14 Constrained Autonomy for a Better Human–Automation Interface

Ø. J. Rødseth
SINTEF Ocean

CONTENTS

SHIP AUTONOMY AND HUMAN CONTROL

The concept of autonomous or uncrewed ships is not new. Japan investigated remote control of ships in the "Highly reliable intelligent ship" project from 1982 to 1988 (Hasegawa 2004). The rocket launching platform *L/P Odyssey*, classified as a mobile offshore unit (MOU), was remotely controlled during the launch phase. Thus, it operated as a de facto uncrewed ship in international waters from 1999 to 2014 (Tass 2018). The first large-scale study on uncrewed and autonomous merchant ships was the EU project MUNIN, running from 2012 to 2015 (Rødseth & Burmeister 2012). Since then, there has been a steady increase in new investigations and concept studies. *M/S Yara Birkeland* is probably the best known and is at the time of writing planned to operate autonomously and uncrewed from 2022 (Yara 2018). A major benefit of ship autonomy is that the ship can be uncrewed, although uncrewed operation can also be achieved through remote control as for *L/P Odyssey*. Uncrewed ships save capital cost when removing the living quarters and life support systems from the ship; it can save crew cost and it allows new and innovative designs of the ship (Rødseth 2018).

The word autonomy comes from the Greek roots *autos*, "self," and *nomos*, "law," and literally means the freedom to make one's own laws. For an autonomous mobile robotic

system like an autonomous ship, there are several suggested definitions of autonomous and the subject will be discussed later in this chapter. An autonomous ship can be classified as an industrial autonomous system. This is an autonomous unit, or a collection of such, that can operate safely and efficiently in a real-world environment while doing operations of direct commercial value and which can be manufactured, maintained, deployed, operated and retrieved at an acceptable cost relative to the value it provides (Grøtli et al. 2015). When operating in general seaways together with other ships and leisure crafts, this puts a high demand on safety and reliability that is difficult to achieve with automation systems today. Furthermore, merchant ships have a high capital value, and it is expected that most autonomous and uncrewed ships will be continuously supervised from a remote control centre (RCC) to keep a close watch on the ship and the corresponding investment. However, when an RCC is in place, it also makes sense to let the RCC operators participate in the control of the ship. This avoids the need for the automation system to be able to handle all possible operational cases as the operator is available for the cases that are too complex for the automation to handle reliably.

This means that most autonomous ship systems will involve both an automation system and a human operator. Thus, the question of a how to design a high-quality human–automation interface (HAI) is a central one for autonomous ships. This chapter will discuss some possibilities for the design of the automation system for autonomous ships that may enable a better HAI to be designed. In the following, the term autonomous ships will be used for an automated ship where human operators are available but are not continuously attending to the control positions. The operators may be on the ship or in the RCC.

CONSTRAINED AUTONOMY IN THE LITERATURE

The form of constrained autonomy discussed in this text was first published as a concept in (Rødseth & Nordahl 2017). Here, it described a designed-in limitation on the possible action of an automation system in a mixed autonomy/human operator context. The objective is to create a more deterministic behaviour as seen from the human designer or operator, and by that make the allocation of tasks and responsibilities between human and automation more efficient and safer.

Other writers have used a similar terminology for other concepts such as in Al-Rifaie et al. (2012) where it applies to Gaussian constrained autonomy in swarms, where constrained refers to a limited random behaviour by swarm members. In Jha et al. (2018), the term chance-constrained temporal logic is used on a variant of temporal logic adapted to perception uncertainty. Both these uses of constrained autonomy are very different from the concept as it is described here.

The terms limited autonomy and partly autonomous have also been used frequently in the literature, but this normally refers to emergent and generally unwanted limitations in the automation system and not to a design feature.

AUTOMATION, AUTONOMY, RESPONSE TIME AND DEADLINE

Automation and autonomy has a wide range of definitions in the literature (see, e.g., Vagia et al. 2016). For the purposes of this chapter, a relatively simple definition will be used. Here, automation can be defined as "pertaining to a process or device that,

under specified conditions, can function without human intervention". Furthermore, this can be used to define autonomy as "in the context of ships, autonomy e.g. as in 'Autonomous Ship', means that the ship uses automation to operate without human intervention, related to one or more ship processes, for the full duration or in limited periods of the ship's operations or voyage." These are definitions that have been proposed to the International Maritime Organization (IMO) by ISO (2020a). These definitions simply say that automation can be used to provide autonomy and that autonomy emerges when the system is *designed, approved* and *deployed* to be operated without human intervention or supervision for certain periods. This could also be used to describe levels of autonomy by how long the operator can stay away as indicated below.

For a relatively simple automation system that is able to detect the danger of a collision, but not to make reliable corrective actions, this could be illustrated along a time axis as in Figure 14.1. Here, the danger of collision is measured in Closest Point of Approach (CPA – typically measured in nautical miles, nm) and Time to CPA (TCPA).

When a danger of collision is detected, the automation system needs to alert the crew to this so that they can take evasive actions. The crew have been organized to arrive and be ready at the control position, at the latest at T_{MR}, which is defined as the crew's *maximum response time*. The deadline for the crew's response is given by the actual situation and is defined as the *response deadline*, T_{DL}. A minimum requirement for safe operation is that T_{DL} is longer than T_{MR}. Some examples of different crew organizations are given below, where response times are only indicative and given as examples:

- *Operator in control:* The operator is directly in control of the ship. Hand-over time is not relevant ($T_{MR} = 0$).
- *Operator supervision:* Automation is used to assist operator, and operator is overseeing the operation and needs only a short time to gain situational awareness when actions are needed ($T_{MR} \approx$ 10s of seconds).
- *Operator at site*: An operator is at the control position but is working with other tasks and will need time to gain situational awareness. This could be on the order of a minute or so ($T_{MR} \approx$ minutes).
- *Operator available:* The operator is available, but is in another location, possibly sleeping, and will need several tens of minutes to reach the control position and to regain safe control ($T_{MR} \approx$ 10s of minutes).
- *No operator:* There is no operator and automation must be able to handle all operations by itself (T_{MR} is the duration of the operation or the voyage).

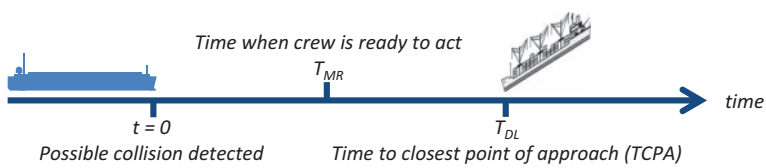

FIGURE 14.1 Simple autonomy with crew support.

Note that this form of characterization of response time is independent of the crew residing on board or in an RCC, although the times will likely be different in the two situations due to differences in equipment and access to detailed sensor or situation information.

Today, most ships have an autopilot or a track pilot. Open sea with no ship or other objects in the vicinity allows the officer of the watch to be away from the bridge for relatively long periods. With the above definition of automation and autonomy, this makes a ship controlled by an autopilot autonomous with respect to the process of keeping a steady speed and course in open sea. Autonomy in this context may seem counter-intuitive but is related to the low abstraction level on the involved functions. In the work presented here, four levels of functional abstraction are used:

1. **System objectives**: This is the highest abstraction level and is associated with generating the objectives for the design of the autonomous ship system. This may in some cases be static or at least have a long horizon, e.g. transport available cargo between ports A and B. This will be the basis for the design of the control system.
2. **Planning:** This is also a high abstraction level and will normally be an external input to the ship control system. It is related to the overall planning of the voyage or mission within the constraints of the system objectives. In most cases, this is expected to be supplied by the ship operators.
3. **Goal based**: These can be seen as a sequence of process goals for the autonomous ship system. Each goal is expected to be associated with one or more processes or tasks. This is also the most likely abstraction level for commands to the autonomous ship system.
4. **Functional**: This is specific instructions to a function, such as an autopilot. This is normally on a low abstraction level and will not normally be used as commands to the autonomous ship system.

As exemplified above, autonomy on the functional level already has been developed and is used in well-controlled environments such as autopilots on high sea or car cruise controls on highways. The goal of the work presented here is to contribute methods to extend autonomy to higher abstraction levels while giving human operators a better understanding of the capabilities and limitations of the automation system.

SENSEMAKING AND TRUST IN AUTOMATION

One basic issue in the HAI is its ability to support a proper level of operator's trust in the automation system (Lee & See 2004). This should not be too low, leading to disuse of the automated functions and neither should it be too high, leading to over-reliance and misuse of the automation. In addition, the operator must be able to make sense of the relationship between automation, his or her responsibilities and the situation at hand. The latter could be called "sensemaking", which can be defined as "a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively" (Klein et al. 2006).

This chapter will not try to link the concept of constrained autonomy to human–machine interface (HMI) research as that is clearly outside the author's expertise. However, the proposal put forward here is that both trust and sensemaking to some degree may be linked to the relationship between T_{MR} and T_{DL}. If the system consistently is able to determine T_{DL} and alert the operator before T_{MR} elapses, one can argue that the operator should get more consistent trust in the automation system's ability, both to control the process under normal conditions and to warn the operator when something requires the operator's attention. It will be left to experts in the HMI field to validate this proposal and investigate what consequences these ideas have for the operator and for the design of the HMI. This issue is complex and it is difficult, if not impossible to draw any clear conclusions on what is the best strategy for building a suitable level of operator trust (Hoff & Bashir 2015). However, some issues that seem to give positive effects are determinism in automation responses, minimizing false alerts and making it as clear as possible what the automation system is able to do, what it actually does and where the operator's intervention is required. Again, it can be argued that a higher emphasis on the deadlines and response times may be important to achieve these objectives.

The remaining part of the chapter will concentrate on the technical aspects of constrained autonomy and how it can be implemented.

THE OPERATIONAL ENVELOPE

The operational envelope (OE) can be defined as "The specific conditions under which a given autonomous ship system is designed to function, including, but not limited to, its environmental conditions and the different mission or voyage phases, as well as all anticipated failures." The definition of OE is based on the concept of the "Operational Design Domain" that was defined in SAE J3016 (2016) and developed further for use on autonomous ships in Rødseth (2018). The name has later been changed to operational envelope (OE) during the work on a standard terminology for autonomous ships (ISO 2020b).

The OE will be directly linked to the Ship Control Tasks (SCT) which will specify the details of the different tasks or processes to be performed. The OE and SCT will also specify the division of responsibilities between humans and automation. The OE and SCT can be defined on basis of a "concept of operations" document, or CONOPS. Figure 14.2 is a simplified object diagram that illustrates objects and relationships related to the OE and SCT.

The two large boxes at the bottom left are the constraints on the OE given by operational limitations in the ship system as well as the properties of the environment. Additional constraints will be added by the concept of operation, e.g. the ship cannot operate at night or during wintertime (phases and functions). The same factors that define the constraints will also play a role in determining the dynamic conditions for SCT.

The darker boxes represent the OE itself as well as the additional fall-back space which contains minimum risk conditions (MRCs). MRCs will be activated when the limits of the OE are exceeded. The OE will normally be divided into subdivisions, usually based on the mission phases and the relevant ship processes. As an example,

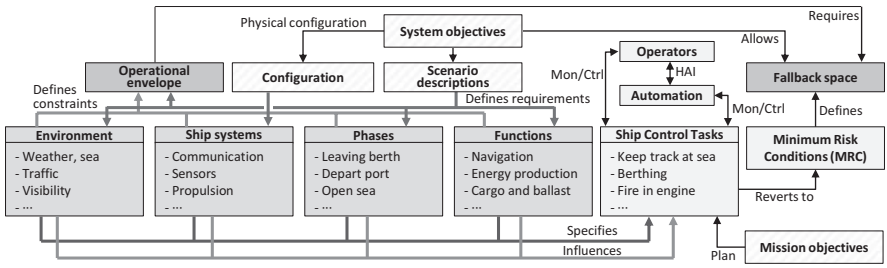

FIGURE 14.2 A simplified ER-diagram showing relationships to operational envelope.

there will normally be different subdivisions for navigation in open sea and navigation in port areas. In addition, non-controllable constraints, such as weather or anticipated technical problems may also require further subdivisions in the OE. Different MRCs can be associated to each of these subdivisions.

The lighter parts of the diagram to the right represent the functional part of the ship system. This will be realized through the SCT. SCT may be executed by operators or automation, and if both are involved, there also needs to be a Human–Automation Interface (HAI) between them. The crew and automation will execute the SCT based on the mission or voyage plan, taking ambient conditions into consideration. For automated operations, it may also be necessary for the operator to specify dynamic constraints for SCT e.g. do not exceed 12 knots or do not allow a cross-track deviation from planned route by more than one nautical mile.

THE OPERATIONAL ENVELOPE AS A STATE SPACE

The OE and the SCT exist in a multidimensional state space that will be called $\boldsymbol{S}$ in the following. Any condition that the ship can end up in is a state vector $\boldsymbol{c}$ in $\boldsymbol{S}$. As has been indicated and as will be discussed later, OE will normally be discretized into a finite number of smaller subdivisions or sub-spaces as illustrated in Figure 14.3. In the following, OE will be denoted as $\boldsymbol{O}$ and a sub-space in $\boldsymbol{O}$ will be denoted as $\boldsymbol{O_n}$. It is important that $\boldsymbol{O}$ covers exactly the same state space as $\boldsymbol{S}$, i.e. it can be said to be congruent with $\boldsymbol{S}$. This is necessary to ensure that any condition $\boldsymbol{c}$ the ship can be in can be mapped to an appropriate number of OE states, so that any individual state variables in $\boldsymbol{c}$ map to one $\boldsymbol{O_n}$. These relationships are presented formally in Eq. (14.1).

$$\begin{aligned} &\boldsymbol{O} \cong \boldsymbol{S} \\ &\boldsymbol{c} = [c_1, c_2, \ldots, c_n]^T \\ &\forall \boldsymbol{c} \in \boldsymbol{S}\ ,\ \forall ci \in \boldsymbol{c}\ , \exists \boldsymbol{O}_n \subset \boldsymbol{O} : c_i \in On \end{aligned} \tag{14.1}$$

The active part of $\boldsymbol{O}$ will normally vary over time, as not all states are relevant in all mission phases or for all ship processes. Thus, $\boldsymbol{O}$ may have to be subdivided into separate components to reflect voyage phases (L: leaving berth, D: depart port; C: coastal, etc.) and different processes (V: voyage planning; S: sailing; O:

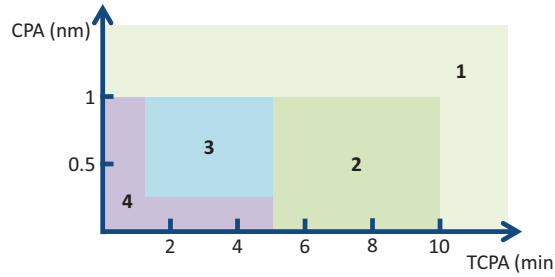

FIGURE 14.3 A discrete number of states over continuous state variables.

observation; F: Fire, etc.). This is shown in Eq. (14.2) where the different subdivisions of $\boldsymbol{O}$ have been given process as subscript and phase as superscript. The actual number of subdivisions will depend on the case at hand. Other principles for subdivisions can also be used. Each of the component spaces may have different number of dimensions.

$$\begin{aligned} \boldsymbol{O} &= \boldsymbol{O}^L \cup \boldsymbol{O}^C \cup \cdots \cup \boldsymbol{O}^X \\ \boldsymbol{O}^X &= \boldsymbol{O}^x{}_V \cup \boldsymbol{O}^x{}_S \cup \boldsymbol{O}^x{}_O \cup \boldsymbol{O}^x{}_F \cup \cdots \cup \boldsymbol{O}^x{}_Y \end{aligned} \tag{14.2}$$

The dimensions of each $\boldsymbol{O}_n$ subdivision are defined by a number of continuous state variables such as CPA and TCPA (see Figure 14.1). $\boldsymbol{O}_n$ can be seen as a state space consisting of a number of possibly multi-dimensional states $\boldsymbol{s}$, where each $\boldsymbol{s}$ is defined over a range of one or more state variables. This is illustrated in Figure 14.3, where four states are suggested for various combinations of the two state variables TCPA and CPA. These states are also the same as the states and corresponding variable ranges shown in Figure 14.4.

For a state $\boldsymbol{s}$, it may be possible to determine T_{DL} for a given environmental and ship condition $\boldsymbol{c}$. As $\boldsymbol{c}$ can vary while $\boldsymbol{s}$ is active, the value of T_{DL} will generally also vary inside $\boldsymbol{s}$.

It may not always be possible to define T_{DL}, e.g. when various forms of artificial intelligence (AI) technologies are used, where one cannot say a priori that the task always will find a useable solution to a problem or in what time frame the solution will be found.

Figure 14.4 shows an example of a simplified state transition diagram for part of the sailing process, corresponding to the states in Figure 14.3 and the illustration in Figure 14.1. In this example, the state space vector consists of the variables TCPA and CPA. The figure also shows the value ranges of T_{DL} for each state.

States 1 and 2 allow autonomous operation if the crew's maximum response time T_{MR} is 10 minutes. States 3 and 4 will require operator assistance before TCPA goes below 1 minute, otherwise a fall-back state F1 will be activated, e.g. ordering the ship to stay still in the water. To allow the crew time to reach the bridge, state 3 must be defined so that it will have a T_{DL} of 10 minutes at the time state 3 is entered. In this case, one would have to alert the crew at the latest in the transition between states 2 and 3. State 4 is a state where the automation leaves the control responsibility to the

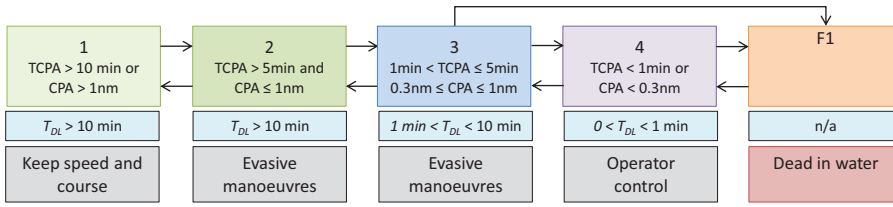

FIGURE 14.4 Examples of operational envelope (OE) states and ship control tasks.

human operator. Also, here the fall-back can be activated, e.g. if the operator does not respond.

The figure also shows how to establish a one to one relationship between the OE states and the SCT. The task associated with the fall-back state F1 is called a MRC as F1 is not part of the OE. The tasks and MRC are illustrated by the boxes at the bottom, below the state boxes. Note that the tasks corresponding to states 2 and 3 is identical in description but are still considered two different tasks as operational parameters are different.

One can also create a corresponding set of states for daytime sailing when crew is active and T_{MR} normally is shorter, e.g. 1 minute. One could keep the same pattern but reduce CPA and TCPA correspondingly. To keep the amount of code and specifications lower, this could be implemented as parameterized states and SCTs. This could be done by comparing TCPA to T_{MR} and distances to a relationship between speed and T_{MR}.

The above discussion shows that it is possible to split $\boldsymbol{O}$ into different partitions, where each partition can define different relationship between automatic and crew control. Note that this partitioning is not the same as the partitioning defined in Eq. (14.2). The different relationships between automation and crew are illustrated in Figure 14.5, with the following $\boldsymbol{O}$ components:

1. $\boldsymbol{O_{FA}}$ – Fully Autonomous: States that the automation system is designed to handle alone, without human interaction. T_{DL} is not relevant as long as the system remains in these states.
2. $\boldsymbol{O_{AC}}$ – Autonomous Control: States where the automation system can sustain automatic operation for at least a known T_{DL} without human support.
3. $\boldsymbol{O_{OA}}$ – Operator and Automation: States where automation can handle some situations, but where T_{DL} cannot be defined. A human needs to supervise automation and be ready to take over control.
4. $\boldsymbol{O_{OE}}$ – Operator Exclusive: States where direct operator control is required. The automation may assist the operator but is not generally able to control the ship in a safe manner.
5. $\boldsymbol{F}$: Fall-back states containing the MRCs that are used in cases where events take the system out of $\boldsymbol{O}$. This may happen due to unanticipated failures, environmental conditions outside $\boldsymbol{O}$ or failure of an operator to respond when the system requires human intervention.

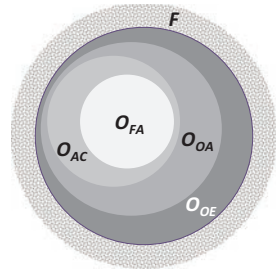

FIGURE 14.5 Operational envelope classification.

The partitions are drawn enclosed in each other to illustrate that restrictions on human presence gradually becomes more stringent as one enters the next level outwards. Fall-back states $\boldsymbol{F}$ are drawn outside $\boldsymbol{O}$ as they are not considered part of $\boldsymbol{O}$.

THE PLAN STATE SPACE AND THE SYSTEM RESPONSE DEADLINE

$\boldsymbol{O}$ is a static description of the capabilities of the autonomous ship system. The mission objectives or the plan (see Figure 14.2) together with environmental conditions will specify what parts of this overall envelope will be used in a particular mission. $\boldsymbol{O}$ has previously been specified as a state space and a corresponding plan state space ($\boldsymbol{P}$) can be defined. This will be a subset of $\boldsymbol{S}$. The corresponding subset of $\boldsymbol{O}$, here denoted $\hat{\boldsymbol{O}}$ with subdivisions $\hat{\boldsymbol{O}}_{\boldsymbol{n}}$, may restrict the states and SCT that are used during the execution of the plan. The plan can be looked at as a discrete function of time $\boldsymbol{p}(t)$ that returns a state vector from $\boldsymbol{P}$ corresponding to the condition the ship should have at the specified time step. These relationships are shown in Eq. (3).

$$\begin{aligned} \boldsymbol{P} &\subseteq \boldsymbol{S} \\ \hat{\boldsymbol{O}} &\cong \boldsymbol{P} \\ \boldsymbol{p}(t) &\in \boldsymbol{P} \end{aligned} \tag{14.3}$$

During the voyage or mission, $\boldsymbol{P}$ will restrict $\boldsymbol{O}$ and by that the "freedom" of the control system and crew, and this could also influence on the T_{DL} calculated from $\hat{\boldsymbol{O}}$. One example is that additional geographic limitations such as a restricted fairway limits the ship's ability to avoid obstacles, and this would obviously have an effect on T_{DL} in state 2 of the system described in the previous section.

If necessary, the automation system needs to cater for this by recalculating T_{DL} for any new constraints that can apply. However, it should in many cases be possible to define subdivisions of $\boldsymbol{O}$ that has states which can be made independent of $\boldsymbol{P}$ restrictions. Typically, this would be used to factor out, e.g. geography-independent subdivisions, as geography is likely to be part of $\boldsymbol{P}$ restrictions. Then, the updated T_{DL} could be linked to the remaining geography related states.

At any point in time, an autonomous ship will have a number of active SCT with different degree of automation and different timing requirements. Examples are SCT related to navigation, energy production, and stability. This means that the operator's response deadline for the full system will be the minimum of the individual T_{DL} values for all active SCT and states. If a function *FDL* is defined to determine T_{DL} for a given state and condition, the system-wide T_{DL} can be defined as in Eq. (14.4).

$$\begin{aligned} &\forall\, c_j \in \boldsymbol{c}, \exists \hat{\boldsymbol{O}}_{\boldsymbol{n}} \subset \hat{\boldsymbol{O}} : c_j \in \hat{\boldsymbol{O}}_{\boldsymbol{n}} \\ &T_{DL} = \min_{n} \left(FDL\left(\hat{\boldsymbol{O}}_{\boldsymbol{n}}, \boldsymbol{c}\right)\right) \end{aligned} \tag{14.4}$$

CONSTRAINED AUTONOMY

The concept of constrained autonomy for ships was proposed by Rødseth and Nordahl (2017). Here, it was defined as the automation system having defined limits to the options it can use to address these conditions, e.g. maximum deviation from planned track or arrival time. The automation system needs to request assistance from human operators if these constraints are exceeded. This definition was linked to the concept of operational envelope in (Rødseth 2019). The latter paper also defines five degrees of automation:

- *DA0 – Operator controlled*: Limited automation and decision support is available, as on most of today's merchant ship. The human is always in charge of operations and need to be present at controls and aware of the situation at all times.
- *DA1 – Automatic*: More advanced automation, e.g. dynamic positioning, automatic crossing or auto-berthing, is used. Crew attention is required to handle problems such as object classification and collision avoidance. The human may use own judgement as to how long he or she may be away from the control position. For automated fjord crossing in good weather, little traffic and in sheltered water, the operator may be away from the controls for several minutes.
- *DA2 – High automation*: The degree of automation is higher than for DA1 and may include certain "cognitive" functions, such as object detection and classification or collision avoidance. However, there are inherent and unknown limits to the automation system's capabilities.. These limits are not defined or constrained (see DA3), so the human operator must still use his or her judgement as to the required attention level. However, it is assumed that the need for attention is lower than for DA1.
- *DA3 – Constrained autonomy*: The degree of automation is similar to DA2, but system capabilities are now constrained by programmed or otherwise defined limits. The limits are set to enable the system to detect when limits are exceeded and to alert the operator in time before operator intervention is required.

- *DA4 – Full autonomy*: The ship automation can handle all states in $\boldsymbol{O}$ without any intervention from a crew.

These levels are related to the concepts of T_{DL} and T_{MR} and one can infer a link to the abstraction level of the crew–automation interface. This is shown in Table 14.1 where the automation levels are listed together with other relevant parameters.

By using the definitions of the operational domain's division into spaces for human and automatic control, it is possible to define constrained autonomy by the following more formal requirements:

1. All system states $\boldsymbol{c}$ that can be reached when the system is constrained autonomous, i.e. in $\boldsymbol{O}_{AC}$, must have a known response deadline.
2. For all states $\boldsymbol{c}$ in $\boldsymbol{O}_{AC}$, the function to determine T_{DL} *(FDL)*, when called after an interval Δt must always return a T_{DL} that is not shortened by more than Δt.
3. For all states $\boldsymbol{c}$ in $\boldsymbol{O}_{AC}$, the operators must be alerted to take command, i.e. requested to intervene (RTI), when T_{DL} is reduced to T_{MRS}.

The above requirements can be formulated as in Eq. (14.5), where $\boldsymbol{c}(t)$ means the system state at time t and *RTI()* means activation of an RTI. This specifies the necessary requirements to constrained autonomy and the corresponding subdivision of the operational envelope $\boldsymbol{O}_{AC}$.

$$\forall \boldsymbol{O}_n \subset \boldsymbol{O}_{AC}, \forall\, c \in \boldsymbol{O}_n : \begin{pmatrix} FDL(\boldsymbol{O}_n, \boldsymbol{c}) > 0 \\ FDL(\boldsymbol{O}_n, \boldsymbol{c}(t)) - FDL(\boldsymbol{O}_n, \boldsymbol{c}(t+\Delta t)) < \Delta t \\ FDL(\boldsymbol{O}_n, \boldsymbol{c}) \le T_{MRS} \Rightarrow RTI() \end{pmatrix} \quad (14.5)$$

The original definition of constrained autonomy (Rødseth & Nordahl 2017) only required constraints on the decision capabilities and the output of the automatic control functions. This was later supplemented by specifications of response deadlines and maximum response times (Rødseth 2019). This text has updated the definitions of the deadline and the response times as well as of the different control spaces of $\boldsymbol{O}$.

TABLE 14.1
Five Degrees of Automation

Degree of Automation	T_{DL}	Abstraction Level	Crew Attendance
DA0: Operator controlled	0	Functional	Continuous
DA1: Automatic	?	Functional	By judgement
DA2: High automation	?	Goal	By judgement
DA3: Constrained autonomy	$>T_{MR}$	Goal	Periodically unattended
DA4: Full autonomy	∞	Goal	Uncrewed

This has made it possible to have a more formal definition of constrained autonomy as provided above.

SUMMARY AND CONCLUSIONS

The concept of constrained autonomy as it has been presented here has been developed over several years and is probably not fully completed yet. However, the concepts and definitions presented in this text are now being used in several development projects as it is expected that the basic ideas are reasonably stable. There may still be some changes in the details of the definitions and in how it will be implemented in actual automation systems. This will be reported on in future publications.

The initial proposal of this text was that a more deterministic automation system in industrial autonomous systems may increase the operator's trust in the automation and by that improve the efficacy of the HAI for periodically unattended autonomous operations. A central element is to have a verifiable response deadline that can be matched to the crew's maximum response time as determined by the organization of watch-keeping on the ship and in the RCC. The theory is that this is likely to help in alerting the crew in time to establish a sufficient situational awareness before the crew is forced to act on situations that the automation system is not able to handle itself. Thus, the use of constrained autonomy should be a useful way to let the operators make better sense of the interaction between constrained autonomous systems and the operators. However, the analysis of the human–automation effects is beyond the scope of this text and will have to be addressed by experts in the human factors field. While the initial proposal seems logical, it will be up to researcher in the area of human factors to see what actual implications constrained autonomy has on the operators' ability to make better sense of the interaction between human and automation

The main purpose of this text is therefore to describe the technical concept of constrained autonomy and give it a more formal definition. Some examples of consequences for implementations of automation systems have also been given.

Independent of the human-factor angle, it is also believed that the concept can be used to improve testability and eventually also formal acceptance of autonomous control systems (Rødseth 2019). The concept of constrained autonomy may be particularly important for industrial autonomous systems, where the systems are costly, need to operate in a commercial business model and where the consequences of system failures may have significant and even catastrophic consequences. Industrial autonomous systems are also very relevant for the concept of constrained autonomy as many of them will need an operator in the loop in any case, mainly to oversee the operation and to safeguard large investments. The examples in this text are from the maritime domain and the work presented has been focusing on autonomous ships. However, the concept of constrained autonomy should also be applicable to other industrial autonomous systems.

The work presented in this text has been partially funded by the Norwegian Research Council project SAREPTA. It has also received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815012 (AUTOSHIP).

REFERENCES

Al-Rifaie, M. M., Bishop, J. M. & Caines, S. (2012). Creativity and autonomy in swarm intelligence systems. *Cognitive Computation*, *4*(3), 320–331.

Grøtli, E., Reinen, T., Grythe, K., Transeth, A., Vagia, M., Bjerkeng, M., Rundtop, P., Svendsen, E., Rødseth, Ø. & Eidnes, G. (2015). Seatonomy design development and validation of marine autonomous systems and operations. In *Proc. MTS/IEEE OCEANS'15*. Washington, DC.

Hasegawa, K. (2004). Some recent developments of next generation's marine traffic systems. *IFAC Proceedings of Computer Applications in Marine Systems (CAMS'04)*, Ancona, Italy, 2004.

Hoff, K. A. & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

ISO (2020a). *Input Document to Maritime Safety Committee, Session 102, Agenda Item 5: Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS), Proposed terminology for MASS*, Submitted by International Organization for Standardization (ISO), February 2020.

ISO 23860 (2020b). *Ships and marine technology — Terminology related to automation of Maritime Autonomous Surface Ships (MASS)* – Internal Committee Working Draft 2.

Jha, S., Raman, V., Sadigh, D. & Seshia, S. A. (2018). Safe autonomy under perception uncertainty using chance-constrained temporal logic. *Journal of Automated Reasoning*, *60*(1), 43–62.

Klein, G., Moon, B. & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70–73.

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

Rødseth, Ø. J. & Burmeister, H. C. (2012). Developments toward the unmanned ship. In *Proceedings of International Symposium Information on Ships–ISIS*. Vol. 201.

Rødseth, Ø. J. & Nordahl, H. (2017). Ed. *Definition for autonomous merchant ships*. Version 1.0, October 10, 2017. Norwegian Forum for Autonomous Ships. Retrieved February 2020 from http://nfas.autonomous-ship.org/resources-en.html

Rødseth, Ø. J. (2018). Assessing business cases for autonomous and unmanned ships. In *Technology and Science for the Ships of the Future. Proceedings of NAV 2018: 19th International Conference on Ship & Maritime Research*. IOS Press.

Rødseth, Ø. J. (2019). Defining ship autonomy by characteristic factors. In *Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships*. SINTEF Academic Press.

SAE J3016 (2016). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.* Revision September 2016, SAE International.

Tass (2018). *S7 Space to modernize Sea Launch floating spaceport for reusable rocket*. Retrieved February 2020 from https://tass.com/science/1029619.

Vagia, M., Transeth, A. A. & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53, 190–202.

Yara, A. S. (2018). *The first ever zero emission, autonomous ship*. Retrieved February 2020 from https://www.yara.com/knowledge-grows/game-changer-for-the-environment/.