Design-Strategy Planning to Handle Exogenous Uncertainty[1]

Author: Stud. Tech. Morten A. Strøm

Supervisor: Professor Bjørn Egil Asbjørnslett

NTNU – Trondheim
Norwegian University of Science and Technology

(I) Background

Even though exogenous[2] uncertainty imposes considerably commercial, operational and technical vulnerabilities (risk) to ocean engineering system (ES), it might as well lead to unforeseen opportunities. Unfortunately, it seems like decision-makers in the maritime industry primarily focuses on preventing the negative aspects of uncertainty. This thesis seeks to develop methods that support life cycle management of ocean engineering systems with the means to both mitigate the vulnerabilities and exploit the opportunities.

(II) Method

Based on a comprehensive review of literature related to systems engineering, product development, strategy, management, finance, operations research and reinforcement learning/artificial intelligence, we developed the Value-Aptitude-Design-Strategy (VADS) framework[3] and Design-Strategy Planning (DSP) to support life-cycle management of ES. Further, we present Markov decision processes (MDP) as a quantitative method for supporting DSP, and Q-learning (i.e. Approximate Dynamic Programming) for solving MDPs. The Q-learning algorithm is coded in MATLAB. To illustrate its use, VADS, DSP and MDP are applied on a offshore case extending from Rehn et al. (2017).

(III) Design-Strategy Planning (DSP)

Design-Strategy Planning (DSP) is a structured approach for managing uncertainty. Building on the VADS framework, DSP focuses on identifying, implementing and monitoring *Strategic systems*[4] with the means to handle exogenous uncertainty. We recognize *real options*[4] as such means. Thus, stakeholders should strive to identify, evaluate and incorporate real options as a part of their *contingency plan* to face the exogenous uncertainty. In the commercial level, such real options can be to expand the fleet in the case of an expected market upturn. In the operational level, such real options can be to lay-up the vessel when the markets are low. In the functional levels, such real options can be to incorporate changeability enabling the ES to change its form to alterations in operating contexts and stakeholders' needs. Figure 1 presents the iterative four-step procedure of DSP.

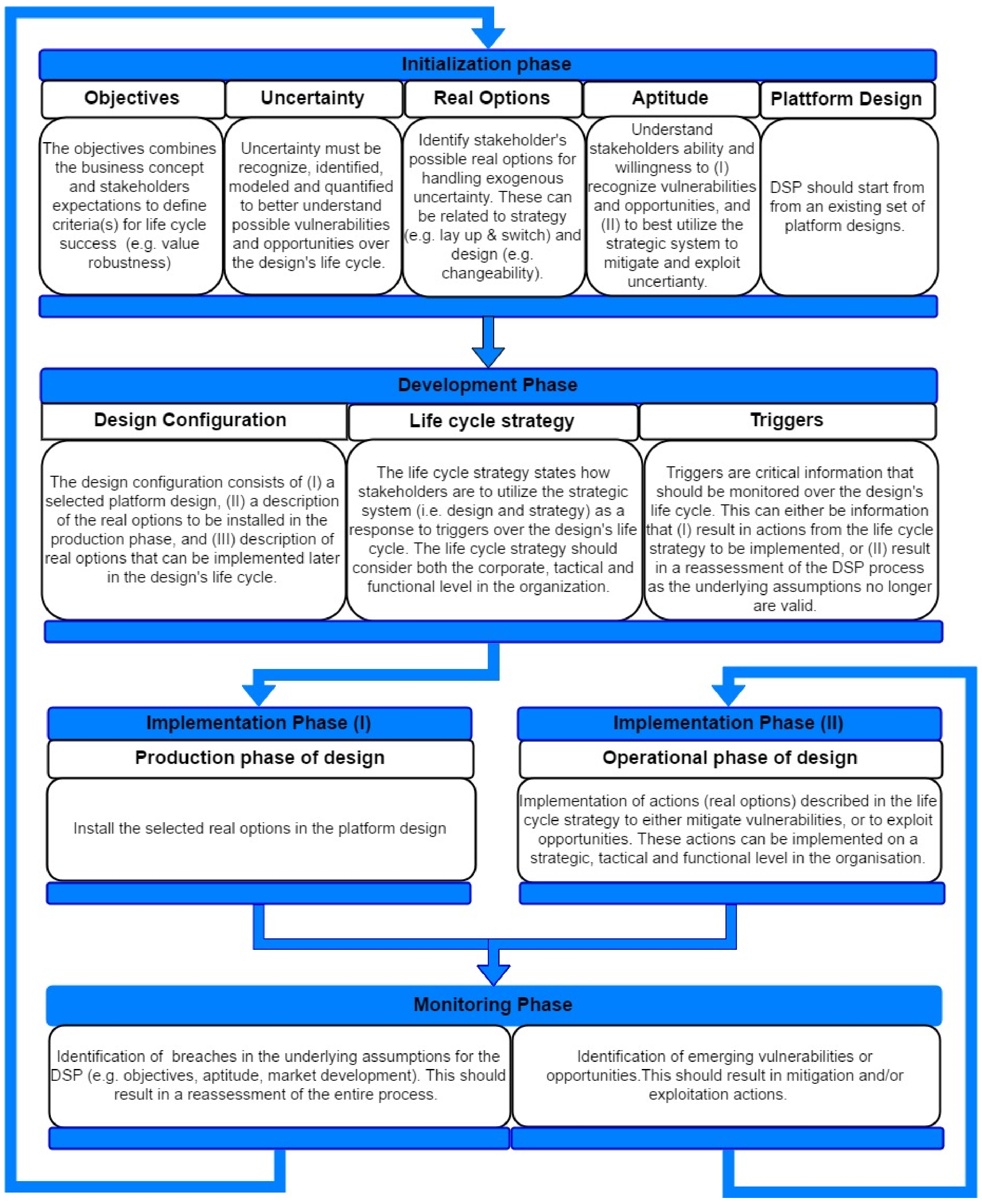

Figure 1: Framework for Design-Strategy Planning

(IV) Markov Decision Processes (MDP)

Unfamiliar to (most) students and professors at NTNU, this thesis seeks to motivate further studies on Markov decision processes and methods for solve it. Generally, MDP (Bellman, 1954) is a state-based method for modelling sequential decision making problems (SDMPs) under uncertainty. We present MDP as a quantitative tool for supporting the development phase in DSP (ref. fig. 1), as it can identify *strategic systems* and *triggers*.

Figure 2 presents the symbolic representation of DSP, supported by the notation described in table 1. For each point in time, t ∈ T, an decision-maker (agent) finds himself in a decision epoch, $s \in S_t$, where he, based on the state of the system, chooses an decision, x_t, from a set of available decisions, X_t. When making the decision, it is assumed that the current system is fully known to the decision-maker. The consequence of the decision is two folded: first, the decision-maker transits on to a new state, S_{t+1}, in the next time step. Which state the process enters is determined by the transition function, $S^M(S_t, x_t, W_{t+1})$, which depends on the current state, the decision made and the exogenous information, W_{t+1}, revealed first after the decision is made. Thus, the decision-maker is not in full control of the transition. Secondly, the decision-maker receives an contribution, $C_{t+1}(S_t, X_t, S_{t+1})$, which can be both positive, negative or zero. Afterwards, the procedure is repeated. The action made by the decision maker is bases on a *decision rule*. A sequence of decision rules is called a *decision policy*, π. The goal of SDMPs is to find the optimal policy which maximizes (or minimizes) the contribution of the system over its lifetime. For the optimal policy, the benefit of the decisions might not be immediate clear, but it is the one that ensures the highest expected contribution over the system's entire lifetime.

This quite simple framework is applicable on a broad range of various decision problems. In this thesis, it is used in the illustrative case to model the lifecycle of an offshore construction vessel and the decisions its stakeholders makes to influences it (on both the commercial, tactical and operational level). By analysing the optimal policy the strategic system and triggers in DSP can be found (ref. fig. 1).

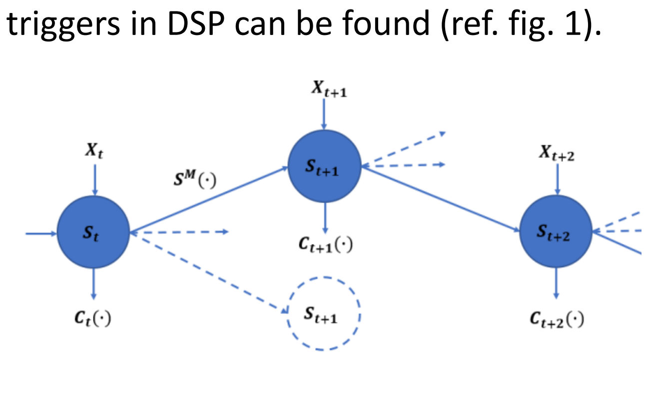

Figure 2: Sequential decision making and Markov Decision Processes

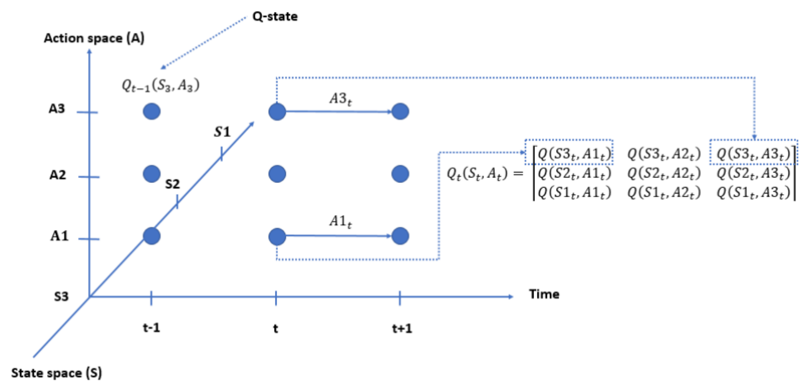

Figure 3: Illustration of Q-states (Post-decision states)

(V) Q-learning (Approximate Dynamic Programming)

A well-known method for solving MDPs is backward dynamic programming (BDP), however, BDO suffers under the *course of dimensionality* and the *course of modelling* when modelling the life cycle of ES. This motivated the authors to find new, untraditional methods which are able to cope with these challenges.

We ended up choosing Q-learning (Watkins, 1989), a model-free reinforcement learning algorithm. As illustrated in figure 3, the Q-learning strategy created post decision states (S,x) and an associate value function Q(S,x). The Q-values presents the expected discounted (γ) contribution for taking decision x_t in state S_t, at time t:

$$Q_t(S_t, x_t) = \mathbb{E}\left\{C_t(S_t, x_t) + \gamma \max_{x_{t+1} \in X_{t+1}} Q_{t+1}(S_{t+1}, x_{t+1})\right\} \qquad \text{(eq. 1)}$$

Based on Powell (2008) and Gosavi (2014), the following generic Q-learning algorithm is presented:

Algorithm 1: Generic Q-learning Algorithm (based on Powell (2008) and Gosavi (2014))

Step 0. Initialize

Step 0a. Set $Q^0_t(S_t, X_t) = 0$ for all $t \in T$.

Step 0b. Set N = Max number of iterations

Step 0c. Set n = 1

Step 0d. Initialize S^1_0

Step 1. Choose a sample path ω^n

Step 2. For t = 0,1,2,...,T do:

Step 2a. Choose which decision, x_t, to make

Step 2b. Simulate the outcome of the decision, $S_{t+1} = S^M(S^n_t, x^n_t, W^n_{t+1})$

Step 2c. Estimate the immediate contribution, $C_t(S_t, x_t, S_{t+1})$

Step 2d. Update the Q-value approximation, using equation 2

Step 3. Increment n by 1. If $n \le N$ go to Step 1. Otherwise, go to step 4.

Step 4. Create the policy, π, by finding $x_t = argmax_{x_t \in X_t} Q^n_t(S_t, x_t)$

As see, the optimal policy is found in step. 4. Equation 2 (referred to in step 2c.) is presented below, where α_n is the learning rate[6]:

$$\underbrace{\bar{Q}^n_t(S^n_t, x^n_t)}_{\text{New estimate}} = (1-\alpha_{n-1})\underbrace{\bar{Q}^{n-1}_t(S^n_t, x^n_t)}_{\text{Old estimate}} + \alpha_{n-1}\overbrace{\left[C_t(S^n_t, x^n_t, S^n_{t+1}) + \underbrace{\max_{x_{t+1}\in X_{t+1}} \bar{Q}^n_{t+1}(S^n_{t+1}, x^n_{t+1})}_{\text{Estimated optimal future value}}\right]}^{\text{Learned value}} \qquad \text{(eq. 2)}$$

(VI) Conclusion

The thesis highlights the importance of expanding the traditional system boundary in engineering, recognising stakeholder's role in managing the system over its life time. The authors states that paring DSP with MDP is an advanced approach for life cycle management, which is essential to identifying value robust strategic systems. We believe that the knowledge from this thesis can be important in life-cycle management of high-value, complex, engineering systems, with long-lifetime, facing high degree of exogenous uncertainty

(VII) References

Bellman, R. (1954). The theory of dynamic programming. Bull. Amer. Math. Soc, 60:503–516.

de Weck et al. (2007). A classification of uncertainty for early product and system design. MIT..

Gosavi, A. (2014). Reinforcement Learning: A Tutorial Survey and Recent Advances.

Powell, W.B. (2008) Approximate Dynamic Programming – Solving the Curse of Dimensionality. Willey.

Rehn et al. (2017): Quantification of changeability level for Engineering Systems (DRAFT). NTNU

Watkins, C.J. (1989) Learning from Delayed Rewards. PhD thesis, Kings College, Cambridge.

(VIII) Acknowledgements

This thesis would not have been possible without the support of my supervisor Bjørn Egil Asbjørnslett[7], and the guidance for Sigurd S. Pettersen[7], Carl F. Rehn[7], Jose J.A. Agis[7], and Dr. Austin A. Kana[8].

(IX) Footnotes

1) Working title.
2) A class of uncertainty that is independent of system designs and development plans (de Weck et al., 2007)
3) Due to space limitation the VADS framework is not presented in this poster.
4) We propose the term Strategic system as a set of distinct devises used to handle uncertainty. A strategic system comprises a selected design configuration and life cycle strategies. This expands the traditional system boundary in engineering to also include the managerial dimension.
5) Quite similar to financial options, *real* options is the right, but not the obligation, to exercise actions or to make specific project decisions at a future time.
6) Representing how fast the algorithm learns the approximated Q-value
7) Norwegian University of Science and Technology
8) Delft University of Technology