*: 579; my emphases). Evidential significance by *OC-E* concerns only the *admissibility as evidence* and remains silent about the resulting confirmation of hypotheses, *whichever* relation of support one may prefer as an appropriately defined one. (Which one to prefer, whether any version of enumerative induction like 'positive instances', 'straight rule', or others like best explanation, epistemic probability, or whatever, is a topic of its own and belongs to the theory of 'support', as Schramm calls it repeatedly, for instance, *ibid.*: 572.)

3.1 Dorst's criticism of Schramm: Counterexamples

Dorst (2016) criticizes Schramm's proposal extensively on several accounts. One of them involves the symmetric answer of the 'grue-speaker', another one the question of how to evaluate counterfactual conditionals. As these issues concern also Dorst's criticism of Freitag, I defer their discussion to the final section.

There are some inaccuracies or ambiguities in how Dorst represents Schramm's proposal; each of them may be negligible if taken individually, but put together they result in neglecting the more salient differences between Schramm's *OC-E* and Jackson's *CC-S*. Dorst claims that Schramm's proposal, by employing a counterfactual condition, '… inherits a number of outstanding problems from Jackson's proposal, which [Schramm] has not shown us how to handle' (*ibid*.: 144), and, although 'it is (almost) a special case of Jackson's proposal …, [Jackson's] is able to rule out some problematic inferences that Schramm's is not.' (*ibid*.: 151) Summarizing all this, Dorst diagnoses that 'there are serious counterexamples to Schramm's proposal, some of which derive from its similarity to Jackson's, and some from its differences.' (*ibid*.: 152)

For substantiating this claim, Dorst discusses (among others) Jackson's red lobster example that was already mentioned above. He invites us to suppose that he had *observed* no other than red lobsters and that, had they not been observed, they would still have been red. This describes correctly how Schramm would pin down the evidence as significant by *OC-E* and, thus, as admissible. But then Dorst continues:

> Thus, according to Schramm, the observed lobsters confirm the hypothesis that all lobsters are red. This is the case even if I know that they are red because they are cooked, and that there are uncooked lobsters. That is, even if I know (or have good evidence) that the redness of the observed lobsters is not counterfactually robust with respect to whether or not they are cooked, Schramm's proposal does not direct me to take that into account when evaluating whether the observed red lobsters confirm the hypothesis that all lobsters are red. All we need to be concerned with, he says, is whether their redness is counterfactually robust with respect to whether or not they are *observed*. (*ibid*.: 151)

Indeed, all we need to be concerned with, says Schramm *referring to the evidence*, is 'counterfactual robustness' with respect to observation. But this concerns a different question from what gets *confirmed by* significant evidence. Thus, Dorst's argument fails because it takes wrongly the *OC-E* for being imposed on confirmation (like Jackson's *CC-S*) instead of on the evidence. If we put this right we get as significant evidence that

> $\mathcal{E}$: All lobsters $a_1, \ldots, a_n$ examined by now are red and would have been red if they had not been examined: $\forall x((Lx \wedge Ox) \rightarrow (Rx \wedge (\neg Ox \,\Box\!\!\rightarrow Rx)))$.

Let us suppose that there exist more, as yet unobserved, lobsters and let us (instead of Dorst's general hypothesis that all lobsters are red) first consider support for the predictive hypothesis

h_1: The next, as yet unobserved, lobster a_{n+1} is red: $(La_{n+1} \wedge \neg Oa_{n+1}) \rightarrow Ra_{n+1}$.

Obviously, $\mathcal{E}$ *may*, but *need not*, confirm h_1 right away, because this depends on *which further* evidence and additional background hypotheses we presuppose when we apply any (appropriately defined) support relation. If neither the background hypothesis

$\boldsymbol{b}_1$: Lobsters are red only if cooked

nor any other relevant evidence or belief is presupposed, then h_1 *will* be confirmed by $\mathcal{E}$; if, in addition, $\boldsymbol{b}_1$ is available, then h_1 *will not* (or very much less) be supported. But this is not yet the end of all this. Consider background hypotheses

$\boldsymbol{b}_2$: I expect (my subjective probability is high) that I'll see the next lobster in my favoured restaurant.

$\boldsymbol{b}_3$: I expect (my subjective probability is high) that I'll see the next lobster during my vacation by the sea.

Whether or not we include $\boldsymbol{b}_1$, in either case h_1 will still be well supported if we include $\boldsymbol{b}_2$, but not so if we settle on $\boldsymbol{b}_1$ and $\boldsymbol{b}_3$ in our calculation of support, etc.

Generally: Let *support* for a hypothesis h, conditional on arguments $\mathcal{E}$ (for observational evidence) and $\boldsymbol{b}_1$, …, $\boldsymbol{b}_n$ (for background hypotheses), be symbolized by '$h/\mathcal{E},\boldsymbol{b}_i$', and let '$\leq_s$' be a weak ordering on the respective strengths of support (such that $\alpha =_s \beta$ iff $\alpha \leq_s \beta$ and $\beta \leq_s \alpha$; $\alpha <_s \beta$ iff $\alpha \leq_s \beta$ and $\alpha \neq_s \beta$). What we then get will, (very) roughly, be like this:

$h_1/\mathcal{E},\boldsymbol{b}_1 =_s h_1/\mathcal{E},\boldsymbol{b}_1,\boldsymbol{b}_3 <_s h_1/\mathcal{E} \leq_s h_1/\mathcal{E},\boldsymbol{b}_2 =_s h_1/\mathcal{E},\boldsymbol{b}_1,\boldsymbol{b}_2$. Furthermore, depending on the probabilities in $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$, with $prob(\boldsymbol{b}_2 \neq_s \boldsymbol{b}_3)$, we will have either $h_1/\mathcal{E} <_s h_1/\mathcal{E},\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3$ or $h_1/\mathcal{E},\boldsymbol{b}_1,\boldsymbol{b}_2,\boldsymbol{b}_3 <_s h_1/\mathcal{E}$, etc.

In short: On Schramm's account, *significance* of evidence is subject to observation *alone* (by *OC-E*), while *confirmation* (i.e. support by significant evidence) is subject to significance of evidence *and* background hypotheses. (And the evaluations of background hypotheses themselves, we may add here, are subject to observation etc. in turn.)

Considering now Dorst's general hypothesis

h_2: All lobsters are red: $\forall x(Lx \rightarrow Rx)$,

it should be clear that h_2 will be supported on the basis of $\mathcal{E}$ (whether or not we add $\boldsymbol{b}_2$ or $\boldsymbol{b}_3$) but will not be supported on the basis of $\mathcal{E}$ and $\boldsymbol{b}_1$. Thus, Schramm's restriction of 'counterfactual robustness' to observation is not a 'mistake', as Dorst deplores (*ibid*.: 152), but a merit. I take up this finding again in the final section. For the time being, we note that being red, green, a lobster, an emerald, a rose, or what not, are (by *OC-E)* *counterfactually robust* with respect to being observed, examined, determined to be such-and-such, etc..

For a further counterexample, Dorst takes the 'emerose'-case (in the version of Godfrey-Smith 2003: 576). We may keep the discussion of it shorter because there the different character of Schramm's and Jackson's counterfactuals is even more obvious.

An emerose is defined as being either an observed emerald or an unobserved rose. The 'parallel' statement one gets from this definition after the observation of no other than green emeralds is 'All emeroses that are observed are green', which by Jackson's *CC-S* confirms the hypothesis 'All emeroses are green'; but then any as yet unobserved rose would have to turn out green. Dorst claims that 'the same case likewise works against Schramm's proposal' because there the significant evidence would be 'Emerose *a* is both observed before *t* and is green … [and] If emerose *a* had not been observed before *t*, it still would have been green.' (2016: 151) But this does not square with what we really get from Schramm's *OC-E*: having observed a green emerald *a* (that may well be *called* a green emerose by 'appropriate definition') the *significant evidence* must read correctly '*a* is a green emerald observed (by *T*) and would be a green emerald if it had not been observed (by *T*).' Thus, we conclude as we did before with respect to the grue-case: the 'parallel' *statement* is not a significant *evidence* statement. The root of this confusion may be that Dorst arrived too rashly at his opinion that '… Schramm's proposal is almost a special case of Jackson's' (*ibid*.: 150).

This observation may also contribute to explaining why, from Schramm's view-point, Dorst looks like merely begging the question by pointing out that

> 'Upon seeing a green emerald, we can of course describe the evidence however we want … But the question is which of these descriptions licenses the right predictions. That is, which of these evidential propositions are we justified in projecting to future cases?' (*ibid*: 147).

Schramm's *overall* argument was that Goodman's argument for the need of distinguishing the projectible from the non-projectible must fail because admissible (significant) observational evidence *e cannot* be turned into 'parallel' equally admissible evidence *e´*

by introducing 'appropriately defined' predicates. Now, Schramm's point is that 'upon seeing (observing, etc.) a *green* emerald' the only *admissible* evidence (as already pointed out) is what we can describe in terms of green/blue or grue/bleen ('however we want') *either* as $\mathcal{E}_1$: 'All the observed emeralds are green and would have been green if they had not been observed' *or* as $\mathcal{E}_1'$: 'All the observed emeralds are grue and would have been bleen if they had not been observed'. What we are *not* justified in, 'upon seeing a *green* emerald', is to describe the evidence either as $\mathcal{E}_2$ 'All the observed emeralds are grue and would have been grue if they had not been observed' or as $\mathcal{E}_2'$: 'All the observed emeralds are green and would have been blue if they had not been observed'. Upon seeing a *green* emerald only $\mathcal{E}_1$ and $\mathcal{E}_1'$ are significant, while $\mathcal{E}_2$ and $\mathcal{E}_2'$ are not.

Of course, one may try to turn the tables by asking which of the evidential propositions we would be justified in projecting to future cases upon seeing (observing, etc.) a *grue* emerald because in *that* case $\mathcal{E}_2$ and $\mathcal{E}_2'$ would be significant while $\mathcal{E}_1$ and $\mathcal{E}_1'$ weren't. But remember that this was not Goodman's example: it started from the observation of no other than *green* emeralds, or, rather, from 'determining' that all the observed ones *are* green; so, contrary to Goodman's claim, they *cannot* be 'determined' to *be* grue as well. One cannot change the world by means of definitions; one can merely switch the sentences describing it in different terms. We return to this issue in the final section when we deal with the 'grue-speaker'.

4 *Freitag: Derivative defeat of projections depending on discriminating predicates*

Besides its argumentative content, Freitag's (2015) is also a masterly piece of playwriting. Lacking (unfortunately) Freitag's wit, I will not follow here the dialogue between a 'Professor' and a 'Mysterious Stranger' and remain content with concentrating on the proposed solution of Goodman's Riddle contained in the plot.[10]

Leaving the steps of our schema from *SO* to *PES* undisputed, Freitag explains *GC* (why only the green-hypothesis is *genuinely* confirmed) by a sophisticated defeasibility argument against *CIP*: the incompatible green- and grue-hypotheses are not equally well confirmed because, after having observed no other than green emeralds, a projection to the grue-

[10] After reading it I mused that the 'Professor' might as well be Dr. Faust who, differently from Goethe's original, doesn't identify the true nature of the visiting 'Mysterious Stranger': it is Mephisto!

hypothesis is '*derivatively defeated*'. Unlike Schramm's proposal, the argument for establishing such derivative defeat is presented *without* essentially involving any counterfactual conditionals; Freitag merely *illustrates* the force of his argument by employing counterfactual formulations. A further point to be noted in advance is that Freitag uses the colloquial 'know' that is often allowed instead of 'believe', 'accept', or similar doxastic terms (*cf ibid.*: 257, fn. 7).

Freitag defines first a useful structure by partitioning a set ***K*** of *objects of a certain kind* (e.g. emeralds) into two non-empty sets: the *sample set* $\boldsymbol{I}_a = \{a_1, \ldots, a_n\}$ and the *target set* $\boldsymbol{I}_b = \{b_1, b_2, \ldots\}$ onto which to project. The ordered pair $\boldsymbol{I} = (\boldsymbol{I}_a, \boldsymbol{I}_b)$ is called *induction set*. Enumerative *inductions* are then 'inferences from the proposition that all a_i's have property *P* ... to the general hypothesis that all ***K***'s, the b_i's included, have *P*.' (*ibid.*: 256) A statement about some $a_i \in \boldsymbol{I}_a$, for instance 'Pa_i', is called an '*inductive evidence proposition*'; all inductive evidence propositions are summarized as *inductive evidence* $\mathcal{E}$; eventual knowledge concerning the b_i's belongs, thus, not to $\mathcal{E}$ but, as *background knowledge* $\mathcal{b}$, to the *total evidence* $\{\mathcal{E}, \mathcal{b}\}$. Based on this, Freitag calls

> '... a predicate *P discriminating* ... with respect to induction set ***I*** if and only if (i) for all elements of $\boldsymbol{I}_a$ we know that they are *P* and (ii) we know that there is at least one element of $\boldsymbol{I}_b$ which is not *P*. (*ibid.*: 257)

Obviously, if *P* is discriminating in this sense, then we cannot infer 'all ***K***'s are *P*' from 'all *a*'s are *P*', because by (ii) of the definition we know that at least one $b_i \in \boldsymbol{I}_b$ is not *P*. Thus, a discriminating predicate (with respect to induction set ***I***) is not projectible, it is undermined by the *defeating* background knowledge $\mathcal{b}$ that (at least) some b_i is not *P*.

Depending on the respective background knowledge $\mathcal{b}$, there can be many and diverse predicates that are discriminating with respect to some chosen induction set ***I***. The most general of these, because discriminating with respect to *any* induction set ***I*** he calls Δ, designating 'the property of being an element of $\boldsymbol{I}_a$ ... [because] with respect to any induction set we know this predicate to apply to all and only the samples ...' (*ibid.*) Thus, although '... the projection of Δ is always perfectly supported by the inductive evidence [by condition (i)] ... it is never valid [by condition (ii)]' (*ibid.*: 258), that is, an according hypothesis is never *genuinely* supported.

Obviously, all this does not yet immediately concern our 'grue'-problem. But it leads to a problem that must be solved beforehand.

Based on our knowledge that 'observed', 'examined', 'sampled' etc. are Δ-kind predicates (therefore discriminating with respect to $\boldsymbol{I}$), we might always construct the following absurd argument: Start with the knowledge that there are objects a_1, …, a_n that are $\boldsymbol{K}$'s and all of them are Δ-kind, say P, such that Pa_1, …, $Pa_n \in \boldsymbol{I}_a$. Allow that there are more $\boldsymbol{K}$'s $b_i \in \boldsymbol{I}_b$, and take *any non-discriminating* predicate Q, then you gain on basis of the disjunctive weakening $Pa_i \vee Qa_i$ a valid inductive argument that *all* the $\boldsymbol{K}$'s are $P \vee Q$, which includes that also all the $b_i \in \boldsymbol{I}_b$, are $Pb_i \vee Qb_i$. However, as you know that P is discriminating, wherefore it is true that $\neg Pb_i$ (at least for some b_i), it follows that Qb_i. By such kind of reasoning 'We would be able to infer wild, even contradictory, hypotheses from arbitrary sample sets.' (*ibid.*)

This leads to a dilemma: if we *disallow* all disjunctive weakenings of discriminating predicates to serve as bases for inductive confirmations, then we forgo all *sound* ('valid') confirmations of this kind together with the absurd ones. But *allowing* all of them means accepting indiscriminately the 'wild' together with the sound ones. We must find some middle-ground: 'the challenge is to separate the projectible wheat from the invalid chaff and to explain what's what. We must identify the conditions for *derivative defeat*.' (*ibid.*: 260) This is what Freitag proposes as his solution:

> 'A disjunctive weakening of a discriminating predicate is derivatively defeated only if it is unsupported by 'genuine' evidence. Evidence is genuine if it is independent of, not grounded in, the evidence concerning a discriminating predicate *P*.' (*ibid.*: 262)

The explanation is as follows: *prima facie*, as the discriminating Pa_i is, and the non-discriminating Qa_i can be, evidential, their (shared) disjunctive weakening $Pa_i \vee Qa_i$ has evidential status as well. However, if all you know is that Pa_i such that you *arbitrarily* choose Qa_i for weakening (you might as well have taken $\neg Qa_i$ such that $Pa_i \vee \neg Qa_i$), then the defeat of Pa_i carries over to $Pa_i \vee Qa_i$ and defeats it as well. On the other hand, if you have evidence of Q then your evidence of $Pa_i \vee Qa_i$ is *genuine* because it is 'independent of, not grounded in' the defeated evidence Pa_i. Consequently:

> We may project, e.g., the property "Δ-or-green" if we have knowledge that the samples are green. By contrast, in those cases in which the evidence for "Δ-or-green" (or "Δ-or-blue") is *derived from* or *grounded in* the evidence for Δ, the disjunctive weakening is as unprojectible as Δ itself, because the defeater for the Δ-hypothesis also undermines the projection of "Δ-or-green".' (*ibid.*: 262).

How does this help us with the grue-example? Observe that Goodman's definition of grue can also be given equivalently as a conjunction of two disjunctions by $GRx =^{df} ((Ox \vee Bx) \wedge (\neg Ox \vee Gx))$, furthermore, that based on our *observations* the sample set consists of no other than green emeralds $Ga_1, \ldots, Ga_n \in \boldsymbol{I}_a$, and, finally, that we must accept as a logical truth that $Gx \leftrightarrow ((Ox \vee Gx) \wedge (\neg Ox \vee Gx))$. It follows that we cannot project from all *observed* emeralds being grue to *all* emeralds being grue because one of the conjuncts of that evidence, namely $Oa_i \vee Ba_i$ is *derivatively defeated*, while the projection from all observed emeralds being green to *all* emeralds being green remains undefeated:

> 'Knowledge that future emeralds are unexamined affects only the projection of "grue", as a grue-belief requires not only a perceptual belief about the colour, but also a belief about the time (and the fact!) of the examination. ... the grue-hypothesis enjoys "full inductive support" by evidence that is however robbed of its inductive potential. It is, to use Goodman's words, fully, but not genuinely, confirmed. Being dependent on the projection of "examined before *t*", the projection of "grue" is derivatively defeated by knowledge of the fact that future emeralds are not examined before *t*. … [However] … That the samples are green is *known by perception alone*.' (*ibid*.: 264; my emphasis)

Before discussing Dorst's criticism, we should pause and compare principal differences and common ground of Schramm's and Freitag's proposals.

1. Schramm employs counterfactual conditionals for distinguishing admissible (because significant) from inadmissible (because not significant) evidence. Freitag's solution does not resort to conditionals except for illustrative purposes.
2. Schramm cuts off Goodman's argument for the need of distinguishing the projectible from the unprojectible right from the start: green-observations allow only significant green-evidence and no significant grue-evidence, such that there is no base for grue-inductions anyway. Freitag allows 'parallel' grue-evidence but eliminates grue-inductions by derivative defeat.
3. Predicates denoting 'experiential procedures' (as Schramm referred summarizing to 'observing', 'examining', etc.) are Δ-kind predicates in Freitag's sense, while 'green', 'blue', 'emerald', etc. are non-Δ-kind. Applying this distinction on Schramm's proposal, we can observe that non-Δ-kind predicates are *counterfactually robust* with respect to Δ-kind predicates: 'If *a* is green (non-Δ-kind) and is observed (Δ-kind), then *a* would

be green (non-Δ-kind) as well if it were not observed (Δ-kind). This explains Schramm's resolution of the Riddle by involving counterfactuals.

Conversely, we may ask in Schramm's terms why projecting a disjunction that 'depends' (in Freitag's sense) on a predicate denoting an experiential procedure is derivatively defeated. Here the answer is that if we know no more of some *a* than that it is 'experienced', then the affirmation of any counterfactual with that antecedent would be arbitrary: For any predicate ϕ, including complex predicates like $\phi =^{df} (\psi \vee \gamma)$, that does not denote an 'experiential procedure' like *O*, any affirmation based on the sole premise *Oa* that $\neg Oa \;\Box\!\!\rightarrow \phi a$ is as arbitrary as affirming that $\neg Oa \;\Box\!\!\rightarrow \neg\phi a$.

So, Schramm's and Freitag's approaches, although fundamentally different, can be used to *explain* each other. They both solve the Riddle in that they converge at the *epistemological* aspect of experiential (perceptual) knowledge or belief: for Freitag, as quoted, it is *known by perception alone* that the actually examined samples are green. For Schramm, one gains objectivity (or 'significance') of that belief *together with*, or *by*, observing the emeralds to be green.

4. Apart from the grue-debate, both accounts contain innovations that are of more general interest. While Freitag remains content with merely referring to 'perception' as gaining us experiential outside-world knowledge of singular facts, Schramm's inclusion of counterfactuals for 'objectifying' experiential beliefs is an interesting thesis in the epistemology of perception. On the other hand, while Schramm dodges a more detailed discussion of confirmatory relations by merely demanding that they must be 'appropriately defined', Freitag's solution for 'derivative defeat' is a noteworthy contribution to the theory of enumerative induction.

I do not intend to arbitrate between the two approaches; we have seen that both have advantages of their own. But I am convinced that both of them succeed in resolving Goodman's 'New Riddle'.

5 Dorst's all-in-one-criticism of Freitag and Schramm: The 'grue-speaker' and the evaluation of counterfactual conditionals

I took care to present Freitag's solution without involving counterfactual conditionals, and we have seen that he does not need them for his *argument*. Nevertheless, Dorst takes exception to Freitag's explanatory remark that 'In the situation Goodman describes, we would not believe the samples to be grue if we did not hold them to be examined.' (Freitag

2015: 264) For Dorst, this gives rise to object '… that Freitag's proposal implicitly makes the incorrect assumption that is fair game in this dialectic to appeal to claims about counterfactual dependence.' (Dorst 2018: 179) The 'dialectic' he refers to is Goodman's well-known account of the mutual definability of green/blue and grue/bleen. Dorst's objection, based on this symmetry, is that '… the grue-speaker will have a precisely symmetric justification open to him in favour of *his* predicates' (2018: 180; *cp.* also 2016: 153), because 'Counterfactual dependence, just like time-dependence, is language relative.' (2018: 181; *cp.* also 2016: 148) By this he repeats more or less identically what he has already claimed in his criticism of Schramm's proposal.[11] So we can examine these issues in one. The next sub-section deals with how much Dorst can gain from the grue-speaker's symmetry argument. (Part of the answer we know already from our discussion of Schramm.) In the final sub-section, we discuss Dorst's (respectively David Lewis's) views on how to 'evaluate' counterfactuals.

5.1 'Grue'- vs. 'Green'-speakers: Symmetry of predicates or of evidence?

Dorst takes a *grue-speaker* to be a person who has 'grue' and 'bleen' as basic predicates of his language (call it $\mathcal{L}_{GR}$) and defines 'green' and 'blue' in their terms, while *we green-speakers* have it the other way round by defining 'grue' and 'bleen' in terms of our basic predicates 'green' and 'blue' ($\mathcal{L}_{G}$)[12] such that

> '… if we appeal to [the time-dependence involved in the respective definitions] in justifying *our* choice of predicates, the "grue"-speaker will have a precisely symmetric justification open to him in favor of *his* predicates. The symmetry of the situation thus renders our attempt at justification ineffectual.' (2016: 148, and 2018: 180; my emphases)

Now, to 'justify' a preference for using one pair of predicates instead of another pair is, as I see it, a matter of practicality; in some situations, it may be less awkward to use the one instead of the other for expressing a proposition, so why use the other? The *relevant* concern is whether we *believe* what we express by sentences containing these predicates, and what we are justified in *believing*. If we (green-speakers) believe it to be a fact that some object *a* is green, we may express this belief either by '*a* is green' or, faithful to the

11 Most of the relevant paragraphs in Dorst (2015: 148, 152-3) and Dorst (2018: 180-2) are identical or nearly identical word by word.

12 This is just in repetition of Goodman (1983: 79-80)

definition, express it by '*a* is grue if observed by *T* or is bleen otherwise'. The grue-speaker, *if he believes the same*, has the same choice as we have for expressing that belief. Dorst, however, appears to argue that the justification for believing a certain proposition, say, that *a* is green, would change with changing the predicates we use for expressing that proposition. (This is just a different guise of Goodman's claim of 'parallel' but 'equally legitimate' evidence statements.) Granting that *we* would not believe the sampled emeralds to be grue if we did not think they were examined, Dorst asks 'But does the *grue-speaker* believe that?' (2018: 180) and then proceeds to answer that question in the negative. In a similar manner, he objects to Schramm's proposal that it does not tell '… why we are justified in treating the 'green' evidence, and not the 'grue' evidence, as significant.' (2016: 149). The reply to this is, that accepting a proposition as (observational) evidence depends on what we *observe* or *perceive*, not on the *terms* we use for expressing the believed proposition. The believed fact is such that both, the green-speaker *and* the grue-speaker, can express it in either language, in $\mathcal{L}_G$ just as well as in $\mathcal{L}_{GR}$. In this sense, the *predicates* are, indeed, perfectly 'symmetrical'. But this does not make the *evidence* symmetrical in that sense that the grue-speaker and the green-speaker were forced to believe different *propositions*.

The same is true if the green-speaker and the grue-speaker do *not* agree on the evidence. Suppose person *X* believes that *a* is green, while person *Y* believes that *a* is grue. Here I leave it intentionally open who of them is the green- and who is the grue-speaker because they can also express their *differing* opinions either way: for instance, *X* says in $\mathcal{L}_G$ '*a* is green whether or not observed by *T*' or in $\mathcal{L}_{GR}$ '*a* is grue if observed by *T* or is bleen otherwise', while *Y* says in $\mathcal{L}_G$ '*a* is green if observed by *T* or is blue otherwise' or in $\mathcal{L}_{GR}$ '*a* is grue whether or not observed by *T*'. Consequently, 'symmetry' of the *predicates* does not force the green-speaker and the grue-speaker to believe the same propositions either. Thus, as the symmetry of the employed predicates neither enforces believing *the same* propositions nor enforces believing *different* propositions, we cannot draw any conclusions from their use of symmetrically defined *predicates* about which evidential *beliefs* the green-speaker or the grue-speaker might hold.

What *is* of relevance, however, is which evidential beliefs result from our *observations* or *perceptions* (however we may describe them, in $\mathcal{L}_G$ or in $\mathcal{L}_{GR}$). Remember *OS* from our schema: Goodman started his argument from having observed no other than *green* emeralds. There is no argument that therefore the grue-speaker would (or should) believe that he had observed grue ones.

Nevertheless, Dorst claims that the grue-speaker would believe counterfactuals different from the ones we green-speakers believe (*cf.* 2018: 181 and 2016: 148) and that counterfactuals can be 'gruified' (*cf.* 2018: 180 and 2016: 152). So, let us finally examine Dorst's proposal for evaluating counterfactual conditionals.

5.2 On evaluating counterfactual conditionals

That the evaluation of counterfactuals (i.e., deciding on their believability) is generally seen as a delicate matter to the present day requires no separate discussion. A recent example for highlighting the involved difficulties is Schwarz (2018). Dorst refers to David Lewis's theory of counterfactuals as his sole and primary source; so let us discuss the matter in these terms.[13]

Lewis's (1973) theory concerns general *truth conditions* for counterfactuals at possible worlds w_i. As this includes the actual world $w_@$,[14] we can restrict our discussion to the realm of inductive reasoning in $w_@$, the world we believe to inhabit. For example, in terms of Lewis's system of comparative similarity (*ibid.*: 48ff), a counterfactual $\phi \Box\!\!\rightarrow \psi$ is then true at $w_@$ according to similarity system ($\leq_@$, $S_@$) iff either (1) no ϕ-world belongs to $S_@$ (vacuous case), or (2) there is a ϕ-world w_k in $S_@$ such that, for any world w_j, if $w_j \leq_@ w_k$ then $\phi \rightarrow \psi$ holds at w_j. This means for a counterfactual like our $\neg Oa \Box\!\!\rightarrow Pa$, that it is true at $w_@$ if and only if, if there is an $\neg Oa$-world accessible from $w_@$, then the consequent Pa holds at every $\neg Oa$-world at least as close to $w_@$ as a certain $\neg Oa$-world.[15] In plain (but imprecise) words: Under the supposition that an actually observed object *Pa* had not been observed, the proposition 'If object *a* had not been observed, then *a* would be *P*' is true (if

---

13 Schramm (2014: 576) took care to motivate and formulate his objectifying condition neutrally with respect to any particular theory of counterfactuals, and Freitag (2015) used counterfactuals merely in an explanatory, non-systematic, manner. But Dorst refers exclusively to David Lewis's theory as belonging to 'the best' in regard of 'evaluating counterfactuals' (2016: 152 and 2018: 180-1); so I think it is fair to discuss these matters along Lewisian lines.

14 Lewis uses the actual world designator already in his (1968), but not in his (1973). He does not need it there, because his theory aims at the more general conditions for 'truth at *i*' from the perspective of any world *i*.

15 Lewis later adapted this for the special context of his (1981) by introducing propositions (i.e., sets of worlds) for antecedents and consequents instead of sentences. This would mesh more naturally with his account of belief as developed in his (1979), but for our purpose we need not delve that deeply into Lewisiology.

there is some world at all where *a* had not been observed) if in the closest world where *a* had not been observed, *a* is *P*.

So far, this concerns the *truth conditions* of counterfactuals, whereas Dorst considers the *evaluation* of counterfactuals, in particular, Schramm's objectifying condition and Freitag's 'implicit' counterfactual. Put in Lewis's usual terminology, we are then concerned with the *assertability* of counterfactuals, i.e. whether we may *believe* or *accept* them to be true at $w_@$, and not whether they *are* true at $w_@$. (This requires nevertheless that we have a clear understanding of their truth conditions). Lewis deals with the assertability of counterfactuals only occasionally and where of interest or appropriate. His general view of the matter was that 'Assertability goes by subjective probability.'[16] In (1973: 71-2) he states, details aside, that the assertability of a counterfactual depends essentially on the respective (subjective) *probabilities* of the relevant matters in one's most probable worlds; in our case, these are the worlds closest to $w_@$. However, Dorst doesn't mention the involved probabilities at all. Instead, he claims that

> '… we evaluate a counterfactual $P \Box\!\!\rightarrow Q$ by (i) imposing a similarity metric on the space of possible worlds, (ii) going to the nearest world where the antecedent *P* of the given counterfactual holds, and (iii) checking whether the consequent *Q* of the counterfactual also holds there. The counterfactual is true iff *Q* holds in the closest *P*-world' (Dorst 2016: 152-3 and 2018: 180-1).

Let us go through these three points.

Concerning (i), I object that a 'metric' would have to involve some kind of measure of similarity. Lewis's possible worlds, however, are *weakly ordered* by $\leq_i$ (in our case by $\leq_@$) with respect to comparative similarity or (equivalently) with respect to their closeness within their spheres of accessibility ($S_@$ in our case). But a weak ordering gives no metric in a technical sense, wherefore a 'metric' as Dorst requires is simply *not available*. As Lewis (1973: 50-2) points out, an interest in a quantitative instead of his comparative concept might derive from the intuition of paralleling closeness of worlds with *spatial distance* (which would afford at least an axiom of symmetry $d(x,y) = d(y,x)$ in the regular case). And after discussing briefly the option of non-symmetric measures for similarity, Lewis resumes 'But why bother? The appeal of a numerical similarity measure comes from the analogy between similarity 'distance' and spatial distance. To the extent that the

[16] Lewis 1976: 297. Although he wrote this in a different context, the whole paragraph is worth reading.

analogy breaks down, *the point of having a numerical measure is lost*.' (*ibid*.: 52; my emphasis)[17] Dorst, in contrast, would need for his 'metric' a nearest world approach which is not in line with Lewis's views.[18]

Points (ii) and (iii) can be dealt with together: within the framework of Lewis's conception of counterfactuals can we neither 'go to' any world different from ours, nor can we 'check there' anything. Lewis's various writings are strewn with remarks opposing trans-world identity; we are bound to the actual world and must evaluate there our counterfactuals. A reliable source is *Sect. 1.9 Potentialities* of his (1973: 36-43). I take just a few sentences from there, adapting them to our needs by interpolating emerald-talk:

> 'Things that do inhabit worlds – people, flames, … concrete particulars – inhabit one world each, no more. Our [observed emerald *a*] is an [emerald] of our world, [that] does not reappear elsewhere. Other worlds may have [emeralds] of their own, but none of these is our [observed emerald *a*]. Rather, they are counterparts of our [observed emerald *a*]. … In general: something has for *counterparts* at a given world those things existing there that resemble it closely enough in important respects of intrinsic quality and extrinsic relations, and that resemble it no less closely than do other things existing there.' (*ibid*.: 39)

Thus we have, contrary to Dorst's claims, neither a 'similarity metric on the space of possible worlds' at our disposal nor can we 'go to the nearest world' for 'checking' any counterfactuals like our $\neg Oa \Box\!\!\rightarrow Pa$. All this is not available to us.

What then can we do for the needed evaluation of counterfactuals? The answer is as plain as straightforward: the subjective probability that some *a* has *P* that we have gained from observing *a* in our world (and from how much we trust our observation under the given circumstances of observation) gives us already the subjective probability that *a* has *P* *objectively*, that is, that *a* has *P whether observed or not*. In other words: the proposition that *a* would have had *P* if we had not observed it, is to be evaluated by our actual-worldly probability of *a* having *P*.[19]

---

[17] Cp. also the last six paragraphs of his (1973: 94-95) discussing the futility of '… any humanly possible definition of comparative similarity of worlds.'

[18] Ironically, Dorst (2018) appeared in the very same issue of this journal as Blumson's excellent paper on the difference between metric-spaced and property-related accounts of similarity.

[19] For a more general discussion of the 'interaction' of probabilities and counterfactuals from a non-similarity perspective, see Hájek (2014).

Summarizing all this: After referring to Lewis (1973) as his primary source, Dorst proposes a procedure or 'method' (*ibid.*: 152) for evaluating counterfactual conditionals that isn't in line at all with that of Lewis. A correct procedure for the evaluation of counterfactuals ought to be in terms of our *actual-worldly* probabilities. In this case, however, there can be no doubt that green-*perceivers*[20] (be they green-*speakers* or grue-*speakers*), given their actual-worldly observations of green emeralds, will assign higher probability to the counterfactual that these emeralds (or their counterparts) would have been green (respectively bleen) if they had been left unobserved, than to the counterfactual that they would have been grue (respectively blue) under the same supposition.

6 Conclusion

Schramm and Freitag have proposed different, although converging, solutions of Goodman's New Riddle of Induction that appear to be successful. In criticisms of their views, Dorst claims wrongly that all counterfactual-based proposals for solving the Riddle would fall prey to analogous objections as have been put forward by Roskies against Jackson. But Schramm's counterfactuals, differently from Jackson's, are restricted to predicates that are counterfactually robust with respect to suppositions concerning experiential procedures, while Freitag's proposal does not systematically involve counterfactuals anyway, such that Dorst misses the points of either solution. Neither Dorst's device of the 'grue-speaker' nor his method for evaluating counterfactuals are suited to refute the discussed solutions of Goodman's New Riddle.

References

Blumson, B. (2018) 'Two Conceptions of Similarity', *The Philosophical Quarterly*, 68/270: 21-37.

Dorst, C. (2016) 'Evidence, Significance, and Counterfactuals: Schramm on the New Riddle of Induction', *Erkenntnis*, 81: 143-54.

[20] A ϕ-perceiver is a person who believes, on account of having examined some object *a*, that ϕa. This does not exclude the fallibility of such belief, but is a condition on the justification of observational belief.

Dorst, C. (2018) 'Bet Accepted: A Reply to Freitag', *The Philosophical Quarterly*, 68/270: 175-83.

Freitag, W. (2015) 'I Bet You'll Solve Goodman's Riddle', *The Philosophical Quarterly*, 65/259: 254-67.

Godfrey-Smith, P. (2003) 'Goodman's Paradox and Scientific Methodology', *The Journal of Philosophy*, 100/11: 573-90.

Goodman, N. (1946) 'A Query on Confirmation', *The Journal of Philosophy*, 43/14: 383-5.

Goodman, N. (1983) *Fact, Fiction, and Forecast,* 4th edn., Cambridge: Harvard University Press.

Goodman, N. (1972) *Problems and Projects*, Indianapolis and New York: Bobbs-Merrill.

Hájek, A. (2014) 'Probabilities of counterfactuals and counterfactual probabilities', *Journal of Applied Logic*, 12: 235-51.

Jackson, F. (1975) 'Grue', *The Journal of Philosophy*, 72/5: 113-31.

—— and Pargetter, R. (1980) 'Confirmation and the Nomological', *Canadian Journal of Philosophy*, 10: 415-28.

Lewis, D. (1968) 'Counterpart Theory and Quantified Modal Logic', *The Journal of Philosophy*, 65/5: 113-26

—— (1973) *Counterfactuals*, Malden: Blackwell (reprint 2005).

—— (1976) 'Probabilities of Conditionals and Conditional Probabilities', *The Philosophical Review*, 85: 297-315.

—— (1979) 'Attitudes De Dicto and De Se', *The Philosophical Review*, 88/4: 513-43.

—— (1981) 'Ordering Semantics and Premise Semantics for Counterfactuals', *Journal of Philosophical Logic*, 10/2: 217-34.

Roskies, A. (2008) 'Robustness and the New Riddle Revived', *Ratio* (n. s.) 21: 218-30.

Schramm, A. (2014) 'Evidence, Hypothesis, and Grue', *Erkenntnis*, 79: 571-91.

Schwarz, W. (2018) 'Subjunctive Conditional Probability', *Journal of Philosophical Logic*, 47: 47-66.