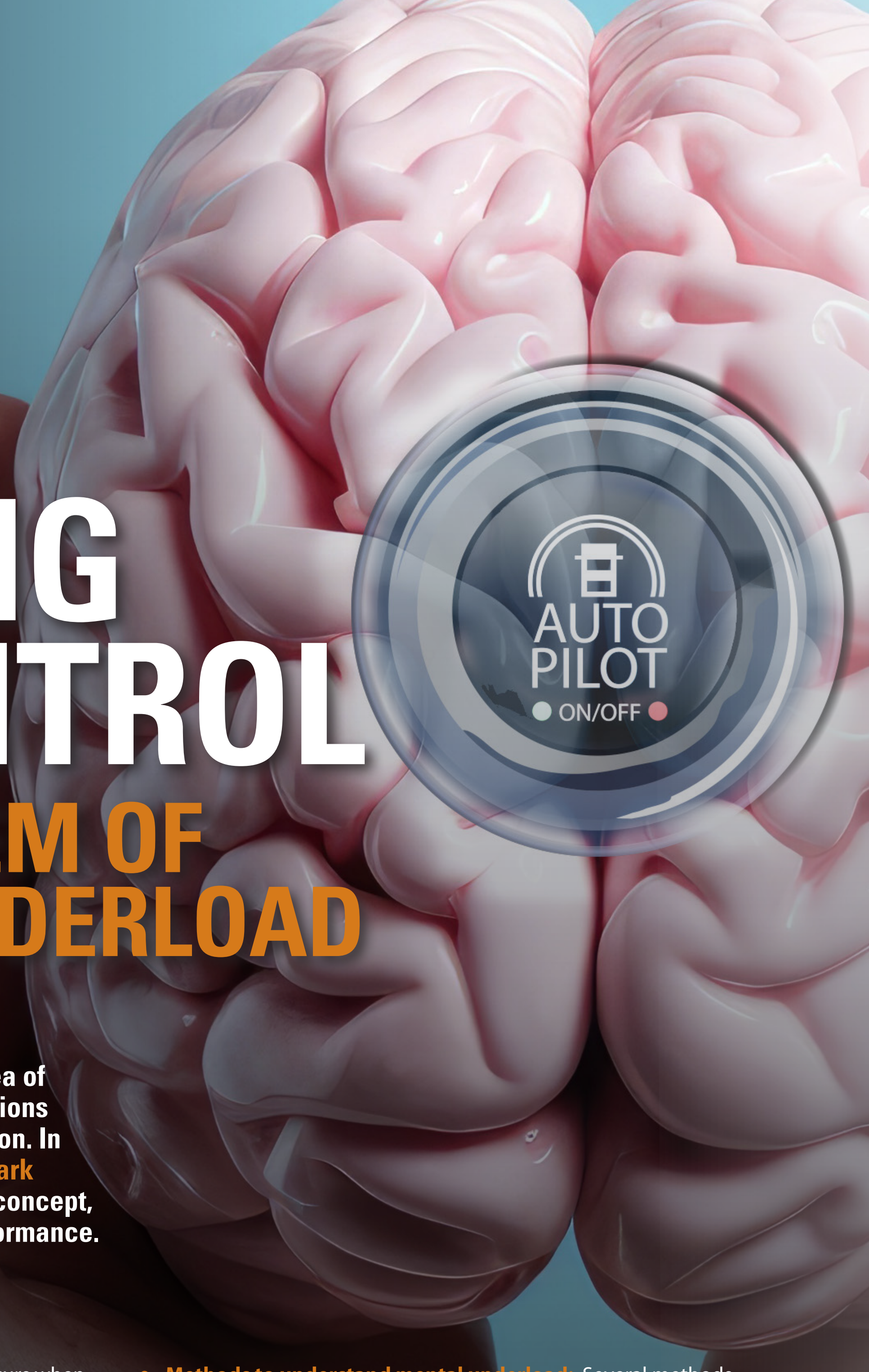
Mark Young



STAYING IN CONTROL

THE PROBLEM OF MENTAL UNDERLOAD

Mental underload is closely tied to the idea of staying in the loop, and is critical in situations where people are a backstop for automation. In this article, based on a recent webinar, Mark Young explains this often-misunderstood concept, and some implications for safety and performance.

KEY POINTS

- **Low cognitive engagement:** Mental underload occurs when tasks are continuous and essential but offer very little demand, resulting in insufficient cognitive engagement.
- **Impaired performance:** Underload can significantly impair performance. When underloaded, attention degrades, monitoring is affected, and reactions slow down, increasing the risk of missing information and responding inadequately.
- **Passive monitoring:** Automation often leads to underload by relegating people to passive monitoring roles. Prolonged periods of low cognitive engagement can leave individuals ill prepared to handle sudden spikes in demand, such as system failures or situations requiring human intervention.
- **Methods to understand mental underload:** Several methods help assess mental underload. These include monitoring performance on the primary task and secondary tasks, subjective ratings of perceived workload, and physiological measures.
- **Mitigation strategies:** Strategies to mitigate the risks of mental underload include periodically reintroducing manual control, incorporating related secondary tasks to maintain engagement, and redesigning systems to minimise prolonged periods of low workload. A more radical proposition involves waiting for fully autonomous systems to be viable.
- **Future research and practice:** Future research and practice should focus on understanding the dynamics of attention decay and recovery during underload, developing more precise measurement tools, and designing systems that balance automation with meaningful human engagement.

INTRODUCTION

Mental underload is something that many operational *HindSight* readers will have experienced, and a concept that I've explored since the start of my career in Human Factors nearly 30 years ago. My own experience mainly comes from two sources: research on driving automation, and practice as a railway accident investigator, concerning train automation. This combination of experience has shown me how underload can leave individuals ill prepared to detect and perceive critical information, or to handle surprises in critical moments.

Underload is closely related to the ideas of people in control and staying in the loop, especially in environments that are high tempo and demand constant monitoring, like transportation. But the concept remains widely misunderstood. In this article, I'll explore what mental underload really is, how it affects performance, and, most importantly, how we can address it from individual and organisational perspectives.

WHAT IS MENTAL UNDERLOAD?

Before we get into the theory, consider these three accidents which brought mental underload into the public eye in the space of two years. In 2016, a tram derailed on a sharp curve in Croydon, South London, tragically resulting in seven deaths. The driver had just navigated a long, straight section of track that required minimal interaction. The tram entered the curve at 73 km/h – well over the 20 km/h speed limit – and overturned. The investigation suggested that the monotony of this part of the journey created an underload state that may have caused the driver to lose awareness, with disastrous consequences.

In 2018, a passenger's bag became caught in the doors of a Central Line underground train in London, leading to them being dragged along the platform. The train operator did not notice the trapped bag. While the Central Line is largely automated, drivers are still responsible for opening and closing doors and monitoring the platform through CCTV before departure. The repetitive nature of this work, with frequent stops and highly automated operations, contributed to underload. While the passenger survived, the incident showed how repetitive tasks can reduce attention, even in highly experienced operators.

During Uber's 2018 test of autonomous vehicles in Tempe, Arizona, a vehicle equipped with sensors designed to detect objects failed to classify a pedestrian walking a bicycle across the road. Although the system detected an object, it couldn't decisively identify whether it was a pedestrian or cyclist. By the time the system responded, 1.2 seconds before the collision, it was too late to avoid the accident, resulting in a fatality. A critical element of this scenario was the presence of a 'safety driver', whose role was to monitor the automated system and intervene if necessary. However, this task had become so undemanding that the driver disengaged, reportedly watching a TV show on their phone.

What the three examples have in common is that mental underload usually occurs in tasks that require some constant engagement, such as driving, but in which the demands are excessively low, leading to a lack of mental stimulation and consequently affecting our attention.

But to understand mental underload, we need to step back and consider its relationship to mental workload more generally. Mental workload refers to the cognitive resources we dedicate to a task, and this depends on our attentional capacity. It's the balance between the mental effort we exert and the demands of the task. While overload results in an overwhelming cognitive burden, underload results in cognitive disengagement. Effectively, our attention 'shrinks' when it is not being used.

"While overload results in an overwhelming cognitive burden, underload results in cognitive disengagement. Effectively, our attention 'shrinks' when it is not being used."

It might seem counter-intuitive, but underload can be just as dangerous as mental overload. The effects of underload can be subtler, however, potentially leading to a decline in performance over time. This might include difficulty in detecting, perceiving or understanding what's going on in a situation, and slower reaction times or inappropriate responses.

Several related concepts are frequently confused with underload. Here are some of the key things that underload is not:

1. Doing Nothing: Underload doesn't mean inactivity. A classic example is when individuals supervise automated systems, such as in flying or driving. Automation may handle most of the workload, with the human operator having to monitor and intervene if necessary. In these cases, the operator is facing a very low demand – but there is still a need to stay engaged.

2. Boredom: While underload can feel unstimulating, it's distinct from boredom, which is defined by the American Psychological Society as "*a state of weariness or ennui resulting from a lack of engagement with stimuli in the environment*".

3. Automatic processing: As individuals become highly proficient in certain tasks, their actions can become automatic, like driving a familiar route without much conscious thought. While this may require little mental effort, it's not the same as underload. Skilled performance still allows for rapid, effective responses to changing conditions, whereas underload tends to reduce the ability to respond.

4. 'Complacency' and over-trust: This often occurs when someone becomes overly reliant on automation or believes a system is so reliable that they no longer need to monitor it carefully. This is a natural response to highly reliable systems.

WHY AND WHEN DOES UNDERLOAD HAPPEN?

Research has shown that the underload 'problem' is predominantly tied to automation, as tasks without automation – even easy ones – often still require some active engagement, making it harder to fully disengage mentally. Automation is often designed to handle repetitive or routine tasks, leaving the operator in a supervisory role. This reduced level of task engagement can lead to mental underload. The operator's job becomes one of passive monitoring, which may lead to periods of low mental activity and a potential drop in alertness and readiness to intervene.

Many automated systems are designed to function at a very high level of reliability, and rarely require human intervention. This reliability further deepens the underload state, because interventions – which increase workload and can restore attention – are few and far between. When technical malfunctions occur, or when a system encounters a situation beyond its capability, there is a sudden transition from passive monitoring with low cognitive engagement to active problem-solving with high cognitive engagement. This sudden shift is particularly dangerous because it can overwhelm the operator.

"When technical malfunctions occur, or when a system encounters a situation beyond its capability, there is a sudden transition from passive monitoring with low cognitive engagement to active problem-solving with high cognitive engagement."

WHY DOES UNDERLOAD AFFECT PERFORMANCE?

To understand this, we need to look at the relationship between stress, arousal, and performance. This is often depicted as an inverted U-shaped curve. The basic concept dates back to 1908, and shows that performance is optimal when stress and arousal levels are in a balanced, moderate range. However, both excessive stress (overload) and insufficient engagement (underload) affect performance negatively.

When workload is too high, demands exceed cognitive resources. But when workload is too low, as in underload, performance declines due to lack of stimulation. In low-demand scenarios, our attention declines, affecting monitoring and engagement, leading to missed cues and slower reactions. To balance overload and underload scenarios, it is important to maintain a state where attentional demands are sufficient to keep operators mentally engaged without overwhelming their capacity.

Traditional models treat our attentional capacity as a fixed and finite resource. Picture it as a bucket with a fixed volume; as task demands increase, the bucket fills, but once it overflows, performance drops off. These models don't account for how underload, or low task demands, can also lead to performance issues.

I developed the 'malleable attentional resources theory' in response to this (Young and Stanton, 2002). It proposes that attentional capacity can expand or contract in response to the demands of a situation. In low-demand situations, our brain may artificially lower its ceiling when it comes to attention. As a result, our performance capacity decreases, even though we are not being overwhelmed by external demands. In higher demand situations, our attentional resources can expand to meet the task, but under low demand, attentional resources shrink, making it harder to respond to unexpected spikes in task difficulty. What might be within our capacity to cope under normal circumstances soon becomes out of reach when demands reduce.

This theory explains why underload, especially in highly automated environments, can impair performance. For example, if a driver or pilot in a high-demand scenario faces a sudden system failure, their attentional capacity may be high enough to respond effectively. However, in a low-demand, highly automated scenario, the same person's attentional capacity may have diminished, leaving them unprepared to handle the same event. The task demands remain constant, but the operator's ability to cope has dropped, leading to performance failure.

HOW CAN WE MEASURE UNDERLOAD?

There are various methods commonly used to assess mental workload. These approaches help us understand how much cognitive capacity is being used during a task and how much spare capacity remains, particularly when tasks are too easy, or automation reduces human involvement. The following four methods are the main types used in research and practice.

Primary Task Performance

The simplest way to assess workload is by monitoring performance on the main task. For driving, this could involve metrics like lane position, speed control, and steering stability. The problem is that primary task performance alone cannot always detect subtle differences between moderate workload and underload. Performance may remain stable at each of these levels of demand because they are both within the operator's capacity. So we need a way of distinguishing these tasks by measuring leftover capacity.

Secondary Task Performance

To capture 'spare cognitive capacity', secondary tasks are often introduced. These tasks are only performed when participants have leftover attentional resources. In driving studies, an example secondary task involves mentally rotating figures and determining via a button press whether they are the same or different. This task competes for the same visual and spatial resources as driving, and so helps to assess how much cognitive capacity is left. If fewer responses are made on the secondary task, it indicates a higher workload on the primary task. In underload situations, more responses on the secondary task are expected because more spare capacity is available.

Subjective Ratings

Subjective measures like the NASA Task Load Index (NASA-TLX) are often used in Human Factors to assess workload. Participants rate their perceived workload on various dimensions after completing a task.

Physiological Measurements

Various physiological metrics provide data on mental workload. For instance, heart rate is a measure of physiological arousal, and can be linked to workload. As workload decreases, so does arousal, and vice versa. More advanced methods are emerging as potential ways to measure brain blood flow, offering a possible direct measurement of attentional capacity. While still developing, these tools could help detect when attentional resources are diminishing due to underload.

Attention Ratio and Malleable Resources

In my research, I've used a combination of secondary task performance and eye tracking to develop an attention ratio measure. This ratio reflects how much time participants spend on the primary task versus the secondary task. By comparing the time spent and the number of responses on the secondary task, we can infer the degree to which attentional capacity has diminished in underload conditions.

Some researchers have proposed a 'red line'. This is a hypothetical boundary beyond which underload or overload begins to affect performance. Defining this precisely remains a challenge. Each person's cognitive limits vary, making it difficult to pin down a universal threshold. However, it's clear that once mental workload drops below a certain point, performance suffers.

Workload is influenced by various factors, such as task difficulty, teamwork, automation, and individual skills or experience. This can make it difficult to understand which aspects of workload we are measuring when conducting research in this area.

DECAY AND RECOVERY OF ATTENTION

A critical aspect of underload is how quickly attentional capacity decays during periods of low demand and how rapidly it can recover when task demands increase. My analysis has shown that attentional capacity decays quickly, typically within the first minute, after a period of low demand. This decline is critical, especially in tasks like driving, where a relatively short span of low workload can leave people unprepared for sudden, urgent and critical demands.

"A critical aspect of underload is how quickly attentional capacity decays during periods of low demand and how rapidly it can recover when task demands increase."

In one of my studies conducted using a driving simulator, participants experienced two driving conditions: one with partial automation, where only the speed and distance to the car in front were controlled by adaptive cruise control, and another with full automation, where both speed and steering were automated. In the fully automated condition, the driver's role shifted to that of a supervisor, monitoring the system's performance rather than actively controlling the vehicle.
The problem arose when the system encountered a failure. In this scenario, the car in front began to slow down, but the automated system failed to adjust the vehicle's speed accordingly. The driver had to recognise the failure quickly, take over manual control, and brake to avoid a collision.

The simulation revealed, not surprisingly, that skilled drivers were able to respond more effectively compared to less experienced drivers. Even though both groups had been in an underload state due to automation, skilled drivers had an automatic, unconscious response to hit the brakes, developed from years of driving experience. This response was less likely in less experienced drivers, resulting in a higher likelihood of collisions.

Recovery from periods of low demand is an area still under investigation. Research in driving suggests that while technology aims for quick recovery times (ideally 10-15 seconds), full re-engagement in a task can take up to a minute. This delay poses significant safety challenges, particularly in scenarios where automation temporarily hands control back to a human operator; in semi-automated driving, even a few seconds is too long. Understanding the dynamics of both decay and recovery is crucial for designing systems that ensure operators remain sufficiently engaged and ready to act when needed.

HOW CAN WE GUARD AGAINST MENTAL UNDERLOAD?

Mental underload can be just as dangerous as overload, particularly in automation-heavy environments. When someone becomes disengaged, they are more prone to missing critical cues or responding too slowly when something unexpected occurs. The challenge, then, is to ensure attentional resources are maintained at an optimal level. Here's how we can guard against underload and even explore how it might be exploited in specific contexts.

"I'm very much an advocate of designing out these problems in the first place. This avoids putting the onus on front-line personnel to deal with underload."

First, and most importantly, I'm very much an advocate of designing out these problems in the first place. This avoids putting the onus on front-line personnel to deal with underload, and is consistent with an ergonomics-oriented approach of fitting the task to the person. Underload shouldn't be their problem.

A common method of maintaining attentional engagement involves periodically reintroducing manual control in highly automated environments. This approach was recommended following investigations into accidents. Periods of manual control help to keep operators engaged, while also allowing automation to relieve cognitive demands when appropriate. Used carefully, it can also help to stabilise mental workload rather than cycling through peaks and troughs (although it is not certain whether people need variety or consistency in workload).

A natural response to underload is to increase task demands by introducing additional activities. However, these tasks should be related to the primary task, particularly in safety-critical tasks and environments. The key is to maintain a cognitive connection. For example, in semi-automated driving, providing tasks that enhance situational awareness (such as, say, a concurrent verbal commentary) can keep the driver engaged. Rather than allowing total passivity, we can encourage actions that maintain a certain level of cognitive engagement while still benefiting from automation's support.

"While full automation is still a distant goal, the intermediate stages, where operators go from minimal engagement to needing to take sudden control, are fraught with risks."

A more radical idea to tackle underload is to rethink how we introduce automation. At the moment, automation is advancing in stages. While full automation (which the automotive industry refers to as 'Level 5') is still a distant goal, the intermediate stages, where operators go from minimal engagement to needing to take sudden control, are fraught with risks.

Instead, we might consider waiting until full automation is achievable, avoiding intermediate phases altogether. While this is a more extreme suggestion, until technology is capable of fully autonomous operation, the issues associated with underload will continue to pose safety challenges.

CONCLUSION

Mental underload is a classic problem in Human Factors research and real work. Addressing it requires evidence-based system design and behavioural interventions. As automation continues to evolve, it's essential to maintain a balance that keeps people meaningfully engaged enough, without overloading them. Ultimately, tackling underload is about keeping people in the loop so long as they have to be able to take control.

Watch Professor Mark Young's webinar Mental underload...what it is and what it isn`t, hosted by EUROCONTROL on 25 June 2024 at https://skybrary.aero/webinars/mental-underloadwhat-it-and-what-it-isnt.

This article was prepared with support from Dr Steven Shorrock.

FURTHER READING

Stanton, N. A., & Young, M. S. (1998) Vehicle automation and driving performance. Ergonomics, 41 (7), 1014-1028. doi:10.1080/001401398186568

Young, M. (2021). Human performance in the spotlight: Underload. *HindSight*, 32. The New Reality. Brussels: EUROCONTROL. https://skybrary.aero/articles/*hindsight*-32

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015) State of science: mental workload in ergonomics. Ergonomics, 58(1), 1-17. doi:10.1080/00140139.2014.95615

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A New explanation for the effects of mental underload on performance. Human Factors, 44(3), 365-375.

Young, M. S., & Stanton, N. A. (2023). Driving automation: A human factors perspective. CRC Press.

Young, M. S., & Stanton, N. A. (2023) To automate or not to automate: advocating the 'cliff-edge' principle. Ergonomics, 66 (11), 1695-1701. doi:10.1080/00140139.2023.2270786

MARK YOUNG is Professor of Human Factors in Transport within the Transportation Research Group at the University of Southampton. Mark has nearly 30 years' experience working in human factors across transport modes in both academia and industry. He has a degree in psychology and a PhD in human factors. Mark's early career was largely based in academia until 2012, when he became an Inspector at the Rail Accident Investigation Branch, applying his human factors expertise to the investigation of railway incidents and accidents. He continued in this role until rejoining the University of Southampton in June 2023. Mark has written over 70 peer-reviewed journal papers and five books. He is a Chartered Ergonomist, a Fellow of the Chartered Institute of Ergonomics and Human Factors (CIEHF) and, from April 2024, President of the CIEHF.