# Optimizing aid activation in adaptive and non-adaptive aiding systems: A framework for design and validation

| | |
|---|---|
| Citation | Greenlee, E.T., Funke, G.J., Hess, L.J., & Vidulich, M.A.. 2022. Optimizing aid activation in adaptive and non-adaptive aiding systems: A framework for design and validation. Applied Ergonomics, 101. https://doi.org/10.1016/j.apergo.2021.103677 |
| Citable Link | https://hdl.handle.net/2346/92709 |
| Terms of Use | CC BY 4.0 |

Applied Ergonomics 101 (2022) 103677

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

# Optimizing aid activation in adaptive and non-adaptive aiding systems: A framework for design and validation

Eric T. Greenlee [a,*], Gregory J. Funke [b], Lucas J. Hess [a], Michael A. Vidulich [b]

[a] Texas Tech University, USA
[b] Air Force Research Laboratory, USA

ARTICLE INFO

*Keywords:*
Adaptive aiding
System design
Human performance

ABSTRACT

Development of adaptive aids to support human performance in complex systems is a cornerstone of human factors. Research in this area has led to a diversity of ideas regarding potential activation methods. However, little guidance has been provided on how to select among aid activation methods, and this lack of guidance could hinder adaptive aid development and deployment. Within the current paper, we review available methods of aid activation and describe a process for developing and validating adaptive aiding systems. We focus on supporting system designers who wish to select the ideal aid activation method for an intended application. The process that we recommend is an empirical approach to evaluate the feasibility, costs, and benefits of various potential methods of aid activation. This methodological framework will support practitioners making critical decisions about the design of aiding systems.

## 1. Introduction

The limitations of human performance are a critical concern in many operational settings, including aviation, medicine, and military domains. Within these and other settings, human performance has been cited as a causal factor in accidents and errors that have culminated in loss of life and damage to material resources; more broadly, operational efficiency and effectiveness is frequently dependent on human performance (Stephens et al., 2012; Zacharias, 2019). Consequently, methods of human performance optimization are highly desirable because they may increase safety, effectiveness, and efficiency in a wide variety of settings.

Adaptive aiding is a method or system that reacts dynamically to the environment, task parameters, system conditions, and/or the user by qualitatively or quantitatively altering the nature of a user's task to maintain optimal levels of overall system performance (Scerbo, 2007). In principle, if human performance can be monitored directly, or inferred by reference to other variables, then it may be possible to detect suboptimal performance and intervene in some fashion to correct or improve the situation. For example, cognitive overload may be counteracted by a system that adaptively reduces the processing demands placed upon a human user (e.g., by temporarily delegating tasks to automation) when elevated levels of mental workload are detected. Though aids are frequently discussed with reference to supporting a single episode of task performance, they can also be deployed to support performance over longer durations. For example, if an individual's performance on a given task tends to wane over time due to fatigue, real-time monitoring of performance could be used to trigger break periods whenever performance begins to decline, thus adaptively providing aid (e.g., rest) that could restore performance or prevent further decrement during a work shift.

In short, adaptive aiding is a method of delivering supportive aid to a user when that aid would benefit user performance, including correction or prevention of degraded human performance (e.g., due to fatigue or distraction), and also enhancement beyond what would otherwise be "normal" performance. Given that human performance is a critical element in most domains, performance augmentation can enhance overall human-system performance. For example, an adaptive aiding system that monitors driver engagement and automatically warns a driver if they have become distracted may improve driver performance, thereby enhancing the safety and performance of the driver-vehicle system (e.g., by reducing the likelihood of traffic collisions).

### 1.1. Adaptive aiding design

To develop an effective adaptive aiding system, a designer must

---

* Corresponding author. 2810 18th Street Room 119, Department of Psychological Sciences, Lubbock, TX, 79409, USA.
*E-mail address:* eric.greenlee@ttu.edu (E.T. Greenlee).

https://doi.org/10.1016/j.apergo.2021.103677
Received 20 May 2021; Received in revised form 23 December 2021; Accepted 27 December 2021

determine 1) *what* aid to provide the user in a particular situation; 2) an aid activation method that determines *how and when* to activate that aid; 3) an aid deactivation method that determines *how and when* to deactivate aid; and 4) the minimal acceptable interval between aid activation occasions. These decisions will impact the potential benefit of an aiding system (e.g., Chen et al., 2017). The crucial importance of these decisions has been recognized, and researchers have provided useful information about available aid and aid activation options. For example, Feigh et al. (2012) generated a cogent review and taxonomy of types of aid (adaptations) and methods of aid activation (triggers). Their review includes discussion of examples, advantages, and disadvantages of various options, which usefully characterizes many of the options available for aid and aid activation methods.

Indeed, previous adaptive aiding research makes it clear that the options available for aid and aid activation are many and diverse. What is less clear is how a designer should decide among available options, because no concrete guidance has been given to facilitate such choices. Consequently, the current paper was written to complement and build upon existing adaptive aiding research (e.g., Feigh et al., 2012) by providing explicit guidance on how to select among the many available aid activation methods.

To do so, we present a decision framework that is broadly based on principles of system engineering, as an approach to quantify the ratio of benefits and financial costs of various system design alternatives to determine which option(s) will provide a positive return on investment. Although system engineering is a well-established and effective method for making system design decisions, the application of system engineering principles to decisions about adaptive-aiding system design is a novel approach that should provide designers objective and definitive answers about the relative benefits of proposed systems for a given application.

To permit ourselves to provide generalizable guidance on how best to select an aid-activation method, we assume that a designer has already selected a type of aid, a deactivation method, and an aid interval that they believe will improve performance. Explicit guidance for how to make decisions about those other components of an aiding system is beyond the scope of the current paper and will need to be addressed in the future.

Critically, we also assume that designers have identified a target application for which they intend to implement an aiding system. The steps within this guide should be completed while considering all relevant task contexts within which aid will be proffered.

### 1.2. Aid activation selection

Designers should carefully consider the potential costs and benefits of all reasonable aid activation methods, relative to each other, when designing a system. This evaluative comparison should permeate the design process in order to maximize the efficiency of system development and ensure that the efficacy of the resulting aiding system justifies the costs of its deployment. Consequently, we suggest that consideration, development, and deployment of aiding systems be guided by four overarching issues that represent crucial steps for selection of an appropriate aid activation method. Those steps include:

- *Feasibility Assessment*: A designer must determine what methods of aid activation would feasibly be expected to augment human-system performance in the target application.
- *Benefit Quantification*: Feasible aid-activation alternatives must be subjected to validation testing to quantify the relative performance benefit of each alternative design.
- *Cost Quantification*: Feasible aid-activation alternatives must also be examined to quantify the expected costs to the deploying organization and monitored user.
- *Cost-Benefit Analysis*: Final selection of an aid-activation method should be based upon evaluation of the cost-benefit ratio associated with each alternative.

These steps represent sequential challenges that must be addressed to fully demonstrate that adaptive aid activation is the best option for augmenting user performance. The remainder of this paper provides an in-depth, sequential exploration of these issues, as well as decision aids and recommendations that we hope will guide future adaptive system designers toward informed choices about the expected value of such systems.

## 2. Feasibility assessment

Initial deliberations should be focused on determining which adaptive and non-adaptive activation methods are likely to be feasible, reasonable solutions for performance augmentation. The feasibility of adaptive aiding will vary by application, as will the feasibility of alternative, non-adaptive systems (i.e., scheduled aid, static aid). To be considered feasible, a method of aid activation must be 1) possible, and 2) the designer should consider it reasonable to expect the activation method to enhance performance.

To support these initial forecasts regarding the feasibility of alternative methods of adaptive and non-adaptive aid activation, we have created a checklist that outlines key elements of the decision process required to evaluate feasibility of alternative system designs (see Fig. 1). The early portion of the checklist (alternatives 1–2) present non-adaptive methods of aid activation. The remainder of the checklist (alternatives 3–6) includes adaptive methods of aid activation. The order of items in the checklist corresponds with our view that development and fielding of an effective aiding system will tend to become more challenging (and potentially more expensive) as the focus of aid-activation shifts from overall human-system performance towards internal user state and the aid-activation method grows more complex.

Because human-system performance is the end result of human-system interaction, variables that are immediately related to overall human-system performance are likely to provide a relatively holistic assessment of the situation and likelihood that user performance, and human-system performance in-turn, would be augmented by aid delivered to the user. As the variable used for aid activation becomes more focused on the user state or performance within specific tasks, rather than overall human-system performance, the efficacy of an aid activation method becomes increasingly dependent on the degree of correlation between task performance or user state and overall human-system performance. Consequently, effective aid-activation based upon user state requires an understanding of the complex relationship between user performance or state and overall human-system performance. The complexity of various aid-activation alternatives also tends to increase as they are made more flexible, or adaptive, especially when adaptation is triggered automatically, or automated.

While successful development of an adaptive-aiding system is possible with complex user state driven aid-activation, and in some cases such systems may be ideal, we would encourage designers to initially consider the feasibility of simpler aid-activation alternatives, especially when alternatives provide aid based upon more direct measurement of human-system performance. The order of alternatives presented within the checklist (Fig. 1) reflects this recommendation. System alternatives that appear earlier in the checklist are relatively simple, comparatively focused on overall human-system performance, or both. Later alternatives rely upon increasingly complex methods of aid activation that are driven by assessment of a user's performance and/or state.

We formulated this decision sequence with the view that the potential benefit of each option differs on situational factors that are likely to be application-specific. Thus, relative benefit cannot easily be generalized and is not represented within the presented decision sequence. Instead, we present a template for feasibility assessment that a

**Static Aid** (see section 2.1)
Feasible If: Continuous aid would result in effective performance augmentation

**Scheduled Aid** (see section 2.2)
Feasible If: A predetermined schedule of aid would augment performance

**System/Environment Activated Adaptive Aid** (see section 2.3)
Feasible If: A measure of system and/or environment could be monitored often enough and with sufficient lead time to activate aid that would enhance performance

**Performance Activated Adaptive Aid** (see section 2.4)
Feasible If: Human performance is a meaningful index of human-system performance that can be monitored often enough and with sufficient lead time to activate aid that would enhance performance

**User-Initiated Adaptive Aid** (see section 2.5)
Feasible If: 1) User can conduct valid, reliable self-appraisal of their own performance, functional state, and/or cognitive state to determine when they need aid, 2) they can self-appraise and activate aid without interfering with human-system performance, and 3) user requests for aid will allow enough lead time to enhance performance

**User State Activated Aid** (see section 2.6)
Feasible If: There is a valid, reliable, and acceptably non-intrusive index of user state that is predictive of performance and provides state-assessment often enough and with sufficient lead time to activate aid that would enhance performance

**Fig. 1.** Feasibility Checklist for Aid Activation Methods. Numbers after each entry refer to specific subsections appearing in this manuscript.

designer should use in combination with their own expertise to determine whether each aid-activation option is likely to be worthwhile. Specifically, we encourage designers to follow the depicted decision sequence by evaluating, in turn, each alternative aid activation method for feasibility. A short description of each aid activation method is presented below (Sections 2.1-2.6; and also see Feigh et al., 2012). In addition, further information regarding the relative benefits of each method is presented in Section 3.

We suggest that designers evaluate all types of aid activation methods and identify any that are feasible for the intended application. All feasible alternatives will need to be evaluated for the performance benefits, cost, and cost-benefit ratio (Sections 3-5 of this manuscript). The remainder of the current section follows the decision sequence depicted within Fig. 1, and each item within that checklist is represented by a single subsection, labeled to match checklist labels.

### 2.1. *Static aid*

Continuous, static aid is a method of constant aid delivery. This approach evaluates the importance of the *adaptive* component of aiding and suggests that a designer consider whether the aid would be effective if provided continuously, or if there would be any negative consequences to static aid. One example of static aid is power steering in motor vehicles. Power steering systems employ intermediate hydraulic or electric devices to augment the force applied by the driver through the steering wheel to turn the wheels of a vehicle from side to side. While a vehicle with functional power steering is operating, it is always enabled – it provides static aid to the driver.

The primary appeal of static aid is the simplicity provided by constancy. It requires no considerations of *when* to provide aid because the aid is ever-present. However, there may be instances or settings where static aid would have a detrimental impact on human-system performance, safety, or cost. For example, when automation is provided as static aid, individuals interacting with the automation tend to become complacent (Parasuraman and Manzey, 2010), lose situational awareness (Endsley and Kiris, 1995), and suffer from skill degradation (Haslbeck and Hoermann, 2016); these issues can reduce overall human-system performance. Constant aid may be prohibitive for other reasons. For example, augmentation via psychoactive stimulants might augment performance, but exposure to such substances also entails risks of addiction and other associated health consequences.

### 2.2. *Scheduled aid*

These systems deliver aid at times that are predetermined to occur at specified intervals. Developing a successful schedule of aid requires an understanding of the temporal dynamics that influence system performance. This knowledge can then be used to judge *when* aid should be delivered. In instances where a target application requires augmentation of a temporally predictable system or phenomenon, a system of scheduled aid is likely to be effective. For example, from evaluations of the rate of fatigue onset in long-haul truck drivers, legislators have mandated that truck drivers take a 30-min break after 8 h of driving (Federal Motor Carrier Safety Administration, 2015). Similarly, scheduled aid is provided to Transportation Security Agency security personnel in the form of forced task switching (e.g., from baggage screening to ID checking) every 30 min to prevent time-on-task related declines in performance that have been observed (Transportation Security Administration, 2008). As these examples illustrate, if the temporal dynamics of a system are predictable, then a fixed schedule of aiding can be effective.

Unfortunately, not all systems are predictable enough to allow formulation of a fixed schedule. Rather, the variability inherent in many systems reduces the feasibility of scheduled aid. For example, it may be difficult for a designer to develop a predetermined schedule that can predict when an individual will become distracted from their task.

### 2.3. *System/Environment Activated Adaptive Aid*

This method uses a measure of the environment, system state, or system performance that can be monitored and used to automatically trigger aid. The key advantage of aid driven by environmental variables and system state is a focus on directly measuring overall system state and performance. The effectiveness of system/environmental activation is dependent on an accurate understanding, by an automated system, of the “ground truth” regarding critical environmental variables, system state, or both. Thus, effective system/environmental activation is precluded in operational settings that prevent an accurate account of these variables by an automated system.

An example of an environmentally-activated adaptive aid is an automatic braking system; vehicles equipped with these systems use sensors capable of monitoring speed and headway to assess the risk of collision with obstacles in the roadway. In cases where a collision is imminent, aid, in the form of emergency braking, can be triggered automatically, reducing response time and increasing safety. Similarly, many modern fixed-wing aircraft include a Ground Collision Avoidance System (GCAS) that monitors a plane's position relative to environmental terrain and triggers automatic ground avoidance maneuvers (e.g., altitude climb) to prevent a crash. Both of these systems manage high-risk situations that can occur due to degraded human performance, and they each do so by reacting to the relationship between the system and environment, momentarily reallocating functional control of the system (car or plane) to automation.

Effectiveness of these aids is also dependent upon the ability to reliably predict the presence of suboptimal states before/when aid needs to be delivered. In applied settings, this can be a challenge because system performance degradation and concomitant potential harm can rise sharply with little warning. Effective deployment of supportive aid requires the system-environment state to be measured (i.e., sampled) frequently enough to detect or predict the change from normal, safe operating conditions to dangerous conditions with sufficient time to prevent onset of the suboptimal state (or at least reduce its severity).

At the very least, the interval between measurements can be no longer than the duration of the suboptimal event or state that aid is meant to prevent. If the time between measurements is greater than the duration of critical events, the window for effective aid may pass undetected. Further, it has long been established that sampling must occur at a rate that is at least twice as frequent as the temporal function being sampled (i.e., the Nyquist rate) to avoid distortions (aliasing) due to sampling rate (e.g., Rompelman et al., 1977). That is the bare minimum required to *detect* a suboptimal system state, but this will not guarantee enough time to execute aid.

In addition, the processing time required to convert sensor data into a meaningful assessment of performance, and the time required to execute aid must also be considered, because both combine to determine the lag between state detection and aid delivery. An adaptive aiding system must be designed to allow enough lead time between detection of a suboptimal state and execution of supportive aid to maintain or return to desired levels of performance. Generally, adequate lead time is made more likely as the sampling rate of sensor measurements increases, as the time required to assess the need for and execute aid decreases, and as the "tolerable" duration of a suboptimal or risky system state increases.

In sum, the monitored variable(s) must be measured often enough to reliably detect or predict suboptimal system states with enough lead time to allow effective, timely aid. This limitation is not unique to System/Environment Activated Adaptive Aid but is a concern for all automatically-initiated forms of aid activation.

### 2.4. *Performance activated adaptive aid*

For the purposes of this framework we have defined human performance somewhat narrowly, as a direct measure of performance (e.g., accuracy or speed) or behavioral inputs (e.g., button presses) in one or more primary task. This definition includes human actions, but excludes the impact of those actions on system state or system performance, which were covered in section 2.3. For example, measures of a driver's manual steering inputs would be included in this category, but measures of a vehicle's lateral position in a lane would not be. The present definition of human performance also excludes behaviors such as posture or physical fidgeting. For our purposes here, we will regard such behaviors as examples of physiobehavioral variables (described further below). Finally, our definition excludes performance on secondary tasks that may be added to an operational setting to assess human cognitive or functional state (e.g., workload; Ogden et al., 1979); such measures are also represented later in this framework.

Broadly, human performance on a user's primary task(s) may provide an objective estimate that is related to system-wide performance. Thus, if a human user's performance can be measured objectively (e.g., via reaction time, accuracy, behavioral inputs) against a predetermined threshold, these measurements may be used to infer overall human-system performance and predict the need for aid. For example, some modern driver assistance systems generate a profile of driver performance based on a driver's typical pattern of steering wheel, gas pedal, and gearshift inputs; when a driver's inputs deviate from their profile (e.g., too little steering input) aid is provided in the form of auditory and/or visual alerts (Audi, n.d.). Performance activated adaptive aid has also been recommended as a potential method for improving performance of image analysts involved in surveillance (Calhoun et al., 2011) and air traffic controllers (Kaber et al., 2006), for example. However, human performance activated adaptive aid is generally impossible in settings where human performance cannot be measured in real time, where meaningful thresholds cannot be established (e.g., due to variability), and where there is no ground-truth against which to compare performance.

There are other settings where human performance could be measured in real time, but the temporal dynamics of the task may preclude useful performance-based activation of aid. Tasks that provide few opportunities to respond also provide few opportunities to assess performance. For example, many monitoring duties (e.g., automated vehicle driving) involve long periods of time during which no response opportunities arise, unless a critical and potentially hazardous event occurs (e.g., automation failures) that requires swift and accurate action from the human user (e.g., Greenlee et al., 2018). In those instances, performance could be measured from whether or how well a user is able to respond to the critical event (e.g., imminent vehicle collision). However, the risk associated with suboptimal human performance in such instances can be extreme and can include harm to the user, the system, or both. Consequently, in settings with few opportunities to assess human performance, aid based on performance may be aid that comes too late.

Additionally, there may be more than one way to measure human performance in some settings (e.g., speed vs. accuracy). In such cases, designers should select the variable or combination of variables that best predicts overall human-system performance. If the most predictive variable(s) are unknown, it may be necessary to conduct pilot studies to identify which of the available performance measures best predicts overall human-system performance in the chosen application; such pilot studies should precede the steps of this guide.

### 2.5. *User-initiated adaptive aid*

Aid that is activated intentionally by the user. User-initiated aid circumvents the need for a system designer to determine when aid should be applied because that responsibility is assigned to the user. Effectively, users are asked to monitor themselves and introspectively assess their own performance, capabilities, and/or functional state relative to an internalized standard so that they can request or activate aid when it would be beneficial or necessary. To the degree that a user can provide reliable, valid, and timely self-assessment and activation of aid, user-initiated aid can be a highly flexible, efficacious method of activating adaptive aid.

There are many examples of user-initiated adaptive aiding systems in the real-world. For instance, many current motor vehicles include cruise control, a function that automatically maintains vehicle speed at a value set by the driver. Similarly, some current motor vehicles include automation technologies that are capable of maintaining safe headway and lane position in some environments (SAE International, 2016). In both cases, the driver is solely responsible for activating the automation, meaning that the driver decides when to receive aid (vehicle automation) within these contexts. Outside of the driving context, the aid provided by virtual assistants such as Apple's Siri and Amazon's Alexa are also user-initiated, given that these devices idle until called upon to provide information or complete tasks for their users.

However, user-initiated aid is limited by a user's willingness to use aid appropriately and their ability to conduct self-appraisal and ask for aid when necessary. Automation research has made it apparent that human users do not always use automated aids appropriately and may instead ignore them, rely too heavily upon them, or use them in situations where they are inappropriate (Parasuraman and Riley, 1997). These same issues are a concern for user-initiated aid, as the user is granted agency to decide when and whether to activate aid. Consideration of user-initiated aid should also include evaluation of the burden associated with activating aid. Research suggests that user-initiated aid is, in effect, a secondary task that requires users to dedicate cognitive resources to determining whether or not they need aid (Bailey et al., 2006). In some cases, this added demand could have deleterious effects on primary task performance, and in others, the demand of a primary task may be too great for a user to properly evaluate whether they need aid – meaning that a user may neglect to request aid in high workload moments where it could be most beneficial. Given the information processing requirement associated with user-initiated aid, the reliability, validity, and timeliness of such systems is likely to vary as a function of task demands and user state (e.g., cognitive load, fatigue). This variability may be exacerbated by individual differences in capability as well as differences in how individuals evaluate their need for aid.

### 2.6. User state activated aid

The final alternative is a class of adaptive-aiding systems that are automatically driven by measures of the human users' cognitive or functional states. The fundamental rationale for this type of system is the notion that human functional and cognitive state is closely related to human performance. For example, a well-rested individual tends to perform better than someone who is tired (Matthews et al., 2012); individuals perform more poorly when cognitively overloaded than when given a more moderate cognitive workload (Tsang and Vidulich, 2006); and a user's degree of attentional engagement with a task tends to correspond positively with their readiness to react to task-critical events (Matthews et al., 2010). The correlation between task-relevant user states and human performance can be leveraged into a viable adaptive aiding system that employs measurement of user state to predict current or future performance and the need for aid (Loft et al., 2018). This potential has been acknowledged by many who have suggested user state monitoring may be particularly useful as a means of activating adaptive aid (Galster and Johnson, 2013; Parasuraman and Galster, 2013; Zacharias, 2019).

Broadly, state measures can be categorized based on measurement modality; common modalities of state assessment include secondary task performance measures and measurement of physiobehavioral indices of user state. Secondary task performance measures infer an individual's state based on their performance (e.g., accuracy, reaction time) on one or more secondary tasks (Ogden et al., 1979). Physiobehavioral measures have been developed to index operator state based on neural (e.g., electroencephalographic; Gevins and Smith, 2007), physiological (e.g., electrocardiographic; Mulder et al., 2004), or behavioral variables (e.g., posture; Huxhold et al., 2006).

One potential benefit of state assessment is the potential for diagnosticity; while direct measures of performance may provide excellent sensitivity to suboptimal system states, human state assessment may be helpful in explaining *why* suboptimal states occur. Comparably suboptimal performance states can result from very different user states. For example, both underload and overload have been linked with poor user performance, yet the aid necessary to correct each of these states is drastically different (Greenlee et al., 2019; Young and Stanton, 2002). Another potential advantage of state-based adaptive aid is the lead time that it may afford relative to direct performance measurement. In some cases, user state degradation may be measurable long before performance changes are evident. For example, users may compensate for cognitive fatigue by exerting additional effort, delaying the onset of a performance decrement (Funke et al., 2010). If cognitive fatigue and/or compensatory effort could be assessed directly, these measures of state could be used to predict performance changes before they occur.

Of the measurement modalities, physiobehavioral measurement of user state may have the greatest potential for use in adaptive aiding systems because it can provide near-real-time state estimates with a relatively high degree of temporal resolution, and monitoring can be relatively non-intrusive (e.g., Greenlee et al., 2021; Kramer, 1991; Matthews et al., 2015). Secondary task performance measures of user state can intrude upon primary task performance (Eggemeier, 1988), require attention and input from the user in order to assess user state, and draw a user's attention and limited information processing resources away from a primary task(s), potentially degrading performance. However, despite arguments and evidence for the non-intrusiveness of physiobehavioral monitoring, intrusive effects may be possible with untested physiobehavioral systems and/or applications (Greenlee et al., 2021).

User state activated aid is currently implemented in a variety of domains. For instance, some manufacturers of partially automated motor vehicles (e.g., Tesla, Nissan, Audi, etc.) have implemented behavioral systems to monitor driver engagement while vehicle automation is engaged. Such systems are capable of detecting whether a driver is gripping the steering wheel, and steering wheel grip is used to make a binary inference about a driver's engagement and readiness to resume manual driving should the need arise (e.g., automation fails). If the driver ceases gripping the steering wheel, the driver is considered to be disengaged, and automated aid is delivered in the form of warnings to encourage driver reengagement. If the driver does not reengage, aid escalates to deactivation of automated driving functions (i.e., car may automatically pull over and stop).

Physiologically-based user state monitoring has also occasionally been implemented to drive adaptive aid. For example, some exercise-focused video games have been developed that utilize postural tracking (e.g., Mueller et al., 2007) or heart rate (e.g., Stach et al., 2009) to dynamically adjust task difficulty. The goal is to provide users with an effective workout. To that end, if the game detects that a user is exerting too much or too little effort, game difficulty is automatically adjusted to maintain optimum effort. Similar concepts have been proposed to automatically control workout difficulty with exercise machines such as treadmills (e.g., Zhao et al., 2020).

One of the primary challenges in the development of a state-based adaptive aiding system is selecting a measure of user state; there are many options, and not all may be equally suited in a specific application. A full account of task-relevant user states (e.g., cognitive workload, fatigue, vigilance, etc.) is beyond the scope of this paper, as is a comprehensive review of the many measures that have been developed to assess those states. However, designers will need to determine which functional or cognitive state(s) is likely to be most predictive of performance in the target application, and how it would be best to measure that state (s). These decisions may require preliminary pilot studies to verify that the chosen user state measure(s) is a meaningful predictor of performance in the chosen application; such pilot studies should precede the steps of this guide.

## 3. Benefit quantification

Once feasibility decisions are made, they must be validated empirically. Empirical validation provides experimentally-derived, quantitative data that can be used to establish the degree of augmentation provided by each of the tested aid activation methods, which can be used as an index of the relative benefit of tested methods. It should be noted that design decisions, such as specific temporal intervals and thresholds for aid activation, are likely to influence the sensitivity and accuracy of aid systems, and may influence validation comparisons; designers must use a combination of existing research, their domain knowledge, experience, and best judgment when determining these characteristics for inclusion in their validation testing.

Our aim in this section is to describe methods for effective validation. We will continue to refer to the decision framework described in Section 2 because this framework helps to determine the alternatives against which selected aid activation methods should be compared during validation. As a rule, we suggest that each feasible activation method should be compared against the alternatives earlier to it in the framework; the later the activation method is within the framework, the more alternatives it must be compared against to demonstrate that it is superior to simpler alternatives. The rationale for this, again, is that aid activation methods become more complex and less directly related to human-system performance as a designer progresses through the decision framework, so each feasible method should be compared against methods that rely upon relatively simple and/or comparably direct measures of human-system performance.

These comparisons can be accomplished through experimentation. The only necessary independent variable in validation experiments is *aid activation method*, with levels (groups) corresponding to activation methods that were identified during the feasibility assessment stage and the systems below those in the feasibility framework. The effectiveness of each method should be assessed by directly measuring human-system performance, and performance comparisons among conditions should be conducted.

A carefully controlled research environment is needed for validation testing so that a designer can be certain that the only systematic difference between the tested conditions is the method by which aid is activated. Also, as mentioned previously, some operational environments do not permit measurement of a user's performance (e.g., performance is not quantifiable); in such cases, it will be necessary to create a simulation of a user's duties that will allow direct performance measurement. Actual or simulated testing situations should be designed to include the full breadth of intended use cases for the aiding system to ensure that measures of benefit are reflective of any variability in the intended application. Table 1 presents a full account of the conditions that we would prescribe for each of the alternative methods of aid activation.

The table contains some control conditions (e.g., no aid) that will be discussed later in this section. The table also reflects the relative complexity of validation across aid activation types. For example, validation of static aid requires only one comparison condition, whereas user state activated aid requires eight comparison conditions to validate. Inclusion of all recommended conditions is necessary to *fully* validate a selected method. If a designer chooses to exclude a recommended condition, they are excluding an alternative aid activation method from further consideration. This could lead a designer to believe a complex system (e.g., performance activated) is the best option, when a much simpler alternative (e.g., scheduled aid) could provide comparable (or better) benefit and/or comparable (or better) cost-benefit ratio.

However, we understand that there may be some scenarios where these risks must be accepted to reduce financial or temporal costs of development. Likewise, designers may choose to exclude conditions where an aid activation method is impractical to implement or their chosen aid would be incompatible with an aid activation method. For example, some forms of aid that are naturally periodic in function (e.g., information or decision support from a virtual assistant, automatic braking, etc.) may be difficult to realize, or perceptually indistinguishable from normal function, if rendered as static aid. In addition, as noted previously, it may be difficult to implement system/environment and performance activated aid methods in the "real world," where it can be difficult to establish ground truth outside of a controlled testing environment. Though we encourage designers to think creatively about alternative measures that may serve as useful surrogates in those environments (e.g., time to complete a standard task, rather than reaction time), if no such measures are available, designers may choose to eliminate them from testing.

Finally, we acknowledge that in some circumstances an industry standard for aid may already be established. Under those circumstances, it may make sense for designers to omit some conditions (e.g., the "no aid" condition) provided that previous validation has already established the benefit of that industry standard relative to those omitted conditions, or that safety or regulatory measures dictate that those conditions should be omitted. For example, as mentioned previously, federal regulations require truck drivers to take a 30-min break after 8 h of driving (Federal Motor Carrier Safety Administration, 2015). Given this mandate, it would make little sense for designers working on aids for truck drivers to consider testing a "no aid" condition.

Within this section, we will present validation requirements as if user state activation has been selected as feasible by the designer and an industry standard for aid has not yet been established. We do so because user state driven aid is the terminal option within the decision framework, and as a consequence, such a system should be validated against all previous alternatives within the framework. That is, the requirements for validating state activated systems are the most extensive. Note that we assume that a designer who selects user state as a feasible method of aid activation has identified a single variable or variable set that they plan to employ to monitor user state and activate aid. We assume that all deliberations about what measure of user state to employ have ended and that comparisons of various user state sensors is not an aim of validation testing. If that assumption is violated, meaning a designer wishes to compare different methods of state-activation within a single validation study, the number of required conditions will increase by two for each additional state measure tested (+1 state activated experimental condition, +1 yoked control condition; described below).

The remainder of this section is organized by alternatives, such that each activation option that appears within the decision framework presented in Section 2 is represented by a separate subsection here. These subsections follow the order represented in the decision framework and each outlines the purpose and procedure of validation testing that compares one alternative option to user state activated aid. Although we have focused our discussions on validation testing of a state

**Table 1**
Control conditions recommended based upon selected aid activation method(s).

| | Conditions Recommended | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Aid Activated By | No Aid | Static Aid | Random Schedule | Rule of Thumb Schedule | Yoked Schedule | System/ Environ. | Performance Activated | User-Initiated | State-Activated |
| User State | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| User-Initiated | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Performance | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| System/ Environ. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | |
| Scheduled | ✓ | ✓ | ✓ | ✓ | | | | | |
| Static | ✓ | ✓ | | | | | | | |

activated system, it is equally important to validate the other types of aiding systems that are considered to be feasible.

### 3.1. *No aid*

To begin, we highly recommend that validation studies – no matter the method of aid selected – should compare the performance resulting from the developed adaptive aiding system against performance in the absence of aid (given an industry standard has not been established). Previous efforts to validate adaptive aiding systems have frequently included a "no aid" control condition, wherein users do not receive any aid (e.g., Prinzel et al., 2003; Wilson and Russell, 2007). If performance is better in the state activated aid condition than in the no aid condition, this provides initial evidence that the developed state activated adaptive aiding system is beneficial. However, additional comparisons will be needed to determine whether the state-based adaptive activation provides any added benefit beyond the aid itself or if the aid would be better activated by a different method. If performance in the state activated aid condition is not superior, it may not be readily apparent whether this is because the activation method is poor, the aid is poor, or something else about the adaptive aiding system is problematic (e.g., intrusiveness). If this happens, the designer will need to decide whether to conduct supplemental experiments to disentangle these potential causes, but such investigations are beyond the scope of this manuscript.

### 3.2. *Static aid*

Another important control condition in validation testing of state activated adaptive aiding is a condition where aid is presented continuously. In this "static aid" condition, aid is never activated or deactivated – it is constant. The static aid control condition allows designers to test the importance of the *adaptive* feature of adaptive aid. For user state activated adaptive aiding to be worthwhile, its application must, at the least, result in either superior performance outcomes, or fewer potential negative drawbacks, relative to application of static aid.

### 3.3. *Scheduled aid*

As discussed in Section 2, there are some situations that are well-suited to predetermined schedules of aid, and it is important when attempting to validate state activated adaptive aiding to evaluate the possibility that scheduled aid could be effective. There are two primary methods available to designers to create scheduled aid. Each should be represented by a separate control condition.

The first is to assess the temporal dynamics of the task, user performance, or user state to determine whether there is a clear temporal pattern or periodicity that could be used to predict the need for aid. For example, if users, on average, can perform a challenging task for 20 min before beginning to show signs of fatigue (e.g., errors), perhaps providing aid (e.g., rest break) every 20 min would be effective. We would suggest that if a designer can detect temporal regularities in the to-be-optimized phenomenon, the designer should develop this type of "rule-of-thumb" schedule. This schedule should be tested as a control condition against state activated aiding.

An alternative method of schedule development is made possible by adaptive aiding validation testing. A temporal pattern of activation is necessarily created by the behavior of an adaptive aiding system as it activates and deactivates aid based on a user's state during a task. Adaptive aiding research often involves a yoked control condition (e.g., Prinzel et al., 2003), in which participants receive the same schedule of aid that participants in the state activated aid condition receive. This is accomplished by recording the schedule of aid that is delivered to one participant in the state activated aid condition, then delivering that schedule of aid to one participant in the yoked control condition. Each participant in the yoked control condition receives the schedule from a different participant in the state activated aid condition, meaning that the yoked participants are paired with state activated aid participants. Whereas participants in the state activated aid condition receive aid that is based upon their own state (e.g., changes in workload), participants in the yoked control condition receive aid that is based on someone else's state. If performance in the "yoked" control condition is worse than performance in the state activated aid condition, this is evidence in favor of the value of personalized aid that is tailored to the user in real time. Conversely, if "yoked" performance is no worse, this may suggest ineffectiveness of the real-time adaptations afforded by the state activated system. Note that if the "yoked" schedule is the most effective option, development of a fixed schedule that can be applied in operational settings will likely require some type of aggregation across a pilot sample of user's who received adaptive aid to create a singular schedule.

In addition to "rule-of-thumb" and "yoked" schedules, we would recommend one final schedule-based control condition, a random schedule. Both the "rule-of-thumb" schedule and the "yoked" schedules are created based on the idea that the schedule of aid matters, i.e., that if the temporal pattern of the to-be-optimized phenomenon can be identified or approximated and matched against a schedule of aid, then performance will be optimized. Alternately, however, it could be that scheduled aid provides performance augmentation simply because it turns aid on for periods of time. The same criticism can be levied against state-based adaptive aiding (and all other adaptive aiding systems). Perhaps the schedule of "when" that aid activation occurs does not matter. To test that possibility, we would highly recommend that designers include a random aid condition. In this condition, the total duration of aid and the number of times aid is activated should be yoked to the state-based adaptive aiding condition, but the aid should be activated at random moments within the task for those in the random aid condition. If performance within the random aid condition is comparable to or better than performance in the state activated aid condition, this may be evidence that the scheduling function provided by the adaptive aid is unnecessary and perhaps not beneficial.

Note that if a designer is attempting to validate any adaptive aiding system other than a state activated system, the designer will need to base a yoked control on that activation method. For example, if a performance activated adaptive aiding system is developed as the preferred method being validated, the schedule of aid activation should be recorded in that performance activated condition. Then, the schedule created by aid activation for each participant in that condition should be paired with each participant in a yoked control condition, and the total duration of aid and number of aid activations in the random aid condition should similarly be derived in a pairwise fashion from the performance activated aid condition. If multiple adaptive aiding systems are considered feasible, comprehensive validation would require that each be tested against a yoked control condition.

### 3.4. *System/environment activated aid*

If it is possible, we would recommend inclusion of a system/environment activated aiding system in validation testing. We recognize that the system and environment may not be usefully measurable in some settings. However, given that system/environment activated aid is likely to provide the most direct assessment of human-system performance, we would suggest that designers test the best reasonable system/environment activated aid against state activated aid, if possible.

### 3.5. *User-initiated aid*

When validating a user state activated adaptive aiding system, it is very important for designers to include a control condition that allows users to toggle and control aid, rather than receiving aid that is automatically initiated by an external system that monitors user state. Performance resulting from a volitional, user-initiated control condition should be compared against the performance that results from automatic, state activated aid. Bailey et al. (2006) provide an excellent example of such comparisons, which are important because they help to evaluate the value of automatically-initiated state assessment relative to user-initiated aid. If performance is greater with the support of automatic, state activated aid, this supports the notion that state activated aid may be beneficial.

### *3.6. Performance activated aid*

If human performance is measurable and potentially useful within the target application, we would also suggest including a control condition that tests the efficacy of a performance activated adaptive aiding system. If it is measurable, performance in the performance activated control condition should be compared against performance in the state activated condition to determine the relative benefit of each in terms of performance outcomes.

### *3.7. Minimal control condition testing*

As mentioned previously, while we advocate for designers to consider all feasible aid activation options so that they may arrive at the most efficacious solution available, we also acknowledge that full validation testing that includes all suggested control conditions may not be possible or desirable (for reasons of cost, time, etc.). In that situation, we suggest that designers consider including a minimal set of comparisons that should provide adequate, but incomplete, information from which to make informed decisions. Designers should keep in mind, however, that this may lead to system designs that are more costly, complex, and/or inefficient than what could be achieved following the full validation schedule.

Proceeding from the assumption that a user state activated aid is under consideration, we suggest designers include the no aid condition (or the industry standard, if it exists), the scheduled aid or rule-of-thumb condition, and the user-initiated aid condition. These conditions will, respectively, allow the designer to assert that the aid provides benefit relative to no aid (or standard aid), and that the aid provides better benefit than a temporally driven program of aid and aid provided on demand, thereby providing some assurance that the relatively more complicated methods of activation are justified.

### *3.8. Evaluation of relative-benefit*

Determination of whether a method (e.g., state-activation) is “best” should be based primarily upon performance comparisons between the state activated aid condition versus each control condition (i.e., alternate activation methods). If state activated aid results in the best performance outcomes, then a designer may proceed with reasonable confidence that their state activated system is the most effective of tested options. Further, the resultant performance estimates inform the designer about how much benefit a selected method may provide, relative to alternatives.

## 4. Cost Quantification

In an ideal world, designers might simply select the aid activation method that demonstrated the greatest capacity for performance augmentation in validation testing (3. Benefit Quantification). Yet, in reality, implementation of any aiding system is likely to come with costs to the user and/or the overarching organization. In order to make fully informed decisions about the value of a novel aiding system, designers must be cognizant of both benefits and costs.

Below we provide a brief, non-comprehensive discussion of the costs associated with programs of adaptive aiding based on an index of operator state. This case was adopted because user state monitoring may incur unique costs. Other methods will likely require some, but not all, of the costs discussed below. For example, scheduled rest breaks may be perfectly acceptable to users, whereas use of heart rate monitoring to automatically prescribe rest might not be acceptable.

We have grouped costs into two categories for discussion – *direct costs* (costs associated with development, implementation, and maintenance) and *indirect costs* (costs associated with users' time, intrusiveness, acceptance of an aiding system, and hiring and retention).

### *4.1. Direct costs*

Implementation of an adaptive aiding system is likely to be a costly endeavor for any organization. Below we present a short synopsis of direct costs that implementation of an aiding system will likely require. Our discussion here is not meant to be a comprehensive accounting of all possible financial costs, or to recommend specific models of cost estimation; rather, it is a reflection of the sources we perceive as significant based on our collective experiences conducting research on adaptive aiding systems.

First, and perhaps most obvious, are costs associated with the hardware and infrastructure necessary to monitor users. This includes any behavioral or physiological assessment devices required to derive measurements from users, methods for data transmission from that hardware, and computers to receive, store, and analyze data. Designers should not omit estimates of costs for “consumables,” e.g., single use gels, electrodes, pads, etc., that must be applied each time a device is used. Many behavioral and physiological assessment devices require the use of such consumables, and it is not uncommon for manufacturers of these devices to adopt a “razor-and-blades” business model, wherein the cost of the device is relatively inexpensive, but the costs of the consumables are relatively high (Jones et al., 2010). Designers also need to consider the associated maintenance costs of the hardware and infrastructure they employ. If users are meant to be untethered from their workstations, consideration must also be given to how data will be received while users move about.

Once users' data have been received, those data must be analyzed to derive a measure of operator state. Thought must be given to the minimum processing power and active memory required to compute estimates of operator state with sufficient speed and frequency to allow performance optimization in the target application. Consideration must also be given to whether or not data should be stored for future analysis or use. This should include a plan for the duration that such data must be stored, as failure to do so may result in the consumption of available storage capacity before the full benefit of an aiding system can be realized.

There are also likely to be administrative costs around application of an aiding system. These costs include development and administration of training materials required for successful implementation of an aiding system, including materials designed to persuade users to accept and comply with monitoring protocols. Companies may also need to hire additional personnel to ensure users comply with monitoring protocols, technicians to apply monitoring devices to users, and technicians to maintain and troubleshoot all aspects of the aiding system.

### *4.2. Indirect costs*

Indirect costs may be harder to estimate, but we encourage designers to consider them seriously when they are planning their aiding systems.

Successful implementation of adaptive aiding systems will require consent from users that they will submit to monitoring and comply with monitoring protocols. This likely requires that users agree to appropriately don and calibrate monitoring equipment at the beginning of their work day, and doff and clean equipment at the end. In addition, users may need to charge monitoring devices and engage in periodic maintenance (e.g., for firmware updates, malfunctioning equipment, etc.). These tasks will likely increase organizational costs associated with an aiding system as users must spend part of their work time engaged in such activities. In addition, there may be costs to operator efficacy and efficiency if monitoring devices interfere with task performance. This could occur because of poor fit between device design factors and task requirements (e.g., device wiring that interferes with user activities), or because users find their monitoring devices unwieldy or distracting.

Designers must also contemplate how their intended users will respond to monitoring. At a minimum, users will need to be monitored during task performance. In addition, if operator state factors such as stress and fatigue are to be considered (Caldwell et al., 2008), more extensive monitoring may be required, possibly including during non-work hours. At the most extreme, this becomes *ubiquitous monitoring* (e.g., Moran et al., 2013), wherein data are collected continuously from users.

With regard to monitoring, research suggests that degree of operator acceptance will vary and is likely to fall on a continuum of responses. At the "low" end of acceptance, users may feel that monitoring is intrusive and reduces privacy, or that monitoring indicates distrust of the user (e.g., Frey, 1993). This may create negative affect in users that undermines job satisfaction, commitment, and perceived control (Jeske and Santuzzi, 2015), and increases job stress (Carayon, 1994; Jeske and Santuzzi, 2015). A further concern may be feelings of discomfort or anxiety associated with self-awareness instigated by monitoring (e.g., Baldwin and Holmes, 1987; Fejfar and Hoyle, 2000) and perceptions of the presence of an evaluative "other," such as a superior, colleague, or the monitoring system itself (e.g., Leitenberg, 1990; Zeidner and Matthews, 2005). These concerns may be justified if application of aid is construed (by management, coworkers, etc.) as an index of task aptitude or inferior performance (e.g., Lee, 1997). Furthermore, users may fear the consequences of unwanted disclosures related to health (e.g., Brohan and Thornicroft, 2010; Dray-Spira et al., 2008) or lifestyle (e.g., Ozeren, 2014) as a result of monitoring, particularly amid ongoing concerns regarding data privacy (e.g., Ahamed et al., 2007; Valentino-DeVries et al., 2018).

At the "high" end of the acceptance continuum, users may respond positively to continuous monitoring, particularly if they perceive that the benefits (to performance, safety, workload, etc.) of the technology outweigh the risks (Moran et al., 2013). Users may also endorse monitoring technologies if they are offered the opportunity to utilize the recorded information for their own purposes, such as fitness or health management (e.g., Heron and Smyth, 2010).

A final relevant group are likely to be users who have little practical experience with, or understanding of, monitoring and aiding technologies and therefore may have unrealistic (positive or negative) expectations concerning system capabilities. Designers will need to think carefully about how to educate all users for appropriate mental model formation and trust calibration. Inaccurate beliefs about aid systems may lead users to over/under trust and disuse, misuse, and inappropriate reliance on those systems (Parasuraman and Riley, 1997; Lee and See, 2004). No aid system will be perfectly reliable under all circumstances, meaning that operators may experience both false positives (the system indicates that aid is needed when it is not) and false negatives (the system does not activate when aid is needed) during their usage (e.g., Teo et al., 2020). While false positives appear to be more damaging to users (e.g., Dixon et al., 2007), both types of errors are likely to affect user trust in, and reliance on, aid systems (e.g., Guznov et al., 2016). However, these effects may be mitigated, to some extent, by appropriate education and training regarding a system prior to its introduction (e.g., Forster et al., 2019; Krampell et al., 2020), and feedback during aid-assisted learning (e.g., Chen et al., 2021).

A concluding consideration is additional indirect costs that may develop around hiring and retaining workers. If employees perceive monitoring as undesirable or burdensome, it may contribute to increased employee churn, i.e., turnover in an organization's staff as existing employees leave and new ones are hired (e.g., Lazear and Spletzer, 2012). Monitoring may also depress the potential candidate pool for filling vacancies, as fewer qualified candidates may apply for those positions.

## 5. Cost-benefit analysis

Final decisions regarding the expected value and potential for "return on investment" (ROI) of any aiding system should be based upon considerations of both the benefit (improved performance, efficiency, safety, etc.) and the costs (direct and indirect) associated with that system. In short, designers should be able to positively assert that the selected method of aiding provides sufficient expected improvements to offset associated costs.

To estimate ROI, a designer should first assess the improvement (in performance, efficiency, safety, etc.) of the selected aid system against the current "industry standard" of aiding (or barring that, the "no aid" condition), which should have been quantified during validation research (Section 3) and use that difference to estimate the expected benefit, preferably in monetary units.

Once an initial estimate of the expected benefit of a system has been roughly quantified, a similar estimate of costs must be generated. Designers will have to use their best discretion to determine what costs (Section 4) to include in their estimates. Estimated benefits and costs can then be compared to calculate an initial ROI. While ROI is usually considered in financial terms, we acknowledge that there may be additional lenses through which a company can evaluate its returns, such as environmental or social value (e.g., Arvidson et al., 2013), that may incentivize aiding approaches even when they are less attractive from a purely fiscal standpoint.

Provided that the initial cost-benefit ratio appears to support moving forward with an aiding system, we advise designers to conduct supplemental field validation(s) in the target application environment. The basic aim of field validation is to determine whether the benefit observed in controlled research will also be observed under more ecologically valid testing. Consideration should be given to assessment of the sensitivity, false positive and false negative aid activation rates, and users' impressions of the system.

During field validation, we would suggest, at a minimum, testing the method that appeared to have the greatest ROI during experimental validation against the no aid or industry standard of aid in the application area; beyond that, designers may wish to field test any other aid activation methods that were identified during experimental validation as being particularly competitive. Successful revalidation in field testing will indicate that the benefits of the designed system seen in controlled experimental settings will generalize to the target application.

Finally, after a field validation study (or several), designers should have a fairly good idea about the benefits and costs associated with an aid system. At this point, if the aid system still seems viable, then it may be worthwhile to develop more detailed ROI calculations (including, e.g., social and environmental factors), and draft an implementation strategy.

## 6. Discussion

We highly recommend that designers adhere to the procedures described above when making decisions about what method of aid activation to employ in an aiding system. Doing so will result in effective, relatively objective design choices. There are some additional issues that are worth considering in relationship to the guidelines that we have given.

First, we have adopted a fairly simplistic approach here, presenting each aid activation method separately; we have done so deliberately for the sake of clarity. However, we acknowledge that aiding systems may employ a hybrid of several aids and aid activation methods, and that this may provide useful and necessary flexibility to such systems. For example, a fatigue monitoring system might be designed to recommend a rest break when system performance begins to decline (system/environment activation) or when the user's eyes are closed for a prolonged period (user state activation). Although the guidelines that we have described present alternative activation methods as mutually exclusive choices, the procedures that we have described can be modified to facilitate objective evaluation of hybrid systems. Critically, the basis for decision making remains unchanged; the designer should only select an activation method (or combination of activation methods) if the cost-benefit ratio justifies the choice by suggesting that it is superior to feasible alternatives.

The sequence of steps for arriving at this choice remain unchanged when hybrid designs are involved, but each step becomes more complex. A multi-method activation scheme can only be feasible if each of the combined activation methods are individually feasible. If any one of the to-be-combined aid activation methods is not feasible, the combination is not feasible. Hybrid systems also complicate decisions around the appropriate comparison among aid activation methods for benefit quantification (Section 3). In such systems, it may be useful to quantify the benefits of changes to individual activation methods while holding other aspects of the aid system constant. The benefit of a proposed multi-method activation scheme should be weighed against simpler single-method activation alternates (i.e., those that we have described). In addition, if more than one activation method is used in a system, the added cost of each activation method must also be considered.

Additional consideration must be exercised when designing complex hybrid systems regarding how multiple aids and activation methods will function together. It is important that designers ensure that such systems harmonize with each other so they do not conflict, or that reasonable solutions have been put into place to resolve conflicts when they do occur. Deliberation must also be given to how hybrid systems will be understood by human users to avoid "automation surprises" (e.g., Woods, 1996), situations in which an aiding system acts in a way that is outside the expectations of human users.

Design and validation are also complicated by contextual variability within the target application. Designers should not assume that the results of a cost-benefit analysis will generalize to untested contexts within the target application. Instead, aid activation methods should be assessed within all task contexts that they will be applied to augment performance (i.e., all foreseeable use cases). As we mentioned previously, feasibility and cost-benefit ratio is likely to vary across different applications, but they may also vary across contexts within a specific application. For example, user-initiated activation may be an effective method of activating automated vehicle functions under typical driving contexts, but the benefit and/or feasibility of user-initiated activated automation may be reduced or eliminated in driving situations that require immediate collision avoidance (e.g., emergency braking). Similarly, a user state activated system may be ideal while a user is at their workstation and relatively sedentary, but the benefit of such a system might be diminished by intrusiveness or reduced sensitivity when the user moves away from that workstation. In sum, failure to consider all relevant contexts could lead to unexpected and undesirable effects when untested, unconsidered situations arise.

Consequently, designers should clearly define the target application prior to executing the steps within this guide, and they should evaluate the feasibility, benefits, costs, and cost-benefit ratio based on all intended task contexts. If different task contexts are considered throughout the design process, it is possible that feasibility and relative cost-benefit ratio of aid activation methods will be found to differ across contexts, meaning that selecting a single aid activation method might require maximizing cost-benefit ratio in some contexts, but not others, or compromising across contexts. In such situations, a designer has two options. One option is to select one method from among aid activation methods that is feasible in all intended contexts; by considering the likelihood of various task contexts it is possible to develop a holistic, probabilistic estimate of cost-benefit ratio across all task contexts. The alternative option is to select different aid activation methods for different task contexts to maximize the cost-benefit ratio within each, but this also entails consideration of how the system will correctly identify the task context so that the appropriate aid activation method is active at any given moment. As mentioned above, hybrid systems development is more complicated, but designers can modify our recommended procedures for evaluation of the relative cost-benefit ratio of hybrid systems.

Another reasonable modification to our guidelines would be to alter the order of benefit and cost assessments. As it stands, our guidelines suggest that evaluation of benefit come before evaluation of cost, but it could be useful in some instances to assess costs before benefits to streamline the decision process by reducing the number of alternatives that must be considered. If cost is determined first, methods that are prohibitively expensive can be excluded from benefit quantification and cost-benefit analysis stages.

Designers should also consider the possibility that the cost-benefit ratio of any given aid application method may shift over time. For instance, technological development of physiobehavioral monitoring devices could result in application-oriented hardware and software that is cheaper and simpler than currently available technology, reducing the direct costs associated with acquisition and the indirect costs associated with setup time, data processing, etc. Assuming performance benefits are maintained with cheaper, simpler systems, the cost-benefit ratio would be increased by the cost reduction. Similarly, as such technology becomes more mobile and miniaturized, intrusive effects on performance may become less likely. For example, if a physiological monitoring system eliminates the need to be tethered to a work station by cords, user performance may improve. Assuming costs of less intrusive solutions are comparable, improvements in performance would increase the cost-benefit ratio. Beyond technological developments, designers should consider whether turn-key solutions exist for their chosen aid activation method. In some cases, it may be cheaper to hire a service to lease necessary hardware, software, and personnel to execute a chosen aid activation method. Turn-key service eliminates the direct costs associated with hardware and software purchases as well as the indirect costs associated with training or hiring technicians that may be required to operate the aid activation system.

Finally, though we suggest focusing on financial cost-benefit analysis for objective decisions about which aid activation method is best for a given application, there are other ways to measure costs and benefits that may be worth considering as a complement to financial cost-benefit. Examples include employee satisfaction, employee safety, public perception of one's organization, and broad societal impact. While these factors have indirect links with finances (e.g., low employee satisfaction may lead to churn), it may also be useful to consider the non-financial impacts that a selected aid activation method could have.

As a concluding consideration, we think it is worth noting that, though our focus has been on selecting an aid activation method, the same system engineering framework can (and should) be applied when developing other aspects of aiding system design. As noted in the Introduction, an effective adaptive aiding system must include 1) *what* aid to provide the user in a particular situation; 2) an aid activation method that determines *how and when* to activate that aid; 3) an aid deactivation method that determines *how and when* to deactivate aid; and 4) the minimal acceptable interval between aid activation occasions. Designers should adopt similarly rigorous, system engineering-based approaches when composing the other aspects of an aiding system.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was supported in part by the Air Force Research Laboratory, 711th Human Performance Wing, Human Effectiveness Directorate, through the Air Force Office of Scientific Research Summer

Faculty Fellowship Program®, Contract Numbers FA8750-15-3-6003 and FA9550-15-0001T.

## References

Ahamed, S.I., Talukder, N., Kameas, A.D., 2007. Towards privacy protection in pervasive healthcare. Paper presented at the 3rd IET International Conference on Intelligent Environments (IE 07), Ulm, Germany. https://www.cs.purdue.edu/homes/ntalukde/papers/IE07.pdf.

Arvidson, M., Lyon, F., McKay, S., Moro, D., 2013. Valuing the social? The nature and controversies of measuring social return on investment (SROI). Volunt. Sect. Rev. 4 (1), 3–18.

Audi. Driver Assistance Systems (n.d.), Audi Media Center. https://www.audi-mediacenter.com/en/even-more-sportiness-and-efficiency-the-audi-a4-is-now-more-attractive-than-ever-2756/driver-assistance-systems-2805.

Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Scott, L.A., 2006. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. Hum. Factors 48 (4), 693–709.

Baldwin, M.W., Holmes, J.G., 1987. Salient private audiences and awareness of the self. J. Pers. Soc. Psychol. 52 (6), 1087–1098.

Brohan, E., Thornicroft, G., 2010. Stigma and discrimination of mental health problems: workplace implications. Occup. Med. 60 (6), 414–415.

Caldwell, J.A., Caldwell, J.L., Schmidt, R.M., 2008. Alertness management strategies for operational contexts. Sleep Med. Rev. 12 (4), 257–273.

Calhoun, G.L., Ward, V.B., Ruff, H.A., 2011. Performance-based adaptive automation for supervisory control. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 55 (1), 2059–2063.

Carayon, P., 1994. Effects of electronic performance monitoring on job design and worker stress: results of two studies. Int. J. Hum. Comput. Interact. 6 (2), 177–190.

Chen, J., Mishler, S., Hu, B., 2021. Automation error type and methods of communicating automation reliability affect trust and performance: an empirical study in the cyber domain. IEEE Trans. Human-Machine Sys. 51 (5), 463–473.

Chen, S.I., Visser, T.A., Huf, S., Loft, S., 2017. Optimizing the balance between task automation and human manual control in simulated submarine track management. J. Exp. Psychol. Appl. 23 (3), 240.

Dixon, S.R., Wickens, C.D., McCarley, J.S., 2007. On the independence of compliance and reliance: are automation false alarms worse than misses? Hum. Factors 49 (4), 564–572.

Dray-Spira, R., Gueguen, A., Lert, F., VESPA Study Group, 2008. Disease severity, self-reported experience of workplace discrimination and employment loss during the course of chronic HIV disease: differences according to gender and education. Occup. Environ. Med. 65 (2), 112–119.

Eggemeier, F.T., 1988. Properties of workload assessment techniques. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier Science Publishers, North-Holland, pp. 41–62.

Endsley, M.R., Kiris, E.O., 1995. The out-of-the-loop performance problem and level of control in automation. Hum. Factors 37 (2), 381–394.

Federal Motor Carrier Safety Administration, 2015. Interstate truck driver's guide to hours of service. https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/Drivers%20Guide%20to%20HOS%202015_508.pdf.

Feigh, K.M., Dorneich, M.C., Hayes, C.C., 2012. Toward a characterization of adaptive systems: a framework for researchers and system designers. Hum. Factors 54 (6), 1008–1024.

Fejfar, M.C., Hoyle, R.H., 2000. Effect of private self-awareness on negative affect and self-referent attribution: a quantitative review. Pers. Soc. Psychol. Rev. 4 (2), 132–142.

Frey, B.S., 1993. Does monitoring increase work effort? The rivalry with trust and loyalty. Econ. Inq. 31 (4), 663–670.

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., Keinath, A., 2019. User education in automated driving: owner's manual and interactive tutorial support mental model formation and human-automation interaction. Information 10 (4), 143.

Funke, M.E., Warm, J.S., Matthews, G., Riley, M., Finomore Jr., V., Funke, G.J., Knott, B., Vidulich, M.A., 2010. A comparison of cerebral hemovelocity and blood oxygen saturation levels during vigilance performance. Proc. Hum. Factors Ergon. Soc. 54th Annu. Meet. 54 (18), 1345–1349.

Galster, S.M., Johnson, E.M., 2013. Sense-assess-augment: A Taxonomy for Human Effectiveness (No. AFRL-RH-WP-TM-2013-0002). Air Force Research Lab. Human Effectiveness Directorate, Wright-Patterson AFB, OH.

Gevins, A., Smith, M.E., 2007. Electroencephalography (EEG) in neuroergonomics. In: Parasuraman, R., Rizzo, M. (Eds.), Neuroergonomics: the Brain at Work. Oxford University Press, New York, NY, pp. 15–31.

Greenlee, E.T., DeLucia, P.R., Newton, D.C., 2018. Driver vigilance in automated vehicles: hazard detection failures are a matter of time. Hum. Factors 60 (4), 465–476.

Greenlee, E.T., DeLucia, P.R., Newton, D.C., 2019. Driver vigilance in automated vehicles: effects of demands on hazard detection performance. Hum. Factors 61 (3), 474–487.

Greenlee, E.T., Lui, T.G., Maw, E.L., 2021. Is physiobehavioral monitoring nonintrusive? An examination of transcranial Doppler sonography in a vigilance task. Hum. Factors 63 (7), 1256–1270.

Guznov, S., Lyons, J., Nelson, A., Woolley, M., 2016. The effects of automation error types on operators' trust and reliance. In: Lackey, S., Shumaker, R. (Eds.), Lecture Notes in Computer Science, 9740. Springer, Cham, Switzerland, pp. 116–124. VAMR 2016: Virtual, Augmented and Mixed Reality.

Haslbeck, A., Hoermann, H.-J., 2016. Flying the needles: flight deck automation erodes fine-motor flying skills among airline pilots. Hum. Factors 58 (4), 533–545.

Heron, K.E., Smyth, J.M., 2010. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behavior treatments. Br. J. Health Psychol. 15 (1), 1–39.

Huxhold, O., Li, S.C., Schmiedek, F., Lindenberger, U., 2006. Dual-tasking postural control: aging and the effects of cognitive demand in conjunction with focus of attention. Brain Res. Bull. 69 (3), 294–305.

Jeske, D., Santuzzi, A.M., 2015. Monitoring what and how: psychological implications of electronic performance monitoring. New Technol. Work. Employ. 30 (1), 62–78.

Jones, Z., Woods, E., Nielson, D., Mahadevan, S.V., 2010. Design of a Pulse Oximeter for Price Sensitive Emerging Markets. Paper Presented at the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina.

Kaber, D.B., Perry, C.M., Segall, N., McClernon, C.K., Prinzel III, L.J., 2006. Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. Int. J. Ind. Ergon. 36 (5), 447–462.

Kramer, A.F., 1991. Physiological metrics of mental workload: a review of recent progress. In: Damos, D. (Ed.), Multiple-task Performance. CRC Press, pp. 279–328.

Krampell, M., Solís-Marcos, I., Hjälmdahl, M., 2020. Driving automation state-of-mind: using training to instigate rapid mental model development. Appl. Ergon. 83, 102986.

Lazear, E.P., Spletzer, J.R., 2012. Hiring, churn, and the business cycle. Am. Econ. Rev. 102 (3), 575–579.

Lee, F., 1997. When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. Organ. Behav. Hum. Decis. Process. 72 (3), 336–363.

Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors 46 (1), 50–80.

Leitenberg, H., 1990. Introduction. In: Leitenberg, H. (Ed.), Handbook of Social and Evaluation Anxiety. Plenum Press, New York, NY, pp. 1–10.

Loft, S., Jooste, L., Li, Y.R., Ballard, T., Huf, S., Lipp, O.V., Visser, T.A., 2018. Using situation awareness and workload to predict performance in submarine track management: a multilevel approach. Hum. Factors 60 (7), 978–991.

Matthews, G., Desmond, P.A., Neubauer, C., Hancock, P.A., 2012. An overview of operator fatigue. In: Matthews, G., Desmond, P.A., Neubauer, C., Hancock, P.A. (Eds.), The Handbook of Operator Fatigue. Ashgate, Surrey, England, pp. 3–23.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J., Abich IV, J., 2015. The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum. Factors 57 (1), 125–143.

Matthews, G., Warm, J.S., Reinerman-Jones, L.E., Langheim, L.K., Washburn, D.A., Tripp, L., 2010. Task engagement, cerebral blood flow velocity, and diagnostic monitoring for sustained attention. J. Exp. Psychol. Appl. 16 (2), 187.

Mulder, L.J.M., de Waard, D., Brookhuis, K.A., 2004. Estimating mental effort using heart rate and heart rate variability. In: Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Henrick, H.W. (Eds.), Handbook of Human Factors and Ergonomics Methods. CRC press, New York, NY, pp. 201–208.

Moran, S., Jaeger, N., Schnadelbach, H., Glover, K., 2013. Using Adaptive Architecture to Probe Attitudes towards Ubiquitous Monitoring. Paper Presented at the 2013 IEEE International Symposium on Technology and Society (ISTAS), Toronto, Canada.

Mueller, F., Stevens, G., Thorogood, A., O'Brien, S., Wulf, V., 2007. Sports over a distance. Personal Ubiquitous Comput. 11 (8), 633–645.

Ogden, G.D., Levine, J.M., Eisner, E.J., 1979. Measurement of workload by secondary tasks. Hum. Factors 21 (5), 529–548.

Ozeren, E., 2014. Sexual orientation discrimination in the workplace: a systematic review of literature. Procedia Soc. Behav. Sci. 109, 1203–1215.

Parasuraman, R., Galster, S., 2013. Sensing, assessing, and augmenting threat detection: behavioral, neuroimaging, and brain stimulation evidence for the critical role of attention. Front. Hum. Neurosci. 8 (273), 1–10.

Parasuraman, R., Manzey, D.H., 2010. Complacency and bias in human use of automation: an attentional integration. Hum. Factors 52 (3), 381–410.

Parasuraman, R., Riley, V., 1997. Humans and automation: use, misuse, disuse, abuse. Hum. Factors 39 (2), 230–253.

Prinzel III, L.J., Parasuraman, R., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T., 2003. *Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design* (Report No. NASA/TP-2003-212442). National Aeronautics and Space Administration (NASA), Langley Research Center, Hampton, VA.

Rompelman, O., Coenen, A.J.R.M., Kitney, R.I., 1977. Measurement of heart-rate variability: Part 1 – comparative study of heart-rate variability analysis methods. Med. Biol. Eng. Comput. 15 (3), 233–239.

SAE International, 2016. *Taxonomy And Definitions for Terms Related to Driving Automated Systems for On-Road Motor Vehicles* (Report No. J3016). Author, Warrendale, PA.

Scerbo, M., 2007. Adaptive automation. In: Parasuraman, R., Rizzo, M. (Eds.), Neuroergonomics: the Brain at Work. Oxford University Press, New York, NY, pp. 239–252.

Stach, T., Graham, T.C.N., Yim, J., Rhodes, R.E., 2009. Heart Rate Control of Exercise Video Games. Paper Presented at the 2009 Graphics Interface Conference, Kelowna, Canada.

Stephens, C.L., Scerbo, M.W., Pope, A., 2012. Adaptive automation for mitigation of hazardous states of awareness. In: Matthews, G., Desmond, P.A., Neubauer, C., Hancock, P.A. (Eds.), The Handbook of Operator Fatigue. Ashgate Publishing, Surry, England, pp. 415–440.

Teo, G., Matthews, G., Reinerman-Jones, L., Barber, D., 2020. Adaptive aiding with an individualized workload model based on psychophysiological measures. Human Intell. Sys. Integrat. 2 (1), 1–15.

Transportation Security Administration, 2008. Screening management standard operating procedures. In: https://www.governmentattic.org/6docs/TSA-SecOpsProcedTOCs_2007-2008.pdf.

Tsang, P.S., Vidulich, M.A., 2006. Mental workload and situation awareness. In: Salvendy, G. (Ed.), Handbook of Human Factors and Ergonomics, third ed. John Wiley & Sons, Inc, Hoboken, NJ, pp. 243–268.

Valentino-DeVries, J., Singer, N., Keller, M.H., Krolik, A., 2018, December 10. Your Apps Know where You Were Last Night, and They're Not Keeping it Secret. The New York Times. https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

Wilson, G.F., Russell, C.A., 2007. Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Hum. Factors 49 (6), 1005–1018.

Woods, D.D., 1996. Decomposing automation: apparent simplicity, real complexity. In: Parasuraman, R., Mouloua, M. (Eds.), Automation and Human Performance: Theory and Applications. CRC Press, New York, NY, pp. 3–18.

Young, M.S., Stanton, N.A., 2002. Attention and automation: new perspectives on mental underload and performance. Theor. Issues Ergon. Sci. 3 (2), 178–194.

Zacharias, G.L., 2019. Autonomous Horizons: the Way Forward. Office of the US Air Force Chief Scientist. Air University Press.

Zeidner, M., Matthews, M., 2005. Evaluation anxiety: current theory and research. In: Elliot, A.J., Dweck, C.S. (Eds.), Handbook of Competence and Motivation. The Guilford Press, New York, NY, pp. 141–163.

Zhao, C., Lin, C.-L., Chen, W., Chen, M.-K., Wang, J., 2020. Visual heart rate estimation and negative feedback control for fitness exercise. Biomed. Signal Process Control 56, 101680.