任务聚焦：模拟无人机系统中眼动追踪器支持的交互式决策支持的混合方法评估

Hyun-Gee Jei [a]，Mustafa Demir [a]，以及 Farzan Sasangohar [a,*]

[a] 得克萨斯农工大学 Wm Michael Barnes工业与系统工程系

利益冲突声明： X

经费：本研究未获得公共、商业或非营利部门资助机构提供的任何专项拨款。

作者贡献： **Hyun-Gee Jei：**概念化、方法论、验证、正式分析、调查、数据整理、撰写——初稿、可视化、项目管理；**Mustafa Demir：**正式分析、数据整理、撰写——初稿、可视化；**Aishwarya Kanaparthi：**软件；**Jacob Kolman：**撰写——审阅与编辑；**Farzan Sasangohar：**概念化、方法论、资源、撰写——审阅与编辑、监督、项目管理

稿件类型：研究论文

准确字数： ~5870 （目标为 <5000 或咨询编辑）

*通讯作者：Farzan Sasangohar，101 Bizzell St，Mail Stop 3131，College Station，TX 77843- 3131。电子邮箱：sasangohar@tamu.edu；电话：+1-(682)-300-8380

MANUSCRIPT

摘要

军事指挥与控制（C2）环境中的监督员面临动态变化的条件。动态变化的信息通过多个显示屏持续传递给监督员。在这种环境中，由于任务和环境的复杂性，重要信息可能会被忽略。本研究考察了基于眼动追踪器的自适应注意力引导决策支持工具（DST）在模拟C2环境中对监督员的效用。该DST实时监测监督员的视觉注意分配，并在监督员遗漏关键变化或事件时显示视觉显著提示。25名军事院校学生参与了一项模拟情报任务。结果表明，启用自适应DST后，绩效显著提升。眼动追踪分析还显示，在关键兴趣区停留时间更长、注视次数更频繁与绩效呈负相关。此外，实验后访谈显示，自适应DST不会造成干扰，并受到参与者的积极评价。这些发现凸显了实时眼动干预在优化监督员决策方面的潜力。未来研究可纳入人工智能驱动的方法，以更好地支持监督员应对复杂任务环境。

关键词：视觉提示、自适应界面、注意力引导

要点

- 基于实时眼动仪的注意力引导提示提升了军事任务指挥官在高保真指挥中心模拟环境中的绩效。
- 注视分析表明，基于实时眼动仪的注意力引导提示不仅能够提升他们的绩效，还能促进更具策略性和效率的视觉加工。
- 基于注视的注意力引导视觉提示因不具干扰性而获得参与者的积极反馈——88% 的参与者没有注意到这些提示，但表现仍优于对照阶段。

MANUSCRIPT

1. 引言

现代指挥、控制、通信与计算机（C4）以及情报、监视与侦察（ISR）领域的任务指挥员（MCs），面临着迅速变化的动态局势，其复杂程度有时会超出他们的认知能力。随着能力不断提升，以及卫星和传感器等复杂技术网络不断发展，ISR 产生的数据量十分庞大。此外，人工智能和增强型传感技术的出现，使信息能够实时传输至指挥中心的界面（Fontes & Kamminga, 2023）。因此，要在监督控制中成功决策并维持态势感知（SA），任务指挥员必须感知并处理显示在各种界面上的大量信息（Demir et al., 2023; Endsley, 2001; Endsley et al., 2003; S. Humr et al., 2025; S. A. Humr et al., 2023）。接触大量信息可能导致信息过载，而视觉注意分配不够理想，也可能使信息未能得到及时关注。及时分析多模态传感器中的动态信息流，对于维持态势感知和有效决策至关重要（Endsley, 1995），有助于部队应对技术进步并更有效地管理战场状况（Rickli & Mantellassi, 2024）。

众所周知，注意是人类信息加工的瓶颈，注意资源十分有限（Rogers，2023；Tsotsos 等，1995；Wickens，1976）。在现代 C4中心，任务指挥员通常被多个任务信息显示器环绕，以维持情境感知并为战术决策提供信息。他们还需要分配注意力，以进行监控和沟通。与信号能够从各个方向被探测到的听觉显示器不同，从视觉显示器中处理信息需要优化注视。然而，鉴于 C4 环境中提供了海量视觉信息，潜在的重要信息可能会因无意视盲而被忽略、未被察觉（Mack, 2003；Mack 和 Rock，1998；Neisser，1979；Simons 和 Chabris，1999）。此外，频繁的中断可能会将注意力从任务显示器上引开，使任务指挥官容易受到干扰，并普遍面临注意方面的问题。另一个例子是变化盲视。鉴于军事任务的动态性，任务指挥员也容易经历一种称为变化盲视的现象（Simons 和 Levin，1997）：即使只是短暂的视觉遮挡，观察者也未能注意到视觉场景中的变化（Rensink 等，1997）。因此，有必要在复杂的 C4环境中支持任务指挥员，使其关注完成当前任务所需的重要信息，从而及时、稳妥地做出决策。

以往研究探讨了注意引导方法，以支持有效的注意分配。在驾驶情境中，Calvi 等人（2021）使用增强现实（AR）的视觉与听觉反馈，帮助驾驶员更安全地执行接管操作。此外，Stahl 等人（2016）在车辆显示器中应用高亮颜色线索来引导注意，从而提升了新手驾驶员的绩效。在教育领域，

MANUSCRIPT

De Koning 等人（2010）、Grant 和 Spivey（2003）以及 Roads 等人（2016）使用引人注目的高亮标记，更好地引导学生理解复杂概念。在航空与国防领域，Hopp 等人（2005）和 Savick 等人（2008）通过振动背心和腰带传递触觉提示，将注意力引向重要事件。在这两种情况下，使用触觉提示引导注意，都能提升复杂任务环境中的绩效。然而，触觉提示也容易引起对提示位置和含义的误解（Nonino 等人，2021；Sklar 和 Sarter，1999）。在高风险、安全关键且动态的环境中，这可能成为一项关键限制。此外，Frame 等人（2022）通过在相关元素上方放置小型彩色球体来引导用户注意，从而提升了对理解场景所需关键元素的识别。尽管提升视觉显著性有助于注意引导，但一些研究发现，视觉显著特征的强度过高可能引发负面情绪（如烦恼），和／或导致绩效下降（Bahr 和 Ford，2011；Lewandowska 等人，2022；Tasse 等人，2016；Ward 和 Helton，2022）。此外，通过注意引导实现的实时自适应辅助仍未得到充分探索：只有一项研究（Fortmann 和 Mengeringhausen，2014）在军事场景中实现了此类系统。虽然该研究并未将注意引导用作决策支持工具，但 Ratwani 及其同事开展了类似方向的研究。他们利用实时眼动追踪来预测并预防后完成错误；若未能妥善处理，这类错误在高风险环境中可能造成严重后果（Ratwani 等人，2008；Ratwani 和 Trafton，2011）。Breslow 等人也采用类似的预测建模，在操作员必须管理多架无人驾驶飞行器（UAV）的动态场景中，利用实时眼动追踪技术预测工作负荷并预防错误（Breslow 等人，2014）。然而，上述研究与我们的工作不同之处在于，这些实验是在简单环境中进行的，未能准确体现军事指挥中心的复杂特性。据我们所知，在复杂的安全关键工作环境中应用实时注意引导，仍是一个研究空白。

为弥补这一空白，本研究旨在通过整合能够响应模拟 C4 环境中任务指挥员注视行为的注意引导功能，探究自适应界面的有效性。本文记录了一项实验的研究结果，该实验在一处仿照 C4 控制室建造的高保真设施中开展，旨在探究这一工具能否提升任务指挥员的决策能力。

2. 材料与方法

2.1 仿真测试平台

模拟任务环境再现了一项地面部队防护任务，具体是保护一支穿越敌对地区的重要政治车队。在该环境中，由一名人类任务指挥官和三名计算机模拟操作员组成的小组控制着三架半自主无人驾驶飞行器，用于搜索

MANUSCRIPT

短程、中程和远程导弹。小组还必须与外部空袭小组协调，摧毁所有发现的、对车队构成威胁的目标。此外，任务指挥官可以调用多种资产来完成车队任务，例如无人机、打击小组以及机载预警与控制系统（E-3预警机）。

运行该情景的仿真测试平台由三台 70 英寸交互式显示屏（Dell C7017T）和一台 55 英寸触控桌面显示屏组成，如图1所示。每台显示屏都提供了多种任务关键信息。任务状态显示屏（图1，左上）显示操作员绩效指标、无人机状态和通信状态，还显示任务指挥官、无人机操作员与空袭小组之间交换的消息历史记录。当无人机花费超过 30 秒仍未能识别已发现的威胁时，任务指挥官可通过远程协助显示屏（图1，右上）请求信息并协助识别威胁。地图显示屏（图1，上方中央）作为主显示器，提供任务的地理空间地图、三名无人机操作员各自的关注区域、车队当前位置、已发现的威胁及其范围、无人机已扫描的区域、车队健康状态、通信状态、已用任务时间，以及用于展示打击计划、潜在和已知威胁及特定无人机被摧毁时间的时间线。最后，任务指挥官显示屏（图1，下方中央）设于桌面上，供任务指挥员进行决策并下达命令。在整个任务过程中，任务指挥官必须保持警觉，监控多台大型显示屏以掌握任务状态，并持续就车队健康状态、各无人机操作员的决策表现以及无人机移动做出决策。成功完成任务不仅需要监控车队和无人机的进展，还要求任务指挥官及时采取行动，例如令车队暂停或继续行进、重新分配无人机，以及将车队改道至更安全的路线。

MANUSCRI

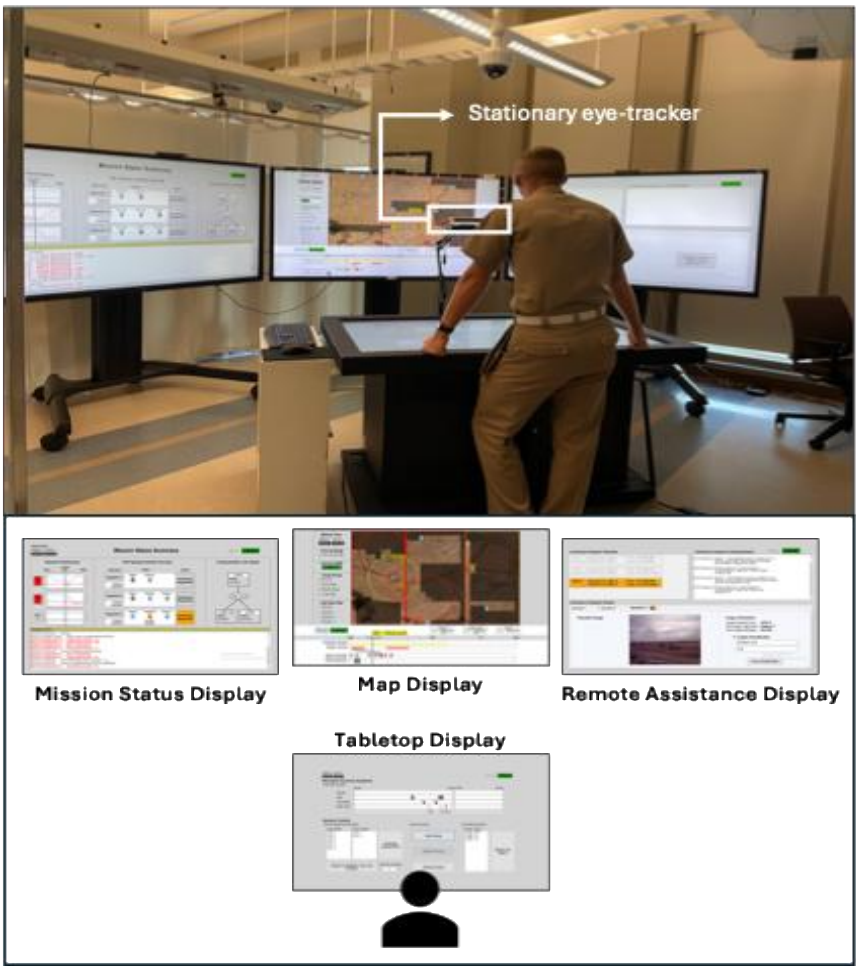


图1。 仿真测试平台。任务指挥官（参与者）与模拟的无人机操作员和 E-3预警机进行沟通，以监督地面部队防护任务。

2.2 基于眼动追踪的自适应决策支持工具

在整个模拟过程中，一台固定式眼动仪（Tobii Pro Nano；采样率为 60Hz，精度为 0.3°）安装在主地图显示界面前方（Tobii AB，2023），以持续监测任务指挥员的注视行为。眼动追踪数据有两个用途：一是提供实时数据流，为自适应注意力引导决策支持工具提供支持；二是用于事后分析任务指挥官的视觉注意模式，以考察视觉注意分配如何影响决策和整体绩效。所有眼动追踪数据均通过 Tobii Pro Lab 软件（版本 1.123）记录并导出。

对于实时注意引导，如果任务指挥官（MC）未能在任务变化发生后的3秒内识别出关键任务变化（例如，无人机被摧毁或车队进入威胁范围），系统便会呈现

一组视觉显著线索，包括警告消息、颜色变化和闪烁，以突出显示相应的关注区域（AOI；图2）。例如，当车队遭到袭击或无人机被敌方炮兵摧毁时，任务指挥员会在地图正下方收到通知消息（图2，框1）。此外，当车队、打击小组或E-3预警机的通信链路断开时，地图下方的状态颜色会变为红色，以吸引任务指挥官的视觉注意（图2，框2）。最后，发生延迟打击时，时间线区域中的红框也会闪烁（图2，框3）。但是，如果任务指挥官的视觉注意已经集中在相关兴趣区，则不会显示注意引导功能，以避免信息与视觉过载。这些显著线索经过策略性布置，位于时间线正上方；时间线是最受关注的兴趣区之一，此布局应用了邻近兼容原则（Wickens & Carswell, 1995）。这种布置旨在尽量减少参与者注意到线索并将注意力转移至线索所需的努力（Wickens，2015）。关键兴趣区和自适应决策支持工具功能见表1。

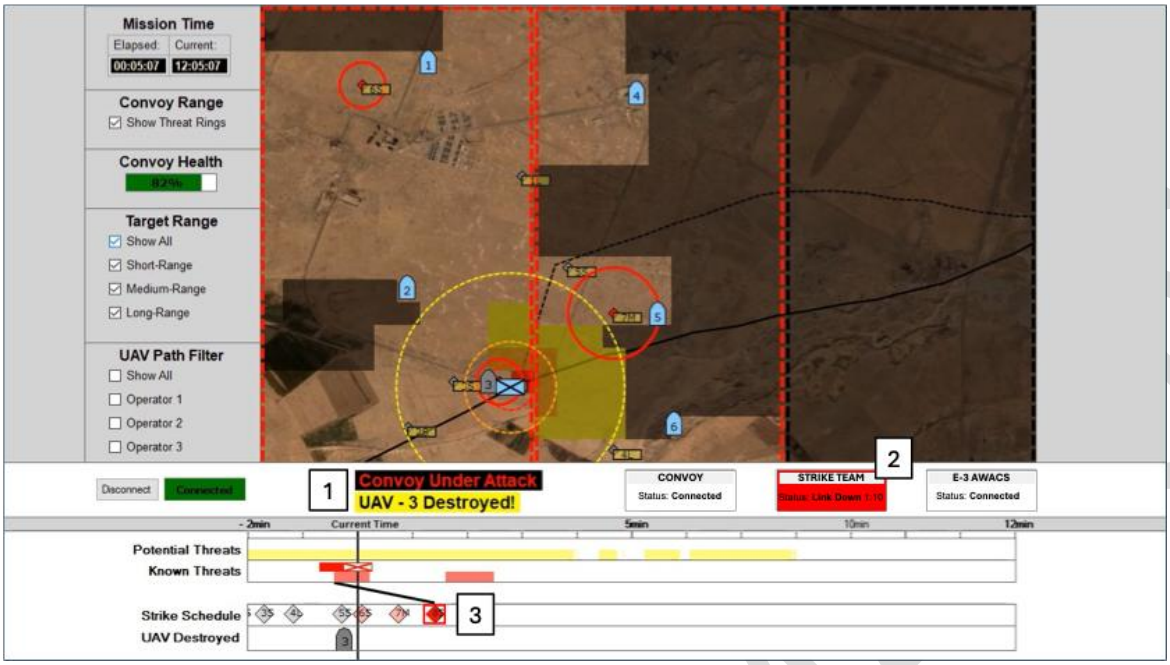


图2。视觉显著的注意引导型自适应决策支持工具示例。时间线上的消息以及时间线中打击计划周围闪烁的红框，都是自适应决策支持工具的示例。

表1。地图显示屏中关键兴趣区的定义及关键自适应功能。

AOI	兴趣区：定义	自适应功能
时间线	地图显示屏上显示中断 恢复辅助时间线的区域。	延迟打击目标周围的红色闪烁框。
Map	地图显示屏的一部分显示了 划分为三个区域的作战区域地图。每个 区域由一名虚拟无人机操作员管理（共 3 名无人机 操作员）。	无自适应功能。但是，自适应 当参与者未能注意到 地图区域中的重要变化时， 其他部分的功能会被激活。
车队 健康	地图上的进度条显示了 车队，直到任务失败。	没有自适应功能。不过，这个兴趣区是 唯一可供使用的实时性能指标 提供给任务指挥员，可能会影响他们的决策 。
消息	地图上的区域显示有关 车队和无人机状态的消息。	当情境感知消息被自适应显示时， 任务指挥官无法识别无人机状态时，消息会自适应显示。
通信 状态	地图区域显示通信连接 任务指挥官、车队、E-3 预警机，和打击小组之间的状态。	闪烁的红色框。

2.3 设计与假设

本研究采用混合2×2×2设计：被试内因素包括条件（对照条件 vs. 自适应决策支持工具）和情景（情景1 vs. 情景2）。被试间因素为试验顺序（参与者先经历对照条件还是干预条件）。条件呈现顺序采用平衡化处理。该设计使我们能够考察条件和情景的主效应，以及呈现顺序是否通过潜在的学习效应影响结果。每位参与者在实验阶段完成两个情景：其中一个处于控制条件下，此时主地图显示界面上不显示自适应决策支持工具；另一个处于干预条件下，此时参与者可以使用自适应决策支持工具。两个情景的设计难度相当。我们假设（H1），与对照条件相比，任务指挥员在干预条件下的决策表现会有所提高。

2.4 样本量与参与者

使用 G*Power3（Faul等，2007年）进行功效分析，以检验 2 个条件（被试内）× 2 个情景（被试内）× 2 种试验顺序（被试间）的均值差异

MANUSCRIPT

采用 F检验，效应量为中等（ηp2 = 0.08；Cohen，2013年），显著性水平（）为 0.05。先验功效分析基于被试内部分（条件 × 情景），因为被试内对比是主要关注的效应，而被试间因素（试验顺序）主要作为区组变量，以控制潜在的学习效应。根据分析结果，要达到 0.80 的功效，需要总计 25 名参与者。

为借鉴军事训练与教育视角（Fletcher，2009），研究人员在美利坚合众国一所大型西南大学招募了25名来自预备军官训练团（ROTC）和退伍军人学生协会的志愿参与者。参与者须年满18岁，英语流利，自述听力和视力正常或经矫正后正常（例如，色盲者不纳入本研究），并能熟练使用标准电脑鼠标和键盘。参与者因参加本研究2.5小时而获得50美元报酬。由于眼动仪无法正确校准，两名参与者在干预条件下的试验结果被排除，因此最终有N = 23 名参与者的数据完整，另有2名参与者的数据不完整。所有纳入分析的参与者均完成了两种情境下的两种条件（对照和自适应DST）；试验顺序（先对照或先干预）经过平衡化。参与者总体上较为年轻（平均年龄 = 22.2，年龄标准差 = 3.57），男性居多（88%自我认同为男性），服役年限为2–4年。25名参与者中，13人曾在空军服役或签约加入空军，3人在陆军服役，1人在海军服役，1人在海岸警卫队服役。另有6名参与者与各军种均无关联，但属于该大学学员团的操练与仪仗队。本研究经该大学机构审查委员会批准（IRB2022-0997D）。参与者均签署了书面知情同意书。

2.5 程序

所有参与者均参加了一次单次实验，时长约为2.5至3小时。首先，参与者阅读了时长20至25分钟的教程幻灯片，内容涵盖实验流程、任务摘要、目标，以及实验中使用的不同显示屏的功能说明。随后，参与者观看了任务处于不同阶段的六张屏幕截图。参与者需要判断该情景下的最佳行动方案，并解释作出这一决策的原因。设置这项小测验是为了确保参与者理解教程幻灯片中的所有材料。研究人员还向参与者说明了将如何衡量其整体绩效。之后，参与者前往实验环境，完成一个模拟测试条件的训练场景。在与测试平台交互的过程中，训练试验也作为基准测试，以确保参与者在开始正式试验前充分了解任务和测试平台。参与者必须使车队至少保有25%的

MANUSCRIPT

健康状况，使用过车队的暂停与恢复、重新分配无人机以及使用E-3预警机等功能，并通过基准测试，之后才能进入正式实验试验。训练试验会重复进行，直到参与者达到设定要求为止。整个教程环节共耗时45至60分钟。

参与者分别完成了对照条件和干预组条件。各条件及情景的顺序在参与者之间进行了平衡化。为避免产生偏差，参与者不知道两种条件的具体内容，也不知道条件的顺序。完成第一次试验后，参与者休息5分钟，再开始第二次试验。每次试验中，参与者在预先设定的时间点两次暂停主要任务，执行中断任务。这些中断任务在认知和视觉上都具有较强的参与性（即根据地理特征进行任务规划，以及根据给定的地理坐标在地图上查找位置）。这些中断任务旨在将参与者的注意力从主要任务上转移开，并模拟现实世界中的动态任务环境，在这种环境中，任务指挥员经常受到打断。车队抵达指定地点后，参与者会接到指示，返回主要任务并作出作战决策。如表2所述，每个情景都包含两个决策点：一个简单决策和一个复杂决策。决策类型的顺序有所不同。情景1中，第一次中断后是简单决策，第二次中断后是复杂决策；情景2中的顺序则相反。在每个情景中，所有参与者都在相同地点返回主要任务。因此，每个人都会遇到相同的作战决策点。总体而言，实验试验耗时约40至50分钟。

表2。 各决策类型及其对应的可选行动。

决策类型	可选行动
简单型	只有一个可用于应对当前情况的行动方案（例如， 车队即将进入威胁区；因此，任务指挥官必须让车队停下）。
复杂	有多种行动方案可用，但只有一种行动方案最适合应对当前 情况（例如，车队即将进入未扫描的主路径，但替代路径 已由无人机完成扫描并清除。任务指挥官可以请求E-3预警机， 重新规划路线或让车队停留。重新规划路线将为车队带来最佳结果）。

在两次试验过程中，均使用 Tobii Pro Nano 眼动仪实时追踪参与者的视觉注意分配。由于眼动仪校准的准确性对于提供精确的 DST 至关重要，因此每次试验开始时，都会使用 Tobii Pro Lab 软件（Tobii AB，2019）内置的九点校准与验证程序对眼动仪进行校准。两次试验结束后，参与者被要求填写一份人口统计学表，并接受访谈，分享他们在实验过程中的体验。访谈采用半结构化形式，预先设定了七个

MANUSCRIPT

指定问题（见附录）。必要时会提出一些追问，以便更好地了解参与者如何收集信息以及他们的决策过程。访谈持续约 10 至 20 分钟。

2.6 衡量指标

绩效。绩效采用综合评分系统（0-100 分）进行评估，该系统旨在反映行动成功的多面性。鉴于多种因素会影响现实任务结果，此方法纳入了三个直接影响任务结果的关键要素。该综合评分方法取自军方通用的绩效测量方法——绩效测量指标（MOP）。根据美国军方的一份联合出版物，MOP 是“用于评估与任务完成情况衡量相关联的友方行动的标准”（联合参谋部 J-7，2011年）。研究选取了三个可量化的 MOP 要素来计算总体绩效，并根据参与者教程中传达的任务优先级进行加权：任务结束时剩余的车队健康状况（50%）、完成每次试验所需的时间（35%），以及报告的延迟打击次数（15%）。每次试验开始时，车队的生命值条为 100%。随着任务推进，如果车队遭到敌方炮兵攻击，或长时间停滞不前，其健康状况就会下降。生命值条持续显示在地图上，使参与者能够监控车队的健康状况。车队健康状况的权重最高，因为保护车队是任务的首要目标。因此，任务结束时剩余的健康状况百分比用于计算 MOP 得分。同样，仿真测试平台也记录了从模拟开始到结束的完成时间和报告的延迟打击次数。如果打击小组消灭敌方目标的时间晚于车队已抵达威胁区域的时间，且车队若未在进入威胁区域前及时拦截就会遭到攻击，则参与者会报告发生了延迟打击。

眼动测量指标。 图3中的标注地图显示屏突出标示了具体的兴趣区（AOI）。这些区域依据任务的重要性加以确定，而任务的重要性取决于从各区域获取的信息量以及这些区域受到的注意程度。表1总结了这些兴趣区及其在四项眼动测量指标中的自适应功能：（1）注视总时长（TDF）：参与者的注视在视觉环境中的某个特定区域内保持不动的总时长；（2）平均注视时长（ADF）：眼睛注视某个特定点的平均时长；（3）注视次数（NF）：视觉环境中某一特定兴趣区内单次注视的总次数；（4）平均瞳孔直径（APD）：进入眼睛的光线平均量，以及大脑状态和其他因素的反映

MANUSCRIPT

（Krejtz 等人，2018年；Van Der Wel 与 Van Steenbergen，2018年；Van Gerven 等人，2004年）。

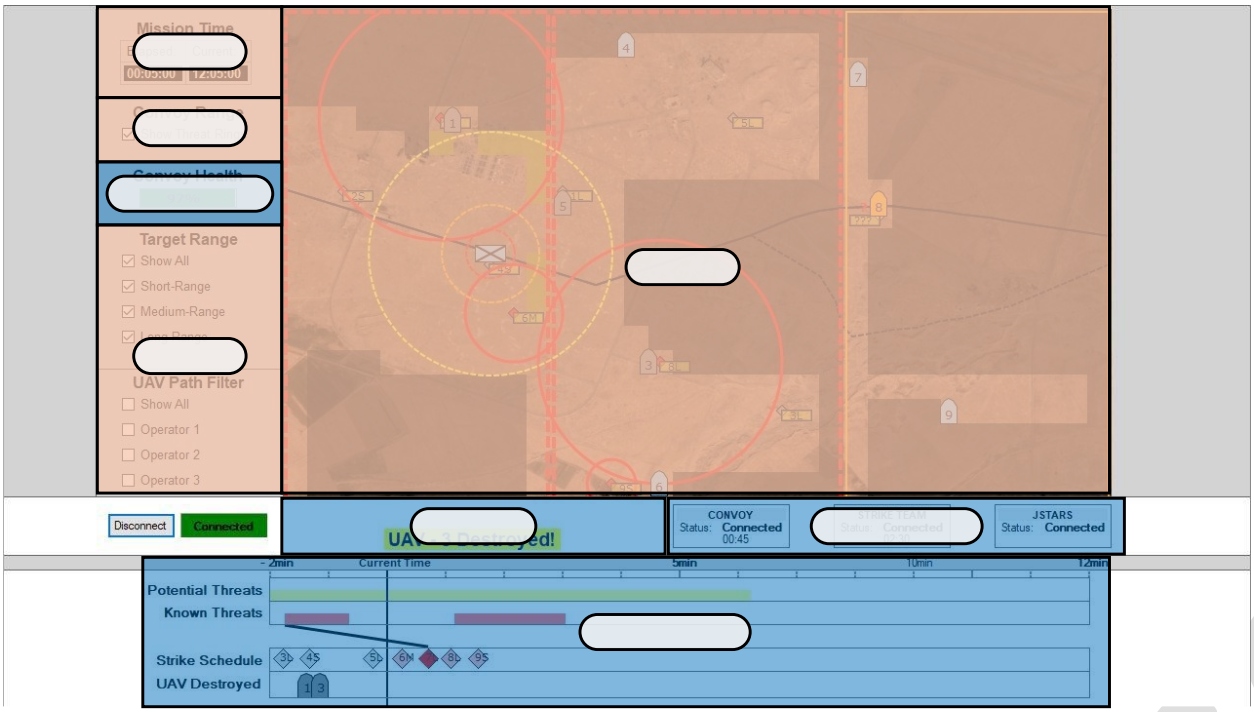


图3。标注地图显示上的兴趣区。在八个兴趣区中，车队健康状况、消息、通信状态和时间线是与任务相关的关键兴趣区。

任务指挥官的决策（首次决策所需时间）。经历中断后做出首次决策所需的时间，也称恢复滞后（Trafton 等人，2005年），至关重要，因为从中断返回主要任务后，决策能力会下降（Monk 等人，2008年）。因此，加入这一指标，以考察自适应注意力引导决策支持工具能否帮助任务指挥员在中断后更及时地做出决策。

2.7 分析方法

采用 2（条件：自适应 DST 或无注意力引导支持）× 2 （情景：简单或复杂决策）× 2 （试验顺序：被试间）裂区方差分析，考察不同条件下 MOP 分数的差异，并将参与者设为随机截距。纳入试验顺序后，我们可以评估暴露顺序是否影响自适应 DST 的有效性，从而考虑潜在的学习效应。我们使用 Shapiro-Wilk 检验评估 MOP 是否满足正态性假设。Shapiro-Wilk检验表明，MOP 数据符合正态

MANUSCRIPT

分布，W(47) = 0.98，p = .452。我们选择了复合对称协方差结构，假设重复测量之间的方差和协方差相等。模型还采用最大似然法进行估计。我们计算了条件、情景和顺序交互作用的估计边际均值，并使用最小显著差异法进行两两比较。

使用 Shapiro-Wilk检验评估眼动响应变量的正态性，结果显示所有眼动测量指标均显著偏离正态分布：TDF (W = 0.79, p < .001)；ADF (W = 0.97, p < .001)；NF (W = 0.82, p < .001)；APD (W = 0.99, p = .016)。鉴于正态性显著违背，且多个变量呈中度至高度偏态，因此采用非参数 Wilcoxon 符号秩检验，比较各兴趣区和指标在对照条件与干预条件下的差异。分析仅纳入两种条件下均有完整配对数据的参与者。

为分析决策时间，首先使用 Shapiro-Wilk检验对潜伏期变量进行正态性检验。分析前，潜伏期数据呈显著正偏态（偏度= 3.19），且偏离正态分布；Shapiro-Wilk检验结果亦显示这一点，W(96) = 0.61, p < .001。由于正态性假设被违背，我们采用2×2 被试内设计的对齐秩变换方差分析（ANOVA），考察条件（对照组与干预组）和决策类型（简单与复杂）对潜伏期的影响。

3. 结果

3.1. 条件与顺序效应对表现指标的影响

方差分析结果汇总于表3，显示条件与试验之间存在显著的交互效应，且条件存在显著的主效应（p = .035 和 p = .013）。总体而言，参与者在干预条件试验中的表现显著优于控制条件试验，但只有在先完成控制条件、再完成干预条件时才观察到这一效应（p < .001）。然而，试验顺序颠倒时，他们的表现相近（p = .770；图4，右）。为评估总体模式是否由离群值驱动，我们还检查了箱线图。未发现极端个案，剔除边缘个案后，结果模式仍保持不变。不过，干预条件下的变异性略大；这似乎反映了自然的表现差异，而非个别观测值的不当影响。

MANUSCRIPT

表3。 用于预测MOP的裂区方差分析结果。

来源：	$df_{numerator}$	$df_{denominator}$	F	p值
条件	1	22.85	7.19	**0.013**
试验顺序	1	22.85	0.38	0.543
情景	1	22.85	0.07	0.797
条件与试验顺序	1	24.97	5.04	**0.035**
条件与情景	1	24.97	0.16	0.689
情景与试验顺序	1	24.97	0.10	0.076
条件与情景与试验顺序	1	22.85	3.46	0.751

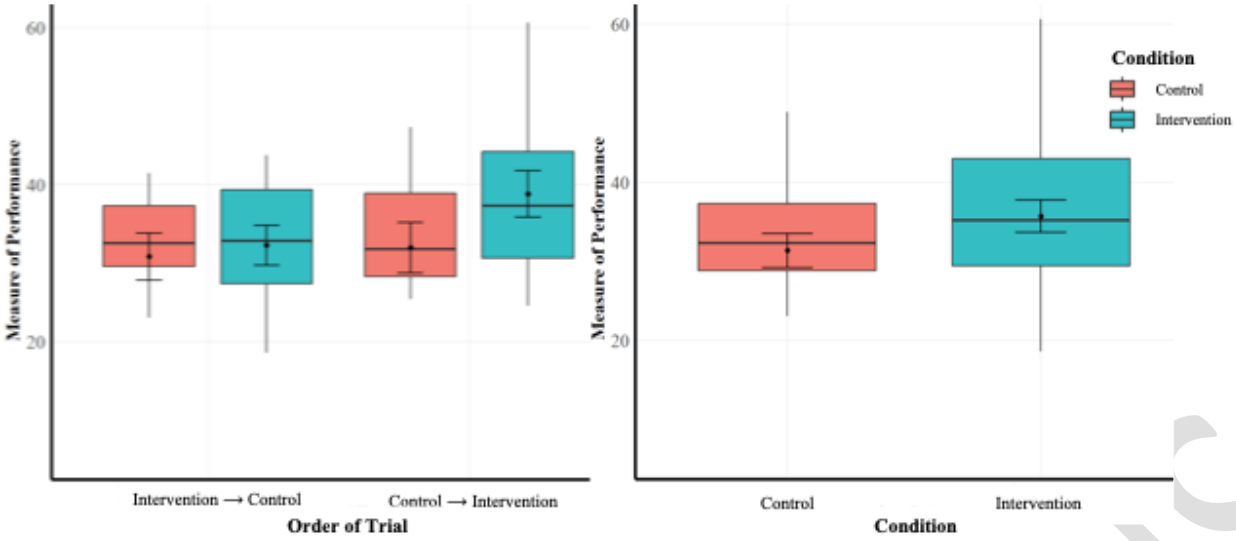


图4。不同条件和试验顺序下（左）以及不同条件下（右）的绩效测量指标（MOP）得分。误差条表示95%置信区间。

3.2. 跨兴趣区的眼动测量指标

表4汇总了Wilcoxon符号秩检验结果。对于每个兴趣区和测量指标，均报告配对观测值数量（n）、检验统计量（V）、p值、标准化z分数及效应量（r）。同时提供了两种条件下的中位数和四分位距（IQR）。在车队健康状况方面，ADF（$p = .012$，$r = .52$）和TDF（$p = .046$，$r = .41$）均发现显著差异；这两项指标在控制条件下的持续时间都长于干预条件（图5）。虽然差异未达到统计显著性，但消息感兴趣区域中的TDF呈现出条件间的趋势性差异（$p = .073$，$r = .37$），控制条件下的参与者注视该区域的时间更长（图5）。

MANUSCRIPT

表 4。比较各兴趣区及眼动测量指标中对照条件与干预条件的 Wilcoxon符号秩检验结果。

							对照		干预组	
AOI	**测量**	***n***	***V***	***p***	***z***	***r***	**中位数**	**IQR**	**中位数**	**IQR**
通信状态	**ADF**	23	126.5	0.738	-0.33	0.07	242.00	55.5	252	71
	APD	23	152.0	0.687	0.40	0.08	3.13	0.32	3.09	0.39
	NF	24	135.0	0.939	0.08	0.02	14.00	11.00	15	15.75
	TDF	24	141.0	0.812	-0.24	0.05	4122	2508.25	3372.5	4326
车队健康	**ADF**	24	237.0	**0.012**	2.53	0.52	335.50	93.50	301	53.5
	APD	24	113.0	0.303	-1.03	0.21	3.12	0.37	3.12	0.45
	NF	24	184.5	0.331	0.97	0.20	54.50	25.50	45.5	44.75
	TDF	24	220.0	**0.046**	2.00	0.41	18969.5	13614.75	13174	11241.5
消息	**ADF**	24	187.5	0.290	1.06	0.22	194.50	60.00	181.5	36.25
	APD	24	146.0	0.922	0.10	0.02	3.20	0.37	3.17	0.33
	NF	24	196.5	0.189	1.31	0.27	32.50	16.75	25	12.5
	TDF	24	197.5	**0.073**	1.79	0.37	6087	4688	4721.5	2652.5
时间线	**ADF**	24	145.0	0.900	0.13	0.03	348	89.00	354.5	98.5
	APD	24	159.0	0.812	-0.24	0.05	3.12	0.41	3.11	0.35
	NF	24	175.0	0.484	0.70	0.14	267.5	144.5	253.5	97.5
	TDF	24	162.0	0.747	0.32	0.07	97587.5	57945.5	81515.5	69155

注。ADF：平均注视时长；AOI：关注区域；APD：平均瞳孔直径；NF：注视次数；TDF：总注视时长。

MANUSCRIPT

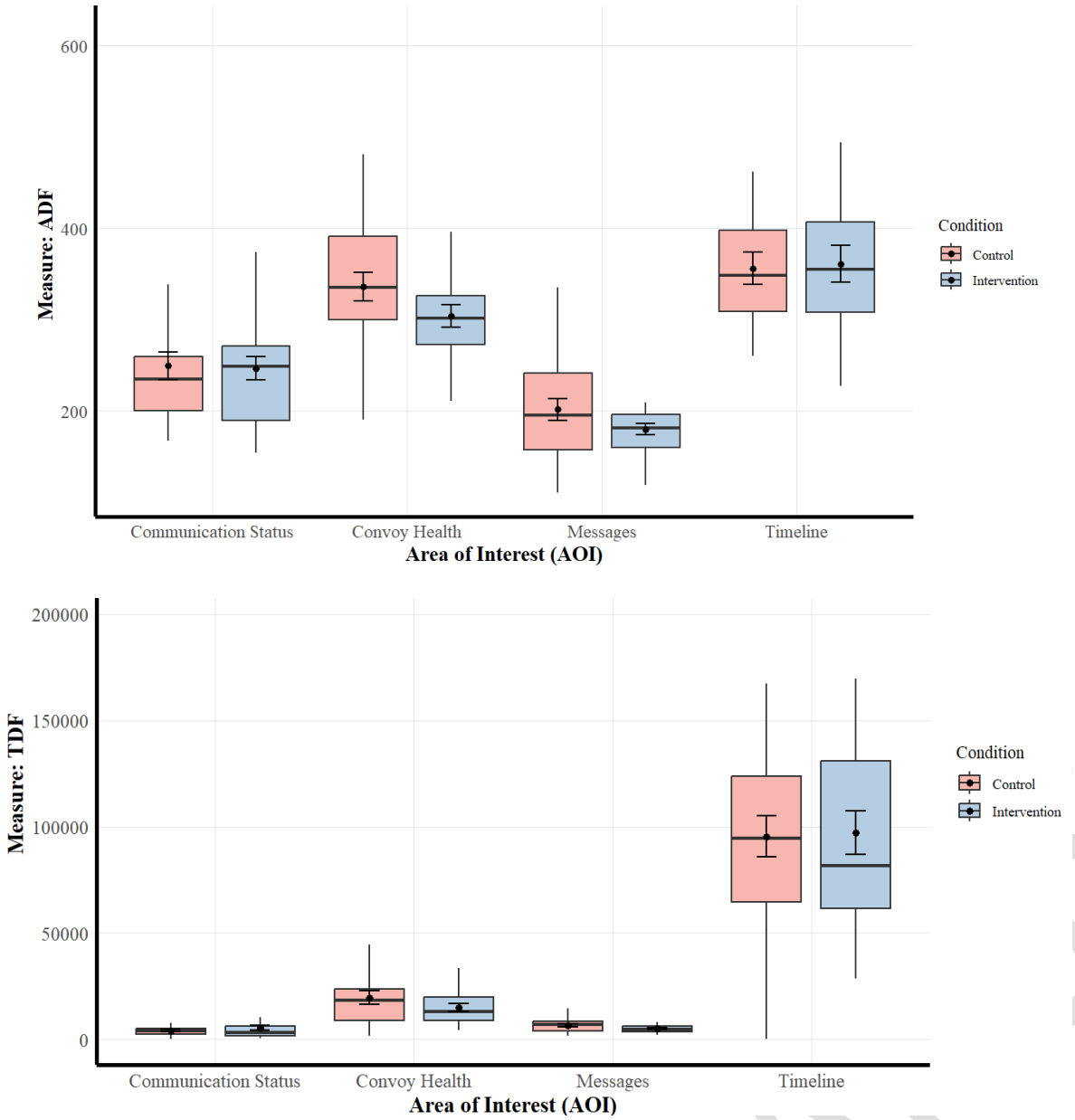


图 5。所选兴趣区中不同条件之间的平均注视时长（上）和总注视时长（下）。误差线表示95%置信区间。

3.3. 决策：做出首次决策所需的时间

方差分析结果显示，决策类型的主效应显著（$F(1, 68.49) = 4.38$，$p = .040$）。然而，条件的主效应以及条件与决策类型之间的交互作用均不显著。具体而言，相较于简单决策（对照组：中位数 = 4 s，四分位距 = 13；干预组：中位数 = 5 s，四分位距 = 9），参与者在复杂决策时的响应时间更长（对照组：中位数 = 11 s，四分位距 = 16；干预组：中位数 = 7 s，四分位距 = 10.5；见图 7）。后续使用

对齐秩转换数据进行的两两比较显示，对于简单决策，对照条件与干预条件之间没有显著差异，$t(70) = 1.00$，$p = .321$。同样，对于复杂决策，各条件下的恢复滞后差异也不显著，$t(70) = 0.48$，$p = .631$。

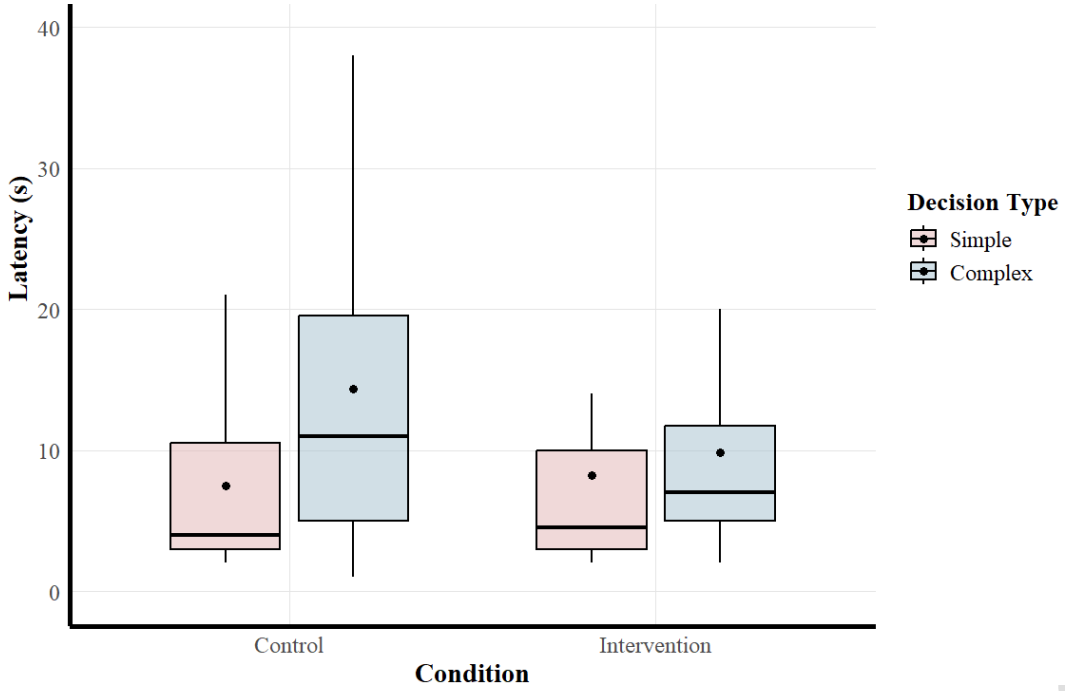


图6。经历中断后首次做出决策所需的时间，按决策类型（复杂与简单）和实验条件（对照组与干预组）划分。误差线表示95%置信区间。

3.4. 实验后访谈结果

我们对参与者进行了访谈，以了解他们使用测试平台及参与实验流程的体验。访谈揭示了与实验相关的三项主要发现：（1）所有参与者都表示使用测试平台的体验良好，尽管实验过程持续时间较长（2.5小时），但总体上都很享受；（2）自适应决策支持工具功能通过鼓励参与者提前规划，而非对正在发生的情况做出反应，提升了决策能力；（3）自适应决策支持工具功能足够隐蔽，因此尽管部分参与者在干预期间注意到了这些功能，但大多数人并未将两次试验之间的差异归因于这些功能。

当被问及整体体验时，所有参与者都给予了积极反馈，其中一些人将实验描述为类似游戏。谈到自适应决策支持工具时，一位参与者表示，其功能让自己在执行任务时感到更加安心。

MANUSCRIPT

“我觉得在玩这个游戏*[*实验*]*，时，我一直努力保持态势感知……*[*这些消息*]* 让我感到更安心，也更有意识。”——*P21*

参与者提到，自适应注意力引导决策支持工具的功能帮助他们制定计划并及时采取行动，尤其是在执行无人机监视时，从而在车队穿越敌对区域的过程中保护车队。

“是啊，所以当我看到无人机被摧毁的*[*消息*]*，时，我会查看地图，看看哪架无人机*[*被摧毁了*]*，再看看它是否处于关键区域。如果不是，我就可以先不管它。但如果是，我就得把握好重新分配无人机的时机……这样就不用等其他无人机之后再赶来。它已经在现场了。”——*P11*

“哦，是啊，我记得*[*看到自适应消息*]* ……在做决策时，我会想：‘好吧，如果情况是这样，如果无人机失效了，那我就得从其他地点调来别的无人机。’是啊，那时我开始更加警觉……这确实很有帮助……没有这些消息的话，我在*[*重新分配无人机*]*时可能会稍微晚一些。”——*P17*

4. 讨论

此前采用同一测试平台开展的研究，重点是在任务指挥员经历中断后为其提供认知支持（Sasangohar 等人，2014年；Scott 等人，2009）。这些研究采用了交互式时间线，任务指挥员可根据时间线上标记的事件快速恢复情境感知。然而，随着任务环境日益复杂，对动态、持续支持的需求也不断增加，而不再仅仅局限于中断后的恢复。因此，本研究开发了一种基于实时眼动追踪的注意力引导自适应决策支持工具，以提升监督层级决策表现。

我们的研究结果表明，基于实时眼动追踪的注意引导能够提升整体绩效。这与先前研究的结论一致：实时自适应视觉提示能够改善监控行为和情境感知——这些因素会显著影响绩效（Fortmann 和 Lüdtke，2013年；Fortmann 和 Mengeringhausen，2014年）。以往研究是在相对简单的环境中开展的，而我们的研究则证明，此类支持在更复杂的任务环境中同样有效。此外，绩效的提升也可能归因于注意选择性的增强。任务指挥员无需扫描所有可用信息来寻找潜在威胁和事件，而是可以专注于优先级较高的信息。这与 Frame 等人（2022年）的发现一致：当最相关的场景元素以视觉方式突出显示时，用户的表现更好。他们的研究比较了

两种用于军事情报中关键要素识别的视觉注意力辅助工具。一种高亮所有人员和车辆，另一种则有选择地高亮最相关的元素，包括武器和毒气罐等不太明显的提示。尽管后一种工具高亮的项目总体更少，但它带来了更好的识别效果。最终，Frame 等人指出，与高亮所有潜在元素相比，更有选择地引导注意力关注相关元素，能够提升理解能力和绩效（Frame 等人，2022年）。这些结果再次表明，在复杂环境中，有效的决策支持并不取决于提供更多信息，而在于提供引导更精准的信息，以优化注意分配。

一个值得关注的发现是，在第二次试验中经历干预条件的参与者，其整体表现得分显著提高。虽然在第一次试验中接受干预的参与者，其表现得分也略高于对照组，但这一差异未达到统计显著性。观察到的条件与试验顺序的交互作用表明，先经历控制条件、再接受自适应 DST 的参与者获益最大。一种可能的解释是，先完成控制条件、随后接受干预的参与者，同时受益于首次试验带来的学习效应和自适应 DST。此模式可能反映了学习效应：参与者在控制条件下建立了基线熟悉度，随后在第二次试验中更有效地运用了 DST。虽然这一发现凸显了 DST 在支持受训操作人员方面的潜力，但也限制了该效应对初次使用者或新手用户的普遍适用性。仍需开展更多研究，并采用组间设计来验证这些发现，以排除学习效应的影响。

在车队健康状况区域，平均注视时长和总注视时长均存在显著差异，且对照条件下的时长均长于干预组。该干预旨在影响参与者的注意分配，显著缩短了他们在车队健康状况区域的 ADF 和 TDF。这可能表明，在对照条件下，参与者必须更频繁、也更长时间地检查车队健康状况，因为长时间停留车队以及更频繁地遭到敌对势力攻击，导致健康状况更频繁、更迅速地下降；这些都是绩效不佳的表现。其他兴趣区均未显示显著差异，这表明干预的影响是特定的，并未广泛波及整个界面。尽管未达到统计显著性，消息感兴趣区域的 TDF 在不同条件间呈现一定差异趋势：对照条件下的参与者在该区域的注视时间更长。这种中等程度的效应可能表明，干预通过引导参与者将注意力转向显示屏上其他更相关的区域，帮助他们更高效地处理消息内容。不过，还需要进一步调查。

MANUSCRIPT

本研究的参与者没有表示对自适应界面的显著特征感到烦恼或不适，这表明该界面在提升绩效的同时，也带来了无干扰的用户体验。这与一些研究结果相反：在那些研究中，尽管高度显著的视觉特征有助于及时引导注意力，参与者却对这些特征感到非常烦恼（Bahr & Ford, 2011年；Lewandowska et al., 2022年；Takahashi, 2017；Tasse et al., 2016；Ward & Helton, 2022年）。事实上，自适应注意力引导线索非常 subtle，以至于 88% 的参与者表示没有注意到它们。Vlassova 等人的研究结果或许可以解释任务指挥员绩效有所提升的原因：无意识（即未受注意）信息仍可能影响人们的决策（Vlassova et al., 2014年）。不过，还需要更多研究来确认，在我们这样的复杂情境中，绩效提升与无意识信息之间是否存在直接关系。这一发现令人鼓舞，因为它表明，自适应界面或许能够在复杂环境中为监督员提供支持，同时避免引发负面情绪反应。这在复杂环境中尤为重要，因为人类操作员可能会承受很大压力。

局限性

本研究存在若干局限性。首先，研究样本量相对较小，且主要采用单一地点的招募策略，参与者以接受过军事训练的学生为主。其次，尽管实验环境高度仿真了现实中的军事指挥中心，但它仍然只是模拟环境。未来有必要开展研究，在真实的 C4环境中与真实的任务指挥官合作，评估此类工具的效用。第三，尽管 MOP 常用于军事任务评估，但它并非《人为因素》领域经过验证的人类绩效指标。要使其成为人为因素研究中可靠的绩效衡量指标，还需要经过更严格的验证流程。最后，尽管参与者在使用自适应决策支持工具时绩效有所提升，但与整个视野相比，显示自适应注意力引导决策支持工具的区域占比相对较小。自适应决策支持工具的位置经过策略性安排，以提高其被注意到的概率。然而，如果显著特征显示在任务指挥官视野之外，注意力引导便无法发挥作用。因此，目前正在开展进一步研究，以检验加入听觉显著性是否能更好地引导注意，即使任务指挥员的视线完全偏离显示视觉自适应决策支持工具的主显示器时也是如此。

5. 结论

本研究展示了自适应决策支持工具在复杂环境中提升任务指挥员决策能力方面的潜力。然而，仍需进一步调查，以评估多模态注意力引导系统在高压环境下能否动态引导注意力，使其聚焦于最关键的信息

MANUSCRIPT

并应用于时间敏感型操作；还需评估主动决策支持工具能否持续监测实时数据，帮助实现有效决策。此类创新有望在未来大幅提升任务成功率和运行效率。

致谢

作者感谢得克萨斯农工大学高级技术专家、休斯顿卫理公会学术研究院高级科学撰稿人 Jacob M. Kolman（硕士，ISMPP CMPPTM）审阅本稿早期版本的语言和表达清晰度。此外，作者还感谢众多本科生和研究生（Christian Perez、Cody Stanford、Michael Lin、Kenny Fisher、Gary Bittick、Omar Hernandez、Jacqueline Hernandez 和 Aishwariya Kanaparthi），他们在设计和开展实验方面提供了极大帮助。

MANUSCRI

参考文献

Bahr, G. S. 与 Ford, R. A.（2011年）。弹出窗口为何不起作用以及如何不起作用：弹出窗口诱发的眼动、用户情感与决策。《人类行为中的计算机》，27(2)，776–783。https://doi.org/10.1016/j.chb.2010.10.030

Breslow, L. A., Gartenberg, D., McCurry, J. M., & Trafton, J. G.（2014年）。动态操作员超负荷：一种预测监督控制期间工作负荷的模型。《IEEE 人机系统汇刊》，44(1), 30–40。https://doi.org/10.1109/TSMC.2013.2293317

Calvi, A., D'Amico, F., Ferrante, C., & Ciampoli, L. B.（2021年）。一项驾驶模拟器研究：评估增强现实技术提升超车操作安全性的潜力。2021年智能交通系统模型与技术国际会议（MT-ITS），1–6。https://doi.org/10.1109/MT-ITS49943.2021.9529339

Cohen, J.（2013年）。行为科学的统计功效分析（第0版）。Routledge。https://doi.org/10.4324/9780203771587

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). 从复杂动画中学习时的注意引导：看见即理解吗？《学习与教学》，20(2), 111–122. https://doi.org/10.1016/j.learninstruc.2009.02.010

Demir, M., Cohen, M., Johnson, C. J., Chiou, E. K., & Cooke, N. J. (2023). 探讨人际沟通与协调动态对人机团队效能的影响。《国际人机交互杂志》，39(9), 1841–1855. https://doi.org/10.1080/10447318.2022.2143004

Endsley, M. (1995年). 态势感知错误分类。《航空运行中的人为因素》：欧洲航空心理学协会（EAAP）第21届会议论文集，1。

Endsley, M.（2001）。复杂系统中的态势感知设计。第二届人类、人工制品与环境共生国际研讨会论文集，1–14。

Endsley, M., Bolte, B., & Jones, D.（2003）。态势感知恶魔：态势感知的敌人。载于《态势感知设计：一种以用户为中心的设计方法》（第1版）。CRC出版社。

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A.（2007）。G*Power 3：一款适用于社会科学、行为科学和生物医学科学的灵活统计功效分析程序。《行为研究方法》，39(2), 175–191。https://doi.org/10.3758/BF03193146

MANUSCRIPT

Fletcher, J. D.（2009）。军事领域的教育与训练技术。《科学》，323(5910), 72–75。https://doi.org/10.1126/science.1167778

Fontes, R., & Kamminga, J. (2023年3月24日). 乌克兰：人工智能战争的实战试验场。《国防》。https://www.nationaldefensemagazine.org/articles/2023/3/24/ukraine-a-living-lab-for-ai-warfare

Fortmann, F., & Lüdtke, A. (2013年). 一种用于辅助监督控制无人机群的智能情境感知自适应界面。2013年 IEEE工业信息学国际会议（INDIN）第11届会议，768–773。https://doi.org/10.1109/INDIN.2013.6622981

Fortmann, F., & Mengeringhausen, T. (2014年). 开发和评估用于辅助多无人机监督控制期间监控行为的辅助系统：D3CoS项目的经验。2014年欧洲认知工效学会议论文集，1–8。https://doi.org/10.1145/2637248.2637257

Frame, M., Maresca, A., Christensen-Salem, A., & Patterson, R. (2022年). 应用监视场景中用于人类意义建构的模拟识别辅助工具评估。《人为因素》。https://doi.org/10.1177/00187208221120461

Grant, E. R., & Spivey, M. J. (2003年). 眼动与问题解决：引导注意即引导思维。《心理科学》, 14(5), 462–466. https://doi.org/10.1111/1467-9280.02454

Hopp, P., Smith, C., Clegg, B., & Heggestad, E. (2005年). 中断管理：注意引导型触觉提示的运用。《人为因素》, 47(1), 1–11. https://doi.org/10.1518/0018720053653884

Humr, S. A., Canan, M., & Demir, M. (2023). 人工智能支持的决策中信任的时间演变。《人因与工效学学会年会论文集》, 21695067231193672. https://doi.org/10.1177/21695067231193672

Humr, S., Canan, M., & Demir, M. (2025). 改进人类与人工智能决策的量子概率方法。《熵》，27(2)，152。https://doi.org/10.3390/e27020152

联合参谋部 J-7. (2011年). 评估规划与执行指挥官手册。国防部。

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I.（2018年）。利用瞳孔直径和固定注视下的微扫视，通过眼动追踪认知负荷。PLOS ONE，13(9)，e0203629。https://doi.org/10.1371/journal.pone.0203629

MANUSCRIPT

Lewandowska, A., Dzisko, M., & Jankowski, J.（2022年）。调查对比度在图形用户界面外围区域习惯化与敏化效应中的作用。《科学报告》，12(1)。https://doi.org/10.1038/s41598-022-16284-2

Mack, A.（2003）。无意视盲：看而不见。《心理科学当前方向》，12(5)，180–184。https://doi.org/10.1111/1467-8721.01256

Mack, A., & Rock, I.（1998）。无意视盲：没有注意的知觉。载于《视觉注意》。

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A.（2008）。中断时长和任务要求对恢复暂停目标的影响。《实验心理学杂志：应用》，14(4)，299–313。https://doi.org/10.1037/a0014402

Neisser, U.（1979年）。选择性观看中的信息获取控制。载于《知觉及其发展》（第1版）。

Nonino, E., Gisler, J., Holzwarth, V., Hirt, C., & Kunz, A.（2021年）。虚拟环境中真实行走的细微注意引导（WOS:000758410300058）。310–315。https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00070

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2008年). 利用眼动预测后完成错误。SIGCHI计算机系统中的人为因素会议论文集，539–542。https://doi.org/10.1145/1357054.1357141

Ratwani, R. M., & Trafton, J. G. (2011年). 用于实时预测和预防后完成错误的眼动追踪系统。《人机交互》，26(3)，205–245。https://doi.org/10.1080/07370024.2011.601692

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). 看见还是看不见：注意是感知场景变化的必要条件。《心理科学》，8(5)，368–373。https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rickli, J.-M. 和 Mantellassi, F.（2024年）。GCSP 出版物 |《乌克兰战争：新兴技术与战争的未来的现实检验》。安全政策中心。https://www.gcsp.ch/publications/war-ukraine-reality-check-emerging-technologies-and-future-warfare

Roads，B.、Mozer，M. 和 Busey，T.（2016）。使用高亮训练注意专长。《PLOS ONE》，11(1)。https://doi.org/10.1371/journal.pone.0146266

MANUSCRIPT

Rogers，T.（2023）。压力下的人类信息加工。《应急管理杂志》，21(2)，141–154。https://doi.org/10.5055/jem.0756

Sasangohar, F., Scott, S. D., & Cummings, M. L. (2014年). 时间紧迫型控制任务中的监督层级中断恢复。《应用人体工程学》，45(4), 1148–1156. https://doi.org/10.1016/j.apergo.2014.02.005

Savick, D. S., Elliott, L. R., Zubal, O., & Stachowiak, C. (2008年). 音频和触觉提示对复杂模拟场景中士兵决策和导航的影响（编号 ARL-TR-4413）。美国陆军研究实验室。

Scott, S. D., Sasangohar, F., & Cummings, M. L. (2009). 面向指挥与控制行动的监督层级活动态势感知显示器研究。《2009年人类系统集成研讨会论文集》。人类系统集成研讨会，安纳波利斯，马里兰州，美国。

Simons, D. J., & Chabris, C. F. (1999). 我们身边的大猩猩：对动态事件持续存在的无意视盲。《知觉》, 28(9), 1059–1074. https://doi.org/10.1068/p281059

Simons, D. J., & Levin, D. T. (1997). 变化盲视。《认知科学趋势》, 1(7), 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2

Sklar, A. E., & Sarter, N. B. (1999). 美妙的振动：通过触觉反馈支持事件驱动领域中的注意分配与人机自动化协调。《人为因素》, 41(4), 543–552. https://doi.org/10.1518/001872099779656716

Stahl, P., Donmez, B., & Jamieson, G. (2016). 通过注意性和解释性车载显示器支持驾驶中的预判。《事故分析与预防》, 91, 103–113. https://doi.org/10.1016/j.aap.2016.02.030

Takahashi, H. (2017). 驾驶辅助系统周边视野中的视觉提示。《高级计算智能与智能信息学杂志》，21(3), 543–558. https://doi.org/10.20965/jaciii.2017.p0543

Tasse, D., Ankolekar, A., & Hailpern, J. (2016). 以讨喜且尽量不令人厌烦的方式吸引用户对网络应用的注意。2016年人机交互计算系统会议论文集，3324–3334. https://doi.org/10.1145/2858036.2858174

Tobii AB。（2019）。Tobii Pro Lab（版本 1.123）[英语]。Tobii AB。

MANUSCRIPT

Tobii AB。（2023 年 9 月）。Tobii Pro Nano [用户手册 v2.5]。Tobii Pro Nano。https://go.tobii.com/tobii-pro-nano-user-manual

Trafton, J. G., Altmann, E. M., & Brock, D. P. (2005年). 咦，我刚才在做什么？人们如何在中断后利用环境线索。《人因与工效学学会年会论文集》，49(3), 468–472. https://doi.org/10.1177/154193120504900354

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995年). 通过选择性调谐对视觉注意建模。《人工智能》，78(1–2), 507–545. https://doi.org/10.1016/0004-3702(95)00025-9

Van Der Wel, P., & Van Steenbergen, H. (2018年). 将瞳孔扩张作为认知控制任务中努力程度的指标：综述。《心理学通报与评论》，25(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G.（2004年）。衰老过程中的记忆负荷与认知性瞳孔反应。《心理生理学》，41(2)，167–174。https://doi.org/10.1111/j.1469-8986.2003.00148.x

Vlassova, A., Donkin, C., & Pearson, J.（2014年）。无意识信息会改变决策准确性，但不会影响信心。《美国国家科学院院刊》，111(45)，16214–16218。https://doi.org/10.1073/pnas.1403619111

Ward, M., & Helton, W.（2022年）。越多越好，还是越少越好？通过降低引导精度，改善单目头戴式显示器辅助的视觉搜索。《应用人体工程学》，102。https://doi.org/10.1016/j.apergo.2022.103720

Wickens, C. D. (1976). 分散注意对手动跟踪中信息处理的影响。《实验心理学杂志：人类知觉与表现》, 2(1), 1–13. https://doi.org/10.1037/0096-1523.2.1.1

Wickens, C. D. (2015). 在视觉工作场所中察觉事件：SEEV 和 NSEEV 模型。载于 R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma（编），《剑桥应用知觉研究手册》（第 1 版，第 749–768 页）。剑桥大学出版社。https://doi.org/10.1017/CBO9780511973017.046

Wickens, C. D., & Carswell, C. M. (1995年). 邻近兼容原则：其心理学基础及与显示设计的相关性。《Human Factors: 人因与工效学学会期刊》, 37(3), 473–494. https://doi.org/10.1518/001872095779049408

MANUSCRIPT

附录

半结构式实验后访谈问题。共有七个预先指定的问题，但必要时会提出追问。

1. 您对这项实验有什么看法？您的体验如何？ 2. 在获取态势感知、处理信息和进行决策时，显示屏上最有帮助或最实用的功能是什么？ 3. 您是否注意到两次试验之间有什么不同？如果有，具体是什么？ 4. 您是否觉得中断任务足以分散您的注意力，使您暂时忘记任务？为什么？ 5. 您是否觉得中断任务的持续时间足够长，能够使您暂时忘记任务？为什么？ 6. 您是否觉得该任务或实验环境足够真实？您是否有任何建议可以使其更加真实？ 7. 对于界面的可用性或任务，您还有其他评论或建议吗？

MANUSCRI