Securing Reliable Evidence

Kent W. Staley

Saint Louis University

October 23, 2008

Abstract: Evidence claims depend on fallible assumptions. Three strategies for making true evidence claims in spite of this fallibility are *strengthening* the support for those assumptions, *weakening* conclusions, and using multiple independent tests to produce *robust* evidence. Reliability itself, understood in frequentist terms, does not explain the usefulness of all three strategies; robustness, in particular, sometimes functions in a way that is not well-characterized in terms of reliability. I argue that, in addition to reliability, the *security* of evidence claims is of epistemic value, where an evidence claim is secure relative to an epistemic situation if it remains true in all scenarios that are epistemically possible relative to that epistemic situation. [1]

[1] I am grateful for helpful comments on earlier drafts of this paper from Jan Sprenger, Joe Salerno, and audiences at Eastern Tennessee State University, University of Frankfurt and University of Wuppertal.

1 Three strategies

Assessments of evidence depend on substantive assumptions that function in an argument from evidence neither as part of the evidence itself, nor as the conclusion. Such assumptions may include claims about the sampling procedure, about the control of extraneous variables, or about the adequacy of a certain mathematical model for representing a substantive scientific claim, as well as theoretical claims about phenomena implicated in the experimental procedures, as in the case of 'theory-mediated' measurements.

These assumptions should not be confused with "mere assumptions." They can be supported by substantial reasons. However, these assumptions are fallible. Heinrich Hertz assumed that his cathode tubes were sufficiently evacuated to yield the deflection of cathode rays, supposing them to be electrically charged. He was mistaken. Galileo assumed that the combined orbital and rotational motion of the earth would producing a sloshing of the oceans that would be observed as tidal phenomena, and concluded that tidal phenomena constituted evidence for such motion. This was not the case. In 1984, Carlo Rubbia of the UA1 Collaboration at CERN believed that, in data collected from high energy proton collisions looking for a decay signature of the top quark, background processes had been sufficiently accounted for to regard the remaining excess of events as evidence of top quark production. A significant source of background had been overlooked, however. There was no evidence for the top quark until ten years later (Staley 2004a).

In describing these examples in this way, I am using a concept of evidence that resembles in several respects what Peter Achinstein calls "potential evidence" (Achinstein 2001, 27–28). First, the kind of evidence concept here invoked is not relativized, either to a person, or to an epistemic situation.[2] Second, these evidence claims are objective in that they are true or not independently of what anyone believes about the data, the hypothesis in question, or the relations between them.[3] Third, the question of whether E is evidence for H, according to this concept, can (and typically does) depend on facts beyond E, a feature that Achinstein denotes by saying that, in this sense, evidence is "potentially empirically incomplete."

An obvious strategy for coping with the fallibility of such evidence claims is to seek further support for any assumptions about which one is uncertain, and to use only those assumptions whose support is thus strengthened. Call this the *strengthening* strategy. Its strongest form would be to rely only on those assumptions for which one has *conclusive evidence*. But the effectiveness of the conclusiveness standard is significantly at odds with its practical value. If we give "conclusive evidence" a sufficiently strong reading to completely remove the threat of false background assumptions (supposing this even to be possible), we will rarely if ever be in a position to make any evidence claims. Furthermore, some of the assumptions on which we might wish to rely might be difficult to establish conclusively. The relevant data might, for example, be too expensive to acquire, or inaccessible with existing technology. Strengthening, therefore, while a valuable strategy, carries a cost, such that we might wish to supplement it with another,

[2]See section five for more on the notion of an epistemic situation.

[3]Here I am setting aside special cases, such as some forms of bias arising from "peeking" at data, in which investigators' beliefs may be *causally* relevant to the reliability of the test procedure used (see Staley 2004a, 278–87) .

potentially less costly strategy.[4]

A second strategy is to keep one's original assumptions, but to alter one's conclusion to a claim logically weaker than the original. The obvious term for this strategy is *weakening*. Instead of claiming that one has discovered evidence for, say, the positron (in the sense of the anti-electron of current theory), one might claim only evidence for a positively charged particle with mass on the order of that of the negatively charged electron.[5] This strategy also has an extreme version, which is to draw conclusions that say no more than the assumptions of which we are already certain.

A third strategy is to appeal to robustness considerations (Campbell and Fiske 1959; Culp 1995; Levins 1966; Staley 2004b; Trout 1993; Wimsatt 1981). In appealing to robustness, one copes with the potential for error by basing one's evidence claims on the convergent outcomes of multiple tests drawing, to some extent, upon independent assumptions. Call this the *robustness* strategy.

The aim of this paper is to articulate and defend a dimension of epistemic assessment that is appropriate to these strategies, particularly robustness. Such an effort is needed because, although we might assume (as I do) that some kind of requirement of *reliability*, in a broadly frequentist sense, of one's inferential procedure is central to the notion of empirical evidence,

[4]As a valuable implementation of the strengthening strategy, consider, for example, the program of mis-specification testing discussed by Mayo and Spanos (2004).

[5]In fact, in announcing the first evidence for the positron in 1933, Carl Anderson did claim to have evidence for the positron (thus coining the term as an abbreviation of "positive electron") but meant by this only to indicate a particle with the properties mentioned in this "weakened" claim.

there are reasons to suspect that reliability considerations are insufficient by themselves to explicate what is at stake in coping with the fallibility of background assumptions, at least insofar as these are addressed by the robustness strategy. A somewhat different perspective can put the strengthening, weakening, and robustness strategies into focus, so that it can be seen clearly how all three strategies contribute in different ways to the same epistemic goal.

In what follows, I will illustrate the methodological issues in question with a recent example of the robustness strategy from the search for dark matter (section two). That example will also serve as the basis for an argument to the effect that what is at issue in at least some uses of the robustness strategy cannot be captured by appealing to reliability considerations alone (section three). In section four, I introduce as a heuristic the notion of a space of epistemic possibilities, drawing upon recent work by David Chalmers. With the aid of that heuristic, I deploy an analogy between geographical space and epistemic space to motivate my definitions of secure evidence and secure inference (section five). Those definitions are followed in section six by general arguments for the epistemic relevance of security, and I conclude with some comments on the analysis of knowledge and on some similarities between security and logical probability.

2 An Example: Evidence for Dark Matter?

To make these preliminary considerations more concrete, consider a recent example of the robustness strategy, used in the context of the assertion of evidence for the existence of weakly interacting massive particles (WIMP's). The primary evidence claim in question is that data revealing an *annual modulation* of the detection rate in particle interactions with nuclei in scintillating sodium-iodide crystals is evidence that those crystals are immersed in a 'WIMP wind' due to Earth's movement through a WIMP halo that pervades the galaxy. That claim has been put forth by the DAMA-NaI group, operating deep underground at the Gran Sasso laboratory in Italy. DAMA's claim has been contested by the negative results of other groups using more conventional experimental methods (i.e., looking for *statistical excesses* in the detection rate beyond expectations from background). The dispute has been discussed in admirable detail in a thought-provoking article by Robert Hudson (2007a).[6] The present paper takes no stance regarding the dispute over DAMA's results; here I attempt only to understand the nature of one argument implicated in that dispute. Thus the present paper will attempt only to sketch very roughly the broad outlines of DAMA's analysis and result.

As Hudson explains, the initial positive result from DAMA was presented both in terms of an estimate of the WIMP mass (m_W) and interaction cross section ($\xi\sigma_p$), and in the form of a "contour" in the space of possible values of m_W and $\xi\sigma_p$.[7] DAMA's basic search strategy is to examine the

[6] I am grateful to Robert Hudson for bringing this example to my attention. I must note that in (Hudson 2007b) he employs the debate over DAMA's results more broadly to argue *against* the methodological value of robustness. I do not propose here to dispute the arguments of that paper, but only note that under Hudson's interpretation of the term, the argument here considered does not exemplify an appeal to robustness.

[7] More precisely, they use the quantity $\xi\sigma_p$ in (Bernabei et al. 1998; 1999; 2000), where

distribution of interaction events with regard to energy, location within the detector, and time, over a one-year data collection period. This information is summarized in the number N_{ijk}, where index i indicates the i-th day, j indicates the j-th detector, and k indicates the k-th energy interval or "bin", each event being recorded as falling into a bin of width 1 keV, from 2-20 keV.

The analysis of this data assumes a theoretical model in which the earth moves through a "halo" of WIMP dark matter with a variable velocity $v_r(t) = V_{Sun} + V_{Earth} \cos\gamma \cos\omega(t - t_0)$ in the galactic frame, where V_{Sun} is the Sun's velocity with respect to the halo, V_{Earth} is the Earth's orbital velocity about the sun, $\gamma = 60°$ is the angle of inclination of Earth's orbital plane with respect to the galactic plane, $\omega = 2\pi/T$ with $T = 1$ year, and $t_0 \simeq$ June 2$^{\text{nd}}$. This model is used then to derive a first order Taylor approximation for the signal rate in the k-th energy interval as $S_k = S_{0,k} + S_{m,k} \cos\omega(t - t_0)$. Consider this, with the parameters, respectively, for the unmodulated and modulated terms, $S_{0,k}, S_{m,k} \neq 0$ for at least some k as the theoretical model for a WIMP annual modulation.

To make the connection between this theoretical model and the results mentioned above, DAMA assumes a model of the data as being generated by a Poisson process with mean value $\mu_{ijk} = (b_{jk} + S_{0,k} + S_{m,k} \cos\omega(t_i -$

$\xi = \sigma_{\text{WIMP}}/\sigma_0$, $\sigma_0 = 0.3\ \text{GeV cm}^{-3}$, and σ_p is the WIMP interaction cross section on the proton. The results presented in (Belli et al. 2002) refer to the quantity $\xi\sigma_{scalar}^{(nucleon)}$, where ξ is defined as the "fractional amount of local nonbaryonic [Dark Matter] density which is ascribed to the WIMP responsible for the effect," and $\sigma_{scalar}^{(nucleon)}$ is the WIMP-nucleon scalar interaction cross section. The following discussion uses $\xi\sigma_p$ for simplicity, though strictly speaking this slightly misrepresents the latter results.

$t_0))M_j \mathit{\Delta} t_i \mathit{\Delta} E \epsilon_{jk}$. Here b_{jk} represents a time independent background, M_j is the mass of the j-th detector, $\mathit{\Delta} t_i$ is the actual running time for the detector on the i-th day, $\mathit{\Delta} E = 1$ keV represents the width of the energy intervals, and ϵ_{jk} is the analysis cut efficiency. A time correlation analysis is then employed that uses a maximum likelihood method to produce an estimate of the WIMP mass and interaction cross section. The likelihood function $\mathscr{L}$ can be written in terms of μ_{ijk} and N_{ijk}, and the maximum likelihood can be determined by minimizing the function $y = -2\ln(\mathscr{L}) - \text{const}$. Such an analysis, carried out on their first year of data, yields the estimates $M_W = (59^{+36}_{-19})$ GeV and $\xi\sigma_p = (1.0^{+0.1}_{-0.4})\, 10^{-5}$ pb. In addition, by taking the ratio of likelihoods between the hypothesis H_1 of annual modulation, with specific values of m_W and $\xi\sigma_p$, and the hypothesis H_0 of no annual modulation as a test statistic λ, they generate a plot of the region in $m_W - \xi\sigma_p$ space where H_1 is favored over H_0 at a 90% confidence level.[8] This latter analysis is the source of the contour mentioned above.

The 2002 paper by Belli et al. that features a robustness analysis advertises itself as an extension of the "previous analyses" (those using the just described methods of data analysis as applied to one (Bernabei et al. 1998), two (Bernabei et al. 1999), and four (Bernabei et al. 2000) years worth of data) "by discussing in detail the implications of the results of the uncertainties on the dark matter galactic velocity distribution" (Belli et al. 2002). More specifically, those earlier analyses adopted the "standard" isothermal sphere model of the WIMP galactic halo. Belli et al. note that, in spite of its simplicity, a number of the assumptions of that model "are

[8] They use the quantity -2lnλ as a χ^2 statistic.

not strongly constrained by astrophysical observations" (ibid., 2). Moreover, the expected rate of WIMP interactions is determined in part by the distribution function for WIMPs in their six-dimensional position-velocity phase space, a function that in turn depends on on the model of the galactic halo. That expected rate, naturally, has in turn consequences for the likelihood functions on which the earlier analyses depended. Thus, they seek to "study in a systematic way possible departures from the isothermal sphere model . . . specifically . . . modifications arising from the various matter density profiles, effects due to anisotropies of the velocity dispersion tensor and rotation of the galactic halo" (ibid., 1). The paper proceeds to examine four general classes of galactic halo models: spherically symmetric matter density with isotropic velocity dispersion, spherically symmetric matter density with nonisotropic velocity dispersion, axisymmetric models, and triaxial models. Contours are presented for a variety of versions of each class of models. The details need not further detain us. Noteworthy, however, is the conclusion drawn: "The hypothesis of WIMP annual modulation, already favored in the previous studies by using an isothermal sphere, is confirmed in all the investigated scenarios, and the effects of the different halo models on the determination of the allowed maximum likelihood region in the WIMP mass and WIMP-nucleon cross section have been derived" (ibid., 16).

3 Reliability, Robustness, and Security

I claimed above that there are reasons to doubt that frequentist reliability considerations alone are sufficient to account for the epistemic value of the robustness strategy as employed in examples like that just described. In this section I defend that claim.

The DAMA group offers their robustness analysis of their evidence claim in order to explain "the implications on [their previous results] of the uncertainties on the dark matter galactic velocity distribution" (ibid., 1). That is, they had previously employed a model of the galactic dark matter halo that, supposing there is such a halo, might not correctly describe it.

Might the correct way to understand their analysis be that they are attempting to demonstrate or enhance the reliability of their results, by employing a procedure that restricts the probability of arriving at an erroneous result? For example, we might attempt to view them as applying a *severe-test requirement*, in Mayo's sense (Mayo 1996) to their earlier evidence claim. That requirement can be framed as follows: Suppose that hypothesis H is subjected to test procedure T, resulting in data E; then H's passing T with E constitutes the passing of a severe test (and hence evidence for H) just in case E fits H, and the probability of H passing T with an outcome such as E (i.e., one that fits H at least as well as E does), given that H is false, is very low (ibid., esp. 178–87).

It is arguable, even plausible, that one could regard DAMA's original evidence claim as resting on the satisfaction of the severe test requirement. Indeed, that some frequentist reliability considerations enter into that argument is ensured by the use of the confidence level construction methodology, the rationale for which lies precisely in the long-run error characteristics ensured by the appropriate use of a test statistic that follows a χ^2 distribution. However, the robustness argument offered by Belli et al. eludes such a characterization in terms of severity or even some more general frequentist notion of reliability. The paper seeks to address an uncertainty regarding an assumption that is used in defending the reliability of their original inference (from the annual modulation data to the contour described above). However, the argument is obviously not meant to give evidence that the original assumption regarding the galactic halo is true, since the paper discusses other possible models and makes no effort to argue against them.

Perhaps more plausibly, we might consider the possibility that DAMA is here attempting to give what Staley has called *second-order evidence* (2004b): they are giving evidence, based on the agreement between the contours generated by different galactic halo model assumptions, that the annual modulation data really are evidence for the existence of WIMP dark-matter. On the severe-testing account, this would require showing, in effect, that, assuming the annual modulation data is not evidence for a WIMP annual modulation, there is a very low probability these different analyses would yield contours that agree as well as these do. However, nowhere in the paper presenting this analysis can such an argument be found. Indeed, it is difficult to see how such an argument *could* be made within the domain of frequentist statistics. To do so would require answering the difficult question: On what would the error rates of such a test depend? Or, more precisely, how would one model the scenario "the annual modulation data are not evidence for a WIMP annual modulation" so as to be able to estimate such error rates, even qualitatively?[9]

[9] I am not claiming that the convergence of results from independent tests *never* constitute a severe test of any hypothesis, but only that some uses of such convergence directed

Rather than awkwardly trying to force this kind of argument into a frequency-reliability framework, I propose that we attend more closely to the kind of problem that such robustness arguments seek to address and what is distinctive about how they address that problem. Specifically, robustness arguments are a response to uncertainty regarding assumptions, and they respond, not by removing that uncertainty (as in the strengthening strategy), but by showing how evidence claims remain valid in spite of that uncertainty. In the next section, I begin the articulation of a framework for making sense of such a strategy.

4 Possibility Spaces

In order to motivate the construction that follows, I will employ as a heuristic the idea of an epistemic space, drawing upon some recent work on epistemic possibility by David Chalmers (2008).[10]

When an investigator puts forth an empirical evidence claim, she does so on the basis of a number of other claims. Typically, some of these are claims that the investigator knows to be true, while other claims are relied upon without being known to be true. (Some claims may even be used that are known to be false, as when errors are intentionally over-estimated in the interest of being conservative in one's conclusions.)

The idea of a space of epistemic possibility gives us an intuitive way

at addressing uncertainties regarding model assumptions cannot be thus interpreted.

[10]For Chalmers, epistemic space is not merely a heuristic, but part of a theory of the semantics of statements of epistemic possibility. That theory may be correct, but need not be for my purposes.

of depicting this situation. The rough picture is that of a space of scenarios that might *for all we know* be actual. (This notion of epistemic possibility, introduced by Hintikka in his (1962) is to be contrasted with the idea of subjunctive possibility, typically described in terms of possible worlds as ways the world might have been or would be if certain counterfactual conditions were to obtain.) This notion has been explicated in different ways – for example, as a contextual notion by DeRose (1991), and non-contextually by Chalmers (2008). As far as I can tell, nothing in the discussion that follows turns on just which analysis we use.

In Chalmers' discussion, a picture of the space of epistemic possibility emerges that, independently of the details of his analysis, captures the right features for the present argument. Chalmers distinguishes what he calls "deep epistemic possibility" from "strict epistemic possibility," where the former is meant to capture "ways the world might be, prior to what anyone knows," and the latter is meant to capture ways the world might be, relative to a particular body of knowledge. While exactly what scenarios are deeply epistemically possible will no doubt be a matter of debate, what is strictly epistemically possible, relative to any body of knowledge K that includes more than what is knowable a priori, will be some subspace of the space of what is deeply epistemically possible. Furthermore, as knowledge is gained, more scenarios are ruled out, and the space of what is strictly epistemically possible shrinks. To state it with a little more precision, though still informally: If (i) Ω is the space of epistemically possible scenarios relative to a body of knowledge K, (ii) Ω' is the space of epistemically possible scenarios relative to K', and (iii) $K \subset K'$, then $\Omega' \subset \Omega$.[11]

5 Security: An analogy and definitions

Using this picture of the space of epistemic possibility as a heuristic, we can depict the situation of the investigator who seeks to make an evidence claim in terms of the investigator not knowing 'where she is' in the space of epistemic possibilities relative to her knowledge. Of course, the assumptions she uses in advancing her evidence claim that she *knows* to be true, will be true throughout that space, but other claims that she uses may be true in some regions of epistemic possibility, and false in others. What is more, the evidence claim itself may or may not be true throughout the entire range of epistemic possiblities.

This raises the question of security. To state it roughly: the investigator has good reason to consider *whether it is possible* that her evidence claim is false. If it is possible, what is the range of possible scenarios in which it is false, and can steps be taken to eliminate some of those possibilities?

[11]Note that at this point all reference to deep epistemic space has dropped out. Henceforth, I will only make use of the idea of a space of epistemic possibilities that is relative to an epistemic situation (taken to include a body of knowledge). That there is a unique space of scenarios that corresponds to "ways the world might be, prior to what anyone knows," is far from obvious, and is in any case not necessary for the view defended here. Thus, all uses of 'epistemic possibility' and its cognates should henceforth be understood in the sense of 'strict epistemic possibility'.

5.1 An analogy: Where am I?

Before introducing a more precise definition of security, I introduce an analogy. The point of this analogy is to make intuitive both the epistemic relevance of security considerations and their distinctness from reliability. It is reasonable for epistemic agents to consider security as well as reliability in drawing conclusions. Furthermore, although judgments about security may be *about* reliability, on the present account, in the sense that the assumptions at issue in assessing the security of an evidence claim might be reliability assumptions, security judgments are not judgments *of* reliability. In order to clarify how reliability and security are related, consider the following analogy between geographical space and the space of epistemic possibility.

Let us suppose that Charlotte is dropped onto the surface of the earth in an amphibious vehicle that gives her no visual access to her surroundings, but over which she has directional control once she is on the surface of the earth (the vehicle automatically navigates around or over objects so as not to collide with them we'll suppose). Charlotte's objective is to determine in which approximate direction is the geographical North Pole, and she is offered a choice between two instruments for use in that task.

Imagine two versions of this story. In the first version, the instruments between which Charlotte must choose are a compass and an AM radio. In the second version, her choice is between a compass and a GPS device.

Consider the first version. You might be wondering what good the AM radio might be. If Charlotte can, by moving around, figure out where the radio stations she is picking up are located, and correlate that with a direction that takes her in the direction of more northerly locations, then she can use the AM radio to determine the approximate direction of the North Pole. However, this will only work if she starts out in a place where she understands the language, or at least can pick out place names when they are uttered on the radio, and she knows which of those named places are further north than others. So the AM radio will help her, but only over a rather limited range of the Earth's surface (it would not, presumably, work over most of the oceans, for example).

A compass is commonly understood to be a reliable indicator of the approximate direction of the North Pole. The compass needle, however, really responds to Earth's magnetic field, and thus points not to the geographical North Pole, but to the Magnetic North.[12] Consequently, the compass also will fail, if Charlotte happens to start out anywhere very close to the magnetic North Pole. However, the region in which it will fail is smaller than the region in which the AM radio will fail. Thus, even if the use of the AM radio to determine direction were not unwieldy, complicated, and somewhat inaccurate, Charlotte would still have reason to prefer the compass.

The decision in the second version seems similarly easy: The GPS device will presumably give accurate information about directions everywhere that the compass will, but also in locations near Magnetic North where the compass will fail, making it the preferable choice.

Suppose, now, that I stipulate that, within the domains in which

[12]In fact, the direction of the compass needle reacts to the local magnetic field, and typically does not point exactly to the magnetic North. More accurate use of the compass requires knowledge of the local magnetic declination, given on some maps.

they function *correctly*, each device works equally *reliably*, in the sense that, provided you are within that region in which a given device is usable (within reach of a number of AM stations with identifiable locations, not too close to magnetic north, etc.), that device will not point you in the wrong direction any more *often* than the other devices will in their regions of usefulness. Our judgments of the relative utility of the three devices, I submit, will be unchanged. This is because the reasoning behind the original preferences was not based on how often the devices would yield errors, but on the geographical range of their usefulness.

To put the point another way: Charlotte has a question, and an instrument (the compass, let's suppose) that can potentially be used to answer that question. Is the direction of the compass needle evidence for the approximate direction of the geographic North? Whether that is so depends on where Charlotte is on the surface of the earth, which she does not know. Since, for all she knows, she could be anywhere on the earth, given two possible sources of evidence regarding the direction of geographic North, all else being equal, she has good reason to prefer the instrument that is reliable over a larger geographic region.

Just as Charlotte does not know where she in geographical space, we do not know where we are in the space of epistemic possibilities. I take this to be true by definition of epistemic possibility. This space simply is, for a given person, the space of all the ways that things might be, for all that person knows.

5.2 Security defined

Thus far I have attempted to argue that there is a category of epistemic appraisal in addition to reliability that is appropriate to consider in the evaluation of evidence. I have appealed to the notion of epistemic possibility and the heuristic notion of a space of epistemic possibilities to characterize this category at an intuitive level. In the interests of engendering theoretical elaboration and methodological implementation of this claim, I next postulate a set of related working definitions for security. My claims for these working definitions at this stage are only that they elaborate the intuitive considerations just discussed and that they may prove a useful starting point for more systematic work to come.

Definition 1 (secure evidence) *Let an* evidence claim *be a claim of the form 'Data E are evidence for the hypothesis that H.' Suppose that Ω_0 is the set of all epistemically possible scenarios relative to epistemic situation K, and $\Omega_1 \subseteq \Omega_0$. An evidence claim C is* secure throughout Ω_1 *relative to K iff for any scenario $\omega \in \Omega_1$, C is true. If C is secure throughout Ω_0 then it is* fully secure.

Definition 2 (secure inference) *An inference[13] made from E to H is*

[13] To make an inference from E to H is to treat E, which is taken to be true, as a good reason for endorsing H as true. I use the notion of 'endorsement' here rather than 'belief' (cp. Achinstein 2001; Staley 2008) so as to indicate that it is the practical dimension of science that is of interest here, as opposed to the mental states of scientists. *Endorsement* here generalizes the notion of *assertion* in such a way as to include such qualified statements of positive epistemic appraisal, as 'H is very likely to be true' or 'it is reasonably certain that H is true.'

secure throughout Ω_1 *relative to K iff the claim 'E is evidence for H' is secure throughout* Ω_1 *relative to K.*

Before proceeding to argue for the epistemic value of security more directly, a few comments on the notion of an epistemic situation and the issue of relativization are in order. The notion of an epistemic situation is borrowed from Achinstein (2001), who describes an epistemic situation as a situation in which "among other things, one knows or believes that certain propositions are true, one is not in a position to know or believe that others are, and one knows (or does not know) how to reason from the former to the hypothesis" (ibid., 20). To this I would add as components of the epistemic situation that one knows (or does not know) how to do things (such as the manipulation of data or instruments, or the performance of speech acts) that facilitate the inference from data and other propositions to the hypothesis of interest.

It will be noted that the idea of an epistemic situation is much richer and multidimensional than the idea of a body of knowledge representable as a set of sentences. I have used the epistemic situation idea in defining security because it does seem that what is epistemically possible for me should indeed depend not only on my propositional knowledge, but also my inferential abilities and other forms of know-how. In DeRose's discussion of epistemic possibility (1991), such considerations do have a place, but only as 'contextual' factors in determining the epistemic possibility of propositions, and not as constitutive elements in the space of epistemic possibility. They have no apparent place in Chalmers' notion of epistemic space (Chalmers 2008).

Finally, it should be noted that, although security of evidence is here treated as relative to an epistemic situation, the truth of an evidence claim is not. Thus the evidence concept here should not be confused with what Achinstein has called "E-S evidence," understood as evidence that is relative to an epistemic situation. As Achinstein emphasizes, although E-S evidence is among the evidence concepts used in scientific contexts, other more important evidence concepts are not thus relativized. It is the latter, non-relativized evidence concept that is employed here. In short, the actual truth of evidence claims is independent of epistemic situation, but security is relative to epistemic situation, because whether an evidence claim is epistemically possibly false is relative to an epistemic situation.

6 Arguments for security

Although the analogy with geography in the Charlotte example should make the epistemic relevance of security plausible, a more systematic argument that employs the definitions just given would make a stronger case. I next present two such arguments, both of which assume that investigators seeking to make evidence claims have certain (plausibly very widely held) epistemic aims. The first argument shows that, all else being equal, such an investigator should prefer to make a given evidence claim under conditions in which it is fully secure to making it under conditions in which it is not fully secure. In the second argument, I show that, all else being equal, an investigator should prefer to make a given evidence claim when it is *more* secure (e.g., in virtue of the application of a robustness strategy) to making that claim when it is less secure (i.e., without having applied the robustness strategy).

Both arguments use the following suppositions: Suppose that S is a person engaged in empirical inquiry with regard to some question Q, to which H is a hypothetical answer. E_1 and E_2 represent two possible bodies of evidence relevant to H. C_1 is the claim: "E_1 is evidence for H." C_2 is the claim: "E_2 is evidence for H." K represents S's epistemic situation at a particular time.

In both arguments, there is also an assumption about S's aims: S seeks to make evidence claims that will not be refuted by subsequent inquiry.

Now for the first argument: Suppose that C_1 is fully secure relative to K. It follows from definition 1 that for all epistemically possible changes in K, C_1 is true. Suppose furthermore that C_2 is not fully secure relative to K. From the same definition it then follows that for some epistemically possible changes in K, C_2 is false. Hence if S makes claim C_1, she makes an evidence claim that, given what she knows, will remain true. On the other hand, if she makes claim C_2, then for all she knows her evidence claim might be false and hence be refuted by subsequent inquiry. So S should prefer to make the fully secure claim C_1 rather than the insecure claim C_2.

The more interesting and realistic case involves partial security. Suppose that Ω_0 is the set of all scenarios epistemically possible relative to K. Suppose that C_1 is secure throughout Ω_1 and C_2 is secure throughout Ω_2, where $\Omega_2 \subset \Omega_1 \subset \Omega_0$. Since both Ω_1 and Ω_2 are strict subsets of the set of all epistemically possible scenarios, if S makes either claim, then her claim might, for all she knows, turn out to be false. But there is an asymmetry between the potential failures of the two claims. For any epistemically possible scenario in which C_1 fails, C_2 also fails. But there are some epistemically-possible scenarios in which C_2 fails, but C_1 does not. Thus the scenarios in which C_1 continues to be upheld as true include all those in which C_2 continues to be upheld as true, as well as some of those in which C_2 is discovered to be false. Since S aims to avoid making evidence claims that are discovered to be false, she should prefer to make claim C_1 rather than C_2.

It will be noted that in this last case, the secure regions for claims C_1 and C_2 were nested, and one might wonder what might be said about cases in which this is not so. For example Ω_1 and Ω_2 might overlap without either being contained in the other, or they might be entirely disjoint. (Charlotte's situation could be a case of overlap, for example, if the AM radio method were actually usable by her in regions in the Arctic Circle where the compass would fail.) It would be tempting here to say that one should then consider the size of the secure regions of the two claims in deciding what claim would be preferable. Indeed, this seems to be the right answer, provided that the crucial condition "all else being equal" were understood in the sense of "there is no more reason to think that one is in any one region of the space of epistemic possibility than in any other." However, this latter condition is almost never satisfied. Although one is aware in an evidential inference that there is a range of possibilities of error, some of these possibilities will be more worrisome than others. Clearly, some means of weighing the relevance of scenarios is needed. Although I attempt no solution to this problem in the present paper, I will comment on it in the concluding remarks.

7 Conclusion: Dark Matter Revisited and Some Challenges

With the concept of security in hand, we can now revisit the robustness strategy and see more clearly how it works to promote security of evidence. Recall the robustness argument regarding DAMA's evidence claim as put forth by Belli et al. The problem confronted by the argument is that, for all that anyone knows, the theoretical model of the WIMP galactic halo used by DAMA in their original analysis, the isothermal sphere model, might be false. So what happens if some other model is the correct one? Belli et al. examine four broad classes of WIMP galactic halo models, any of which might, for all anyone knows, be the correct one, and they find that the WIMP annual modulation hypothesis is "confirmed in all the investigated scenarios" (Belli et al., 16). In other words, even if the isothermal sphere model is false, so long as the correct model falls into one of the investigated classes, DAMA's annual modulation data are still evidence for WIMPs.

The effect of this argument, then, is to show, in a way that DAMA's original analysis did not, the extent to which the evidence for WIMP annual modulation is secure. Note that the argument does not seek to establish that the evidence is *fully* secure. To do so would require at least establishing that the WIMP annual modulation hypothesis is confirmed in all possible theoretical models, and Belli et al. do not attempt this.

Finally, let me address one worry and point to one interesting connection. The worry has to do with the analysis of knowledge. The connection is with logical probabilities.

The emphasis on *knowledge* in the above discussion might prompt the worry that our grasp of the notion of security is going to prove as elusive as has the correct analysis of knowledge in the interminable debates among epistemology's various warring camps. That would indeed be a troubling outcome, but happily this worry can be easily set aside. The real value of the concept of security is methodological, and the methodological implementation of the idea is largely independent of the correct analysis of knowledge. Methodologically speaking, what is important is not primarily "how secure is this inference?" but "how do I make this inference more secure than it might otherwise be?" Since the latter question only involves judgments of relative security, it does not require that one be able, at any time, to identify what one knows, but only to apply the strategies that enhance security. The point can be stated succinctly: The robustness strategy and weakening strategy enhance security under any plausible analysis of knowledge. The same holds for the strengthening strategy, under the additional assumption that strengthening adds to our knowledge. (If an analysis of knowledge leads to complete skepticism, then strengthening will not add to our knowledge, thus cannot restrict the space of epistemic possibilities, and hence cannot enhance security.)

Finally, it may have been noted there is at least a family resemblance between security and logical probability. Just as an evidence claim is secure to the extent that it is true over the space of what is possible relative to an epistemic situation, logical probability has often been framed in terms of the satisfaction of a formula by a class of models consistent with a certain body of background knowledge. Why not, a logical probabilist might ask, dispense with frequency-reliabilism (and frequency statistics) entirely (or at least pursue a two-probability approach like Carnap 1962) and get at the relevant desiderata more directly within a logical probability framework? Perhaps there, like in Dr. Seuss's land of Solla Sollew, one will find that "they never have troubles, or at least very few."

This challenge raises problems that cannot be satisfactorily addressed in a brief discussion. Let me here simply respond by articulating the stance here adopted. The evidence concept at issue here constitutes a kind of generalization of the idea of a "reliable indicator," (such as a compass!) and thus to rest crucially on considerations of diachronic performance with regard to errors. The severe test requirements amount to a codification of such concerns, and although conceptual difficulties and debates plague every philosophical approach to the foundations of probability, frequency-probabilities have greatly facilitated the articulation of methodologies for assessing evidence in this sense.

So, although one might use a logical probability approach to articulate the concerns here subsumed under the concept of security, for the purposes of the present account, this would still need to be accompanied by some kind of frequentist approach to reliable evidence such as the error-statistical account. Thus one inherits all the challenges facing logical probabilities in addition to those that the error-statisician must confront. Fortunately, such a multiplication of troubles (like those encountered on the way to Solla Sollew) is not necessary for what I have sought here – namely, a unified understanding of how strengthening, weakening, and robustness strategies enable scientists to confront uncertainties regarding the assumptions underlying their primary evidence claims.

REFERENCES

Achinstein, Peter (2001), *The Book of Evidence*. New York: Oxford University Press.

Belli, P., R. Cerulli, N. Fornengo, and S. Scopel (2002), "Effect of the Galactic Halo Modeling on the DAMA-NaI Annual Modulation Result: An Extended Analysis of the Data for Weakly Interacting Massive Particles with a Purely Spin-Independent Coupling," *Physical Review D* 66: 043503.

Bernabei, R., P. Belli, F. Montecchia, et al. (1998), "Searching for WIMPs by the Annual Modulation Signature," *Physics Letters B* 424: 195-201.

Bernabei, R., P. Belli, F. Montecchia, et al. (1999), "On a Further Search for a Yearly Modulation of the Rate in Particle Dark Matter Direct Search," *Physics Letters B* 450: 448–55.

Bernabei, R., P. Belli, R. Cerulli, et al. (2000), "Search for WIMP Annual Modulation Signature: Results from DAMA/NaI-3 and DAMA/NaI-4 and the Global Combined Analysis," *Physics Letters B* 480: 23–31.

Campbell, Donald T. and Donald W. Fiske (1959), "Convergent and Discriminant Validation in the Multitrait-Multimethod Matrix," *Psychological Bulletin* 56: 81–105.

Carnap, Rudolf (1962), *The Logical Foundations of Probability*, second edition. Chicago: University of Chicago Press.

Chalmers, David (2008), "The Nature of Epistemic Space," URL (cited on

June 26, 2008): http://consc.net/papers/espace.html.

Culp, Sylvia (1995), "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles," *Philosophy of Science* 62: 430–50.

DeRose, Keith (1991), "Epistemic Possibilities," *The Philosophical Review* 100: 581–605.

Hintikka, Jaakko (1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press.

Hudson, Robert G. (2007), "Annual Modulation Experiments, Galactic Models and WIMPs," *Studies in History and Philosophy of Modern Physics* 38: 97–119.

Hudson, Robert G. (2008), "Robustness vs. Model-Independence." Paper presented at LSE/Pittsburgh Conference on Confirmation, Induction, and Science, March 2007. Available at http://philsci-archive.pitt.edu/archive/00003209/ (accessed on September 26, 2008).

Levins, Richard (1996), "The Strategy of Model Building in Population Biology." *American Scientist* 54: 421–31.

Mayo, Deborah (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Mayo, Deborah and Aris Spanos (2004), "Methodology in Practice: Statistical Misspecification Testing," *Philosophy of Science* 71: 1007–1025.

Staley, Kent (2004a), *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation*. New York: Cambridge University Press.

Staley, Kent. (2004b) "Robust Evidence and Secure Evidence Claims." *Phi-*

losophy of Science 71: 467–88.

Staley, Kent (2008), "Error-Statistical Elimination of Alternative Hypotheses," *Synthese* 163: 397–408.

Trout, J. D. (1993) "Robustness and Integrative Survival in Significance Testing: The World's Contribution to Rationality," *British Journal for the Philosophy of Science* 44: 1–15.

Wimsatt, William. "Robustness, Reliability, and Overdetermination," in Marilynn B. Brewer and Barry E. Collins (eds.), *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass, pp. 124–63.