NASA/TM—20205011433

Characterization of How CO_2 Level May Impact Crew Performance Related to the HSIA Risk

Bettina L. Beard
NASA Ames Research Center

December 2020

NASA STI Program...in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov
- E-mail your question via to help@sti.nasa.gov
- Phone the NASA STI Help Desk at (757) 864-9658
- Write to:
  NASA STI Information Desk
  Mail Stop 148
  NASA Langley Research Center
  Hampton, VA 23681-2199

NASA/TM—20205011433

Characterization of How CO_2 Level May Impact Crew Performance Related to the HSIA Risk

Bettina L. Beard
NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

December 2020

Available from:

NASA STI Program
STI Support Services
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

This report is also available in electronic form at http://www.sti.nasa.gov
or http://ntrs.nasa.gov/

Table of Contents

List of Figures and Tables

Acronyms and Definitions

AI	artificial intelligence
AMO	Autonomous Mission Operations
ASO	Autonomous Systems and Operations (NASA)
C&W	caution and warning
CANTAB	Cambridge Neuropsychological Test Automated Battery
CM-d	crewmember-days
CM-h	crewmember-hours
CO_2	carbon dioxide
ConOps	Concept of Operations
CTA	cognitive task analysis
EMU	Extravehicular Mobility Unit
EVA	extravehicular activities
ExMC	Exploration Medical Capability
FAA	Federal Aviation Administration
HCAAM	Human Capabilities Assessment for Autonomous Missions
HFBP	Human Factors and Behavioral Performance
hr	hour(s)
HRP	Human Research Program (NASA)
HSI	Human-System Integration
HSIA	Human-System Integration Architecture
IFI	Items for Investigation
IS	intelligent systems
ISS	International Space Station
IT	information technology
MCC	Mission Control Center
MER	Mission Evaluation Room
min	minute
mmHG	millimeters of mercury
MSS	Mobile Servicing System
NASA	National Aviation and Space Administration
PI	Principal Investigator
ppm	parts per million
PVT	Psychomotor Vigilance Test
RPD	recognition primed decision
SA	situation awareness
SASO	Safe Autonomous Systems Operations
SMS	Strategic Management System
TC	Technology Challenges
TIM	technical interchange meeting
TLV	Threshold Limit Value
TOCA	total organic carbon analyzer
VOLT	Visual Object Learning Test

Characterization of How CO_2 Level May Impact Crew Performance Related to the HSIA Risk

Bettina L. Beard[1]

Safety and mission critical anomalies are inevitable on NASA exploration missions. Delays and interruptions in communication with Earth-experts drives the requirement that crew resolve these anomalies on their own. The Human Factors and Behavioral Performance (HFBP) element of the NASA Human Research Program (HRP) recently stood-up the Human-System Integration Architecture (HSIA) risk. The specific risk statement is:

> *"Given decreasing real-time ground support for execution of complex operations during future exploration missions, there is a possibility of adverse performance outcomes including that crew are unable to adequately respond to unanticipated critical malfunctions or detect safety critical procedural errors."*

Reliable, on-board capabilities will need to support the crew - not only to resolve anomalies, but also to promote situation awareness (SA) and to reduce workload. Unfortunately, the present level of technological advancement limits the range of achievable intelligent support (Wu & Vera, 2019). The Human Capabilities Assessment for Autonomous Missions (HCAAM) projects were devised to identify how technology may benefit crew situation awareness and enhance crew trust in the intelligent systems. A suite of time-critical and complex exploration tasks have been identified (Holden et al., 2019) that may advance human-integration Concept of Operations (ConOps) development. This ConOps (a future HFBP directed task) along with standards and guidelines (Holden et al., 2019) may be used as the foundation for human-system requirements for a self-reliant crew.

Spaceflight CO_2 levels are elevated relative to terrestrial levels. Above 2.5 mmHg there is an increase in the incidence of crew headaches (Law et al., 2014) and Mission Control reports of mood and performance changes in the crew. The concern is that elevated CO_2 effects may exacerbate the HSIA risk. Unfortunately, effects of higher concentrations of CO_2 on performance decrements are not well established.

The main goal of the current report is to ***identify how elevated CO_2 impacts the HSIA risk****. The approach taken is to trace the relationship between cognitive and motor abilities required for anomaly response with critical mission tasks and the level of technological advancement to support anomaly resolution to identify how elevated CO_2 may impact the HSIA risk. The conclusion is that elevated CO_2 could exacerbate the HSIA risk. Several research gaps are identified.*

---

[1] NASA Ames Research Center; Moffett Field, California.

1. Background

Anomaly: “an unexpected event, hardware or software damage, a departure from established procedures or performance, or a deviation of system, subsystem, or hardware or software performance outside certified or approved design and performance specification limits” (NASA-HDBK-8739.18).

1.1. A Broad View of NASA’s Plans and Concerns

The ultimate goal of this report is to characterize how research-identified cognitive and motor changes resulting from elevated levels of carbon dioxide (CO_2) could affect an exploration crew’s ability to independently respond to time-critical anomalies. The overarching context involves an interplay between three critical and interrelated components: NASA’s Artemis/Gateway plans, increased crew self-reliance and the risk that the crew will not be able to effectively respond to unexpected anomalies without Mission Control Center (MCC) and Mission Evaluation Room (MER) support. How these components relate to each other is shown in Figure 1.

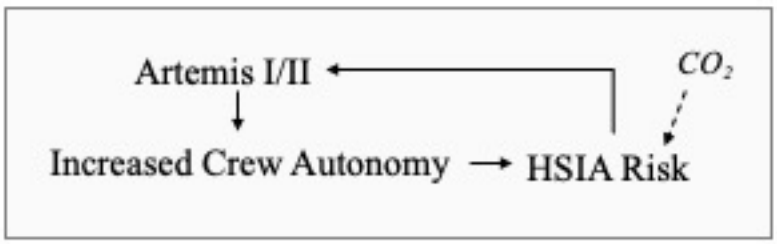

Figure 1. Critical and interrelated components in NASA’s plans.

First, NASA plans to send humans to the Moon as a springboard to Mars missions. The Artemis lunar exploration program will construct a lunar orbit base camp from which men and women will take excursions to the Moon’s Lunar South Pole within the next decade. Further excursions will provide an opportunity to test new technology and better understand spaceflight effects on the human body to enable human-rated Mars missions.

The second relevant component in NASA’s plan is a mandatory increase in crew self-reliance on these missions. Limited communication windows, communication latencies, communication disruptions, limited bandwidth and potential communication failures will constrain access and availability of ground personnel oversight requiring that crew independently identify, interpret and resolve unexpected events.

All spacecraft require a level of autonomy. To date, spacecraft, rovers, satellites and probes execute scripts triggered by events. This restricted form of autonomy is adequate when activity sequences can be determined well in advance, however it breaks down under increased uncertainty. Future manned missions to cis-lunar space, the surface of the moon and possibly further into our solar system, will require unprecedented autonomous operations.

The third component pertains to a critical risk identified by the Human Factors and Behavioral Performance (HFBP) element of the NASA Human Research Program (HRP). The complexity of NASA’s plans for the development of a lunar base, extravehicular activities (EVA), the complement of experiments to be performed and multi-year missions to Mars increases the probability of a wide variety of anomalies. It is unclear if the crew will have the ability to effectively respond without real-time ground support to these unanticipated, acute malfunctions or to detect critical procedure errors. This threat to mission safety and success is referred to as the Human-System Integration Architecture (HSIA) risk. The crew will need their own resources (skills, training) and specialized, on-board support systems to meet the new mission challenges.

1.2. Relevant HFBP Efforts

Since the approval of the HSIA risk into the HRP risk posture, the HFBP Element has analyzed Items for Investigation (IFI) entries, held a technical interchange meeting (TIM), solicited and funded seven research projects and convened a focus group. The purpose of these four efforts are to understand the:

- frequency and time criticality of high priority events. Vera et al. (2019) found that events that could have resulted in loss of crew or loss of mission occurred, on average, once every two months during the first two years of ISS operations. It was estimated that one-third of these were time-critical.
- current state of the art in intelligent systems (IS). Wu & Vera (2019) summarize the presentations (29 speakers) and discussions from a TIM held in August of 2018. The result was a set of Technology Challenges (TC). An effort being undertaken by the NASA Aviation Safety Program under the Safe Autonomous Systems Operations (SASO) Project (a.k.a. element) is of relevance here. Their goal is to:

  > "Develop autonomy that transparently teams with and even supervises human partners to allow a safe, affordable, scalable system. For highly complex systems, maintain human judgment, but allow for scalability through human automation teaming. Develop mixed initiative autonomy incorporating adjustable and adaptive automation that transparently teams with a human partner to allow safe, affordable, scalable monitoring and supervision of one to very many vehicles…"

These two NASA efforts could/should leverage one another to:

- define standards and requirements toward integrated, intelligent systems including advanced autonomous decision-support. HFBP funded seven HCAAM grants, summaries are available at https://humanresearchroadmap.nasa.gov/tasks/task.aspx?i=2244. This effort is ongoing and will be referred to here as the HCAAM Projects.
- identify candidate at-risk tasks for a self-reliant crew. Holden et al. (2019) summarize the discussions of a focus group of spaceflight subject matter experts (e.g., crew, mission control, crew training). Of particular concern to this group was the early detection of, and response to, crew health issues, poor teamwork and system faults. The crew's ability to process information during time-critical anomalies was considered a potential risk as was the absence of relevant data and data analytics about the complex system for crew decision-making. Other concerns revolved around crew training for anomalies and crew forgetting critical aspects of their pre-mission training. The focus group's list of risky tasks will be referred to as the Autonomous Task List. The list from the report is provided in Appendix A.

2. An Anomaly Response Framework

The framework presented in Figure 2 depicts how spaceflight and task-related stressors, training, team behavior, goals and metacognitive processes can influence crew anomaly response. The upper portion of Figure 2 represents intelligent systems composed of an operational Human-System Integration (HSI) Data System and Potential Crew Aiding Systems.

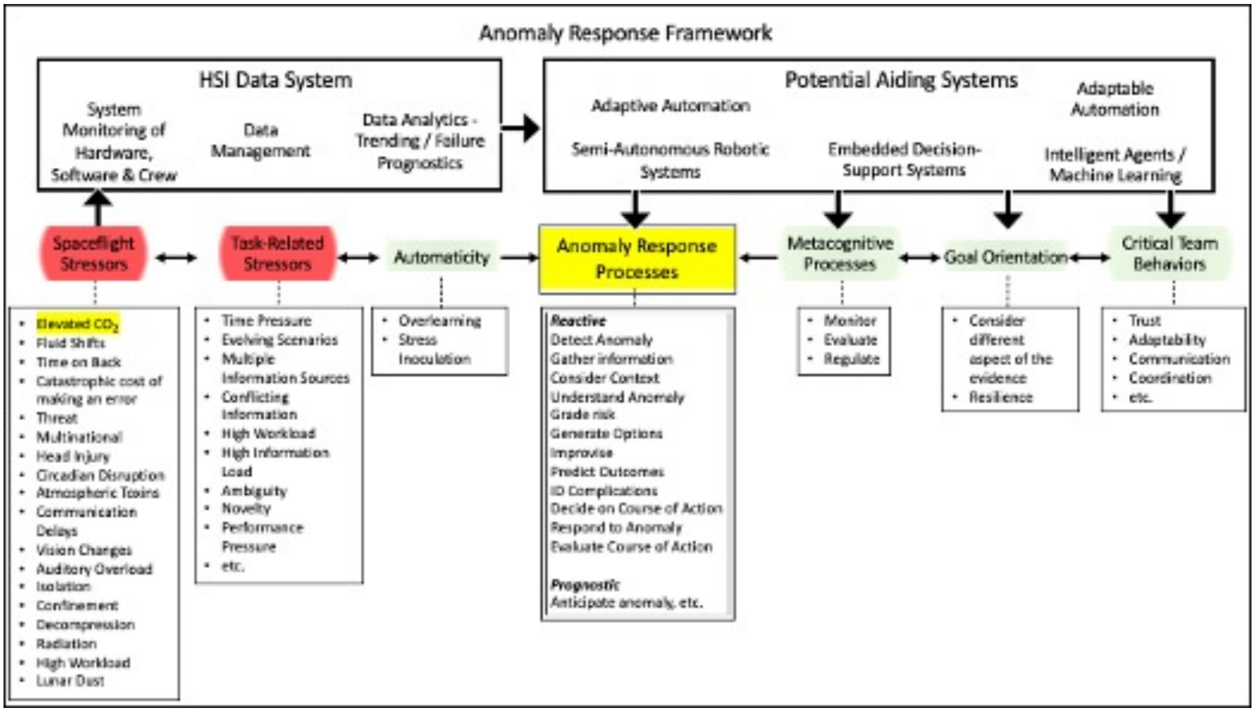

Figure 2. Theoretical characterization of factors that affect the crew's ability to self-reliantly respond to time-critical anomalies.

The envisioned HSI Data System would include: 1) continuous agent (hardware, software and liveware [crew]) monitoring for deviations from system-nominal; 2) a terrestrially-managed data warehouse of spaceflight and terrestrial normative and anomalous evidence; and 3) data analytics, synthesis and interpretation for known anomalies, anomaly occurrence probabilities, anomaly resolution options, resources required for the resolution, resource volume and power consumption, whether those resources are available on-board, where they are located, relevant procedures for anomaly resolution and other supporting data. This data system would incorporate stochastic elements, variable uncertainties and event dependencies. All data would be located on a single, secure platform. The Exploration Medical Capability (ExMC) is developing a medically-focused data architecture (Krihak, 2016). Many of their assumptions and progress could be applied to an HSI Data System. Ideally, a comprehensive data system integrating all HRP elements is envisioned.

Potential Aiding Systems (Figure 2, upper right) would support crew (and ground) situation awareness about key information. Crew information needs would be based on the Autonomous Task List (Holden et al., 2019) and a yet to be performed HFBP cognitive task analysis (CTA) that is based upon the tasks and abilities (Stuster et al., 2019) and generalizable skills and knowledge (Stuster et al., 2018) for exploration missions identified previously. An ExMC-funded cognitive task analysis (Daiker et al., 2020) for medical tasks could be folded into this overall HFBP CTA.

Spaceflight scenarios are often characterized by rapidly evolving and changing conditions, severe time compression, and high degrees of ambiguity and uncertainty. A self-reliant crew will be presented with an overwhelming amount of data that may require the coordinated performance of the team who must gather, process, integrate, communicate, and act on these data in support of a decision. The lower section of Figure 2 lists a variety of other stressors (both physical and psychological) that exist in the operational setting, not the least of which is the catastrophic costs of making an error. These spaceflight and task-related stressors are dependent upon the crew's training and the retention of the training material. Through overtraining emergency events, NASA helps to make the crew more resilient to stress effects. Certain task components become automatized, requiring fewer cognitive resources (Kirlik et al., 1998). Spaceflight and task-related stressors also interact with cognitive and metacognitive processes to influence critical team behavior during time-critical anomaly resolution. Noticing when something is amiss, assessing the anomaly and guiding a solution into place is referred to as a metacognitive loop (Anderson & Perlis, 2005). Metacognitive skills can help with time assessment, can reduce errors and help avoid decision biases (Cohen et al, 1996). This is all coupled with the fact that the spaceflight crew may be multinational.

Anomaly response processes (highlighted in yellow) are positioned in the center of Figure 2. Classic assumptions of how people recognize, diagnose and respond to anomalies were that people map problem symptoms to diagnostic categories. However, field research has shown a much more complex and dynamic interplay between team members requiring coordination, resilience and affordance. Human and machine anomaly response may be broken down into three categories (Woods & Hollnagel, 2006, Chapter 8):

- detect and recognize that an anomaly has occurred
- troubleshoot/diagnose the anomaly
- response contingency management and recovery from the anomaly

Anomaly detection refers to the recognition of data patterns that differ significantly from the majority of the data (Zimek & Filzmoser, 2019). Troubleshooting is a form of problem solving that is logical and involves a systematic search for the source of the problem. Initial anomaly diagnosis or classification depends on expertise. Contingency management consists of an analysis of possible damage or after-effects that may occur as a consequence of the anomaly and the response to the anomaly. Anomaly recovery strategies are used to restore the fault to an acceptable state. These categories are not distinct, sequential stages, but rather interwoven processes that require revised situation assessments. Further information about these categories will be provided in Sections 2.1.1 through 2.1.3.

2.1. CO_2 Impacts on Health and Performance

During International Space Station (ISS) assembly, a tradeoff was made between limits for ambient CO_2 and the increased power and supply required to maintain low levels of CO_2. As a result, spaceflight crew are exposed to CO_2 concentrations higher than Earth normal.[2] It is

[2] The current outdoor Earth concentration of CO_2 is approximately 415 parts per million (ppm; co2.earth/daily-co2).

anticipated that lunar and Mars transit vehicles and habitats will have the same power and supply constraints, and therefore that crews will be exposed to elevated CO_2.

CO_2 concentrations on the ISS hover around 2500 parts per million (ppm; Simon et al., 2018), but can depend on such things as where the sensor is located, how many crewmembers are located in the node or module, crew activity level and experiment off-gassing. In microgravity, localized CO_2 pockets can form around a crewmember's nose and mouth in poorly ventilated areas. Carr (2006) found a positive correlation between reported symptoms, such as headache, and CO_2 level on the ISS. ISS crew report headaches, may miss procedure steps or have more difficulty finishing tasks on schedule when concentrations reach 3950 ppm (Law & Alexander, 2016). Law et al. (2014) reported that CO_2 level, crew age and time in-flight were significantly related to headache probability. How chronic exposure to elevated CO_2 will affect crew anomaly resolution is unknown.

Section 2.2 will address the purpose of this report: how elevated CO_2 may impact the HSIA risk. The discussion is organized around the three anomaly response processes outlined above:

1. Anomaly detection and recognition (Section 2.2.1 and Table 1)
2. Troubleshooting and diagnosis (Section 2.2.2 and Table 2)
3. Response contingency management and resolution (Section 2.2.3 and Table 3)

Each anomaly process will be briefly described followed by a table showing the relationship between:

- **Key Technology Challenges** (Wu & Vera, 2019)
- ongoing **HCAAM projects**
- the **Autonomous Task List** (Holden et al., 2019)
- evidence for a CO_2-related decline in performance

The evidence for a CO_2-related decline in performance will be expanded for each technology challenge.

To err on the side of crew and mission safety, the discussion assumes a cautious approach. Only those scientific publications that have reported statistically-significant evidence for a CO_2-related decline in performance will be considered.

2.2. How Elevated CO_2 may Impact the HSIA Risk

2.2.1. Anomaly Detection and Recognition

The first step in anomaly resolution is to recognize that a problem exists. Spaceflight safety-critical anomalies must be quickly identified so that prompt corrective action can be taken. Currently, mission and program success can be attributed to large teams of engineers who understand the details of hardware design, test history, and operational characteristics of their respective systems. It is typically the ground personnel's expertise that can be credited for successful anomaly detection. See Appendix B for a brief discussion about anomaly detection and recognition during the Apollo Project.

Table 1 presents four technology challenges identified by Wu & Vera (2019) that relate to anomaly detection and recognition. Each will be discussed separately in the following four sub-sections (2.2.1.1 through 2.2.1.4).

Table 1. Anomaly Detection and Recognition				
Key Technology Challenge		*HCAAM Project*	*Autonomous Task List*	*Significant CO_2 Effects*
1	Real-time monitoring of vehicle, environment, team, resources and crew	Crew Task Performance Quantification (PI: Fanchiang)	Monitor … vehicle system performance (6a); Monitor system displays (8c); Trend monitoring (Cc)	Allen et al. (2016) Basner et al. (2017) Kajtár & Herczeg (2012) Manzey et al. (1998) Myhrvold et al. (1996) Satish et al. (2012) Scully et al. (2019) Sun et al. (1996) Yang et al. (1997
2	Link agent state with external sources or context (sensor/data fusion)		Integrate information (Cb)	Allen et al. (2016) Myhrvold et al. (1996) Satish et al. (2012) Savulich et al. (2019) Scully et al. (2019)
3	Anomaly or fault detection		Detect anomalies (Cd)	Allen et al. (2016) Basner et al. (2017) Gill et al. (2014) Kajtár & Herczeg (2012) Maula et al. (2017) Myhrvold et al. (1996) Satish et al. (2012) Savulich et al. (2019) Sayers et al. (1987) Scully et al. (2019) Zhang et al. (2017a) Zhang et al. (2017b)
4	Quantify and provide system reliability to build user trust	Conversation Trust Analysis (Lee)		There is no research on CO_2 effects on trust-in or reliance-on automation

2.2.1.1. Technology Challenge #1 in Table 1

The first Technology Challenge (TC#1) pertains to real-time monitoring and awareness of agent state. Continuous, real-time data monitoring is critical to maintain awareness of the state of the hardware, software and liveware [crew], and to detect an anomaly. Although Wu & Vera (2019) discuss a case where real-time monitoring by technology was successfully used (i.e., Autonomous Mission Operations [AMO]), this remains a technology challenge for most vehicle, environment, team, resource and crew data to effectively result in anomaly detection and recognition. This is because the physical world is inherently variable and it is difficult to know which changes are relevant. Experts are sensitive to event patterns. Interpretation depends upon what has preceded and what is expected to happen next. To recognize an anomaly, experts tune their attention to the future or to what is coming next (Christoffersen & Woods, 2003).

Monitoring displays is impractical for a busy crew for several reasons. First, it is very difficult to maintain attention in intellectually unchallenging, monotonous situations (see Robertson & O'Connell, 2010 for a review of "vigilant attention"). Further, multi-tasking, where only one task is system monitoring, is mentally fatiguing (Stark et al., 2000). Loss of vigilance at the onset of, or during, a time-critical event could compromise crew effectiveness (Molloy & Parasuraman, 1996).

This technological challenge includes vehicle, environment and team monitoring. Fanchiang's HCAAM project partially addresses this challenge by tracking crew task performance psychophysically (speed and accuracy) and biometrically (sensors). For this technology to be of value, the output must provide event patterns that will compel recognition of an underlying anomaly.

Within the Holden et al. (2019) Autonomous Task List (Table 1, fourth column), there are three tasks that best apply to TC#1. The tags listed inside of the parentheses (6a, 8c, and Cc) refer to the labels assigned by Holden et al. Purple font indicates a time critical task rating by the focus group (although the monitoring of vehicle system performance and trend monitoring are arguably also time-critical).

The fifth column of Table 1 lists publications reporting significant CO_2 exposure effects on tasks pertinent to a crew that is required to continuously monitor for anomaly detection (TC#1). Table 1.1 expands upon this research, summarizing details about the studies. The references are shown in the first column, the specific perceptual, cognitive or motor task is provided in the second column, the CO_2 level(s) in parts per million (ppm) at which significant performance changes were reported (compared to baseline) is given in the third column, the fourth column shows the duration of exposure (note that this is not necessarily the time before measures were taken) and the rightmost column provides the number of subjects tested.

Table 1.1. TC#1: Elevated CO_2 Outcomes, Real-time Monitoring				
Publication	*Task*	*CO_2 Concentration (ppm) Compared to Baseline*	*Exposure Duration*	*N*
	Rapid Response			
Myhrvold et al. (1996)	Simple reaction time	1000 to 1499, 1500 to 4000	6 hr	548
Myhrvold et al. (1996)	Choice reaction time	1499, 4000	6 hr	548
	Sustained Attention & Information Seeking			
Allen et al. (2016)	Focused activity	1400	8 hr	24
Basner et al. (2017)	PVT accuracy	5050	26.5 hr	6
Kajtár & Herczeg (2012)	Errors found while proofreading	3000, 4000	4 hr	10
Manzey et al. (1998)	Subjective alertness	12500	26 days	4
Manzey et al. (1998)	Unstable tracking	12500	26 days	4
Satish et al. (2012)	Focused activity	2500	2.5 hr	22
Scully et al. (2019)	Focused activity	1200	3 hr	22
	Perception			
Basner et al. (2017)	Line Orientation Discrimination speed	5050	26.5 hr	6
Sun et al. (1996)	Stereoacuity	25000	30 min	3
Yang et al. (1997)	Coherent motion	25000	30 min	3

Table 1.1 separates the references into rapid response, sustained attention/information seeking and perception capabilities. Myhrvold et al. (1996) found elevated CO_2 exposure within a range of values slowed performance on simple and choice reaction time tasks. These data were collected in a large group of school children in the classroom (age range: 15 to 20 years).

The ability to control attention and to seek alternative candidate explanations has been shown to be affected by elevated CO_2. Using the Strategic Management System (SMS), a scenario-based assessment of problem-solving, Allen et al. (2016) found that subjects performed fewer focused activities when exposed to 1400 ppm as compared to baseline (550 ppm). Scully et al. (2019) and Satish et al. (2012) also reported a decline in focused abilities at 1200 and 2500 ppm, respectively. The Psychomotor Vigilance Test (PVT) measures how quickly a subject can respond to the onset of a millisecond counter. Accuracy on the PVT is a function of responding before the stimulus onset (false start) and missing the stimulus completely (miss). Basner et al. (2017) found that PVT accuracy was affected by elevated CO_2 concentrations in subjects exposed to 5050 ppm while in a -12 deg head down position for approximately a day. Kajtár & Herczeg (2012) reported that exposure to 3000 and 4000 ppm reduced the number of proofreading errors identified compared to ambient air exposure and Manzey et al. (1998) found that subjective alertness ratings were lower and visuo-motor performance declined during 26 days of exposure to 12500 ppm when compared to pre- and post-exposure control data.

Visual perception, which is a requirement for monitoring displays, is also affected by elevated CO_2. Basner et al. (2017) reported that head-down tilt subjects were slower to adjust a line to be parallel to another line when exposed to CO_2 concentrations of 5050 ppm. Other visual perception tests including stereoacuity (Sun et al., 1996), visual motion perception (Yang et al., 1997), critical flicker fusion (Alpern & Hendley, 1952), visual acuity, depth perception, visual fields, color sensitivity, night vision and accommodation amplitude (Weitzman et al., 1969) have all shown significant declines at high CO_2 concentrations.

2.2.1.2. Technology Challenge #2 in Table 1

The second Technology Challenge (TC#2; see Table 1) in the anomaly detection and recognition category relates to associating crew state with external sources. An intelligent system could, for example, correlate Fanchiang's HCAAM crew state data with vehicle, environment and resource data. For technology, this is a sensor and data fusion issue. The data from different devices will either require standardization or, as is the ExMC data architecture strategy, data from various types of data structures can be translated by the system (Krihak, 2016).

Related to TC#2, Holden et al.'s (2019) Autonomous Task List (Table 1, fourth column) identifies information integration as a critical task for an autonomous crew.

The fifth column of Table 1 lists publications reporting significant CO_2 exposure effects on tasks pertinent to associating crew state with external sources (TC#2). Table 1.2 expands upon this research, summarizing details about the studies. Again, the references are shown in the first column, the specific perceptual, cognitive or motor task is provided in the second column, the CO_2 level(s) in parts per million (ppm) at which a significant change was reported (compared to baseline) is given in the third column, the fourth column shows the duration of exposure and the rightmost column provides the number of subjects tested. Seven tasks are separated into two categories. To relate crew state with external sources requires selective attention and ignoring distractions.

Table 1.2. TC#2: Elevated CO_2 Outcomes, Link Agent State with Context

Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Selective Attention			
Allen et al. (2016)	Task orientation	945	8 hr	24
Satish et al. (2012)	Task orientation	1000, 2500	2.5 hr	22
Scully et al. (2019)	Task orientation	1200, 2500 (improvement)	3 hr	22
Satish et al. (2012)	Applied activity	1000, 2500	2.5 hr	22
Scully et al. (2019)	Applied activity	1200, 2500	3 hr	22
	Inhibit Distractions			
Myhrvold et al. (1996)	Color word matching	1000 to 1499, 1500 to 4000	6 hr	548
Savulich et al. (2019)	Go/No Go omission errors	75000	20 min	27

Most cognitive tests involve problems that are decontextualized and well-structured. On the other hand, problems to be faced by crew will be remarkably complex, requiring inventive ways for restructuring the data using what has been called "sense making" (Lave, 1988). Even the Strategic Management System (SMS), used in studies of CO_2 effects on cognition, provides well-structured information with a correct and convergent answer or solution so that study participants may be scored. Using the SMS scenario-based assessment, Allen et al. (2016), Satish et al. (2012) and Scully et al. (2019) found a decline in task orientation, or the ability to "make specific decisions that affect completion of current tasks" with elevated CO_2. Satish et al. (2012) and Scully et al. (2019) reported that making decisions that are relevant to overall goal achievement (i.e., applied activity) was also significantly affected at CO_2 concentrations ranging from 1000 to 2500 ppm.

The ability to conditionally respond to a stimulus is also affected by elevated CO_2. Myhrvold et al. (1996) presented students with color words. The task was to only respond if the color of the word was consistent with the color-word. Savulich et al. (2019) reported that subjects made significantly more errors of omission (to fail to respond to a target before the time-out period has elapsed) and responded slower when exposed to 75000 ppm CO_2 for twenty minutes.

2.2.1.3. Technology Challenge #3 in Table 1

TC#3 relates to comparing current data with nominal configurations in order to identify risk hotspots, faults or anomalies. To detect an anomaly, for either humans or machines, requires finding patterns in gathered data that do not conform to expected behavior (i.e., outliers). This includes removing noisy-data and detecting novel and emergent patterns in the data. This is challenging (for both humans and machines) for four reasons: defining a normal region which encompasses every possible normal behavior is difficult, the boundary between normal and anomalous behavior is typically not precise, normal behavior may evolve and detection critically depends on the unique assumptions made. Recent work on anomaly detection for streaming data include the domain of monitoring sensor networks (Subramaniam et al., 2006), abnormal event detection (Davy et al., 2005) and anomaly detection in evolving data streams (Tan et al., 2011).

Holden et al.'s (2019) Autonomous Task List (Table 1, fourth column) identifies anomaly detection as a critical task for an autonomous crew.

The fifth column of Table 1 lists publications reporting significant CO_2 exposure effects on tasks pertinent to outlier identification (TC#3). Table 1.3 expands upon this research, summarizing details about the studies. Fourteen instances of significant CO_2 effects are separated into three categories. To relate crew state with external sources requires recognizing patterns, anomaly identification and working memory maintenance and manipulation.

Table 1.3. TC#3: Elevated CO_2 Outcomes, Anomaly or Fault Detection				
Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
No study identified	Mental Model Development			
No study identified	Projecting into the Future			
	Pattern Recognition			
Allen et al. (2016)	Information seeking	945, 1400	8 hr	24
Savulich et al. (2019)	Information search	75000	20 min	27
Scully et al. (2019)	Information seeking	1200	3 hr	22
Basner et al. (2017)	Visual obj learning acc	5050	26.5 hr	6
Scully et al. (2019)	Visual obj learning acc	2500	3 hr	22
	Anomaly Identification			
Kajtár & Herczeg (2012)	Errors found while proofreading	3000, 4000	4 hr	10
	WM Maintenance & Manipulation			
Gill et al. (2014)	Auditory n-back (1, 2 or 3-back difficulty adjustments)	75000, 85000	4 min	12
Maula et al. (2017)	WM maintenance & retrieval	2260	4 hr	36
Savulich et al. (2019)	Spatial WM	75000	20 min	27
Sayers et al. (1987) Exp 1	Subtraction speed	65000, 75000	5 min	10
Sayers et al. (1987) Exp 2	Subtraction speed	65000	10 min	21
Zhang et al. (2017a)	Addition	3000	4.25 hr	25
Zhang et al. (2017b)	Addition	5000	2.5 hr	25

To relate current data with nominal configurations, identifying data regularities, and irregularities (i.e., pattern recognition) is key. Theories of naturalistic decision-making and the recognition primed decision (RPD) model, in particular, highlight the importance of previous experience with particular situations, to enable the development of patterns or *mental models*. Mental models draw attention to relevant cues, guide interpretation of the cues, determine plausible goals, suggest typical responses to the situation and project the status into the future (Endsley, 1995). Current technology fails to provide information in a way that fosters mental model development. Although no studies have specifically addressed CO_2 effects on mental model development or on the ability to project into the future, the competence to seek information (Allen et al., 2016; Savulich et al., 2019; Scully et al., 2019) and to recognize learned objects (Scully et al., 2019) are both affected by elevated CO_2. Although not a life-threatening anomaly, the only CO_2 research to specifically tap into anomaly identification are the studies by Kajtar & Herczeg (2012) who measured errors identified while proofreading.

To identify outliers requires executive functions including working memory maintenance and manipulation. Gill et al. (2014) found a decline in n-back test scores (% correct) relating to the working memory requirements that would be required for anomaly detection. Maula et al. (2017)

tested 36, young students in an office setting on seven tasks. The only task affected by 2260 ppm during the 4 hour period was a recall task, with interference. Subjects first had to solve a mathematical problem and then memorize a presented word. The Savulich et al. (2019) spatial working memory task requires subjects to strategically find a target amongst an ever increasing number of options. They found that elevated CO_2 resulted in more errors, particularly for the more difficult conditions. Sayers et al. (1987) ran two experiments. In the first, they found that the speed with which subject's performed subtraction problems slowed at both 65000 and 75000 ppm with reported discomfort at the 75000 ppm level. In the second experiment, subjects were exposed to 65000 ppm. After 10 min the time taken to perform subtraction problems rose to a peak. Finally, Zhang et al. (2017a,b) found that errors increased and speed decreased when subjects were asked to add five two-digit numbers printed in the same vertical column when exposed to 3000 or 5000 ppm CO_2.

2.2.1.4. Technology Challenge #4 in Table 1

TC#4 relates to building user trust in technology. A meta-analytical review identified three factors that influence trust towards automation; team collaboration and tasking, operator abilities and personality and the automation attributes and performance, with the latter having the greatest influence (Hancock et al., 2011). The level of trust influences the operator's reliance on automation. Trust and reliance are influenced by system reliability (Moray et al., 2000; Wiegmann et al., 2001; Bailey & Scerbo, 2007; Ma & Kaber, 2007) although this is not a simple linear relationship (Wiegmann et al., 2001). Improvement in trust calibration might be implemented with a flexible form of automation (e.g., adaptable or adaptive, see Discussion).

Lee's HCAAM project partially addresses this challenge through the development of real-time conversation indicators of trust. Hergeth et al. (2016) demonstrated that higher trust and reliance in automation was associated with monitoring the automation less frequently (measured through gaze behavior).

2.2.2. Anomaly Diagnosis, Troubleshooting and Interventions

The second major process in anomaly response is to identify and isolate the fault. Many laboratory and real-world studies have investigated how individuals respond to time-critical emergencies (see Woods & Hollnagel, 2006; Watts-Perotti & Woods, 2007 for reviews). In realistic conditions, judgements, problem solving and decisions are complex, time-constrained and must be made before all the critical information that might be wanted is available. How a person interprets the information is strongly linked to the context and individual factors, such as expertise.

Development of information technology (IT) for crisis management has received increased attention in recent years. As an example, the RESCUE project has brought together computer scientists, engineers and social scientists to develop IT that quickly gathers, analyzes, disseminates and presents data in disaster situations (Mehrotra et al., 2003).

Table 2 presents two technology challenges identified by Wu & Vera (2019) that relate to anomaly diagnosis, troubleshooting and intervention. Each will be discussed separately in the following sub-sections (2.2.2.1 through 2.2.2.2)

Table 2. Anomaly Diagnosis				
	Key Technological Challenges	*HCAAM Projects*	*Autonomous Task List*	*Significant CO_2 Effects*
5	Provide essential information for diagnosis. Integrate data into information. Display only relevant information. Answer queries. Gather critical data. Problem-solve. Generate new data. Determine time criticality.	Virt Asst for LDEM Spcraft Anomaly (Selva). Resp Multimodal for SA (Stirling). Enhancing SA of Auto Proc (Schreckenghost).	Use medical software for diagnosis (1b); Troubleshoot for unknown issues (4c). Review documentation (6c). Diagnose software problem (6c). Integrate information from existing sources (Cb). Trend monitoring (Cc). Data analysis (Cd). Validate solutions (Cg). Use lessons learned (Ce);.	Allen et al. (2016) Allen et al. (2019) Gill et al. (2014) Lui et al. (2017) Maula et al. (2017) Satish et al. (2012) Snow et al. (2019) Zhang et al. (2017) Zhang et al. (2017)
6	Provide failure response recommendations with rationales. Decision support. Interactive procedures. Provide alternatives. Intelligent tutoring. Self-scheduling. Just-in-time training.	Virt Asst for LDEM Spcraft Anomaly (Selva). Resp Multimodal for SA (Stirling). Enhancing SA of Auto Proc (Schreckenghost).	Determine candidate paths (Cf).	Allen et al. (2016) Allen et al. (2019) Freiberger et al. (2016) Gill et al. (2014) Lui et al. (2017) Maula et al. (2017) Satish et al. (2012) Sayers et al. (1987) Scully et al. (2019) Snow et al. (2019) Vercruyssen (2014) Zhang et al. (2017)

2.2.2.1. Technology Challenge #5 in Table 2

The fifth Technology Challenge relates to the use of existing and new data and information to diagnose the potential sources of the anomaly, to isolate the anomaly and to determine the time criticality of the problem. The problem could be with a crewmember (e.g., medical) or a machine. Fault diagnosis, or the isolation of faults on defective systems, is a task requiring a high skill set. When a system fails, warning messages are often the only evidence available for assessing and diagnosing the underlying cause. The problem increases when there are multiple simultaneous faults.

Automated diagnostic tools using artificial intelligence (AI) techniques are an active area of research. To assist the crew, the system should gather, integrate and make accessible, existing data from various sources (e.g., observed symptoms, current context, domain knowledge). Appendix C provides an example from ISS operations where critical information had not been input into the on-board system. Without MCC assistance, the crew would have misinterpreted the situation.

In addition to gathering existing data, the system should proactively generate data (e.g., acquire vital measurements, run diagnostics). Data mining is also an active research area where patterns or anomalies in the data are discovered, analyzed and then depicted to the user.

Anomalies vary with respect to their urgency and therefore the degree to which the team must immediately respond to mitigate negative consequences. To know the time criticality of a problem, one must determine how fast the problem is progressing. When faced with an urgent event, team resources must shift away from ongoing routines and toward the event, potentially to "safe" the situation. Today's workstations do not provide adequate support for the manipulation of time-critical information.

TC#5 also refers to a system that provides essential information for anomaly troubleshooting. Imagine that the system has detected an unexpected pattern of data and notifies the crew with a caution, warning or alert. Assuming that the technology possesses the needed information, a major concern is averting an avalanche of data presented to the crew (i.e., danger of drowning in data but starving for knowledge; Simon, 1996). Therefore, another requirement would be for the system to fuse the data into actionable knowledge that is relevant for both immediate guidance and prolonged time horizons. Possibilities include a graph of graphs, risk heatmaps or virtual reality that translates formal engineering terminology into crew colloquial descriptions that are aligned with crew mental models and mission goals.

Extracting useful information from data is complicated and difficult for machines (Lee & Siau, 2001) and for humans (Greitzer, 2005). Some common problems for machines include "missing data" management (Brown & Kros, 2003), text classification (Zaghloul et al., 2009) or interpreting rare events (Weiss, 2004). For humans some of the processes required to sift through reems of data include:

- validation: ensuring the supplied data is relevant and correct
- sorting: arranging the data into sets or into some sequence
- summarizing: reducing the data to its main points
- aggregation: combining multiple pieces of data

While experts can perform these processes almost intuitively, a crew of four will not have expertise in all critical areas. They will, however, have received a great deal of training - underscoring the importance of information retention in, and retrieval from, long term memory. A critical experiment would be to assess crew knowledge, acquired throughout their multiple years of training, during exposure to elevated CO_2.

Not only should the technology push information, but it should be capable of answering the crew's queries for additional information. Data mining systems should enable information discovery from each crewmember's perspective (i.e., at different levels of granularity). It is important to display anomaly-relevant information in a manner that facilitates understanding for the individual crew members and for the team. In time-critical, dangerous situations, the time required for the crew to review information can lead to costly delays and errors. The Vista Project was initiated to help MER engineers interpret Space Shuttle telemetry data (Horvitz & Barry, 1995). The idea was to manage what was displayed using probabilistic and decision-theoretic models. Although the engineer could access any data they wish, a software program determined the costs and benefits of proactively displaying certain data.

Partially addressing this Technology Challenge from the HCAAM researchers are Texas A&M's (PI: Selva) HCAAM virtual assistant for Spacecraft Anomaly Treatment, University of Michigan's (PI: Stirling) Responsive Multimodal Human-Automation systems for SA, and TRACLabs' (PI: Schreckenghost) Enhancing SA of Automated Procedures.

Holden et al. (2019) identified nine tasks the crew will need to perform autonomously that are relevant to this Technology Challenge including the need for crew to access information from documentation, to use lessons learned, to proactively communicate, to analyze data and to integrate information.

The fifth column of Table 2 lists publications reporting significant CO_2 exposure effects on the use of existing data and information to diagnose the potential sources of a failure, to isolate the fault and to determine the time criticality of the problem (TC#5). Table 2.1 expands upon this research, summarizing details about the studies. Nine tasks were identified and categorized into three domains: use of existing information, executive functions and information integration.

Table 2.1. TC#5: Elevated CO_2 Outcomes, Provide Essential Information for Diagnosis

Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Use of existing information			
Allen et al. (2016)	Information usage	945	8 hr	24
Satish et al. (2012)	Information usage	1000, 2500	2.5 hr	22
	Executive Functions			
Allen et al. (2019)	Complex flight maneuvers	1500 (sign effect after 120 min), 2500 (sign effect after 80 min)	3 hr	30
Gill et al. (2014)	Auditory n-back (1, 2 or 3-back difficulty adjustments)	75000, 85000	4 min	12
Lui et al. (2017)	Addition, Subtraction	3000	3 hr	12
Maula et al. (2017)	Information retrieval	2600	4 hr	36
Snow et al. (2019)	Visual and verbal recognition	2700	< 60 min	31
Zhang et al. (2017)	Mental rotation	3000	5.25 hr	25
Zhang et al. (2017)	Addition	3000	5.25 hr	25
No study identified	Information Integration			

For humans to examine existing information in order to diagnose the source of a failure requires complex cognitive processing. To troubleshoot the source of an anomaly requires reasoning, working memory, flexible thinking and long term memory recall. "Information usage," which refers to the ability to use information that has been provided or gathered was found to be affected by elevated CO_2 (Allen et al., 2016; Satish et al., 2012).

Executive functions are affected by elevated CO_2 exposure. Allen et al. (2019) recruited 30 commercial airline pilots to perform 21 maneuvers in a flight simulator. An FAA-designated Pilot Examiner rated the pilot's performance. Passing rates were significantly lower for the more difficult maneuvers when the pilots were exposed to elevated CO_2. They found that the higher

CO_2 concentration affected performance more quickly. Gill et al. (2014) found that a 4 min exposure to CO_2 reduced response accuracy. Maula et al. (2017) exposed 36 subjects to 2600 ppm for 4 hours. The Maula et al. (2017) reported that a 4 hour exposure to 2600 ppm significantly affected performance on an information retrieval task that involved answering a specific question such as "Which multilingual country, with area over 100,000 km2, has the highest gross national income?" by inspecting a table of 20 countries containing seven items of information for each country. When asked to recognize visual and verbal items, subjects exposed to 2700 ppm performed more poorly (Snow et al., 2019). It is not clear whether the Snow et al. results were due to a failure of recognition or of learning. Mental rotation and addition response time increased and addition accuracy decreased during exposure to 3000 ppm CO_2 (Zhang et al., 2017). Lui et al. (2017) found that addition accuracy was significantly reduced, but their results were likely attributed to an increase in temperature rather than elevated CO_2.

CO_2 effects on recall of information stored in long term memory was explored by Sayers et al. (1987). The task was to read a short story and to recall information from the story at a later time. They did not find a significant effect of CO_2. This was the only identified study investigating long term memory retrieval.

No research studies have pointedly investigated CO_2 effects on information integration.

2.2.2.2. Technology Challenge #6 in Table 2

TC#6 relates to the provision of failure response recommendations with rationales. Technology that can provide solutions is commonly referred to as decision support. Decision support is crucial for an autonomous crew. The ExMC element of the HRP is currently assessing the feasibility of a clinical decision support system (Lindsey et al., 2016). Landon & O'Keefe (2018) describe the need for crew to have an intelligent tutoring system. Crew autonomy demonstrations include autonomous procedures (Beisert et al., 2013; Stetson et al., 2015) and crew self-scheduling (to be discussed in Section 2.2.3.2). Procedures, for example, can include additional information to help guide their execution or contain automated commands. Autonomous procedures have been successfully developed for non-critical tasks such as treadmill maintenance, ventilation flow measurements, extravehicular mobility units (EMU) loop scrubs and sensor placement activities (Frank et al., 2016). Rader et al. (2013) and Frank et al. (2013) summarize the NASA-sponsored demonstrations on the impact of time delay in analog settings. Over the past 10 years, the NASA Autonomous Systems and Operations (ASO) project has developed and demonstrated many autonomy enabling technologies using artificial intelligence techniques (Frank, 2019). Here, one example will be briefly described.

Currently, the ISS crew has no insight into the current state of water quality or system faults. They rely on MCC to monitor, control and plan water quality analyses and faults with the system. During the 2014-2015 time period, Frank et al. (2016) performed a 7-month demonstration where ISS crew autonomously performed tasks required for the total organic carbon analyzer (TOCA) including fault diagnosis and response. TOCA data were presented to the crew in real-time and software provided recommended responses to actual situations. Although MCC was still involved (the crew provided next-step recommendations to MCC) and only one system was involved during this demonstration, its success is encouraging for non-experts to respond to faults in a complex system. One problem with the demonstration related to the crew's extreme trust in the system. They had not been adequately trained to understand the system's limitations.

A vulnerability that human's display in anomaly response is referred to as premature attention narrowing, where the responder(s) become so fixated on an initial hypothesis that counter-evidence is discounted (Woods et al., 1987; De Keyser & Woods, 1990). Gettys et al. (1987) found that premature attentional narrowing can be reduced with explicit alternative hypotheses. Technology can aid the responder by providing these alternative hypotheses (Fenton et al., 2001). On the other hand, the list provided can narrow the range of data considered and hypotheses they explore (Layton et al., 1994). Therefore, system design characteristics that broaden the solution space to reduce mis-assessments and project possible worst case scenarios are of paramount importance. It is also important to train crew to develop an independent assessment prior to interacting with other crew members or the machine agent to reduce fixation and improve detection of weaknesses in an assessment or plan (Layton et al., 1994; Smith et al., 2009). If more than one anomaly candidate is proposed, the system could automatically perform other tests or employ historical data such as probabilities to discriminate between the alternatives. Otherwise, crew experience or trial and error may be required to determine the most appropriate response.

Rule-based systems, fault trees, model-based and machine learning approaches have been used to diagnose anomalies. Rule-based diagnostics characterize the experience of expert engineers in the form of rules. Fault trees, the most commonly used method, use symptoms or test results followed by a branching decision tree composed of actions, decisions and recommendations. Models represent the actual system, using observations and stored information about the system. Machine learning approaches exploit knowledge of previous successful or failed diagnoses to continually improve system performance or to use available domain data to automatically generate knowledge. Other approaches include fuzzy logic and artificial neural networks. These approaches face enormous challenges dealing with big data (Xu et al., 2017).

Holden et al. (2019) identified only one critical task that the crew must perform that is relevant to this Technology Challenge, to determine candidate paths.

The fifth column of Table 2 lists publications reporting significant CO_2 exposure effects relating to the provision of recommendations (TC#6). Table 2.2 expands upon this research, summarizing details about the studies. Ten tasks are separated into two categories, problem solving and response selection.

Table 2.2. TC#6: Elevated CO_2 Outcomes, Provide Failure Response Recommendations with Rationales				
Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Problem Solving			
Sayers et al. (1987) Exp 1	Logic	65000, 75000	20 min	10
Sayers et al. (1987) Exp 2	Logic	65000	70 min	21
Freiberger et al. (2016)	Planning & Problem Solving	75000	~30 min	42
	Response Selection			
Allen et al. (2016)	Strategy	945	8 hr	24
Satish et al. (2012)	Strategy	1000, 2500	2.5 hr	22
Scully et al. (2019)	Strategy	1200	3 hr	22
Allen et al. (2016)	Crisis Response	945	8 hr	24
Satish et al. (2012)	Crisis Response	2500	2.5 hr	22
Scully et al. (2019)	Crisis Response	1200	3 hrs	22
Vercruyssen (2014)	Response Selection	40000	60 min	6

For a self-reliant crew to effectively use an intelligent decision support system would require logic, strategic processing, planning, problem solving and a response. CO_2 research provides evidence that all of these cognitive requirements are compromised. Sayers et al. (1987) reported that subjects solving 16 logic problems (e.g., "B does not follow A; BA") took longer after a five-minute exposure to elevated CO_2 which peaked at 10 min exposure duration. Freiberger et al. (2016) tested subjects on a planning and problem solving task that involved an aircraft fuel management scenario with failing fuel pumps. The SMS Crisis Response refers to the "ability to plan, stay prepared, and strategize under emergency conditions" and Strategy refers to the "ability to use well-integrated solutions with the help of optimal use of information and planning." In two separate experiments, Vercruyssen (2014) found that elevated CO_2 slowed responses and increased errors on a response selection task, particularly when the response was made more complex and when the stimulus was degraded.

2.2.3. Anomaly Response Contingency Management and Recovery

Anomaly response and contingency management and recovery is the final process in anomaly response identified by Woods & Hollnagel (2006). Anomaly response contingency management is highly interwoven with information management and the diagnostic process. After the initial response to the critical anomaly, a self-reliant crew will need to monitor the system's reaction, continue troubleshooting the problem and react to the consequences of their initial response. In accordance with the HSIA risk, we are assuming a time-critical and possibly complex anomaly has occurred. It is likely that an intervention is required before the nature and source of the anomaly are identified to buy time for further diagnosis. How the crew responded to the anomaly will affect how the event progresses and the crew will need to predict that progression. The intervention may itself pose an additional, unexpected risk or provide further diagnostic information. As the situation evolves, the crew will need to evaluate the contingencies that might arise in light of the anomaly.

Table 3 summarizes three technology challenges identified by Wu & Vera (2019) that relate to anomaly response contingency management and recovery. Each will be discussed separately in the following three sub-sections (2.2.3.1 through 2.2.3.3)

Table 3. Anomaly Response Contingency Management and Recovery

	Key Technology Challenge	*HCAAM Projects*	*Autonomous Task List*	*Significant CO_2 Effects*
7	Adapt to the dynamics of the task	Virt Asst for LDEM Spcraft Anomaly (Selva)	Validate solutions (Cg);	Allen et al. (2016) Satish et al. (2012) Scully et al. (2019) Savulich et al. (2019) Snow et al. (2019)
8	Re-schedule in the case of schedule disruption; Finite resources/logistics management (RFID)	Crew Autonomy through Self-Scheduling (Marquez); Crew Task Performance Quantification (Fanchiang)	Schedule tasks and monitor performance of work … (Ab); Perform inventory and consumables management (15)	Allen et al. (2016) Basner et al. (2017) Freiberger et al. (2016) Manzey et al. (1998) Satish et al. (2012) Savulich et al. (2019) Scully et al. (2019)
9	Update database to reflect anomaly response processes and lessons learned		Document resolutions, new data … (Ch)	Basner et al. (2017) Manzey et al. (1998) Maula et al. (2017)

2.2.3.1. Technology Challenge #7 in Table 3

After responding to an anomaly, the crew will need to continue to monitor for new changes and whether the intervention produced the desired, and expected, results. Any abnormalities will require further interventions, evaluation and re-planning. The initial assessment of the situation may have been appropriate given the evidence available, but people are vulnerable to "garden path" problems. Garden path problems refer to the reduced ability to shift attention from a previous focus to a new one to explore relevant changes. Flexibility is critical. If new stimuli are distractions, they must be identified as so, and ignored.

After diagnosis, crew may follow the steps within a pre-written procedure. Procedures should be written to accommodate the general risk of misdiagnosis. In other words, contain a set of steps to redirect the crew to consider another plausible fault-type. Brittle AI refers to a system, designed for a certain task, that is unable to perform when faced with unanticipated events (Anderson & Perlis, 2005). Humans typically can perform quite well under sudden disturbances, but does elevated CO_2 lower the perturbation tolerance?

People will often overlook the side effects of changes to a plan (Smith et al., 2004). This is a time where premature narrowing can again occur. Anomaly response can deteriorate when the responder loses track of, or doesn't grasp, the implications of cascading events (Smith et al., 1997; Woods & Patterson 2000). Re-planning may focus on working around specific bottlenecks and missing the byproducts of other constraints (Shattuck & Woods 2000). Responses can be poorly synchronized in time leaving gaps that look locally appropriate but that actually work at cross-purposes when considered from a broader perspective (Klein, 2007). Collaborative cross-checks across crew members are critical at this time (Patterson et al., 2004; Fischer & Orasanu, 2000). Crew should be trained to implement broadening checks to help to revise previous assessments. Contingency plans may guide responses to evolving situations.

Partially addressing this Technology Challenge may be Texas A&M's (PI: Selva) HCAAM virtual assistant.

Holden et al.'s (2019) Autonomous Task List identifies a crew task referred to as "validate solutions" that relates to this Technology Challenge.

The fifth column of Table 3 lists publications reporting significant CO_2 exposure effects for adapting to the dynamics of a task (TC#7). Table 3.1 expands upon this research, summarizing details about five studies that have reported a reduction in cognitive flexibility during elevated CO_2 exposure.

Table 3.1. TC#7: Elevated CO_2 Outcomes, Adapt to the Dynamics of the Task

Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Cognitive Flexibility			
Savulich et al. (2019) Exp 1	Extra-dimensional shift errors	75000	20 min	44
Snow et al. (2019)	Shifting Attention Task of the CNS Vital Signs test battery	2700	50 min	31
Allen et al. (2016)	Breadth of Approach	945	8 hr	24
Satish et al. (2012)	Breadth of Approach	1000, 2500	2.5 hr	22
Scully et al. (2019)	Breadth of Approach	1200	3 hr	22
No study identified	Team Coordination			

Cognitive flexibility is a critical executive function that refers to the ability to adapt behaviors in response to changes. It is typically investigated using task-switching paradigms. Savulich et al. (2019) used a complex rule acquisition task from the Cambridge Neuropsychological Test Automated Battery (CANTAB) test battery that they suggested required visual discrimination, attentional set formation, maintenance of attention, set shifting and cognitive flexibility. Using feedback, subjects had more difficulty learning an arbitrary rule that changes after six correct responses when exposed to elevated CO_2. Snow et al. (2019) found that at 2700 ppm, subjects had significantly more errors while shifting from one instruction set to another. Satish et al.

(2012) and Allen et al. (2016) reported that flexibility in approach to the task (i.e., breadth of approach) was affected by elevated CO_2.

No research studies were identified that have pointedly investigated CO_2 effects on teamwork.

2.2.3.2. Technology Challenge #8 in Table 3

Anomalous events will have disrupted ongoing plans that can have cascading effects and change the tempo of operations. The crew may need to manage multiple interleaved tasks and be prepared to revise their assessment as new evidence emerges. It is vital that during critical times and critical operations that alarms and alerts provide informative information (Kemeny, 1979), are discriminable from one another (Patterson, 1990) and selectively control attention (Murray & Cox, 1989) without high false alarm rates (Getty et al., 1995). Alarms must be explicitly designed to function effectively to redirect attention or serve as part of a distributed system that coordinates activities as situations evolve (Woods & Hollnagel, 2006). To do so will require a system that is aware of what the crew is currently doing to judge whether they should be interrupted.

The anomaly may have implications for plans in progress requiring reprioritizing mission goals. During exploration missions, the crew need the ability to self-schedule and to re-schedule. To that end, Marquez et al. (2019) reported on a technology demonstration of a self-scheduling tool referred to as Playbook. Over the course of their mission, one astronaut used the tool with many Lessons Learned for the developers. Marquez's Crew Autonomy through Self-Scheduling HCAAM project directly addresses the need for self-scheduling. Interactions between this project and Fanchiang's Crew Task Performance Quantification HCAAM project will further tackle this need.

Often, others will not have been present during the initial handling of an anomaly and must be updated. The update should provide data and information about the events, the dynamics and temporal flow of the events, other potentially relevant parameters and what actions or interventions have been and are being taken. Technology could aid in this storytelling if the distributed system is coordinated. Otherwise, the attending crewmember will need to communicate the anomaly status while attending to and handling cascading events. It is also important for the crew to project how the anomaly will evolve into the future.

The fifth column of Table 3 lists publications reporting significant CO_2 exposure effects on rescheduling in the case of schedule disruption (TC#8). Table 3.2 expands upon this research, summarizing details about the studies.

Table 3.2. TC#8: Elevated CO_2 Outcomes, Recognizing and Handling Cascading Events				
Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Working Memory			
Savulich et al. (2019)	Spatial WM	75000	20 min	27
	Visuo-Motor Responses			
Basner et al. (2017)	Motor Praxis accuracy	5050	26.5 hr	6
Manzey et al. (1998)	Unstable tracking	12500	26 days	4
	Planning			
Freiberger et al. (2016)	Planning & Problem Solving	75000	~30 min	42
	Focused Activity			
Allen et al. (2016)	Focused activity	1400	8 hr	24
Satish et al. (2012)	Focused activity	2500	2.5 hr	22
Scully et al. (2019)	Focused activity	1200	3 hr	22
No study identified	Projecting into the Future			

As indicated in previous technology challenge tables, elevated CO_2 effects spatial working memory (Savulich et al., 2019), visuo-motor tracking (Basner et al., 2017; Manzey et al., 1998), planning (Freiberger et al., 2016) and focused activity (Satish et al., 2012; Allen et al., 2016; Scully et al., 2019) which would all be required to manipulate the Playbook interface.

2.2.3.3. Technology Challenge #9 in Table 3

TC#9 refers to updating databases to reflect anomaly response processes and lessons learned. Erroneous or unusual observed behaviors must be recorded. Currently, mission operators and integration specialists record problem reports of off-nominal performance, deviations from design and human errors that occur while building and operating these systems. Anomaly, or problem, reports contain descriptions of the anomalous event, a description of its root cause, a risk rating, a characterization of the corrective actions and the residual risk arising from the corrective actions. Layman et al. (2012) analyzed over 14 thousand unmanned mission anomalies and found that approximately 25% of them are software related. Since then, that same research group has focused on problem reports which contain human operator errors in addition to hardware and software defects or environment induced failures (Layman et al., 2016). They report that analyzing these data is an intensive manual process because of the range of forms used, the use of natural language and data quality issues. There are a number of publications addressing the challenge of analyzing unstructured data (e.g., Agrawal et al., 2018).

In addition, lessons learned need to be recorded. This requires recalling actions taken and their consequences during the event, typing into a lessons learned database, verbally updating the ground, projecting how this information may be used in the future and concentrating on the task at hand.

Holden et al. (2019) list document resolutions, new data and lessons learned in their cross-cutting task list.

TC#9 requires many of the cognitive and motor capabilities that were listed under other challenges. Rather than repeat these discussions of the individual studies, the reader is referred to the relevant sections in Table 3.3. There were no studies identified that have investigated CO_2 effects on the task of communication, referring to communication with the ground.

Table 3.3. TC#9: Elevated CO_2 Outcomes, Update Databases to Reflect Anomaly Response Processes and Lessons Learned

Publication	*Task*	*CO_2 Concentration (ppm) Decreased Performance Compared to Baseline*	*Exposure Duration*	*N*
	Recall			
Maula et al. (2017) See discussion for TC#3 and TC#5	Information retrieval	2600	4 hr	36
	Motor Response			
Basner et al. (2017) See discussion for TC#8	Motor Praxis accuracy	5050	26.5 hr	6
Manzey et al. (1998) See discussion for TC#1 and TC#8	Unstable tracking	12500	26 days	4
No study identified	Communication			
See discussion for TC#3	Pattern Recognition/ Projecting			
Allen et al. (2016)	Information seeking	945, 1400	8 hr	24
Savulich et al. (2019)	Information search	75000	20 min	27
Scully et al. (2019)	Information seeking	1200	3 hr	22
Basner et al. (2017)	Visual obj learning acc	5050	26.5 hr	6
Scully et al. (2019)	Visual obj learning acc	2500	3 hr	22
See discussion for TC#2	Selective Attention			
Allen et al. (2016)	Task orientation	945	8 hr	24
Satish et al. (2012)	Task orientation	1000, 2500	2.5 hr	22
Scully et al. (2019)	Task orientation	1200, 2500 (improvement)	3 hr	22
Satish et al. (2012)	Applied activity	1000, 2500	2.5 hr	22
Scully et al. (2019)	Applied activity	1200, 2500	3 hr	22

Two studies were identified that did not find that text typing errors nor number of words typed were affected by elevated CO_2 levels ranging from 3000 to 5000 ppm for 4.25 and 2.5 hours exposure durations, respectively (Zhang et al., 2016; Zhang et al., 2017). However, subjects were copying text and the researchers did not log keystrokes. Keystroke logging provides a detailed record of the process of writing as it unfolds in time. Pauses, bursts and revisions can provide a general interpretation of the cognitive processes involved (Galbraith & Baaijen, 2019). Pauses reflect levels of planning, reflection, rereading and text production. Bursts of keystrokes reflect an initial formulation of thought or improvement to previously formulated text. Revisions reflect semi-automatic correction of errors or a systematic attempt to modify content. Since crewmembers will be creating their own text (not copying text), a valuable experiment would be to measure potential CO_2 elevation effects on keystroke parameters, rather than merely typing speed and errors.

3. Discussion

Substantial resources are required to remove carbon dioxide within spacecraft. As a result average CO_2 levels are elevated relative to Earth normal. The goal of this report was to characterize how research-identified cognitive and motor changes resulting from elevated levels of carbon dioxide (CO_2) could affect an exploration crew's ability to independently respond to time-critical anomalies. Exploration crews to the moon and then to Mars will require increasing self-reliance to resolve the issues that are currently handled by a large team of terrestrial experts. The current assessment of published research revealed statistically significant CO_2 effects on performance in the three main anomaly response categories; detection and recognition that an anomaly has occurred, diagnosis of the anomaly and response contingency management and recovery.

It would be beneficial to have software that could aid the crew through anomaly resolution, however, Wu & Vera reported that the current maturity of technology is inadequate. They refer to these inadequacies as Technology Challenges. Nine technology challenges align with the three anomaly response processes. Eight of the nine technology challenges are associated with published research showing that performance is significantly affected by elevated CO_2 and will require technological support.

3.1. A Threshold Limit Value (TLV)

This report has rationally addressed how elevated CO_2 could affect a crew's response to anomalies. To support the setting of a TLV, Figure 3 plots exposure duration as a function of CO_2 concentration for the research contained in Tables 1, 2, and 3.

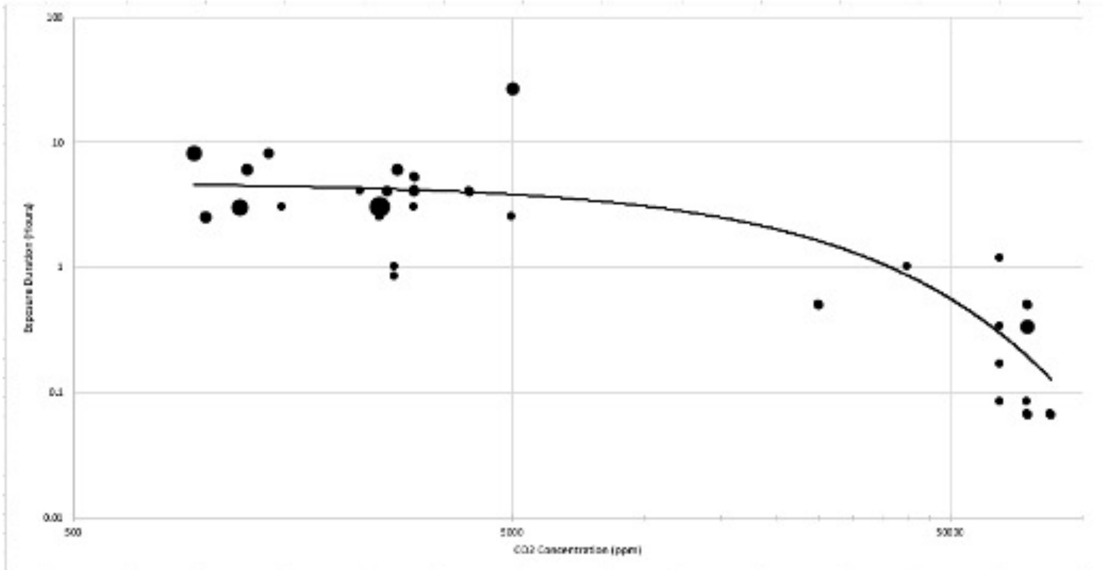

Figure 3. Graphic containing the studies reporting a significant effect of elevated CO_2 on cognitive or motor performance. CO_2 exposure duration (in hours) is shown as a function of CO_2 concentration in parts per million (ppm). Because a wide range of durations and concentrations have been used in the literature, the axes are in log-log coordinates. Point size is an indicator of how many different cognitive or motor tasks were significantly affected by CO_2. The largest points, for example, represents 6 tasks that were negatively affected at that exposure duration and CO_2 concentration. The solid line is an exponential function fit to the data.

ISS CO_2 levels predominantly fluctuate in the 3000 to 5000 ppm range. Unlike the published literature, the crew's exposure is chronic. It can be seen in Figure 3 that the data at CO_2 concentrations within the range relevant to spaceflight were obtained at short-term exposure durations (i.e., less than an hour to up to 26.5 hours). Except for one study (Manzey et al., 1998, not shown in this Figure) where only 4 subjects were were exposed for 26 days (624 hours), higher CO_2 concentrations are affiliated with exposure durations of just over an hour or less. The fitted curve could be used to help define a TLV. Unfortunately, there are missing data from 5050 to 25000 ppm in order to be certain that this curve depicts an accurate threshold (e.g., a linear relationship may be better fit). But, it does represent current knowledge.

NASA's main goal is to send and return crew on exploration missions without compromising safety or the mission. Although there are some studies that have failed to show negative effects of elevated CO_2, many studies *have* established adverse outcomes. A conservative approach therefore dictates that the effects of elevated CO_2 be classified as a risk to anomaly response. Figure 2 bounds the problem of the HSIA risk. The Figure depicts how multiple spaceflight and task-related stressors can affect the crew's ability to respond to anomalies. The effects of these stressors in concert will only exacerbate the effects of elevated CO_2, underscoring the importance of a conservative approach.

3.2. Critique of the CO_2 Research

It is difficult to determine why some research shows no effect of elevated CO_2, while other research show a significant effect. Because of the complexity of human physiological and psychological processes, the range of tasks and participant experience with the task, other factors may be confounding the results. For example, exposure to elevated CO_2 can result in headache and subjective fatigue which are both associated with a decline in sustained attention. However, most studies have not collected this vital information. In addition, wide differences in CO_2 tolerance have been observed. Most result sections provide averaged data, obscuring potential individual differences in response to elevated CO_2. Future studies investigating the performance implications of elevated CO_2 should gather, analyze and publish information about the potential confounds that could help uncover the apparent discrepancies in the literature.

Interestingly, we (ground-based earthlings) are often exposed to elevated CO_2 concentrations. Depending upon the number of people, square footage, and ventilation rate, indoor CO_2 concentrations can reach:

- 6077 ppm in submarines (Mudgett et al., 2018; average of 2600 ppm)
- 1400 ppm in offices (Arendt et al., 2018)
- 2300 ppm in school rooms (Wargocki et al., 2020)
- 771 ppm in underground subway stations (Hwang et al., 2017)
- 1353 to 3000 ppm in airplane cabins (Cao et al., 2018)
- 20000 ppm in motorcycle helmets (Bruhwiler et al., 2005)

How acclimatized the study participants are to elevated CO_2 was not determined in any of the studies referenced in this report. Further, how well the astronauts acclimatize to chronically elevated CO_2 during spaceflight would need to be systematically identified to truly match an astronaut-like participant population with spaceflight crew.

Technical troubleshooting processes differ between experts and novices (Johnson, 1988). Experts not only have extensive knowledge, their knowledge is more effectively organized and better accessible in long-term memory (Ericsson & Lehmann, 1996). Non experts will need to build a mental representation of the system to reason about the system's behavior. Crew on a Mars mission will have been trained to identify most expected anomalies. Although they may not be experts in the specific technology, the amount and organization of their domain knowledge will allow them to build a representation faster, and their strategic knowledge permits the application of effective strategies (Schaafstal et al., 2000).

There is a need to conduct studies on the effects of elevated CO_2 on problem solving (which includes decision making) in a complex environment using scenarios that are relevant to the participant's expertise. Participants who are not trained for the situation in the scenarios will not respond in the same way as experts. To conduct meaningful research, the experimental tasks must be selected or constructed to resemble the sub-population of interest. For example, Cao et al. (2019) asked 30 commercial airline pilots to perform a series of maneuvers *in a flight simulator* at three CO_2 concentrations: 700, 1500 and 2500 ppm. Pilots were less likely to pass a maneuver with increasing CO_2, particularly for the more difficult maneuvers. It is unclear how the results of studies with astronaut-like participants tested within medical or military scenarios that are not time- or safety-critical will generalize to highly-trained astronauts responding to an emergency in space. Taken a step further, reliance on research using context-free psychophysical tests reduces NASA's ability to generalize how a crew will respond to elevated CO_2.

3.3. Technological Support

An important issue for the definition of functional requirements of an exploration spacecraft is the question of how automation could aid the crew given the stressors to which they are exposed. NASA should develop and test principles for decision support, information display, just-in-time and recurrent training system design, and simulation that would mitigate elevated CO_2 stressor effects. To take a practical approach to such a complex problem requires strong theoretical positions such as naturalistic decision making, team performance and effectiveness and shared mental models. These theoretical perspectives describe how highly trained crew would gather and use information, how they perceive the situation, how teammates will interact with one another and how task demands will lead to strategy adjustments. From these theories, hypotheses can be generated which will in turn drive required empirical work.

Harris et al. (1995) found that operator performance is improved, and fatigue and workload reduced, with adaptive automation that performs a tracking subtask while the subject manages resources and performs other tasks. This topic relates to two critical decisions that the HRP must make: 1) Should task allocation be static with a fixed allocation between the crew and the technology or be adjusted over time and tailored to the operators' competences and current workload (Sauer et al., 2012), and if adjustable automation is chosen; and 2) Will spaceflight systems be adaptable, adaptive or both? In adaptable automation, task allocation would be decided by the crew. In adaptive automation the technology makes this decision (Parasuraman & Wickens, 2008). There are benefits to both adaptable and adaptive automation over fixed levels of automation (e.g., Sauer et al., 2012). Moray et al. (2000) found that an event-driven adaptive automation mode did not allow operators to adapt the automation level to their current need, since the level of automation was predefined for each type of fault. In contrast, adaptable automation permitted actual preferences to be determined, according to such factors as level of trust towards the automation. It is likely that some blend of fixed, adaptable and adaptive automation is required, depending on technological maturity. NASA-STD-3001_Vol_2_revb section 10.6.1.6 requires that the operator be able to determine the level of automation. Other critical requirements of adaptable automation are absent in the standard such as providing the remaining time to a change in level of automation, reasons for the level and a preview of ongoing and future actions (Beggiato et al., 2015).

3.4. Research Gaps

Several research gaps were identified during this project:

1. A ***cognitive*** *task analysis* that is based upon the tasks and abilities (Stuster et al., 2019) and generalizable skills and knowledge (Stuster et al., 2018) for exploration missions identified previously.
2. Exploration missions will consist of a team of astronauts. A large gap in our knowledge involves how elevated CO_2 may affect agent teaming performance, particularly for teams in isolation and confinement.
3. Assessment of CO_2 effects on trust-in or reliance-on automation.
4. Assessment of crew knowledge, acquired throughout their multiple years of training, during exposure to elevated CO_2.

5. In preparation for an autonomous Mars crew, characterizations should be performed for other stressors, on metacognitive processes and on team responses, as has been done here for CO_2 effects on anomaly response.
6. There was no published research identified that investigated how elevated CO_2 affects information integration.
7. There was no published research identified that investigated elevated CO_2 effects on keystrokes.
8. There was no published research identified that investigated elevated CO_2 effects on communication.
9. There was no published research identified that investigated elevated CO_2 effects on projecting what will happen in the future.

3.5. Potential Crew Performance Monitoring Measures

Wenzel (2019) described a set of human factors standard measures applicable for research in spaceflight analogs, on ISS and long-duration exploration missions. Table 4 shows critical cognitive and physiological effects of elevated CO_2 exposure and potentially useful diagnostic monitoring measures. These measures differ from some of the Wenzel standard measures because their utility is for anomaly resolution. Many of the performance measures are non-intrusive, whereas others are psychophysically-based tests, such as the Visual Object Learning Test (VOLT).

Table 4. Cognitive, Motor and Physiological Effects of Elevated CO_2 and Potentially Useful Crew Diagnostic Monitoring Measures

Stressor Effects (CO_2-specific)	*Example Measures of Crew Performance*
Cognitive	*Cognitive*
Alertness	Eye Fixations
Vigilance	PVT
Attention	Eye Movements
Serial Choice RT	Key-Strokes
Activity Level	Motion Detectors/Wrist Actigraphy
Visuo-motor	Key-Strokes/MPT
Strategy	SMS-like Measures
Information Seeking	Eye Movements
Object learning	VOLT
Complex maneuvers	Glove or Control Sensors/Lunar Vehicle Steering Sensors
Outlier Detection	Eye Movements
Physiological	*Physiological*
Headache	Monitor Medications
Increased Fatigue	Activity Speed
Increased Cerebral Blood Flow	Diffuse Correlation Spectroscopy
Increased Heart Rate	Wrist Actigraphy

4. Conclusions

Recognizing what is anomalous, diagnosing what is producing the anomaly, analyzing the implications of the anomaly, modifying plans and reprioritizing goals to maintain control are quite difficult tasks vulnerable to many generic forms of failure (e.g., premature narrowing). The available evidence suggests that elevated CO_2 could affect the cognitive processes of detection, diagnosis and recovery used in anomaly response. Without mature technology to aid the crew, elevated CO_2 could exacerbate the risk that crew may not be able to independently respond to these events.

References

Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modeling? and how to fix it using search-based software engineering. *Information and Software Technology*, *98*, 74-88.

Allen, J. G., MacNaughton, P., Cedeno-Laurent, J. G., Cao, X., Flanigan, S., Vallarino, J., ... & Spengler, J. D. (2019). Airplane pilot flight performance on 21 maneuvers in a flight simulator under varying carbon dioxide concentrations. *Journal of Exposure Science & Environmental Epidemiology*, *29*(4), 457-468.

Allen, J. G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., & Spengler, J. D. (2016). Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: a controlled exposure study of green and conventional office environments. *Environmental Health Perspectives*, *124*(6), 805-812.

Alpern, M., & Hendley, C. D. (1952). Visual functions as indices of physiological changes in the acid-base balance of the blood. *Optometry and Vision Science*, *29*(6), 301-314.

Anderson, M. L., & Perlis, D. R. (2005). Logic, self-awareness and self-improvement: The metacognitive loop and the problem of brittleness. *Journal of Logic and Computation, 15*(1), 21-40.

Arendt, K., Johansen, A., Jørgensen, B. N., Kjærgaard, M. B., Mattera, C. G., Sangogboye, F. C., ... & Veje, C. T. (2018, November). Room-level occupant counts, airflow and CO_2 data from an office building. In *Proceedings of the First Workshop on Data Acquisition To Analysis* (pp. 13-14).

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75-84.

Basner, M., Nasrini, J., Hermosillo, E., McGuire, S., Dinges, D. F., Moore, T. M., ... & Donoviel, D. (2018). Effects of -12° head-down tilt with and without elevated levels of CO_2 on cognitive performance: The SPACECOT study. *Journal of Applied Physiology*, *124*(3), 750-760.

Beisert, S., Rodriggs, M., Moreno, F., Korth, D., Gibson, S., Lee, Y. H., & Eagles, D. E. (2013). Development and Execution of Autonomous Procedures Onboard the International Space Station to Support the Next Phase of Human Space Exploration. Paper Presented at the 2013 AIAA Space Conference, 10–12 September, San Diego, CA.

Brown, M. L., & Kros, J. F. (2003). Data mining and the impact of missing data. *Industrial Management & Data Systems*, 103(8), 621.

Brühwiler, P. A., Stämpfli, R., Huber, R., & Camenzind, M. (2005). CO_2 and O_2 concentrations in integral motorcycle helmets. *Applied ergonomics, 36*(5), 625-633.

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science*, *7*(3), 83-87.

Cao, X., Zevitas, C. D., Spengler, J. D., Coull, B., McNeely, E., Jones, B., ... & Allen, J. G. (2018). The on-board carbon dioxide concentrations and ventilation performance in passenger cabins of US domestic flights. *Indoor and Built Environment*, *28*(6), 761-771.

Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J. G., Flanigan, S., Vallarino, J., ... & Allen, J. G. (2019). Heart rate variability and performance of commercial airline pilots during flight simulations. *International Journal of Environmental Research and Public Health*, *16*(2), 237.

Carr, C. E. (2006). Impact of moderate elevations in CO_2 on astronauts during long-duration spaceflight on the International Space Station. *Final Report.*

Christoffersen, K., & Woods, D. D. (2003). Making sense of change: The challenges of events in operations environments. Cognitive Systems Engineering Laboratory, Ohio State University, 30 pgs.

Cohen, M. S., Freeman, J. T., & Wolf, S. (1996). Metacognition in time-stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Cronyn PD, Watkins S, Alexander DJ. 2012. Chronic Exposure to Moderately Elevated CO_2 during Long-Duration Space Flight. NASA Technical Report NASA/TP-2012-217358. Available: http://ston.jsc.nasa.gov/collections/trs/_techrep/TP-2012-217358.pdf

Daiker, R., Harrivel, A., Ghatas, R., Lake, R., & Maddock, S. (2020, July). Cognitive Task Analysis and Knowledge Elicitation to Inform Medical Workstation Requirements for Long Duration Space Missions. In *International Conference on Applied Human Factors and Ergonomics* (pp. 164-171). Springer, Cham.

Davy, M., Desobry, F., Gretton, A., & Doncarli, C. (2006). An online support vector machine for abnormal events detection. *Signal Processing*, *86*(8), 2009-2025.

De Keyser, V., & Woods, D. D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A. G. Colombo & A. Saiz de Bustamante (Eds.), Systems reliability assessment (pp. 231–251). Dordrecht, Netherlands: Kluwer Academic Publishers.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65-84.

Fenton, W. G., McGinnity, T. M., & Maguire, L. P. (2001). Fault diagnosis of electronic systems using intelligent techniques: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *31*(3), 269-281.

Fischer, U., & Orasanu, J. (2000). Error-challenging strategies: Their role in preventing and correcting errors. Proceedings of the International Ergonomics Association 14th Triennial Congress and Human Factors and Ergonomics Society 44th Annual Meeting (pp. 1-30–1-33). Santa Monica, CA: Human Factors and Ergonomics Society

Frank, J.; Spirkovska, L.; McCann, R.; Wang, L.; Pohlkamp, K.; and Morin, L. (2013). Autonomous Mission Operations. In Proceedings of the 2013 IEEE Aerospace Conference. Piscataway, NJ: Institute for Electrical and Electronics Engineers.

Frank, J. D., McGuire, K., Moses, H. R., & Stephenson, J. (2016). Developing decision aids to enable human spaceflight autonomy. *AI Magazine*, *37*(4), 46-54.

Frank, J. D. (2019). Artificial Intelligence: Powering Human Exploration of the Moon and Mars. *arXiv preprint arXiv:1910.03014*.

Freiberger, J. J., Derrick, B. J., Natoli, M. J., Akushevich, I., Schinazi, E. A., Parker, C., ... & Moon, R. E. (2016). Assessment of the interaction of hyperbaric N2, CO_2, and O_2 on psychomotor performance in divers. *Journal of Applied Physiology*, *121*(4), 953-964.

Galbraith, D., & Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In *Observing Writing* (pp. 306-325). Brill.

Getty, D. J., Swets, J. A., Pickett, R. M., & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*(1), 19.

Gettys, C. F., Pliske, R. M., Manning, C., & Casey, J. T. (1987). An evaluation of human act generation performance. *Organizational Behavior and Human Decision Processes, 39,* 23–51.

Gill, M., Natoli, M. J., Vacchiano, C., MacLeod, D. B., Ikeda, K., Qin, M., ... & Vann, R. D. (2014). Effects of elevated oxygen and carbon dioxide partial pressures on respiratory function and cognitive performance. *Journal of Applied Physiology*, *117*(4), 406-412.

Greitzer, F. L. (2005). Toward the development of cognitive task difficulty metrics to support intelligence analysis research. In *Fourth IEEE Conference on Cognitive Informatics, 2005.(ICCI 2005).* (pp. 315-320). IEEE.

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors*, *58*(3), 509-519.

Hilburn, B., Jorna, P. G. A. M., & Parasuraman, R. (1995). The effect of advanced ATC automation on mental workload and monitoring performance: An empirical investigation in Dutch airspace. In Proceedings of the 8th International Symposium on Aviation Psychology (pp. 387-391). Columbus: Ohio State University.

Hilburn, B., Parasuraman, R.: & Mouloua, M. (1995). Effects of short- and long-cycle adaptive function allocation on performance of flight-related tasks. In N. Johnston, R. Fuller, & N. McDonald (Eds.), Aviation psychology: Training and selection (pp. 347-353). Hampshire, England: Ashgate

Holden, K., Munson, B., Russi-Vigoya, N., Dempsey, D., Adelstein, B. (2019a). Human Capabilities Assessment for Autonomous Missions (HCAAM) Phase II: Development and Validation of an Autonomous Operations Task List. Final Report

Hong, L., & Page, S. E. (2002). Groups of diverse problem solvers can outperform groups of high ability problem solvers. Proceedings of the National Academy of Science: Economic Sciences, 101, 16385–16389.

Horvitz, E., & Barry, M. (1995). Display of information for time-critical decision making. In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence* (pp. 296-305). Morgan Kaufmann Publishers Inc.

Hwang, S. H., Park, W. M., Park, J. B., & Nam, T. (2017). Characteristics of PM10 and CO_2 concentrations on 100 underground subway station platforms in 2014 and 2015. *Atmospheric Environment, 167*, 143-149.

Kajtár, L., & Herczeg, L. (2012). Influence of carbon-dioxide concentration on human well-being and intensity of mental work. *QJ Hung. Meteorol. Serv*, *116*, 145-169.

Kemeny, J. G., et al. 1979, Report of the President's Commission on the Accident at Three Mile Island (Pergamon Press, New York).

Kirlik, A., Fisk, A. D., Walker, N., & Rothrock, L. (1998). Feedback augmentation and part-task practice in training dynamic decision-making skills. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions under Stress: Implications for Individual and Team training* (p. 91–113).

Klein, G., Orasanu, J., Calderwood, R. & Zsambok, E.E. (Eds.). (1994). *Decision Making in Action: Models and Methods.* Ablex, Norwood, NJ.

Klein, G. (2001). Features of team coordination. In M. McNeese, M. R. Endsley, & E. Salas (Eds.), New trends in cooperative activities: Understanding system dynamics in complex environments (pp. 68–95). Santa Monica, CA: Human Factors and Ergonomics Society.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2007). A data–frame theory of sensemaking. In *Expertise out of Context* (pp. 118-160). Psychology Press.

Knox, J. (2017, July). Development of Carbon Dioxide Removal Systems for NASA's Deep Space Human Exploration Missions 2016-2017. 47th International Conference on Environmental Systems.

Knox, J. (2018, July). Development of Carbon Dioxide Removal Systems for NASA's Deep Space Human Exploration Missions 2017-2018. 48th International Conference on Environmental Systems.

Krihak, M. (2016). Exploration medical capability: Medical data architecture technology development plan. NASA Document 7082.

Landon, L. B., & O'Keefe, W. S. (2018). Team Training Is a Go: Team Training for Future Spaceflight', Building Intelligent Tutoring Systems for Teams (Research on Managing Groups and Teams, Volume 19).

Lave, J. (1988). Cognition in practice: Mind, mathematics and culture in everyday life. Cambridge University Press.

Law, J., Van Baalen, M., Foy, M., Mason, S. S., Mendez, C., Wear, M. L., ... & Alexander, D. (2014). Relationship between carbon dioxide levels and reported headaches on the international space station. *Journal of Occupational and Environmental Medicine*, *56*(5), 477-483.

Layman, L., Nikora, A. P., Meek, J., & Menzies, T. (2016, May). Topic modeling of NASA space system problem reports: research in practice. In *Proceedings of the 13th International Conference on Mining Software Repositories* (pp. 303-314).

Layman, L., Zelkowitz, M., Basili, V., & Nikora, A. P. (2012, November). Toward Baselining Software Anomalies in NASA Missions. In *2012 IEEE 23rd International Symposium on Software Reliability Engineering Workshops* (pp. 13-14). IEEE.

Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. Human Factors, 36, 94–119.

Lee, Y. H., Beisert, S., Moreno, F., Rodriggs, M., & Eagles, D. (2013). Development and Execution of Autonomous Procedures Onboard the International Space Station to Support the Next Phase of Human Space Exploration. In *AIAA SPACE 2013 Conference and Exposition* (p. 5473).

Lee, S. J., & Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems,* 101(1), 41-46.

Lindsey, T., Shetye, S., & Shaw, T. (2016). Exploration Clinical Decision Support System: Medical Data Architecture. Investigator's Workshop abstract.

Liu, W., Zhong, W., & Wargocki, P. (2017). Performance, acute health symptoms and physiological responses during exposure to high air temperature and carbon dioxide concentration. *Building and Environment, 114*, 96-105.

Ma, R., & Kaber, D. B. (2007). Effects of in-vehicle navigation assistance and performance on driver trust and vehicle control. *International Journal of Industrial Ergonomics, 37*(8), 665-673.

Macatangay, A., Simon, C., Bautista, J., Moses, H., Morency, R., & Misek, W. (2018, July). Personal CO_2 Monitor (PCO_2M)-In-flight Evaluation of the 2x2015 Technology Demonstration. 48th International Conference on Environmental Systems.

Manzey, D., & Lorenz, B. (1998). Effects of chronically elevated CO_2 on mental performance during 26 days of confinement. *Aviation, Space, and Environmental Medicine*, *69*(5), 506–514

Mark, G. (2002). Extreme collaboration. Communications of the ACM, 45, 89–93.

Marquez, J. J., Hillenius, S., Healy, M., & Silva-Martinez, J. (2019). Lessons Learned from International Space Station Crew Autonomous Scheduling Test. Final Report to HFBP.

Marquez, J. J., Hillenius, S., Zheng, J., Deliz, I., Kanefsky, B., & Gale, J. (2019, November). Designing for Astronaut-Centric Planning and Scheduling Aids. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 468-469). Sage CA: Los Angeles, CA: SAGE Publications.

Marquez, J. J., Miller, M. J., Cohen, T., Deliz, I., Lees, D. S., Zheng, J., ... & Hillenius, S. (2019). Future needs for science-driven geospatial and temporal extravehicular activity planning and execution. *Astrobiology, 19*(3), 440-461.

Maula, H., Hongisto, V., Naatula, V., Haapakangas, A., & Koskela, H. (2017). The effect of low ventilation rate with elevated bioeffluent concentration on work performance, perceived indoor air quality, and health symptoms. *Indoor Air*, *27*(6), 1141-1153.

Mehrotra, S., Butts, C. T., Kalashnikov, D., Venkatasubramanian, N., Rao, R. R., Chockalingam, G., ... & Huyck, C. (2003, December). Project RESCUE: challenges in responding to the unexpected. In *Internet Imaging V* (Vol. 5304, pp. 179-192). International Society for Optics and Photonics.

Meyer, C., & Schneider, W. (2018). NASA Advanced Explorations Systems: 2018 Advancements in Life Support Systems. In *2018 AIAA SPACE and Astronautics Forum and Exposition* (p. 5151).

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, *38*(2), 311-322.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*(1), 44.

Mudgett, P., Manney, J., Smith, M., Neal, S. J., & Pilgrim, J. (2018, July). US Navy Submarine Sea Trial of a NASA developed Multi-Gas Monitor. 48th International Conference on Environmental Systems.

Murray, C. and Cox, C. B. 1989, Apollo, The Race to the Moon (Simon & Schuster, New York).

Myhrvold, A. N., Olsen, E., & Lauridsen, O. (1996). Indoor environment in schools–pupils health and performance in regard to CO_2 concentrations. *Indoor Air, 96*(4), 369-371.

Patterson, R. D. 1990, Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London, B 327*, 485-492.

Patterson, E. S., Roth, E. M., Woods, D. D., Chow, R., & Gomes, J. O. (2004). Handoff strategies in settings with high consequences for failure: lessons for health care operations. *International journal for quality in health care, 16*(2), 125-132.

Patterson, E. S., Watts-Perotti, J., & Woods, D. D. (1999). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work (CSCW)*, *8*(4), 353-371.

Patterson, E. S., Woods, D. D., Cook, R. I., & Render, M. L. (2007). Collaborative cross-checking to enhance resilience. *Cognition, Technology & Work*, *9*(3), 155-162.

Parasuraman, R.. Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency." International Journal of Aviation Psychology, 3, 1-23.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1996a). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), Automation and human performance: Theory and applications (pp. 91-115). Mahwah, NJ: Erlbaum

Parasuraman, R., Mouloua, M., & Molloy, R. (1996b). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, *38*(4), 665-679.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors*, *50*(3), 511-520.

Robertson, I. H., & O'Connell, R. (2010). Vigilant attention. In *Attention and Time (Nobre, A & Coull, J., Eds)*, 79-88.

Rader, S.; Reagan, M.; Janoiko, B.; and Johnson, J. (2013). Human-in-the-Loop Operations over Time Delay: NASA Analog Missions Lessons Learned. Paper Presented at the 43rd International Conference on Environmental Systems (ICES). Vail, CO, 14–18 July, p. 3520.

Rice, S. A. (2014). Human health risk assessment of CO_2: Survivors of acute high-level exposure and populations sensitive to prolonged low-level exposure. *Environments, 3*(5), 7-15.

Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., & Fisk, W. J. (2012). Is CO_2 an indoor pollutant? Direct effects of low-to-moderate CO_2 concentrations on human decision-making performance. *Environmental Health Perspectives*, *120*(12), 1671-1677.

Sauer, J., Kao, C. S., & Wastell, D. (2012). A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics*, *55*(8), 840-853.

Savulich, G., Hezemans, F. H., van Ghesel Grothe, S., Dafflon, J., Schulten, N., Brühl, A. B., ... & Robbins, T. W. (2019). Acute anxiety and autonomic arousal induced by CO 2 inhalation impairs prefrontal executive functions in healthy humans. *Translational Psychiatry*, *9*(1), 1-10.

Scully, R. R., Basner, M., Nasrini, J., Lam, C. W., Hermosillo, E., Gur, R. C., ... & Ryder, V. E. (2019). Effects of acute exposures to carbon dioxide on decision making and cognition in astronaut-like subjects. *npj Microgravity*, *5*(1), 1-15.

Shalin, V. L. (2005). The roles of humans and computers in distributed planning for dynamic domains. Cognition, Technology, & Work, 7, 198–211.

Shattuck, L. G., & Woods, D. D. (2000). Communication of intent in military command and control systems. In *The Human in Command* (pp. 279-291). Springer, Boston, MA.

Simon, H. A. (1996). *Models of my life*. MIT press.

Singh. I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *International Journal of Aviation Psychology, 3*, 111-121.

Smith, P. J., Klopfenstein, M., Jezerinac, J., & Spencer, A. (2017). Distributed work in the National Airspace System: Providing feedback loops using the post-operations evaluation tool (POET). In *Human Factors impacts in Air Traffic Management* (pp. 147-168). Routledge.

Smith, P. J., McCoy, E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. *IEEE Transactions on Systems, Man and Cybernetics, 27,* 360–371.

Snow, S., Boyson, A. S., Paas, K. H., Gough, H., King, M. F., Barlow, J., & Noakes, C. J. (2019). Exploring the physiological, neurophysiological and cognitive performance effects of elevated carbon dioxide concentrations indoors. *Building and Environment*, *156*, 243-252.

Son, K. N., Richardson, T. M. J., & Cmarik, G. E. (2019). Equilibrium Adsorption Isotherms for H2O on Zeolite 13X. *Journal of Chemical & Engineering Data*, *64*(3), 1063-1071.

Stankovic, A., Alexander, D., Oman, C. M., & Schneiderman, J. (2016). A review of cognitive and behavioral effects of increased carbon dioxide exposure in humans. *NASA Technical Paper,* 2016-219277.

Stark, J. M., Scerbo, M. W., Freeman, F. G., & Mikulka, P. J. (2000, July). Mental fatigue and workload: Effort allocation during multiple task performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 22, pp. 863-866). Sage CA: Los Angeles, CA: SAGE Publications.

Stetson, H.; Frank, J.; Cornelius, R.; Haddock, A.; Wang, L.; and Garner, L. (2015). Amo Express: A Command and Control Experiment for Crew Autonomy. Paper presented at the Annual International Space Station Research and Development Conference, 7–9 July, Boston, MA. NASA Technical Report 20150016490, Washington DC.

Strangman, G. E., Sipes, W., & Beven, G. (2014). Human cognitive performance in spaceflight and analogue environments. *Aviation, Space, and Environmental Medicine*, *85*(10), 1033-1048.

Storm, W. F., & Giannetta, C. L. (1974). Effects of hypercapnia and bedrest on psychomotor performance. *Aviation, Space, and Environmental Medicine*, 45, 431-433.

Stuster, J. W., Adolf, J. A., Byrne, V. E., & Greene, M. (2018). Human Exploration of Mars: Preliminary Lists of Crew Tasks. NASA/CR-2018-220043

Stuster, J. W., Adolf, J. A., Byrne, V. E., & Greene, M. (2019). Generalizable Skills and Knowledge for Exploration Missions. NASA/CR-2018-220445

Subramaniam, S., Palpanas, T., Papadopoulos, D., Kalogeraki, V., & Gunopulos, D. (2006, September). Online outlier detection in sensor data using non-parametric models. In *Proceedings of the 32nd international conference on Very large data bases* (pp. 187-198). VLDB Endowment.

Tan, S. C., Ting, K. M., & Liu, T. F. (2011, June). Fast anomaly detection for streaming data. In *Twenty-Second International Joint Conference on Artificial Intelligence*. Downloaded on January 19, 2020 from https://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/viewFile/3229/3469.

Thompson, J. F., Bellerjeau, C., Marinick, G., Osio-Norgaard, J., Evans, A., Carry, P., ... & Whiting, G. L. (2019). Intrinsic Thermal Desorption in a 3D Printed Multifunctional Composite CO_2 Sorbent with Embedded Heating Capability. *ACS applied Materials & Interfaces*, *11*(46), 43337-43343.

U.S. Department of Labor, Occupational Safety and Health Administration. 2018. Available from https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.1000TABLEZ1.

Vera, A. H., Holden, K. L., Dempsey, D. L., Russi-Vigoya, M. N., Wu, S. & Beutter, B. R. (2019). Contextual Inquiries and Interviews to Support Crew Autonomous Operations in Future Deep Space Missions: Preliminary Requirements and Proposed Future Research. Final Report to NASA.

Vercruyssen, M. (2014, September). Breathing Carbon Dioxide (4% for 1-Hour) Slows Response Selection, Not Stimulus Encoding. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 914-918). Sage CA: Los Angeles, CA: SAGE Publications.

Wald, G., Harper Jr, P. V., Goodman, H. C., & Krieger, H. P. (1942). Respiratory effects upon the visual threshold. *The Journal of General Physiology*, *25*(6), 891-903.

Wargocki, P., Porras-Salazar, J. A., Contreras-Espinoza, S., & Bahnfleth, W. (2020). The relationships between classroom air quality and children's performance in school. *Building and Environment, 173*, 106749.

Watts-Perotti, J., & Woods, D. D. (2007). How anomaly response is distributed across functionally distinct teams in space shuttle mission control. *Journal of Cognitive Engineering and Decision Making*, *1*(4), 405-433.

Weiss, G. M. (2004). Mining with rarity: A unifying framework. *ACM Sigkdd Explorations Newsletter*, *6*(1), 7-19.

Weitzman, D. O., Kinney, J. S., & Luria, S. M. (1969). Effect On Vision Of Repeated Exposure To Carbon Dioxide. SMRL Report No. 566.

Wenzel, E. M. (2019). Human Factors Standard Measure Recommendations for Spaceflight, Analogs, and other Research Studies. Final Report to NASA.

Wiegmann, D. A., Rich, A., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science, 2*(4), 352-367.

Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, & R. Calderwood (Eds.), Decision making in action: Models and methods (pp. 228–251). Norwood, NJ: Ablex.

Woods, D. D. (1994). Cognitive demands and activities in dynamic fault management: Abductive reasoning and disturbance management. In N. Stanton (Ed.), Human factors in alarm design (pp. 63–92). Bristol, PA: Taylor & Francis.

Woods, D. D. (2005). Creating foresight: Lessons for resilience from Columbia. In W. H. Starbuck & M. Farjoun (Eds.), Organization at the limit: NASA and the Columbia disaster(pp. 289–308). Malden MA: Blackwell.

Woods, D. D., & Hollnagel, E. (2006). Joint cognitive systems: Patterns in cognitive systems engineering. Boca Raton, FL: Taylor & Francis.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). Behind human error: Cognitive systems, computers, and hindsight. Dayton, OH: Human Systems Integration Information and Analysis Center.

Woods, D. D., O'Brien, J., & Hanes, L. F. (1987). Human factors challenges in process control: The case of nuclear power plants. In G. Salvendy (Ed.), Handbook of human factors/ergonomics (pp. 1724–1770). New York: Wiley.

Woods, D. D., & Patterson, E. S. (2000). How unexpected events produce an escalation of cognitive and coordinative demands. In P. A. Hancock & P. Desmond (Eds.), Stress, workload and fatigue (pp. 290–302). Mahwah, NJ: Erlbaum.

Wu, S. C., & Vera, A. H. (2019). Capability Considerations for Enhancing Safety on Long Duration Crewed Missions: Insights from a Technical Interchange Meeting on Autonomous Crew Operations. Downloaded on January 15, 2020 at https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20190027727.pdf.

Xu, Y., Sun, Y., Wan, J., Liu, X., & Song, Z. (2017). Industrial big data for fault diagnosis: Taxonomy, review, and applications. *IEEE Access*, *5*, 17368-17380.

Yang Y, Sun C, Sun M . (1997). The effect of moderately increased CO_2 concentration on perception of coherent motion . *Aviat Space Environ Med, 68*, 187-91.

Zaghloul, W., Lee, S. M., & Trimi, S. (2009). Text classification: Neural networks vs support vector machines. *Industrial Management & Data Systems*, *109,* 708-717.

Zhang, X., Wargocki, P., & Lian, Z. (2014). Literature survey on the effects of pure carbon dioxide on health, comfort and performance. *Proceedings of Indoor Air*, *2014*, 13th.

Zhang X, Wargocki P, Lian Z, et al. (2017a) Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance. *Indoor Air, 27*, 47–64.

Zhang, X., Wargocki, P., & Lian, Z. (2017b). Physiological responses during exposure to carbon dioxide and bioeffluents at levels typically occurring indoors. *Indoor Air, 27*, 65-77.

Zimek, A., & Filzmoser, P. (2018). There and back again: Outlier detection between statistical reasoning and data mining algorithms. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(6), e1280.

Appendix A. Autonomous Task List

Study aim: Identify the set of candidate autonomous mission tasks, focusing on those that are most likely to pose a threat to crew health and performance during an autonomous phase of a long-duration spaceflight mission. Dark purple: time-critical tasks. Light purple: complex tasks. White: nominal tasks.		
GOAL: Maintain crew health		
	Tasks	
	1	**Respond to medical/behavioral health events**
	a	Respond to sudden cardiac arrest: Limited to BVM, chest compressions, AED, IO and epinephrine; treatment lasting < 45 mins.
	b	Use medical software along with vitals/test results to help diagnoseis condition of unconscious injured/ill crewmember.
	c	Respond to behavioral emergency: Treatment period is short (0-3 days) and well defined (acute, organic event).
	d	Respond to unexpected traumatic injury in the context of an interplanetary mission.
	2	**Maintain physical and psychological crew health and fitness**
	a	Manage sleep, evaluating disorders as they arise.
	b	Coordinate exercise device availability among crew to ensure access to maintain cardiovascular conditioning, muscle strength, and bone density.
	c	Manage diet, including nutritional intake, food growth, and meal preparation.
	d	Maintain hygiene, e.g. during dental cleaning, or during waste management.
	e	Perform social and recreational activities during rest hours.
GOAL: Maintain vehicle/habitat		
	Tasks	
	3	**Respond to unanticipated major vehicle/habitat malfunctions**
	a	Respond to hatch failure during docking.
	b	Respond to major fault of the cooling system.
	c	Respond to failure of autonomous software in the avionics system.
	d	Respond to major Electric Power Systems (EPS) failure.
	4	**Perform installation/activation/inspection of vehicle/habitat systems**
	a	Deploy supports/structures, manually, to configure dedicated area of spacecraft for medical procedure.
	b	Perform first-time or dormant activation of vehicle/habitat system.
	c	Anticipate known vehicle/habitat issues through inspection, or troubleshoot for unknown vehicle/habitat issues.
	5	**Manufacture hardware, software, and fuel**
	a	Connect flexible hose to fittings on the fuel production storage tanks and Mars Excursion Vehicle (MEV)/Mars Ascent Vehicle (MAV), manually to prepare for refueling.
	b	Activate pumps on fuel production system, manually to transfer fuel from storage tanks to Mars Excursion Vehicle (MEV)/Mars Ascent Vehicle (MAV).
	c	Manufacture replacement parts to repair oxygen generation equipment.
	6	**Modify, maintain, repair, and replace hardware, software, and procedures**
	a	Monitor (i.e., measure and estimate) and predict vehicle system performance.
	b	Enter control inputs, manually, to load software patch / reload software.
	c	Review documentation / enter control inputs, visually / manually, to diagnose software problem.
	7	**Perform nominal system commanding**
	a	Reconfigure communication system for private calls.
	b	Manually configure habitat settings (i.e., temperature, light, etc).
	c	Adjust surface EVA suit controls, manually to operate mobile communications with Mars habitat personnel.

GOAL: Perform mission-related tasks		
Tasks		
8		**Perform piloting/navigation task**
	a	Operate propulsion controls to maneuver spacecraft for either near- or far-field rendezvous or docking.
	b	Adjust attitude control thrusters, manually, to dock Mars Excursion Vehicle (MEM)/Mars Ascent Vehicle (MAV) to spacecraft in Mars Orbit.
	c	Monitor system displays, visually, to assess proximity and verify docking with spacecraft in Mars Orbit.
9		**Perform space or planetary EVAs**
	a	Conduct surface EVA on unfamiliar terrain as part of a scientific mission.
	b	Descend gully while carrying hand tools and wearing surface EVA suit to conduct geological research.
	c	Conduct space EVA to repair failed equipment.
10		**Perform planetary rover vehicle ops**
	a	Deploy and attach battery cables to surface rover, manually while wearing surface EVA suit, to prepare for recharging rover batteries.
	b	Deploy surface rover vehicle, manually/visually while wearing surface EVA suit, to prepare for use.
	c	Navigate rover to a predetermined research site.
11		**Perform robotics activities**
	a	Deploy and operate robot, remotely during cruise phase, to inspect external features of spacecraft.
	b	Operate robot, remotely on surface, to assemble system elements to prepare field camp for humans.
	c	Operate multiple robotic drones to survey geological sites of interest.
12		**Perform science activities**
	a	Conduct and record Mars observations using onboard equipment and telescopes.
	b	Conduct life science experiments involving crew members, manually using various instruments in the surface habitat, to generate data.
	c	Collect geological samples, manually using Apollo-type scoop (1m handle) and sample bags, while wearing surface EVA suit.
13		**Perform scheduling, planning, and task allocation**
	a	Monitor performance of work to ensure that opportunities and resources are allocated appropriately among crew personnel.
	b	Modify schedules in response to changes in mission priorities.
	c	Balance workload among team members.
14		**Perform in-mission training**
	a	Conduct (training) simulation using spacecraft computer to refresh piloting skills for Mars Orbit Injection.
	b	Perform just in time training for an emergent task (no trained pre-mission).
	c	Conduct emergency refresher training for various possibilities (fire, micro-meteorite impact, hull breech, outgassing, ECLS failure, etc.).
15		**Perform inventory and consumables management**
	a	Remove personal garments from storage in preparation for changing clothes.
	b	Transfer food packages from deep storage to proximal storage, manually, to prepare galley for crew use.
	c	Record and plan nutritional intake.
	d	Manage pharmaceutical supplies.
Cross-cutting tasks (Task enablers)		
A		**Coordination, leadership, and team work**
	a	Coordinate simulator availability among crew to ensure refresher training for all required skills and functions.
	b	Schedule tasks and monitor performance of work to ensure that opportunities and resources are allocated appropriately among crew personnel.
	c	Coordinate multi-teams towards a single objective.
B		**Communication**
	a	Speak with other members of the crew accurately and concisely concerning technical and task-related topics.
	b	Communicate with team members proactively to enhance situation awareness among the team.
	a	Communicate on different modalities with Earth/MCC under different time delay regimes.
C		**Problem solving/decision making**
	a	Define goals and acceptable performance.
	b	Integrate information from existing sources (i.e., hardware, software, human, and operational aspects).
	c	Trend monitoring / data analysis.
	d	Detect anomalies.
	e	Utilize lessons learned/historical data (e.g., Skylab, ISS lessons-learned documents).
	f	Determine candidate paths / solutions, workarounds / alternatives, and consequences / impacts.
	g	Validate solutions.
	h	Document resolutions, new data, and lessons learned.

Appendix B. Historical Anomaly Identification during the Apollo Project

The current practice of using a Mission Evaluation Team (MET) composed of engineering specialists to resolve anomalies started with the Mercury Project (Logsdon, 1999).[3] Each specialist had a thorough understanding of system characteristics, operations, and limitations gained from experience with particular systems from initial design through development and testing of the hardware. According to the *Apollo Experience Report: Flight Anomaly Resolution*, Lobb (1975) describes the MET's sequence of tasks to identify an anomaly. Identification requires constant awareness of total system performance including telemetry data and air-ground voice communication of crew activities. System performance may be determined by comparing flight data with performance predictions (from ground tests or from other flights). Initial indications may not always be evident in real-time data, but rather requires detailed data processing. During Apollo 7, for example, the Command Module entry battery recharging characteristics were below predicted values, but available ground test data was insufficient to explain the anomaly. It was not until postflight tests that it was discovered that plate-divider materials used in the batteries could limit the recharge capacity of the battery in microgravity.

Flight data may also limit anomaly identification. For example, during Apollo 15, a light on the entry monitor system panel was illuminated when it should not have been. The crew were provided with troubleshooting procedures to pinpoint the problem. Once the fault was isolated, a special procedure, written by the MET engineers, was used to resolve the anomaly. Post-flight tests revealed a free-floating strand of contaminating wire had created an intermittent short-circuit.

Anomalies during Apollo were also recognized by a specific component observably not functioning. An example comes from Apollo 12 when the color TV camera (on the lunar surface) lost its picture. Ground engineers observed that the astronaut had inadvertently damaged the image sensor by pointing the camera toward direct sunlight.

Finally, crewmembers have identified anomalies. During Apollo 7 the command module primary floodlights failed. However, it wasn't until postflight that an investigation revealed that the lights had been used indiscriminately in a dimmed mode before the flight and had simply burned out. If LEDs are run well below their rated current, their electrical efficiencies can be quite high.

In conclusion, Apollo Project anomalies were identified from one of four sources: insufficient ground test data, insufficient flight data, obvious component failure and crew observation. The causes were typically of three types: manufacturing quality, hardware design and operational procedures.

For crew to independently identify potential anomalies will require technology that provides constant awareness of total system performance including spacecraft, software and crew health data. Ideally, the software would monitor crew activities, compare flight data with performance predictions, know the time-criticality of the anomalous data, identify potential explanations for atypical performance, process data, prioritize the safety-criticality of the event, and provide an appropriate caution, warning or alert to the crew.

---

[3] To better understand mission control front and back room coordination for Shuttle missions see Patterson et al., (1999), Jones (1995), Mark (2002) and Shalin (2005).

Artemis plans will involve a great deal of assembly work. Lessons learned from International Space Station assembly missions may shed some light into “unexpected” anomaly resolution. To resolve an anomaly after it has been identified has required a team-understanding of the underlying issue. Mission control organizational structure and culture ensures a rapid response.

These teams have access to other specialists and facilities as required by the particular problem.

Appendix C. Lessons Learned from International Space Station (ISS) Assembly Missions

Lessons learned are published for a wide range of ISS operational domains. This discussion will focus on anomaly resolution for the Mobile Servicing System (MSS) since a robotic system will likely be used to assemble Gateway and lunar outposts. The MSS is composed of two workstations, the Canadarm2 robotic arm, the mobile transporter and the mobile base system. The mobile transporter moves the arm and its base to different locations on the ISS truss structure. The entire system is integrated into the power, data and video infrastructure of the US segment of ISS. The Canadarm2 has 7 degrees of freedom, which Nancy Currie described as "tricky to use" particularly when she was manipulating it from a rotated relative position (personal communication, 2005).

To assist the on-orbit crew and the ground team when faced with anomalies, Malfunction procedures (Mals) have been developed. Mals are designed to guide the operator through a series of troubleshooting, failure isolation, and recovery actions to determine the cause of the anomaly, safe the system, and reconfiguration into an operational state. As lessons were learned, the troubleshooting and failure isolation actions were minimized.

Each computer unit of the MSS separately performs health monitoring. The results of these tests are fed into a central computer which performs health diagnostics on its own communication and power electronics. The central computer will Safe the system if any tests fail. Cat-1 failures can result in a hazardous condition, such as uncommanded motion or payload release. Each failed test has a corresponding caution and warning (C&W) text message that is displayed to crew and ground. In 1999, there were a total of 1400 caution and warnings. The crewmember would examine the C&W messages raised by the system and run the malfunction procedure corresponding to the highest priority code seen in the messages. These C&W must be cleared for operations to continue. The system would generate a Failure Detection, Isolation and Recovery worksheet, that was used by ground to develop Mals.

It was initially assumed that Mals would be used by the crew without assistance from the ground. Once executed, operations could resume autonomously. However, about a year after the arm had been deployed, a condition occurred that produced a potential failure cause that had a similar caution and warning signature to a failure that had not been input into the on-board system. The crew was directed to the wrong failed component which pointed them to the wrong Mal. Fortunately, the ground had been simultaneously performing troubleshooting and diagnostics and was communicating with the crew during anomaly resolution. This type of poor integrated behavior can sully crew trust in the system. This example holds a valuable lesson learned about problems that can arise when transitioning to a self-reliant crew.

In conclusion, early MSS anomalies were plagued with inadequate and conflicting information supplied to crew and deficient recover actions. The causes were typically of two types: poor software/hardware integration and operational procedures.

Three EMU's were onboard ISS for an extended duration following the Columbia accident. All three units eventually experienced a cooling failure. The leading theory for the loss of cooling was that the water pump rotor within the Extravehicular Mobility Unit (EMU) cooling loop was jammed. The pump assembly is an intricate, precision device which up until this time had only been serviced on the ground. On orbit maintenance procedures were developed and spare parts were launched on a Russian vehicle. In 2003, in a first of its kind on-orbit operation, the pump of EMU s/n 3013 was disassembled, cleaned and reassembled with a new rotor. The process went smoothly and cooling was successfully restored.

Appendix D. Additional Items that may Exacerbate the HSIA Risk

D.1. Planetary Dust risks to HSIA

There is a risk that lunar and planetary dust will exacerbate the HSIA risk. In 2008, a team of NASA scientists and engineers identified systems *that will be affected* by planetary dust (Wagner, 2008). Many of their concerns are summarized below.

- Air revitalization systems including the following sub-components
  - Ventilation systems
  - Trace contaminant control
  - CO_2 removal
  - CO_2 reduction
  - O_2 generation
  - Particulate control system
- Water recovery system
  - Biological water processor
  - Water quality monitor
- Waste collectors and disposal
- Thermal
  - Radiators
  - Humidity control
- Other life support systems, such as
  - Valves that seal surfaces
  - Moving parts such as pumps
  - Filters will be plugged
- Airlock effects
  - Seal degradation causing leaks and requiring higher maintenance
  - Dust will transfer into habitat/vehicle
- Space suit assembly
  - Outer garment degradation of materials and dust transfer to airlock-habitat
  - Bearings
  - Visor coatings including scratches and severe abrasions, loss of coatings
  - Lighting reduction from dust coating illumination source
- Portable life support system power and communications
  - Charged dust particles could result in static shock to electronics
  - Dust in batteries can cause a power drain and possible short circuit
- Advance food systems contamination
  - Storage, processing equipment and food preparation equipment

In addition, the team identified that dust will affect subsystems critical to anomaly resolution such as power tools, wrenches, sockets, drills and joints.

Engineering efforts to reduce contamination include designing with materials that are smooth and dust and abrasion resistant, or that could even reject dust—when possible, to reduce pockets and folds on the EMU that could trap dust and to incorporate dust covers for sensitive equipment. Anderson et al. (2018) stated that a best practice would be to include human factors experts throughout the design process.

It is critical that continuously active and automated cleaning systems be used to reduce crew time required for particulate management and removal (*Asteroid, Lunar and Planetary Regolith Management: A Layered Engineering Defense*, NASA/TP-2014-217399, identifies the technology capabilities needed).

D.2. Re-directed Crewtime as a Risk to HSIA

Time-critical anomaly resolution requires crewmember-hours (CM-h) or crewmember-days (CM-d). Crew survival requires that they eat, even during anomaly resolution. Food production, food product preparation, meal preparation and waste disposal require a large amount of crewtime (Anderson et al., 2018) and depends upon properties of the food, such as if it is based on a crop or how the food was packaged if it was a resupply. There is a risk that these survival tasks could interfere with anomaly response.