Unfalsifiability of security claims

Cormac Herley[a,1]

[a]Microsoft Research, Redmond, WA 98052

Edited by Moshe Y. Vardi, Rice University, Houston, TX, and approved April 21, 2016 (received for review September 9, 2015)

There is an inherent asymmetry in computer security: Things can be declared insecure by observation, but not the reverse. There is no observation that allows us to declare an arbitrary system or technique secure. We show that this implies that claims of necessary conditions for security (and sufficient conditions for insecurity) are unfalsifiable. This in turn implies an asymmetry in self-correction: Whereas the claim that countermeasures are sufficient is always subject to correction, the claim that they are necessary is not. Thus, the response to new information can only be to ratchet upward: Newly observed or speculated attack capabilities can argue a countermeasure in, but no possible observation argues one out. Further, when justifications are unfalsifiable, deciding the relative importance of defensive measures reduces to a subjective comparison of assumptions. Relying on such claims is the source of two problems: once we go wrong we stay wrong and errors accumulate, and we have no systematic way to rank or prioritize measures.

security | falsifiable | passwords | self-correction

> A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.
>
> K. Popper, *Conjectures and Refutations* (1)

Declaring anything to be "secure" is a risky proposition. This is true independently of how (and whether) the term is defined. The Snowden disclosures and the steady stream of breaches at major institutions make clear that things that have been used for years without incident can turn out to have major flaws (2). Systems with no known vulnerability might be secure, or it may simply be that no vulnerability has been found yet. Thus, although things can often be declared insecure by observing a failure, there is no empirical test that allows us to label an arbitrary system (or technique) secure.

Hence, claims of insecurity are impossible to prove wrong empirically: No observable outcome proves a thing secure. Therein, however, lies the problem; irrefutability of empirical claims is not a strength, but a weakness. If we have no test for security, then statements that any set of things or behaviors is insecure are unfalsifiable. It follows that any claim that a condition is necessary for security (i.e., that everything that does not meet the condition is insecure) is also unfalsifiable, as are sufficient conditions for insecurity. This problem is inherent because attainment of the goal (the avoidance of certain outcomes) is unobservable (because it occurs at an unspecified point in the future). Thus, tweaking our definition of security does not help unless we strip it of reference to the future (which would seem to defeat the purpose).

Much in computer security involves recommending defensive measures; i.e., making statements of the form: "You should do X." A defender may end up with very many such measures (e.g., an Internet user will have dozens of instructions about how to choose and handle passwords, etc). We show that attempts to justify defensive measures using statements of the form "if you don't do X then you are not secure" or "security is improved if you do X" are unfalsifiable for all X. Thus, the inherent asymmetry noted in security means that self-correction operates only in one direction: Whereas acceptance of measures can always be justified based on new information, there is no mechanism whatsoever for rejecting them. Further, if justifications are unfalsifiable, then deciding the relative importance of defensive measures reduces to subjective assessment of different assumptions. Thus, there is no system for detecting or dealing with an accumulation of wasteful, redundant, or outdated measures, and no system for ordering them by importance.

The remainder of this paper is structured as follows. In *Claims of Necessary Conditions for Security*, we show that necessary claims to avoid bad outcomes are unfalsifiable, either by induction or deduction. We then examine three alternative definitions: security by design goals, security as proving the impossibility of bad outcomes, and claims of improved security (i.e., as a nonbinary quality) and show that all of them share the same problem. The discussion examines some of the consequences of unfalsifiability and gives examples.

Claims of Necessary Conditions for Security

Suppose x is a particular system, technique, or object that we use to protect an asset from compromise. For example, the asset might be an online banking account and x the associated password, or the asset might be a computer and x the software configured to protect it. We want to explore the range of values that x can take while protecting the asset. Define the set $\mathbf{Y}$:

$$x \in \begin{cases} \mathbf{Y} & \text{if bad outcomes will be avoided,} \\ \overline{\mathbf{Y}} & \text{otherwise.} \end{cases} \quad [1]$$

We wish to explore to what degree we can reason about $\mathbf{Y}$. Surprisingly, even without committing to what a bad outcome involves, we will be able to find significant restrictions on the claims we can make about $\mathbf{Y}$. We merely assume that we recognize a bad outcome when it occurs (if not, we are arguing about unobservable phenomena and all statements about outcomes are unfalsifiable). This does not require access to x; e.g., we do not need to know anything about the password to determine whether a bad outcome has occurred. In the particular example above, $\mathbf{Y}$ would be the space of passwords which protect the account from

Significance

Much in computer security involves recommending defensive measures: telling people how they should choose and maintain passwords, manage their computers, and so on. We show that claims that any measure is necessary for security are empirically unfalsifiable. That is, no possible observation contradicts a claim of the form "if you don't do X you are not secure." This means that self-correction operates only in one direction. If we are wrong about a measure being sufficient, a successful attack will demonstrate that fact, but if we are wrong about necessity, no possible observation reveals the error. The fact that claims of necessity are easy to make, but impossible to refute, makes waste inevitable and cumulative.

Author contributions: C.H. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Email: cormac@microsoft.com.

bad outcomes, and $\overline{\mathbf{Y}}$ those which do not. We will refer to examining the outcome as "observation."

We can reason about **Y** using induction and deduction. Induction involves generalizing from many observations to infer general properties or rules (e.g., classes of x lie in **Y** or $\overline{\mathbf{Y}}$). Deduction involves proving some property starting from axioms or assumptions.

Necessary Claims Are Unfalsifiable by Observation. The desire to protect our asset stretches into the future. However, there is an unavoidable limitation about statements about the future.

Claim 1: Unless an interval is specified, a claim that an event will occur is verifiable, but cannot be falsified.

The proof is immediate. For example, the claim that a 6-character password for a bank account will be guessed is unfalsifiable, because no amount of event-free use rules out the possibility that a bad outcome has simply not happened yet.

Obviously, if a bad outcome is observed, then we can say that $x \in \overline{\mathbf{Y}}$. However, if a bad outcome is not observed, we cannot say $x \in \mathbf{Y}$. This asymmetry is inherent to any claim that something will occur (unless we commit to a time interval). Even if the predicted event is no more precise than "bad outcome," the claim that it will happen is unfalsifiable.

Definition 1: A set is untestable if we cannot ever observe that something is a member.

Clearly, **Y**, as defined in Eq. **1** is an untestable set. We assert that, to be interesting, a statement about protecting assets must reference the future, which makes the inability to falsify unavoidable: The set of x that will protect the asset is untestable.

We are interested not just in single observations, but in the ability to infer general claims from observations. For example, if we observe that many elements x that have a particular property lie in $\overline{\mathbf{Y}}$, we might conjecture that all elements with that property lie in $\overline{\mathbf{Y}}$ (this would be a sufficient condition for $\overline{\mathbf{Y}}$). Or, we might conjecture that all elements in $\overline{\mathbf{Y}}$ have the particular property (this would be a necessary condition for $\overline{\mathbf{Y}}$).

Consider now a claim of a necessary condition for an untestable set **Y**:

$$X \text{ is necessary for } Y \left(\text{i.e.}, \mathbf{X} \supset \mathbf{Y} \equiv \overline{\mathbf{X}} \Rightarrow \overline{\mathbf{Y}}\right). \qquad [2]$$

A consequence of untestability is immediate.

Claim 2: No possible observation falsifies a claim of a necessary condition for an untestable set.

Proof: Let the untestable set be **Y** and the claim be $\mathbf{X} \supset \mathbf{Y}$. Refuting the claim requires finding a member of **Y** that is not in **X**, i.e., showing that $\overline{\mathbf{X}} \cap \mathbf{Y}$ is nonempty. Because we can never observe $x \in \mathbf{Y}$, this is impossible. □

Thus, claims of necessary conditions for membership of an untestable set, such as $\mathbf{Y}_x \triangleq \{$"Passwords secure against guessing"$\}$ are unfalsifiable. This is true without committing to a particular definition of security: The inability to observe membership is the only assumption, and that clearly holds if "being secure" rules out certain future events without specifying a time interval. Note that sufficient conditions for $\overline{\mathbf{Y}}$ are equivalent to necessary conditions for **Y** and are thus also unfalsifiable.

Note, even if we cannot observe membership of **Y** we might be able to assume it. For example, we might say that $\mathbf{Y}_{40} \triangleq \{$"Random passwords of length $>$ 40 characters"$\}$ is secure against guessing. First, this is not an observation, but a deduction from an assumption about attacker limitations. Second, if $\mathbf{Y}_{40} \subset \mathbf{Y}$, this fact is of no help falsifying $\mathbf{X} \supset \mathbf{Y}$. This is so because if $\mathbf{Y}_{40} \subset \mathbf{X}$, then no member of $\mathbf{Y}_{40}$ can help show that $\overline{\mathbf{X}} \cap \mathbf{Y}$ is nonempty. Alternatively, if $\mathbf{Y}_{40} \not\subset \mathbf{X}$, then the claimed necessary condition is impossible, because there are elements of **Y** not in **X**. This is similar for any other assumed subset of **Y**.

Deductive Claims Say Nothing About Outcomes. Falsification is a standard often applied to claims that we try to establish by induction, i.e., those where we use multiple observations to infer a general claim (1, 3). The alternative approach is deduction, where we prove claims true starting from axioms or premises. We now explore under what circumstances a necessary claim can be proved true. If this can be done then the inability to falsify need not trouble us.

To derive a necessary condition for **Y** we must show that it follows by deduction from the premises (which are axioms or assumptions). However, deductive claims never have greater generality than is contained in the premises. Expressed in set terms: A necessary condition for membership of **Y** is equivalent to a statement that the set **Y** is contained by another set. No number of statements about sets that **Y** contains, overlaps, or does not intersect can be combined to make such a claim. Hence, the premises must implicitly contain a statement about a set that contains **Y**. Thus, to make a necessary claim, the premises already contain an implicit necessary claim on **Y**.

Our premises can be either axioms or assumptions. If we assume a condition is necessary, it is then unfalsifiable by *Claim 2*. If we define a claim W to be necessary, we have altered the problem we started with; that is, **W** is a superset of something, but not of the set **Y** that we began with in Eq. **1**. In this case, we can define security so that certain things are necessary, but this does not allow us to conclude anything about outcomes. That is, divide the population into those who comply with the defined necessary condition $\mathbf{P}_W$, and those who do not $\mathbf{P}_{\overline{W}}$. We cannot, without additional assumptions, state that the average outcome in $\mathbf{P}_W$ will be better than $\mathbf{P}_{\overline{W}}$, that at least one case will be better, or even rule out the possibility that the outcomes of the two groups will be identical.

For example, if we define a password of greater than length 6 to be necessary to be secure, we cannot (without additional assumptions) make any statement about differences in experience between those who comply and those who do not. If we assume that an attacker will (and is in a position to) brute-force the set of all such passwords, then it is sufficient to use a password not in that set. This assumption is, however, by *Claim 1*, unfalsifiable. The author has used a 6-character lowercase password at a major online retailer for 15 y without incident.

Other Approaches to Security

If we wish to speak about avoiding certain events then unfalsifiablity is an unavoidable consequence. The inability to observe that something will not happen appears to impose serious restrictions on our ability to reason about sets such as **Y**. A natural question is whether we can reason about other sets that do not have this difficulty and yet serve as good proxies for what we want. We now examine two such approaches. The first is to pursue a set of defined security goals. The second is to define insecurity not in terms of what will happen, but rather in terms of what can.

Security by Achieving Desired Goals. One approach is to start with a set of security goals that are to be met. We call sets of things that satisfy these individual goals $\mathbf{X}_i$, and define the set of things that meet all goals as $\mathbf{Y}_g \triangleq \cap_i \mathbf{X}_i$. The desire then is to find $x \in \mathbf{Y}_g$. The goals might be arrived at based on assumed or observed attacker capabilities, or a threat modeling exercise (4).

Some of our difficulties now appear to melt away: Whereas we could never observe that $x \in \mathbf{Y}$ we most certainly can observe that $x \in \mathbf{Y}_g$. Thus, $\mathbf{Y}_g$ is not an untestable set, and claims of necessary conditions for membership of $\mathbf{Y}_g$ can be falsified. This takes care of a major problem, but it remains to check how well $\mathbf{Y}_g$ approximates **Y**.

Consider how the set $\mathbf{Y}_g$ relates to the avoidance of bad outcomes (i.e., the set **Y**). The claim that $\mathbf{Y}_g \subset \mathbf{Y}$, i.e., that meeting

the goals is sufficient to avoid bad outcomes, can be falsified by finding $x \in \mathbf{Y}_g \cap \overline{\mathbf{Y}}$. This happens when an attacker "steps outside" the model and uses an attack that has not been considered, or was not previously known. In this case, even though $x \in \mathbf{Y}_g$, a bad outcome is still observed. The response to this problem is generally that we constantly search for attack opportunities that might have been missed, and add them when discovered; security researchers and practitioners are often advised to "think like an attacker" to minimize the risk of insufficient defenses. However, the claim that $\mathbf{Y}_g \supset \mathbf{Y}$, i.e., that meeting the goals is necessary to avoid bad outcomes, cannot be falsified because it requires observing $x \in \overline{\mathbf{Y}}_g \cap \mathbf{Y}$ (and we cannot ever observe $x \in \mathbf{Y}$).

Thus, in this approach, the claim that those goals are sufficient can be falsified, but the claim that they are necessary cannot. Thus, although $\mathbf{Y}_g$ is in many respects easier to reason about than $\mathbf{Y}$, it does not address the central problem. Any attempt to argue that $\mathbf{Y}_g$ rather than $\mathbf{Y}$ is the real goal is difficult, no matter how extensive the set of goals. If $x \in \mathbf{Y}_g \cap \overline{\mathbf{Y}}$ then the goals have been achieved but a bad outcome still occurs. Augmenting the goals when this happens is incompatible with the claim that $\mathbf{Y}$ is not the true aim.

Insecurity Is What Can Happen Rather than What Will. A second approach involves labeling something "insecure" if a bad outcome can happen rather than if it will. An example of this approach is articulated by Schneider (5):

> A secure system must defend against all possible attacks, including those unknown to the defender.

It should be clear that this is a definition, because it is obviously unfalsifiable as a claim. Define the set $\mathbf{K}$:

$$x \in \begin{cases} \mathbf{K} & \text{if bad outcomes cannot happen,} \\ \overline{\mathbf{K}} & \text{otherwise.} \end{cases} \quad [3]$$

Defining $\mathbf{K}$ to be the set of interest is not the only approach. It is common to claim that it is the same set as $\mathbf{Y}$; i.e., $\mathbf{K} \equiv \mathbf{Y}$. For example, a popular textbook writes (6):

> Principle of Easiest Penetration: An intruder must be expected to use any available means of penetration.

The authors elaborate, claiming that what can happen will (6): "the attackers can (and will) use any means they can." Whereas few would dispute that not everything that can happen does, a commonly offered justification for the assumption $\mathbf{K} \equiv \mathbf{Y}$ in computer security is that (in contrast to crimes that occur in the physical world) costs for many computer exploits are small enough to be negligible (7).

If we ensure that bad outcomes cannot happen, then we are guaranteed that they will not happen. Thus, $\mathbf{K} \subset \mathbf{Y}$. This represents one major difference with security by design goals: Whereas we must constantly check, and try to falsify, the claim that design goals are sufficient (i.e., $\mathbf{Y}_g \subset \mathbf{Y}$), we have that $\mathbf{K} \subset \mathbf{Y}$ by construction.

Two consequences are immediate. First, $\mathbf{K}$ is of no help in finding a necessary condition for $\mathbf{Y}$. That is, a subset of $\mathbf{Y}$ cannot help us find a superset of $\mathbf{Y}$. Second, $\mathbf{K}$, as a subset of $\mathbf{Y}$, is also untestable: We can never observe that something cannot happen. Hence, reasoning about $\mathbf{K}$ is no easier than reasoning about $\mathbf{Y}$: We cannot observe that something is a member. Thus, claims of necessary conditions for $\mathbf{K}$ are also unfalsifiable.

When trying to ensure that a bad outcome cannot happen, generally we intend to prove rather than observe this fact. Suppose that we could prove membership of $\mathbf{K}$. That is, might we show that some systems or techniques rule out the possibility of bad outcomes? This would then give a subset $\mathbf{K}_p \subset \mathbf{K}$. Although this might seem to eliminate the untestability problem, this is still of no help in finding a necessary condition for $\mathbf{K}$. Mirroring our earlier demonstration for things assumed to be subsets of $\mathbf{Y}$: A claim of a necessary condition for $\mathbf{K}$ is a claim that we have a superset $\mathbf{X} \supset \mathbf{K}$; this would be falsified by finding an element of $\mathbf{K}$ that is not in $\mathbf{X}$. However, if $\mathbf{X} \supset \mathbf{K}_p$, then no element of $\mathbf{K}_p$ can help show that $\overline{\mathbf{X}} \cap \mathbf{K}$ is nonempty. Alternatively, if $\mathbf{K}_p \not\subset \mathbf{X}$, then the claimed necessary condition is impossible.

That bad outcomes cannot happen is not something we can demonstrate empirically. Nonetheless, proving the absence of failure modes or the presence of security properties is an important part of cryptography, formal methods, etc. (8). How do we reconcile formal proofs of security with the empirical untestability of $\mathbf{K}$? Of course, to prove anything formally we must begin with assumptions about what an attacker can and cannot do. For example, stating that a certain task is computationally infeasible is an assumption about what an attacker cannot do, whereas the ability to access the file of hashed passwords is an assumption about what an attacker can do. If we are wrong about the former the error will surface as soon as we observe a successful attack, but if we are wrong about the latter no possible observation reveals the mistake. So, the first type of assumption is falsifiable, whereas the second is not. Further, assumptions about attacker limitations are used to figure out what is sufficient, whereas assumptions about attacker capabilities are used to figure out what is necessary. Thus, formal approaches offer no escape from our basic problem; only by making unfalsifiable assumptions (about what an attacker can do) will they allow derivation of a necessary condition.

It is worth pointing out that the above analysis does not suggest some previously unknown fundamental deficiency in formal techniques. The impossibility of falsifying a necessary claim in no way affects statements about sufficiency. Often finding a way of doing something is more important than demonstrating that it is the only, or most efficient way. For example, Diffie–Hellman key exchange (9) is proved secure subject to assumptions on the computational hardness of the underlying primitive. There is no claim that this is the only way of sharing a key or that anything about it is necessary. Similarly, some properties are so important that having a formal guarantee that they have been achieved is vital. In this case proof of sufficiency is all that is needed. We will revisit the question of the consequences of unfalsifiability in *Discussion*.

Confusing Sufficient for Necessary: $\mathbf{X} \Rightarrow \mathbf{Y} \not\equiv \overline{\mathbf{X}} \Rightarrow \overline{\mathbf{Y}}$

We have seen that claims of necessary conditions for an untestable set are unfalsifiable. However, the same is not true of sufficient conditions. Consider a sufficient condition for an untestable set $\mathbf{Y}_i$:

$$\mathbf{X}_i \text{ is sufficient for } \mathbf{Y}_i \ (\text{i.e. } \mathbf{X}_i \Rightarrow \mathbf{Y}_i \equiv \mathbf{X}_i \subset \mathbf{Y}_i). \quad [4]$$

This can, for example, be falsified by finding a single element of $\mathbf{X}_i$ that is also in $\overline{\mathbf{Y}}_i$. It is corroborated by finding elements common in $\mathbf{X}_i$ that are not in $\overline{\mathbf{Y}}_i$.

Observe that the same evidence corroborates the sufficient condition as the necessary one; and, whereas the claim that $\mathbf{X}_i$ is sufficient for $\mathbf{Y}_i$ is falsifiable, the claim that it is also necessary is not. Thus, there is an extremely easy upgrade path: If we find a sufficient condition for an untestable set, and assert that it is also necessary, there is no possible evidence that can refute the upgraded claim.

Claim 3: No possible observation falsifies a claim that a sufficient condition (i.e., $\mathbf{X} \subset \mathbf{Y}$) for an untestable set, $\mathbf{Y}$, is also necessary (i.e., $\mathbf{X} \supset \mathbf{Y}$).

Proof: Immediate from *Claim 2*.

Consider, for example, $\mathbf{Y}_i \triangleq \{\text{Passwords secure against guessing}\}$ and $\mathbf{Y}_{40} \triangleq \{\text{Random passwords of length} > 40 \text{ characters}\}$. Clearly, $\mathbf{Y}_{40}$ is a sufficient condition for $\mathbf{Y}_i$. However, the claim that $\mathbf{Y}_{40}$ is necessary for $\mathbf{Y}_i$ cannot be falsified unless we have

means of identifying members of $\mathbf{Y}_i$ that do not lie in $\mathbf{Y}_{40}$ (i.e., are in $\overline{\mathbf{Y}}_{40} \cap \mathbf{Y}_i$).

Simultaneous Sufficient Conditions

The problem with treating sufficient conditions (or in general nonnecessary ones) as though they were necessary becomes clearer when we consider not one such condition but several. We have seen that we often have conditions which are sufficient to protect against different particular attacks, that is we have a series of sufficient conditions:

$$\mathbf{X}_i \subset \mathbf{Y}_i. \quad [5]$$

However, suppose we mistakenly interpret these as necessary conditions:

$$\mathbf{X}_i \supset \mathbf{Y}_i. \quad [6]$$

Systems that simultaneously meet several conditions lie in the intersection of the constraint sets: $\mathbf{X} \triangleq \cap_i \mathbf{X}_i$. Let us denote systems that are secure as those that are secure against all of the attacks: $\mathbf{Y} \triangleq \cap_i \mathbf{Y}_i$. Clearly, this means that $\overline{\mathbf{Y}}_i \Rightarrow \overline{\mathbf{Y}}$, and (if we believe [**6**]) $\overline{\mathbf{X}}_i \Rightarrow \overline{\mathbf{Y}}$.

If the conditions are indeed necessary then [**6**] gives

$$\mathbf{X} \triangleq \bigcap_i \mathbf{X}_i \supset \bigcap_i \mathbf{Y}_i.$$

Thus, being in $\mathbf{X}$ (i.e., satisfying all of the conditions) is necessary to be in $\mathbf{Y}$ (i.e., being secure against all of the attacks): $\mathbf{X} \supset \mathbf{Y} \equiv \overline{\mathbf{X}} \Rightarrow \overline{\mathbf{Y}}$. The intersection of several supersets of $\mathbf{Y}$ contains $\mathbf{Y}$. Hence, as expected, we must impose all of the necessary conditions to be secure.

Consider however what happens when [**5**] rather than [**6**] holds: We have sufficient conditions that we mistakenly consider necessary. Rather than contain $\mathbf{Y}$, the intersection of several independent subsets of $\mathbf{Y}$ can be empty: $\cap_i \mathbf{X}_i = \emptyset$. Thus, if we have sufficient conditions which we mistakenly believe to be necessary, imposing many claims can lead to an overconstrained space. There is no solution to the system of conditions that we (mistakenly) believe to be necessary. Obviously this is a risk mainly if we mistake sufficient conditions for necessary ones. An ensemble of sufficient conditions is not inherently problematic so long as we recognize it as such.

Claims of Improvement Rather than Necessity

Speaking of necessary conditions implies a view of security that is binary: Things are either secure or not, and a necessary condition is a universal generalization about the things that are. Although influential, this is not the only approach; indeed, its shortcomings and contradictions have been increasingly noted recently (7, 10). Thus, the idealized, binary view is often abandoned in favor of a more graduated approach. For example, practitioners tend to view actions which make things better or worse rather than an all-or-nothing affair.

Thus, rather than claiming that a measure X_i is necessary for security (i.e., $\overline{X_i} \Rightarrow \overline{Y}$) it is common to argue that X_i is a worthwhile improvement, or that X_i is better than $\overline{X_i}$. An example might be "security is improved if passwords are changed regularly." It does not claim that all security is lost if they are not, but simply that security will be better if they are. In an abuse of notation let us write this claim as

$$\text{Security}(X_i) > \text{Security}\left(\overline{X_i}\right), \quad [7]$$

where Security(.) is the as-yet-undefined state that is to improve. Returning to the question studied earlier, how might we falsify [**7**]?

Let us denote the observed outcomes of a population that uses measure X_i as Outcome(X_i) and those of the rest of the population as Outcome($\overline{X_i}$). Outcomes might include observable features that capture the experience of the user appropriate to the type of harm that X_i tries to reduce (e.g., levels of hijacking, fraud, abuse, and so on). Clearly, if

$$\text{Outcome}(X_i) > \text{Outcome}\left(\overline{X_i}\right), \quad [8]$$

then we might say the claim is established. That is, we can agree that better observed outcomes for the population that uses the measure establishes [**7**]. If $\text{Outcome}(X_i) < \text{Outcome}(\overline{X_i})$ then the reverse of the claim is shown; X_i makes things worse, not better. The only other possibility is that no effect is observed:

$$\text{Outcome}(X_i) \approx \text{Outcome}\left(\overline{X_i}\right). \quad [9]$$

(We use approximate rather than exact equality to accommodate the fact that testing outcomes is likely statistical, and failure to find a statistically significant difference is the closest we can get to determining equality.)

So, does failure to observe a difference, as in [**9**], refute [**7**]? There are many reasons why observing no effect between two complementary populations X_i and $\overline{X_i}$ might not be regarded as proof that the measure does not improve security. First, if X_i is part of a defense-in-depth measure then we do not expect a difference in outcomes unless the main defense fails. For example, the experience of those who travel on a ship without lifeboats will be the same as those who travel on one with lifeboats unless the ship sinks; the fact that the experiences are the same does not mean the measure has no value. Second, we often face adaptive attackers; a vulnerability might not be exploited if it is undiscovered or if an alternative path to the same resource can be found at lower cost. For example, shoulder-surfing might be a far more expensive way of acquiring passwords than guessing or keylogging, but might remain a viable vector in certain circumstances. Third, an observation over some population might not have the statistical power to show significant difference if the base rate of a particular attack is low (11). For example, if one in a million users per year falls victim to a certain attack type, a statistically significant difference in outcomes for any counter measure would likely require observing millions of users for several years.

Thus, the fact that outcomes of X_i and $\overline{X_i}$ are not observed to be significantly different is not necessarily a demonstration that X_i does not improve security. However, if the observation $\text{Outcome}(X_i) \approx \text{Outcome}(\overline{X_i})$ does not refute the claim $\text{Security}(X_i) > \text{Security}(\overline{X_i})$ and we have no direct measure of security, then the claim is unfalsifiable: No conceivable event proves it wrong. Thus, the null hypothesis (that security is unaffected by X_i) is never accepted.

As before, we can define security as a way to evade the problem. For example, we can say that the more guesses a password withstands the more secure it is; thus, an 8-character password with upper, lower, and special characters would in general be more secure than a 6-digit personal identification number (and this might be verified using a cracking tool). However, the claim now says nothing about outcomes. We can prove that the more guess-resistant a password is the more secure it is, but only if security is defined in terms of guess resistance. This may indeed improve outcomes if such a guessing attack occurs, but the claim that one will is unfalsifiable, by *Claim 1*. We can make true statements about improvement if security is defined circularly; but, if the security of a system is to be tied to observed outcomes then we must be able to describe

the evidence that would prove a claim wrong in terms of those outcomes.

A partial answer is that we can modify an unfalsifiable claim to produce a falsifiable one if we explicitly state the conditions under which the measure should make an observable difference to outcomes. Thus, we seek the conditions < cond > under which (if no difference in outcomes is observed) the claim is refuted. That is, we want conditions such that the observation

$$\text{Outcome}(X_i|\langle\text{cond}\rangle) \approx \text{Outcome}\left(\overline{X_i}\middle|\langle\text{cond}\rangle\right) \quad [10]$$

necessarily implies

$$\text{Security}(X_i|\langle\text{cond}\rangle) = \text{Security}\left(\overline{X_i}\middle|\langle\text{cond}\rangle\right). \quad [11]$$

If we can find such conditions, then the claim is falsifiable: If the condition holds, then similar outcomes means the claim that X_i improves security is false. If the conditions cannot be determined then the claim is unfalsifiable. Stating the conditions that make [**10**] true is the same as describing the evidence that proves the security claim false.

Discussion

Types of Claims We Can Make. We return to the question posed in the Introduction: What justifications can we offer when we recommend a defensive measure X? A general approach to describing something as necessary is statements of the form

$$\text{if } (\langle\text{cond}\rangle\text{AND you do not do X})\text{then } \langle\text{claim}\rangle, \quad [12]$$

where < claim > is a statement about the consequences of failing to do X when conditions < cond > hold. We have seen that if < claim > is "you are not secure" or "a bad outcome will occur" then [**12**] is unfalsifiable for all X and all < cond >. If < claim > is "a bad outcome can occur" then it is tautological (saying only that anything not made impossible by X can happen). If either < claim > or < cond > is vague, then it is not possible to be sure what evidence counts as refutation. For example, if < cond > is "given a sufficiently motivated attacker" the conditions are elastic enough that we can never convincingly argue that they have been met. Finally, to relabel claims as suggestions, best practices, or recommendations is simply to make no claim at all. For example, saying "it is suggested that you do X" in place of [**12**] makes no attempt to justify the measure. Thus, all of our attempts to justify security measures as being necessary appear to be empirically unfalsifiable.

Offering provable instead of empirical claims as justifications does not help. A claim can be proved true, if it says nothing about experience. A claim can describe experience, if it runs some risk of being wrong. What a claim cannot do is have it both ways: be immune to contradiction while making useful statements about experience. If it cannot be contradicted by some possible observation a claim is consistent with every possible observation. Thus, it is worthless, on its own, as justification of a measure to influence anything observable. Only when it is combined with some assumption about how the formal statements model reality can a proof make claims about outcomes. Because a proof cannot add anything that was not implicit in the assumptions, a proof of a necessary condition always begins with an unfalsifiable assumption. To have confidence that a measure indeed influences outcomes it must be supported by a claim that is both corroborated (so we have good reason for believing it true) and contradictable (so we have a means of knowing if it is false).

We remind the reader that it is only claims of necessity, and claims that security is improved (without an observable improvement in outcomes), that are unfalsifiable. The evidence that contradicts a claim of a sufficient condition is clear: observing a successful attack. The claim that observable outcomes improve significantly, i.e.,

$$\text{Outcome}(X|\langle\text{cond}\rangle) > \text{Outcome}\left(\overline{X}\middle|\langle\text{cond}\rangle\right), \quad [13]$$

can be contradicted by observing no effect.

Consequences of Unfalsifiability. Whereas Popper famously argued that falsifiability marks the boundary between the scientific and nonscientific (1, 3), we need not take a side in that debate to note serious drawbacks to making unfalsifiable claims. Unfalsifiable claims attempt to evade or reverse the burden of proof; it is the null hypothesis (i.e., the claim that X is not necessary or has no effect) that is taken to be refuted by default. Whereas this may violate some abstract sense of what is appropriate for scientific claims, a much more concrete problem is that it restricts self-correction, means that we cannot identify waste, and we lack the means to decide which measures to accept and reject.

The inability to test claims means that if they are in fact wrong we will not be able to discover it. If we mistakenly accept that measure X improves or is necessary for security, no possible subsequent evidence reveals the error. Hence, the set of defensive measures that we accept evolves in a one-sided way. Because there is no mechanism for rejecting measures, waste is inevitable, and cumulative, unless the process for accepting them is error-free. If wasteful measures accumulate, there is also a considerable risk that we get an unsolvable system: When we upgrade sufficient claims to necessary, we end up with a system of constraints which may not have a solution. Because something cannot be both necessary and impossible, it is easy to be blind to the danger: We can be lured into thinking that everything which we (falsely) believe to be necessary is, as a consequence, possible.

Finally, how can we decide which unfalsifiable claims to accept and which to reject? We lack a mechanism for ordering unfalsifiable claims by importance. If they were justified by a testable claim like [**13**], we might perhaps order a collection of measures by the effect size of the improvement that each delivers [although this is only one input to a sensible cost-benefit decision (12)]. However, if they are justified by untestable claims like [**12**], there is nothing quantitative to compare. For example, if X_a is justified using one set of assumptions, and X_b by another, there is little we can do beyond subjective assessments about which set of assumptions seems most plausible. A criticism of risk analysis approaches (13) in security is that we lack probability estimates for many attacks. However, we now see that when we use unfalsifiable claims as justifications we end up making subjective assessments of plausibility anyway. A further justification for treating attacks probabilistically is that attacker adaptation, which complicates the question of assigning probabilities to attacks, is seldom cost-free. Although attackers with perfect knowledge and zero switching costs are hard to model, assuming realistic limitations on their abilities, knowledge, and costs makes probabilistic approaches very useful in practice (14, 15).

The idea of allowing all unfalsifiable claims seems unworkable, as it is incompatible with a limited budget for counter measures. However, if we allow only some then the question of an acceptability criterion becomes important. Unfalsifiable claims are used to justify inconveniences such as password policies, but also to claim that National Security Agency spying and backdoors in cryptoalgorithms are necessary to prevent terrorism. The basis on which some unfalsifiable claims are to be accepted and others rejected seems worth serious consideration.

Examples of Waste and Inability to Rank. Unfalsifiable justifications carry a risk of waste that does not apply to claims of sufficient

conditions, or claims of improvement that are supported by data. In certain circumstances the risk of waste may be more tolerable than in others. Suppose, for example, that we believe Diffie–Hellman to be a necessary method for key exchange. The consequence of being wrong is waste if a simpler alternative exists. However, because much of the cost is the one-time effort of formally analyzing and implementing the technique, there is little ongoing waste. This is also the case when formally verifying many desired security properties: Upfront costs are larger than ongoing ones, so the waste is less serious (even if we believe the property to be necessary rather than sufficient). By contrast, when measures have recurring costs waste can be very significant. Measures that involve human effort, such as those involved in the choosing and maintaining of passwords, are ready examples, but the problem is by no means limited to those cases.

Surprisingly then, none of the common recommendations that user passwords should be long, strong, contain certain characters, kept unique to each account, never written down, and changed regularly appears to be supported by a corroborated contradictable statement. Although numerous organizations give password guidance, none that we can find supports them with evidence of improved outcomes or testable claims. For example, the Cyber Emergency Response Readiness Team of the US Department of Homeland Security (US-Cyber Emergency Response Team, Cybersecurity Tips; https://www.us-cert.gov/ncas/tips) and the Open Web Application Security Project (https://www.owasp.org) describe their recommendations as "tips" and "best practices," respectively. The National Institute of Standards and Technology (16) details a set of assumptions under which some of these password measures become necessary, but none of the assumptions is falsifiable and the report makes clear that they are not based on empirical support. Thus, whereas a credible justification should both be corroborated by evidence and falsifiable, a majority of recommended password measures are neither. This is also true for numerous other areas of user security advice (17). This does not, of course, mean that these measures have no value; it simply means that we receive no feedback on whether they are accomplishing any of the hoped-for improvement in outcomes.

Real examples of ending up with unsolvable systems also exist. Choosing a unique password per account, for example, is sufficient to protect against a breach at one account having consequences for another. However, as Florêncio et al. (18) point out, following this rule over a portfolio of 100 distinct 40-bit passwords requires remembering 4,525 random bits (e.g., equivalent to memorizing the first 1,362 places of π). This appears a clear case where confusing $\mathbf{X} \Rightarrow \mathbf{Y}$ for $\overline{\mathbf{X}} \Rightarrow \overline{\mathbf{Y}}$ 99 times leads to the absurd conclusion that something clearly impossible is actually necessary.

An example of the consequences of the inability to rank a collection of measures is that implementing anything short of all of them must be done in an unsystematic way. Whereas neglecting any defense might represent an unacceptable risk for very high value targets, doing everything is neither possible nor appropriate for most Internet users. However, this acknowledgment does not help us decide which measures to neglect. For example, is it more important that users not write their passwords down or that they change them regularly? Is examining emails for suspicious links a better use of effort than enabling two-factor authentication? Because these measures are justified by untestable claims we can do no better than make subjective assessments of which assumptions are more plausible. The subjective nature of these assessments is corroborated by Ion et al. (19), who in a survey of 231 computer security experts found great variation in the importance they attached to different recommendations targeted at end-users. The net effect of being confronted with overly long unordered lists of security measures appears to be that a majority of users simply tune out (10, 17, 20).

ACKNOWLEDGMENTS. The author thanks Shuo Chen, Baris Coskun, Dusko Pavlovic, Wolter Pieters, and the anonymous reviewers for comments and suggestions.

1. Popper K (1959) *Conjectures and Refutations: The Growth of Scientific Knowledge* (Routledge, London).
2. Landau S (2013) Making sense from Snowden: What's significant in the NSA surveillance revelations. *IEEE Security & Privacy*, (4):54–63.
3. Godfrey-Smith P (2009) *Theory and Reality: An Introduction to the Philosophy of Science* (Univ of Chicago Press, Chicago).
4. Shostack A (2014) *Threat Modeling: Designing for Security* (John Wiley & Sons, Hoboken, NJ).
5. Schneider FB Blueprint for a science of cybersecurity. *National Security Agency: The Next Wave*, 19(2):6–16, 2011.
6. Pfleeger CP, Pfleeger SL (2003) *Security in Computing* (Prentice Hall Professional, Upper Saddle River, NJ), 3rd Ed.
7. Odlyzko AM (2010) Providing security with insecure systems. *Proceedings of WiSec'10* (ACM, New York), pp 87–88.
8. Pavlovic D (2015) Towards a science of trust. *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security* (ACM, New York), p 3.
9. Diffie W, Hellman M (1976) New directions in cryptography. *IEEE Trans Inf Theory* 22(6):644–654.
10. Lampson B (2009) Usable security: How to get it. *Commun ACM* 52(11):25–27.
11. Axelsson S (2000) The base-rate fallacy and the difficulty of intrusion detection. *ACM Trans Inf Syst Secur* 3(3):186–205.
12. Anderson R (2001) Why information security is hard–An economic perspective. *Proceedings of ACSAC* (IEEE Computer Society, Los Alamitos, CA), pp 358–365.
13. Adams J (1995) *Risk* (Routledge, London).
14. Mohler GO, Short MB, Malinowski S, Johnson M, Tita GE, Bertozzi AL, Brantingham PJ (2015) Randomized controlled field trials of predictive policing. *J Am Stat Assoc* 110(512):1399–1411.
15. Florêncio D, Herley C (2011) Where do all the attacks go? *Economics of Information Security and Privacy III* (Springer, New York), pp 13–33.
16. Burr WE, Dodson DF, Polk WT (2004) *Electronic Authentication Guideline*. National Institute of Standards and Technology (NIST) Special Publication 800-63 (NIST, Gaithersburg, MD).
17. Herley C (2009) So long, and no thanks for the externalities: The rational rejection of security advice by users. *Proceedings of NSPW 2009* (ACM, New York), pp 133–144.
18. Florêncio D, Herley C, van Oorschot PC (2014) Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. *23rd USENIX Security Symposium* (USENIX Association, Berkeley, CA), pp 575–590.
19. Ion I, Reeder R, Consolvo S (2015) No one can hack my mind: Comparing expert and non-expert security practices. *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)* (Usenix, Ottawa, Canada), pp 327–346.
20. Adams A, Sasse MA (1999) Users are not the enemy. *Commun ACM* 42(12):41–46.