To what extent are identified psychological processes created in laboratories? The present work addresses this issue with reference to one particular realm of behavior: vigilance. Specifically, I argue that the classic vigilance decrement function can be viewed more realistically and advantageously as an "invigilant" increment function. Rather than characterizing the transient decrease in detection capability that is evident on exposure to enforced monitoring as a diminishment in capacity, it may be more usefully seen as an appropriate scaling by the designated observer to adapt to the nonoptimal circumstances that he or she is forced to endure. This proposition emphasizes the dynamic response characteristics of the observer and locates the origin of the phenomenon and the onus for practical improvements in the design of operational displays with designers rather than apportioning blame for performance decrements to the operator. This perspective reinforces the recognition of a crucial presence of the necessary but often unrecognized external arbiter in the vigilance paradigm and the extrinsically imposed imperative to sustain attention. Explicit recognition of this fact also helps explain the stress involved with extended vigils. In identifying the traditional vigilance decrement as a form of iatrogenic disease, I argue that modern design of work systems should alleviate the need for either the acute or the chronic expressions of such enforced human monitoring activity. It is possible that the case of vigilance is itself representative of a modern propensity to create new psychological phenomena in the face of human exposure to modern, evolving technical environments.

Keywords: vigilance, sustained attention, iatrogenesis, displays, semiautonomous systems

Almost 70 years ago, assiduous young British servicemen headed out over the Bay of Biscay in wartime Royal Air Force aircraft, intent on spotting and destroying German U-boats. At the time, this was a perilous but critical endeavor because each U-boat sunk would represent a meaningful reduction in the damage that such craft were wreaking on the crucial lifeline of the North Atlantic convoys. That vital umbilicus between the needs of the British people and the source of supplies in the United States could hardly have been a more important one. In consequence, these missions were not practice exercises but, rather, a literal matter of life and death. Frustratingly—depending on one’s perspective, of course—these fit, young, and highly motivated profession-
of modern, electronic forms of work in so many diverse operational realms. Also, it is an effort to understand whether other dimensions of laboratory-based psychological research are iatrogenic reflections of adaptive response as opposed to naturally occurring and spontaneous human behavior. It is to the fundamental issue of the configuration of modern work systems, however, that I first turn to establish the real-world foundation of and validation for each of the observations that follow.

Vigilance and Sustained Attention in Modern Work Environments

The phenomenal growth in the contribution of automation to virtually all realms of human work continues unabated. One could reasonably argue that this rate is actually increasing, and some postulate that it could be increasing at an increasing rate (see, e.g., Kurzweil, 2005). The remnant role for the human performer in such situations is often one of automation monitor, and this transition has been the emblem and, indeed, the epitome of work evolution for at least the past four decades or more. This developmental vector has progressively emphasized the need for greater vigilance or, nominally, the sustenance of attention in the human operator. The problem with this trend is that the human operator is arguably magnificently disqualified for this particular form of sustained attentive response (Hancock, 1991). Most designers of systems that emphasize automation wrestle predominantly (and sometimes exclusively) with the technical challenges of creating the computer-based automation itself. From the perspective of the manufacturer, this is a reasonable and, indeed, natural point of focus. Thus, relatively little attention has traditionally been paid to the human role until the human performer ended up “right of bang,” that is, after the disaster has occurred. However, after disasters do actually occur, questions as to the failure attendant on the human role subsequently abound (see, e.g., Casey, 2006). Confirmation of this cycle can be had with reference to the media coverage following almost any major disaster. The great sadness is that this cycle of a priori ignoring and a posteriori blame has now continued unabated since at least the end of the Second World War. Those advocating for human-centered systems design, which is based on the fundamental knowledge of psychological theory, persist in their valiant but apparently insufficiently influential effort to change the social mind-set concerning the importance of such human operator issues. One of my goals in this present work is therefore expressly polemical in nature, in that I seek to question the basic assumptions concerning the phenomenon of vigilance and the traditional and persistent view that the decrement associated with sustained attention is a result of intrinsic limitations in human capacity. As I endeavor to show, this issue is, however, not limited to vigilance alone. To begin this disputation, it is necessary to return to the origins of the vigilance concept and the decrement function itself, which has portrayed the human capacity to sustain attention for close to three quarters of a century of its scientific legitimacy.

The Decrement Function

There can be little doubt that human beings have been aware of the putative failings of personnel engaged in long but uneventful periods on watch since the very onset of the first episodes of military conflict. Because violent skirmishes and battles sadly characterize the human condition, one may conclude that vigilance has been around as long as human beings themselves. Nor were these watches solely for military purposes, for, as Alluisi (cited in Warm, 1993) recognized, this was a critical role on all of the “sailing ships of yore” (p. 143). Indeed, one can argue that in the persistent and perennial contest between predator and prey, the critical capacity to wait patiently for the critical signal to attack or, alternatively, escape was an essential element of success in either role. Animals today exhibit behavior consistent with notions of sustained attention, so it appears possible to extend the previous conclusion and posit that, perhaps, the capacity to sustain attention (which I later distinguish from the modern phenomenon of vigilance) actually predates human existence entirely. Despite a number of references to diminished capacities in religious texts and other early narratives and literature over the years, the first formalized recognition of the nature and state of vigilance is most frequently attributed to the English neurologist Sir Henry Head. Head was largely interested in clinical conditions and associated failures in normal functional capacity (e.g., Head, 1920). He represented the term vigilance as describing a “state of maximum physiological and psychological readiness to react” (see Head, 1923; Warm, 1984, p. 3). Although Head is thus associated with the identification of the scientific and conceptual basis of vigilance, there can be little doubt that the name of Norman Mackworth will be forever associated with the fundamental
foundation of vigilance research. This is especially the case because Mackworth conducted the seminal, empirical attack on the issue. His splendid and comprehensive efforts still stand up well to modern scrutiny more than six decades later. It is supportable to say that many, if not most, of the factors that influence the identified phenomenon of vigilance were first explored in Mackworth’s comprehensive initial efforts (see, e.g., Mackworth, 1948, 1950).

Of all of the findings in his collective works, Mackworth’s report of the vigilance decrement function is perhaps the best known and most replicated. This is illustrated in its prototypical form in Figure 1 (from Mackworth, 1948). As this curve is so well-known, it is only necessary to briefly describe it here (for a detailed treatment, see See, Howe, Warm, & Dember, 1995). As is evident, the capacity of the individual begins to fail under the constraints of the vigil, sometimes not even 20 to 30 minutes into the period on watch. After this initial decrease in capacity, which has traditionally been expressed in terms of rates of missed signals, increasing response time to successfully identify targets (e.g., Buck, 1966), or sometimes their co-occurrence, there is a relative plateau in which the capacity to respond does not change substantively. Akin to the functional form of many learning curves, the decrement function purports to describe human capacities to sustain attention over relatively extended intervals of tens of minutes to several hours (although for an abbreviated form of vigilance, see Helton et al., 2007; Nuechterlein, Parasuraman, & Jiang, 1983). Mackworth must be saluted for his experimental exploration of this effect, but there is one throwaway line about nonresponse in his work that I emphasize here. The basis of this observation lies in the fact that the Applied Psychology Unit at Cambridge University in England was, at that time, greatly advantaged by having access to military personnel who were assigned to the unit for testing as an ongoing rotation (Broadbent, 1980). These Naval ratings would arrive at the facility and be available for performance evaluation on any one of the behavioral experiments in progress. It was some of these individuals who first performed in Mackworth’s controlled experimental procedures.

In one of his most crucial observations, Mackworth noted that, on occasion, he had entered the testing chamber only to find one of these individuals asleep. Of course, this immediately brings up the methodological question of what to do with the data from such nonperformers, but such pragmatic concerns block the central but simple fact that they were asleep. This is arguably the only crucial observation that Mackworth fails to exploit in his extended exposition. That is, when the compulsion to persist with the vigil is insufficient or certain intrinsic capacities of the individual are so depleted, then the task itself falls away. Then, in the boring and understimulating environment created by the experimenter, the adaptive behavior is indeed to go to sleep. As I discuss later, the nature of the compulsion is critical to the occurrence and persistence of this change in response capacity. In essence, one has to engage for the vigil to become a task in the first place (see Hancock & Caird, 1993). The phenomenon of falling asleep is not simply a historic oddity but persists in vigilance studies conducted in the present day (F. Durso, personal communication, August 2011). The issue of compulsion is one that is central to my argument, but before exploring that issue at length, it is important to first evaluate the precise nature of the decrement function and especially the fact that it emerges from the performance of groups of people whose own individual response patterns may vary extensively from this average curve.

**Average Decrement: Sleepers Versus Superstars**

One of the most important facts of the vigilance decrement function is that it derives from average performance across individuals whose own personal performance can vary greatly. Those who are asleep frequently miss signals and have long response times but happily also generate very
few false alarms (Parasuraman, 2011). To balance the poor and nonresponders and produce the average decrement curve, experimenters must also have exceptional participants. This observation is certainly true, and some vigilance superstars perform the task in either near to or actual optimal fashion. These apparently rare persons do not miss signals but respond rapidly and accurately while suppressing any false alarms. In essence, there are large individual differences in vigilance capacity, but in the on-going battle to understand the overall nomothetic pattern, these individual variations have been studied only to a limited extent (but see Shaw et al., 2009; Szalma, 2009; Szalma, Hancock, Dember, & Warm, 2006).

Some recent important efforts have begun to identify those who exhibit close to flawless performance in vigilance, and some promising early results have been reported (see Parasuraman, 2009). There is also some recent evidence that superstar performance is not confined to the vigilance realm (see Watson & Strayer, 2010) but may indeed be a property of specific individuals in conjunction with specific task types (see Foer, 2011). Given that superior performance in differing everyday response capacities (e.g., sport skills) are distributed in this fashion, it should be no surprise that certain basic psychological capacities are distributed in this manner also. Whether these individuals are endowed by nature with the innate capacity to perform a flawless vigil, this is a learned behavior, or this pattern of response is mediated by an intervening process, such as the particular individual’s strict adherence to external instructions, has yet to be fully distinguished. However, it is sufficient for the present argument to note that the prototypical decrement function is something of an illusion derived from performance averaging across groups of individuals. This fact will become important as the argument is reversed to focus on an invigilant increment function.

**On the Nature of the Compulsion**

Perhaps the most important dimension of vigilance revolves around the fundamental question concerning the source of compulsion. That is, who or what is mandating that the individual undertakes the vigil in the first place? In general, we can dichotomize sources of compulsion into two fundamental categories. Motivation can be either intrinsic or extrinsic in nature. Understanding this difference is extremely important. If the motivation is intrinsic, one can then control the situation oneself and with it the associated level of experienced stress. From previous work, it is now clear that the stress of vigilance is considerable (Hancock & Warm, 1989). If one can control the circumstances of the vigil and particularly the choice to either continue or suspend it, then associated levels of stress are much diminished (see Gunn et al., 2005; Karasek & Theorell, 1990). To reiterate, that one has the option of abandoning the vigil with few, if any, subsequent ramifications is absolutely critical. It is this dimension of control that, I argue, distinguishes vigilance from sustained attention. The latter opportunity to suspend one’s participation is certainly not true when the vigil is externally imposed (witness, e.g., the harsh punishment imposed by the military for being caught asleep on watch). I claim that the phenomenon that is recognized today as vigilance and its associated decrement is exclusively derived from the external imposition of the need to sustain attention. Thus, let us turn to these conditions.

When recalling Mackworth’s initial source of interest in pursuing research into vigilance, one must remember the context of his personal motivation: His investigation was driven explicitly by the pragmatic needs of combat in the Second World War. Manifestly atheroretical, his initial explorations were part of the war effort specifically designed to destroy German submarines. For the monitors of the radar screens whose tiny, masked, and barely recognizable blips could represent life or death for many of their military comrades, the compulsion was both explicit and severe. That is, as military personnel, they were duty bound to follow their assigned orders, but as conscientious and patriotic individuals, they would certainly have understood the wider importance and implications of their task. Of course, it was also possible that the U-boat itself or some attacking German aircraft could shoot down their own aircraft, so there was a background of intrinsic personal motivation (although some would argue that this would still be primarily extrinsic in nature; see Ryan & Deci, 2000). This combination of primarily extrinsic and marginally intrinsic sources of compulsion would have pushed the stress of this task to extreme levels. Indeed, it is reasonable to assume that it was the outcome of these highly maladaptive work conditions and level of associated failure that brought the problem to Mackworth’s attention in the first place. I therefore further contend that it is the combination of such compulsion and specific task difficulty that acts to create vigilance conditions and the associated decrement. I conclude that there are three necessary conditions:

1. The primary compulsion for the task has to be external in nature and to create a vigilance decrement, the extrinsic compulsion must exceed any intrinsic compulsion toward superior performance.
2. The display to be observed must be designed (overtly or inadvertently) to suppress the signal noise level to near the threshold of possible sensory observation, thereby creating high levels of task difficulty.
3. The imposed frequency of signal occurrence must be rare (e.g., Parasuraman, 1979) and the background must be distracting of attention.

Although the specification of these latter two conditions are not new in the sense that they have been commented on before, it is the combination of the three that induces what might be considered an artificial or iatrogenically induced decrement condition. Further, although formal vigilance research is largely a post-World War II enterprise that is vastly magnified in modern, electronically mediated workplaces, the requirement for vigilance has been one that has persisted through large swaths of human
history (see Hurst, 2010). These necessary conditions explain why one observes sustenance of attention in the natural world while rarely encountering the classic decrement function in other animals in anything but highly artificially constrained conditions (cf. Jerison, 1965; Krasnegor & Brady, 1972). Of course, this latter assertion compounds an epistemological conundrum with a necessity for knowledge of ground truth in the wild; however, the explicit enforcement of a social compulsion is rare in anything but human society. Although these comparisons are explored in more detail below, I suggest a causal account here of the sequence of events. Imposed vigilance constrains attention and thus serves to induce stress in the individual. Stress increases across a vigil (Szalma et al., 2004), especially one with infrequent or even no appearances of a target. Intrinsically motivated vigil permits the dissipation of this stress by periodic or permanent abandonment of the vigil, whereas extrinsically imposed vigils do not allow such dissipation to occur. Consequently, the observers adjust their associated observing response (Jerison & Pickett, 1964) to the residual meaning that they can attach to the task. Such a proposition is highly consistent with empirical observations of phenomena such as end spurt in vigilance (see Bergum & Lehr, 1963), but it also very much relates to the wider issue of paced versus unpaced work (see Parkes, Styles, & Broadbent, 1990).

At first, it might appear that the nature and issue of compulsion is largely a theoretical one and thus of associated theoretical concern; the reality is that this identification is very much of pragmatic relevance and open to experimental attack. Indeed, some important, pioneering work on this very issue has already been conducted by Scerbo and his colleagues (Scerbo, Greenwald, & Sawin, 1993). In one particularly innovative experiment, Sawin and Scerbo (1995) argued that the initial instructions provided by the experimenters themselves set critical constraints and expectations with respect to subsequent performance outcome. To evaluate this proposition, these authors asked their participants to watch a purpose-built piece of apparatus that, although at heart a standard computer screen, did not appear to be so. After telling participants that they were watching some new form of color display, the authors cleverly embedded into the situation a traditional vigilance task but without the usual form of compulsion. In fact, the tested individuals fell into two groups, one that received the traditional vigilance instruction set for the new apparatus and another group that were just asked to simply relax and respond if they saw the new display flicker (indeed, they were told that most people found the new display itself a source of relaxation)! Despite the stern admonitions to pay attention in one group and to simply relax in the other, the authors found no difference in the level of objective performance. It is important to note that capacity declined equally across both groups, but the level of associated perceived workload was significantly higher in the group with the traditional instruction set (see also Scerbo, 1998). Similar conclusions concerning this effect of choice can be drawn from the equally important and insightful work on event rate by Scerbo, Greenwald, and Sawin (1992). Here, one group could choose the signal event rate, whereas a yoked control group had that equivalent rate imposed on them. Consistent with the prior study, the group with the choice showed no decrement while the yoked control exhibited the standard vigilance decrement. These are important findings because they confirm that the stress associated with vigilance is strongly mediated by the nature of the compulsion, whereas the nature of the decline is more associated with the physical characteristics of the task itself. The precise apportionment of these specific effects has yet to be fully distinguished. However, it is important to note here that even though, in the first cited experiment, these differing instruction sets resulted in an equivalent decrement, they both still represent extrinsic compulsion. That is, the vigil (in its two distinct forms) was imposed by an external agency.

The notion of the effect of differential compulsion is also especially evident in some trenchant observations by Jerison (1965):

"... I work with animals and people. The initial excuse for working with animals was to find out why well-trained animals are so much better than people on vigilance tasks. The answer was so obvious it is painful for me to report it. . . . The animals had to show me."

The point turned out quite simply. Animals do better than people on vigilance tasks, because the value of detecting or the cost of failure to detect or both are much greater for all animals. When an animal is well-trained in a discrimination learning situation, it avoids a severe electric shock or gets its only ration of food by responding appropriately when signals appear. (p. 580)

In essence, when the external compulsion is sequentially increased in severity, the animal’s behavior changes in accordance and the vigilance decrement can, to a degree, be extinguished. However, when the stress is elevated beyond a tolerable level, the whole order of performance breaks down. Jerison (1965) goes on:

"... When the intensity of shock reinforcement was raised they responded very frequently and stopped attending to (facing) the signal source. We call this a free-response strategy. In signal detection theory terms this is essentially the adoption of so low a criterion that the signal is treated as if it were on all the time. By penalizing the animals for false-alarms, we could produce a dramatic shift to a strategy involving attending to the signal source. Animals shifted from one strategy to another simply when the cost of false-alarms and missed signals was varied. (p. 580)"

Here again we see that the animal adapts according to the external constraints placed upon them. The circle is finally complete. The experimenter exerts the original compulsion, delimits the constraints of the required task demands, and infers the nature of behavioral causation from the output performance. The whole circumstance thus proves to be iatrogenic by very definition. If one is to both parse and evaluate the inherent nature of compulsion, testing differing levels of external compulsion is useful, of course. However, intrinsically motivated sustenance of attention still needs to be evaluated against
any version of its external imposition. Such naturalistic studies have yet to be adequately conducted.

**Vigilance as a Technology-Induced Iatrogenic Disease**

Given the foregoing arguments, I suggest that “vigilance,” as it has come to be known, is an iatrogenic or self-generated disease derived (predominantly) from the socio-technical organization of work and is increasing as the demands of modern work emphasize such relevant task characteristics more and more. The foregoing statement is both involved and polemic, so perhaps by unpacking its respective elements, the case can be argued both incrementally and more persuasively. I have placed the term vigilance in quotes because, as I have noted previously, people have hunted for prey and been hunted essentially as long as human beings have been around. I do not argue, therefore, that intrinsically motivated sustained attention is a new thing derived from the circumstances of modern work environments. Rather, I am arguing that the decrement function associated with vigilance derives from the set of circumstances in which some external agency imposes the imperative to observe; quite often, that same agency dictates the nature of what is observed and the format of the equipment and methodology through which such observations are made (i.e., the design and form of the task display). It is this latter combination of external and essentially indisputable mandate (remember individuals caught sleeping on watch during war have occasionally been executed) and poorly designed instruments of watchkeeping that generates the disease (see Szalma, 2011). But why call it a disease?

There are a number of reasons for using the disease label. First and foremost, calling vigilance a disease is a polemic act that implicates and, I hope, stimulates the need for responsive action. Second, calling vigilance a disease implies there could be a cure; indeed, although external imperatives are unlikely to disappear from modern work environments, appropriately designed related work interfaces can and should ameliorate much of the deleterious effect. It should be readily acknowledged that others (e.g., Mackie, 1987; Wiener, 1987) have also emphasized such problems of vigilance, and although they do not express it in these precise terms, they have also advocated for practical solutions. Third, diseases can be of acute or chronic form, which in the present case can be contrasted as the occurrence of one single vigil versus the task of those who engage in monotonous watchkeeping as a profession. I argue that if an individual has little in the way of control of the task, the nature of the observation process, or how and what is observed, then the outcome is likely to be highly stressful epochs of work. Repeated over many exposures, such levels of stress are liable to induce actual illness through stress-related suppression of immunological function, among other damaging avenues (see Cooper & Smith, 1985). Indeed, a more general argument would encompass the way in which the nature of work itself is conceived and how this perspective has evolved across time (as sequential revolutions in work structure have been experienced; Hancock, 1997). It may therefore be that variously attributed modern dimensions of work—for example, its aversive and arduous nature—are themselves actually artificial and can, with a change of perspective and work system design, become more amenable to fruitful manipulation.

**The Vigilance Myth?**

Given what has been discussed, I can now examine whether vigilance, as epitomized by the decrement function, is actually a myth. That is, do people create the circumstances that induce this specific pattern of performance and then search, somewhere within the brain, for the putative neurological source of this decrement? If this is true, then it is literally the case that vigilance is invented by the brain but that invention is a function of the design of the greater social order and the technology that is used to support it and not a fundamental failing of the underlying neurological processes per se. This possibility is supported by the manner in which vigilance was first identified. Mackworth’s first observations of so-called vigilance derived solely from empirical experience. That is, as a practice-driven phenomenon, the observed behavior did not represent a distinct and unique aspect of human response capacity but was, rather, a reflexive identification by Mackworth of the prototypical but average pattern that he himself observed. In essence, he had an apparently emergent phenomenon in search of a name. Vigilance was an indeterminate but convenient term that he extracted from the clinical neurological literature at hand. After all, in a critical sense, what does “a state of maximum physiological and psychological readiness to react” actually mean? But as the philosopher Hume has rightly noted, a name has power, and that power has persisted now for almost a century. However, iatrogenic or not—or, indeed, myth or not—the progressive failure of certain individuals to detect mandated signals for response in poorly designed detection tasks certainly is evident in many modern world situations. Perhaps if this diminution of capacity is recast, the whole problem can be reconceptualized. Thus, at this point I consider not the vigilance decrement but rather the invigilant increment function.

**The Invigilant Increment Function**

Let us now turn the known vigilance world upside down, for a few moments at least. The traditional response pattern is seen as a decrement. That is, the individual starts off at some specified rate of responding and then falls from that initial condition to a new and lower level of performance. As noted earlier, this pattern is most often expressed in terms of missed signals, but as speed can, on occasion, be traded for time. The traditional vigilance decrement is also thought to be reflected in elongated response times, although this indeed is an arguable proposition. Regardless of how it is represented, this decrement is inevitably viewed as an expression of human incapacity and failure. In a general sense, of course, this traditional perspective is true. But I argue that the fundamental failure is not that of
the proximal observer whom we immediately see at work in this situation but is rather that of the human designer of the task circumstances that the proximal observer has to tolerate and operate within. Thus, although technology makes possible actions that are unachievable by the individual alone, the inadvertent creation of maladaptive work, which persists and even proliferates in the modern world, is interpreted as a personal and not a systemic failure. This perspective persists today in many practical realms where the missed signal is attributed to the watchkeeper and not the system designer.

As we have seen, the actual impetus for the whole genre of vigilance research really emerged because nascent radar technology itself was so poor. That is, the capacity to detect signals by the radar technology itself was highly limited. Added to this, the displays associated with these poor initial technical detection capacities were similarly highly impoverished. Although this was all state-of-the-art at that time, these problematic technological capacities were what first induced the effect. Then, for Mackworth himself to replicate this intrinsic design failure, he had to devise a similarly fiendish clock task that, although legendary to vigilance researchers, is equally vacuous in terms of conveying crucial response information to observers. Scientists interested in exploring vigilance have, since that time, had to devise equally poor displays and task regimens to drag their unwilling and untrained and/or unpracticed observers over the precipice of the vigilance decrement.

In the foregoing discussion, I have placed the “blame” for vigilance failures squarely on the shoulders of the task designer. However, by implication, I have, to a degree, exonerated the individual observers. This might seem somewhat perverse given that the actual performance emanates from individuals themselves. However, what these experiments show is the adaptive adjustment of the observer to the situation he or she encounters. Thus, the transient portion in which performance changes, as illustrated in Figure 1, is the volatile region in which such adjustment is made. It is known from the vigilance literature (Craig, 1978; Holland, 1958) and indeed from the wider spectrum of behavioral understanding that people adjust their effort to the immediate and prospective demands of tasks that are put before them (see Hancock, 1989; Hancock, Williams, Miyake, & Manning, 1995). Here, I argue that the vigilance decrement, which I have relabeled the invigilant increment, is truly a transient effect due to the relative novelty of the situation as the observers learn to adjust their responding to the new, constraining, and most frequent uninteresting situation. The illustration of this process is shown in Figure 2.

In Figure 2, the ordinate is reversed from its normal connotation of good performance ascending the vertical axis. In the present illustration the converse is true. Thus, performance actually gets poorer the higher one goes up the axis. Here we can see three curves (labeled A, B, and C) such that Curve A asymptotes to the worst performance, whereas Curves B and C represent sequentially better response capacity. Across time, each reaches a stable level represented by the dotted line. Each curve shows how the invigilant increment grows across time, and I claim that this growth (shown by the three different curves) is actually proportional to how poorly the respective observing task is designed. Always remember that individuals could actually get better across time (see Smith, Valentino, & Arruda, 2003), so each of these curves is still representative of some putative form of failure, yet it is also the case that in this conceptual example, none reach the sad condition of chance performance, as shown at the top of the illustration. I claim here that the invigilant increment is proportional to the iatrogenic level of display poverty. That is, the relative poverty of the design of the display to be observed dictates the average level of responding across a group of observers (see Holland, 1958). I emphasize the average here because the previously noted differences between individuals mean that some exceptional observers are able to maintain completely efficient response capacity despite highly impoverished displays, whereas other observers will fail to succeed in their observations despite relatively well-designed circumstances. These levels of individual difference are thus proportional to the way in which each specific individual seeks novel information and/or is constrained by the imperative to search the impoverished design to hand.

Fortunately, the decades of extant vigilance research can tell us much about the character and nature of the factors that contribute to such display poverty (Teichner, 1974). For example, the insightful observations of Parasuraman (1979) showed that one crucial feature was the absence on the display of a ready target comparator, which consequently mandated that the observer had to hold a long-standing representation of the searched for target in their memory. Without this necessity for elevated memory load, which of course is easily remedied by good display
design, one had a much reduced chance of recording a vigilance sensitivity decrement. Parasuraman (1979) also showed that the unfortunate designer had not only to neglect the presence of a simultaneous comparator but also had to produce a rate of searchable items that exceeded 24 per minute; a rate of one every 2.5 s. Indeed, it was only with the creation of this pernicious combination that the designer could induce the observer into a manifest invigilant increment. Further, Fisk and Schneider (1981) confirmed that a vigilance decrement is also facilitated when the signal-to-noise ratio is unpredictable and the associated processing of the stimulus cannot be automated (Schneider & Shiffrin, 1977). Sadly, in the real world, many designs that incorporate these fundamental flaws are scattered around practical and important operational systems. Of course, these are not the only dimensions of display poverty that have to be visited on the poor observer for the decrement to emerge. Absence of current knowledge of performance and subsequent knowledge of results also foster invigilance. Also, the presence of distracting sources of stress aids the designer in his or her contrarian efforts to thwart the usually competent capacities of the alert human beings (see Hancock & Warm, 1989; Szalma & Hancock, 2011). In addition to each of these sources of designed incipient failure, I can add the deterrence to performance induced by the monotonous, boring, and repetitive nature of displays presented to the unfortunate individual (Davies, 1970; Helton et al., 2005). To articulate all of the other variables that inculcate and exacerbate traditional forms of vigilance is essentially to recount the psychophysics of the area, and numerous summaries are readily available (see Warm, 1984; Warm, Dember, & Hancock, 1996).

Another window to the vigilance issue is of exceptional practical importance but, unfortunately, rarely enters into the realms of psychological theorizing. This concern revolves around the evidence derived from professional watchkeepers. It turns out that, in the applied world, many people are employed in professions such as security, maintenance, medical image processing, and the like who watch for a living (Czaja & Drury, 1981; Drury & Addison, 1973). These professions span multiple application areas, from aircraft inspection through general industrial inspection to radiographic analysis and evaluation as noted. Although such professionals are by no means perfect at these tasks, one important difference between the watch employees and the participants from the vigilance research of psychology is that these individuals frequently do not exhibit the transient change in their level of response efficiency at the beginning of the watch. For example, Drury (in press) recently observed that “remarkably few documented examples of the vigilance decrement in industrial inspection have been reported in the literature.” There are many reasons why this might be so. For example, many industrial inspection tasks are performed in a social context, whereas most laboratory vigilance tasks isolate the tested individual from all external contact. Be this as it may, from the present perspective, it is suggested that the professionals long ago adjusted their adaptive response to the constant nature of the inspection task that they are required to perform on a daily, monthly, and even yearly basis. Thus, in some ways, the transient decrement, which gives the fundamental character to vigilance itself in the psychological literature, is often missing when one encounters real-world examples of such inspection tasks (see also Mackie, 1987; Wiener, 1987). In essence, the adjustment has already been made and internalized some days, weeks, or months earlier. This observation alone should give pause to those for whom vigilance is predominantly identified with the decrement function. However, the essential point to reiterate is that each individual observer adapts, probably as best they can, to the extrinsic mandate to sustain attention and then to the intrinsic limitations of each specific task design (see also Craig, 1978, 1987; Drury, Holness, Gylin, & Green, 2009; Holland, 1958). When concatenated across groups of observers, a prototypical curve of failure emerges, which reflects the average human capacity to voluntarily adapt to whatever suboptimal task design is foisted on them by the uninspired designer or the fiendishly inventive experimenter. It is thus the case that researchers have to titrate their respective laboratory tasks to create conditions amenable to the emergence of the traditional decrement. In essence, they know and control ground truth and so have to both hide and pervert this truth to elicit the pattern of behavior that they then putatively study. Indeed, I can argue on the basis of studies on what has been termed the cognitive vigilance increment that when experimenters tinker with the conditions such that individuals need to make repeated cognitive decisions, as opposed to sensory discriminations, an actual increment in performance can be derived, although evidence for this latter increment is itself not without dispute (cf. Loeb, Noonan, Ash, & Holding, 1987; Lysaght, Warm, Dember, & Loeb, 1984; Warm, Howe, Fishbein, Dember, & Sprague, 1984).

The Origins of the Signal

One important dimension that I have not yet considered in detail concerns the origin and transformation of the signal itself. Before looking to reach some final conclusions about the artificial nature of vigilance, it is important to consider the genesis and manipulation of this target signal, especially because it is here that a number of sources of literal and figurative confusion arise. Even the word signal itself is a relatively arbitrary one because it acts to identify a specific relationship between the observer and what is to be observed. Thus, one must be aware that in nature, what is purported to be a target for one organism is noise for another organism. Nor is it truly sufficient to parse any world into pure, exclusive categories of signal versus noise because in many situations where a signal is at or near threshold conditions, the evolution of the ongoing circumstance actually serves to resolve the presence of a signal over space and time. For example, what might appear to be an ambiguous source of prey at 100 m becomes unequivocally identifiable as such at 10 m. In other work, my colleagues and I have sought to address this aspect of signal resolution through the combination of fuzzy set theory (Zadeh, 1965) and traditional signal detection theory (Tan-
The outcome of this synthesis, termed fuzzy signal detection theory (FSDT; Hancock, Masalonis, & Parasuraman, 2000; Parasuraman, Masalonis, & Hancock, 2000), refers explicitly to “degrees of signalness” in any environment presentation. FSDT represents a more generalized version of target detection than traditional signal detection theory which provides only one specific subset in which the target is already collapsed into either the zero or one state of the absent or present cases only. I do not go into greater detail concerning nondiscrete signal status here because this concept has been explored elsewhere (see Hancock et al., 2000), and it is evident that in many professional search tasks, the individual is looking for more than one form of signal. However, these observations about the indeterminacy of any putative signal underlie the following observations on signal genesis.

In so-called natural environments, all displays are direct in nature. That is, they are not, by definition, transformed or mediated in any way by an intervening technological system. It is in this way that Gibson (1966, 1979) emphasized the adaptation of the organism to the environment to which it is tuned to react. So, for example, a lookout on a ship or an individual searching for shooting stars in a night’s sky are using only direct perception in a vigilance context. Although it is true that some artificial displays (so-called ecological interface designs) can be purposively created to take extensive advantage of these intrinsic processing characteristics (see Vicente, 2002; Vicente & Rasmussen, 1990, 1992), in most working circumstances, designers have yet to understand the subtle nuances of evolved perception and exploit these inherent capacities to any extensive degree. As a result, in most working conditions, the display presenting the target that the human operator must detect imposes a series of transformations, a number of which do not enhance detection but actually serve to materially inhibit effective performance.

In many practical realms, the signal itself is first strained through a technical detection system. So, for example, even the individuals who motivated Mackworth’s original interest were not engaged in directly trying to view submarines in the sea out of the aircraft’s windows. Neither did such individuals hang out of the sides of the aircraft using the unaided eye to spot the key cues for detection. That is not to say that some people undertaking tasks such as air-ship rescue do not apply this method, at least as a part of their detection strategy. Rather, it is simply the case that most modern detection tasks first pass through the filter of a technological medium. In essence, this initial transformation is itself a form of signal detection. So, whatever eventually gets to the human observer is necessarily constrained by this first technical filtering system, whose aspiration is, assumedly, to improve the overall detection process. This first filtering of the detection environment usually produces some electronic representation of the signal and the noise against which it is set. Now, the challenge for the work designer is to take this electronic stream of information and to display it most effectively on some sort of sensory display. This step necessitates a second transformation of the original data. Contingent on the level of understanding and creativity of the designer, the “to be viewed” display is now either more or less facilitative. As Drury (in press) noted, in the real world, there never exists the complete certainty and control that characterizes the experimental world. If there were, the initial technological step of detection would be sufficient for completely effective response (see also Mackie, 1987).

A second concern in respect of the signal and noise derives from their inherent constitution. The question arises, are the signal and/or the noise discrete or continuous events in space and time? In a somewhat trivial way, noise is always continuous in the natural environment, such that as soon as one has specified the distinguishing characteristics of the signal, then all other ambient stimulation becomes, by definition, noise. Further, because the world is in a state of constant evolution, the noise itself is an ever-present continuous stream of distraction. Targets, by contrast, are often but not always individually punctate events in space and time. For example, drivers on a roadway have a spectrum of distractions ranging from the overall driving setting to specific sources of in-vehicle and extravehicular distraction (see Hancock, Mouloua, & Senders, 2009). As drivers search for a specific target (e.g., a particular street name), each of the other environmental elements acts as a continuous source of distraction in the prolonged visual search. Whereas other street signs are highly confusable sources of distraction, background foliage is less so but nevertheless still a constituent component of the noise spectrum. Industrial inspection, radiograph analysis, and airport security screening tasks are each searches for punctate targets in displays that are themselves artificially designed to be discontinuous and limited frames of search. There remains, however, some contention over whether certain forms of visual search are examples of vigilance or whether they represent a somewhat differing order of behavior (cf. Wolfe, Horowitz, & Kenner, 2005).

These natural sources of signal genesis and a self-motivated imperative to sustained attention compared with derived, technical displays of previously processed information and an artificial, imposed imperative to engage in detection lie at the heart of the vigilance paradigm. In the latter circumstance, what is witnessed is the adaptive response to the genesis of a technologically originated iatrogenic disease. The two practical questions that now emerge are, first, is this an important disease and, second, can it be cured?

Iatrogenic Versus Important?

Even if vigilance itself is iatrogenic in origin, does this automatically mean that it is unimportant? The answer to this question is a resounding no. The pragmatic basis of this denial emanates from the obvious fact that there remain many tasks out in the real world in which vigilance is still mandated. Thus, understanding more of the factors that affect vigilance capacity will help people deal with those vestigial systems that still, unfortunately, impose this burden on their operators. Similarly and tragically, because
many systems designers remain unaware of this flaw in their creations, there continue to be many interfaces produced that require people to engage in enforced sustained attention performance. Indeed, as was noted initially, it is reasonable to assert that, in this technically mediated world, such conditions are actually growing rather than diminishing. Thus, vigilance is becoming more of a practical problem in the real world rather than being extinguished. So, importance is measured by practical, real-world impact, then vigilance continues to bedevil the modern world of electronic work. Whether there is a case to support the idea that vigilance should persist and to what degree future systems can be altered are not simply operational issues but are actually considerations that spill over into the teleological and moral dimensions of design (Hancock, 2009).

**Bringing Things to a Head: Is There a Cure for Vigilance?**

One of the most insightful commentaries on vigilance has been given by Scerbo (2001). He observed that vigilance is stressful because of the need to remain alert and combat boredom over extended periods of time. Vigilance tasks typically require individuals to work at a highly repetitive activity, in an under-stimulating and homogeneous environment, and to remain attentive for intervals determined by someone else. Simple changes to psychophysical task parameters can make the activity more or less difficult and thereby affect one’s perception of workload, but they do nothing to relieve the monotony. (Scerbo, 2001, p. 276)

The empirical question is whether one can design displays that people freely choose to observe and thereby negate many of the factors that Scerbo (2001) identified that induce vigilance failure (see also Scerbo, 1998). Or, in terms presented in the current work, can one create environments that encourage individuals to adapt their performance levels to actually express performance gains rather than performance losses? It will come as little surprise that I suggest that there are. After all, I believe that I can call on existence proofs that already occur in society. Quite simply, there are many forms of entertainment, such as television programs, movies, video games, and so on, with which individuals freely and voluntarily engage in prolonged periods of sustained attention and show not only little to no failure in this capacity but rather exhibit manifest interest and thus potential performance increments, such as recalling the nature of events in the presented narrative. Such behaviors are most especially obvious when individuals play modern video games (Robertson & O’Connell, 2010). Here, players become involved in many of the factors that Scerbo (2001) identified that require people to engage in enforced sustained attention performance. Indeed, as was noted initially, it is reasonable to assert that, in this technically mediated world, such conditions are actually growing rather than diminishing. Thus, vigilance is becoming more of a practical problem in the real world rather than being extinguished. So, importance is measured by practical, real-world impact, then vigilance continues to bedevil the modern world of electronic work. Whether there is a case to support the idea that vigilance should persist and to what degree future systems can be altered are not simply operational issues but are actually considerations that spill over into the teleological and moral dimensions of design (Hancock, 2009).

In the present work, I have sought to suggest that the vigilance decrement, and thus vigilance itself, derives from the activation of a form of self-regulated effort application (see also Robert & Hockey, 1997; Williams, 1986). The effort adjustment that this represents is contingent on the impoverished designs that accompanied certain burgeoning technologies of the Second World War. I have further suggested that as designers become familiar with this pattern of behavior that derives from inherently poor design, they can make appropriate changes that will alleviate much of this iatrogenic disease. But can the disease be eradicated altogether? That is, could vigilance be made a thing of the past? I predict not. Partly, this derives from people’s uncertainty of the world and their ignorance within it, for as Drury (in press) has also noted, “The problem with inspection tasks is that if you know ground truth in all cases no inspection would be needed so that the task itself could be eliminated.” If all displays were under the active control of beneficent designers who each had the best of intentions of creating operator (human)-centered designs, then this goal might possibly be achieved. However, sadly, this is not the case (see Dekker, 2011). The reason for this is that some actions in the world are fundamentally Manichean in nature, and it is this aspect that I now address.

In today’s world and throughout human history we encounter a continuing story of conflict. One result of the persistence of conflict is the obvious fact that certain of the involved parties must inevitably act to oppose. One expression of this opposition is that each group seeks to design conditions that deceive and defeat their enemies. With respect to the present observations, this means that conditions that create vigilance need not accrue simply from a particular designer’s unawareness of certain human propensities or even the vagaries of an unpredictable display in the natural environment; rather, the vigilance decrement may well come from those who specifically aspire to induce this failure in others. In short, enemy combatants are well advised to fabricate conditions exactly like those approximating traditional vigilance circumstances, hiding signals among proven distractors. In philosophical terms, people are thus often faced by Manichean forces whose manifest purpose is to oppose. Therefore, they will perhaps always have this source of display designers who will look to create conditions that very much promulgate the invigilance increment. We can always improve displays within

---

106 February–March 2013 • American Psychologist
our own control but not those within the power of those who seek to oppose. To complete this shortfall from the idea of extinguishing the vigilance disease completely, these displays will also occur in nature (e.g., the exploration of representations in astrophysics and quantum physics) where human powers of resolution are barely sufficient to raise the searched for signal above the naturally occurring background noise. Here, nature does not seek to oppose, but inherent physical characteristics of the limits of people’s technical powers to observe, induce the same conditions that were seen in operation above the Bay of Biscay so many years ago. In consequence and somewhat sadly, vigilance does not look like it is going away any time soon. Thomas Jefferson is often credited with observing that “eternal vigilance is the price of freedom,” but the careful application of the science of psychology can, perhaps, reduce the price just a little.

After all is said and done, a disquisition such as the present one should provide a number of simple summary points that serve to advance the positive state of the world. The following statements therefore summarize the present argument.

- Humans have encountered the problem of detection failure during long periods of enforced watch for many centuries.
- This incapacity was highlighted by the design shortfall of emerging electronic detection technologies under the driving pressures of worldwide conflict in the middle of the previous century.
- Transferred to the laboratory for decomposition and comprehension, a transient decrement was observed in the performance of largely task naive subjects.
- Encouraged to pursue this observation, Mackworth termed this phenomenon vigilance and found historic precedent for the term in the work of Head.
- It is arguably the case that such decrement was born of laboratory-induced conditions and may not be a fundamental characteristic of professional watchkeepers.
- It is also arguable that both the laboratory-induced decrement and the general failure in professional watchkeeping are primarily, if not exclusively, iatrogenic in origin.
- Such conditions can be considered a technologically induced disease.
- In this, the asymptotic level of average performance can be considered diagnostic of the impoverished state of the designed display to be monitored.
- Although intrinsically motivated sustained attention occurs in nature, it is only human society that imposes a rigid, extrinsic imperative to persist in vigilance.
- Targets can be hidden by nature, by poor technological display design, or by active (and even malevolent) intention. To a degree, detection capacity is mediated by each of these divergent sources.

From these observations, the vigilance decrement can be seen to reside as much in the way people encounter, organize, and convey work to the people that do it as it does in any intrinsic human failing. So vigilance and the vigilance decrement function are forms of technologically induced and socially mandated illness, composed of an individual’s reaction to certain maladaptive work circumstances. There are potential cures for this illness, although like all illnesses, not everyone falls prey to them. Some find it extremely damaging while others are barely disturbed, if at all, by the challenges such a disease poses. Whether psychologists are better off identifying those immune to this illness and thus limiting exposure to the least affected or trying to cure the basic disease is a matter of collective social determination.

**A Final Thought**

There is, however, perhaps a larger issue here that the iatrogenic nature of vigilance now exposes. That is, is the phenomenon of vigilance unique in its iatrogenic origins, or are there (many) other psychological phenomena sharing this same etiology? Indeed, how many other fundamental aspects of human performance are actually expressions of the amenable and adaptive human subject adhering or complying with the situational constraints and contexts laid out by the avid and enthusiastic but eventually misguided researcher? Although it cannot be denied that many investigators explore the fundamental limits of human capacity, how many paradigms purport to identify human shortfalls that are not embedded in any natural incapability but rather represent limits to the conceptual vision of the experimenters involved? This issue is especially important in applied psychological domains such as human factors, where the answers garnered to questions about nominally basic human functions are now let loose in a highly contextually driven world of complex systems operations (see Wagemakers, van der Maas, & Farrell, 2012). Little wonder that if many such basic processes are even partially iatrogenically created, there is highly limited transfer to real-world situations in which the adaptive human operator now responds very differently to the more complex portraiture of situational constraints laid before him or her. Although in the present work, I do not develop a fully exhaustive exploration and discourse on such matters, it is essential to lay down the basis of this challenge to psychology’s often unquestioned dogmas. Deeply bound up in the necessary, sufficient, and exclusive standards of evidence, the fact that this general form of iatrogenic phenomena may be heavily involved in clinical situations has not escaped my attention.

**REFERENCES**


